Structure and dynamics of Smoothened (SMO)– a potential target for regulation of the Hedgehog (HH) signaling pathway

Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy
in
Bioinformatics

by

Shweta Kumari
20172078
shweta.kumari@research.iiit.ac.in

International Institute of Information Technology
Hyderabad - 500 032, INDIA
December, 2023

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled "**Structure and dynamics of Smoothened (SMO)– a potential target for regulation of the Hedgehog (HH) signaling pathway**" by Shweta Kumari, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Supervisor: Prof. Abhijit Mitra

Co-Supervisor: Prof. Gopalakrishnan Bulusu

To my mother, sisters & dearest ones

Acknowledgements

Completing a Ph.D. journey is a wonderful achievement, and it would not have been possible without the encouragement, support, and affection of many people.

गुरु गोविन्द दोऊ खड़े, काके लागू पाय |
बलिहारी गुरु आपने, गोविन्द दियो बताय | |

Kabir Das emphasizes the value of the Guru in our life in this couplet. He suggests that when the Guru and the Divine are together, it is the Guru who should be honored first. The Guru holds a position of great significance as they are the ones who introduce us to the Divine.

First and foremost, I would like to thank my mother, my first Guru, who has been a constant source of strength and support for me. Your love, care, encouragement, sacrifices, and belief in God have been the foundation upon which I have built my academic aspirations. Your belief in me has given me the confidence to pursue my dreams, and I am forever grateful to you. I am greatly indebted to my Guru(s) supervisors, Prof. Abhijit Mitra and Prof. Gopalakrishnan Bulusu, for their unwavering support, guidance, and encouragement, without which this work would not have been possible. Their invaluable insights, patience, and belief in my abilities have shaped my research journey. Prof. Mitra is an excellent source of ideas and inspiration who oversaw any project with enormous patience and kindness. His ability to think in various directions is genuinely remarkable. He advised me throughout the research and thesis writing process. I am deeply grateful to Prof. Bulusu for his remarkable patience in not only guiding me but also for his understanding and support in dealing with my writing. His expertise and constant dedication have played a significant role in honing my skills and improving the quality of my work. Both have been my guardians, mentors, and easily accessible persons throughout this journey. I am extremely fortunate to have had the opportunity to benefit from their intelligence and dedication.

It would not have been possible to pursue my research at CCNSB with Prof. Mitra and Prof. Bulusu if I had not met Dr. Arjun Ray and Dr. Nidhi Jatana, my former colleagues at CSIR-IGIB, Delhi. Dr. Arjun Ray, who calls me Basanti, introduced me to Prof. Mitra and motivated me to do my doctorate here. Dr. Nidhi Jatana, who adores me like her sister, booked my flight tickets so that I could attend the Ph.D. interview. I would like to express my

gratitude to Dr. Lipi Thukral (CSIR-IGIB, Delhi), Prof. Ravindra Singh Rathore, and all the Central University of South Bihar professors for sparking my interest in computational biology.

I would like to express my heartfelt gratitude to Prof. U Deva Priyakumar, Prof. Marimuthu Krishnan, Prof. Prabhakar Bhimalapuram, Prof. Nita Parekh, all the professors in the CCNSB department and Mrs. Sriranjani K for their invaluable support and guidance throughout my academic journey. Thank you for sharing your knowledge and expertise, which has contributed significantly to my growth and learning. I am deeply grateful to Dr. Semparithi Arvindan for his invaluable support in providing outstanding high-performance computing facilities and many resources for efficient storage management. I also want to thank DST, the Government of India, and Prof. Mitra for funding my research. I want to make a special thanks to Prof. Amitabha Chattopadhyay for his detailed discussion, valuable comments on my research, warm compliments, and enriching my knowledge. I am grateful to Mr. Balsantosh, Mrs. Sirisha, Mrs. Nagamani, Ms. Prathima Mandapati, Burugu Pushpalatha and all administration staff who have facilitated the smooth functioning of the university and provided assistance whenever needed. I am grateful to Umesh and his team for keeping the workspace clean and organized.

I would like to sincerely thank Dr. Navneet Bung, Dr. Siladitya Padhi, Dr. Bipin Singh, and Dr. Antarip Halder for generously dedicating their time to engaging in fruitful discussions and providing valuable suggestions.

A special thanks to Murtuza, the closest person with whom I have lived and shared my campus life. Your strength and belief in me have empowered me to be confident in my abilities. I also want to convey my gratitude to my dearest people at IIIT, Rohit Modee, Nishtha Pandey, Nirmal Surange, and Tirusha, for being my circle of support and trusted companions during this challenging journey. Your friendship and unwavering support have been a constant source of strength, laughter, and comfort, and I am truly grateful for the wonderful memories we have shared. Thank you, Mr. Modee, for always being there to listen to me. Whether discussing work-related issues or engaging in lighthearted conversations, your attentive ear has been invaluable. Additionally, I am extremely grateful for your generous act of lending me your laptop for my thesis writing. Nishtha, I deeply appreciate your caring nature and sisterly affection, which have made me feel safe and loved. Your support and assistance have played a significant role in my academic journey.

I would like to thank my group members Antarip, Preethi, Sohini, Ramya, Saurabh Vemuri, Rohit Roy, and Dedeepya for their invaluable input during our lab meetings. Additionally, I would like to express my heartfelt gratitude to my fellow lab mates, both senior and junior, Broto, Kartheek, Prashanthi, Mohan, Sanchari, Parijaat, Tanashree, Vijay, Manisri, Pradeep, Cyrin, Shruti, Suba, Srilakshmi, Akansha, Rami, and Vishal, for their companionship, knowledge-sharing, and unwavering support. I was fortunate to have Nitin, Ashish K. Abhinav, Pulkit, Kuntal, Ruchi, Priyanka, Bhanu, Aniket Bhagat, and many more in my Ph.D. journey.

I would like to extend my heartfelt thanks to Mrs. Mitra (kaki-maa) and Mrs. Bulusu (Sudeshna Ma'am) for their motherly care and support. Your compassion and understanding have greatly impacted my life, and I appreciate your nurturing presence. I am grateful to Prof. Aniket Alam and his family (Prof. Manjari Katju, Rewa, and Sara), whom I met during the difficult moments of the COVID-19 pandemic. Your kindness and support gave me a feeling of family during a period of uncertainty, for which I am thankful. I am also grateful to the families of my friends, especially Murtuza, Rohit, Nishtha, and Pradeep, for their love, care, delicious meals, and the memorable homely environment they provided.

I would also like to thank Prof. Aniket Alam and Prof. Radhika Krishnan for giving me an opportunity to volunteer in the community kitchen during the pandemic. It is always less to thank Prof. Bhimalapuram for safely managing the situation on campus during the lockdown, whether by arranging weekly Covid-19 testing or organizing the vaccination drive for us. I am grateful to my IIIT family (including faculty, their families, staff, and students) for their efforts to keep the campus safe. I would like to thank Prof. Mitra, Prof. Bulusu, and their families again for every possible support during COVID-19.

I want to thank my non-campus pals, Divyani, Ashish, Abhishek, Vinayak Tiwari, Asad, Kritika, and Karim, for regularly checking in on me. Despite the physical distance, your friendship has not changed. My apologies that I couldn't make it to Kritika, Divyani, and Abhishek's wedding. Thank you for being such an important part of my journey.

Last but certainly not least, lots of love goes to my sisters, Swati (my twin) and Shilpi (my younger one). Thank you for being my pillars of support. Your presence, understanding, and unconditional love have been a constant source of comfort and motivation. I am grateful for the shared memories, laughter, and moments of respite that we have cherished together.

A Ph.D. is a demanding endeavor that requires dedication, perseverance, and a thirst for knowledge. It involves countless hours of research, analysis, writing, and self-reflection. This intellectual endeavor has not only broadened my grasp of my field but has also developed me as a person, providing me with vital skills and experiences that will serve me well throughout my career and life.

I am grateful to everyone I have had the privilege of interacting with, who has helped me improve academically and personally during this life-changing journey. I am extremely thankful to those who have brought positivity and goodness into my life, and I appreciate the lessons and strength I have gained from the experiences with those who may not have been as positive. Truly, I am a completely different person now than I was when I started my Ph.D. adventure. This thesis would not have been possible without all of you, including someone I mistakenly overlooked. Your contributions have made me the person I am today, and I carry your influence with gratitude and appreciation as I move forward into the next chapter of my life.

Thank you all from the bottom of my heart.

Abstract

Hedgehog (HH) signaling pathway is the cellular pathway for patterning and morphogenesis. It is controlled by two membrane proteins, the HH receptor Patched1 (PTCH1), a 12-pass transmembrane (TM) protein, and Smoothened (SMO), a 7-pass transmembrane (TM) signal transducer. SMO transduces the signal from the extracellular (EC) region to the cytoplasmic (CP) region. Binding of HH ligand to PTCH1 receptor allows SMO entry to the primary cilium (PC) and activates the downstream signaling process. The malfunctioning of SMO results in misregulation of the HH signaling pathway, which contributes to birth defects and various cancers. The story has evolved from initial conjectures involving direct communication of PTCH1 with SMO to indirect communication between the two TM proteins. In the process, some of the old theories have been discarded in recent literature. However, the underlying mechanism of how PTCH1 controls SMO is still not very clear, and thus constitutes an important area of research. Based on available evidence, the consensus appears to be limited to the assertion that, driven by HH signaling, PTCH1 regulates the activity of SMO activity through the intermediacy of some small molecules (such as cholesterol). In addition, it appears that the lipid bilayer composition affects SMO orientation and function. In this context, several earlier hypotheses involving the primary role of cholesterol are being disputed. At the same time, several HH pathway inhibitors viz., cyclopamine, taladegib, and vismodegib, which target SMO have been reported. Notwithstanding their potential utility in the treatment of cancer, long-term administration of these drugs was reported to lead to the development of resistance.

The broad aim of my research is to build upon the existing molecular-level understanding of the functioning of the HH pathway, with a view to facilitate the formulation of newer approaches towards therapeutic intervention. The focus of my research is to investigate the structure and domain organization of SMO, which is a class F GPCR. This thesis attempts to provide some insights into the functioning of the SMO receptor, using molecular dynamics-based approaches. It is shown that SMO is activated by cholesterol molecules and that its dynamics is influenced by other membrane lipids such as PI4P. The specific objectives include 1. To identify important motifs, especially including, but not limited to, the lipid-binding sites/motifs. 2. To study the conformational dynamics of SMO, to help understand its molecular behavior and also the role of the composition of the primary cilia. 3. To study the regulation mechanisms in SMO by exploring the different cholesterol-bound structures of SMO.

Contents

List of Figures

List of Tables

Chapter *1*

Introduction

1.1 Overview of Hedgehog pathway

The Hedgehog (HH) signaling pathway is one of the most important evolutionarily conserved cellular pathways associated with embryonic development and cancer throughout the animal kingdom [57]. The HH pathway is a developmentally and therapeutically significant pathway responsible for the expression of glioma-associated oncogenes (GLI) that orchestrate cell morphogenesis and cell proliferation. It controls the development of patterning, tissue growth, mitogenesis, homeostasis, and tissue repair. The name of the pathway is derived from its signaling molecule, HH, which facilitates communication between cells. HH pathway activation is controlled by two transmembrane (TM) membrane proteins, the HH receptor protein, Patched1 (PTCH1) (a 12-pass transmembrane protein), and Smoothened (SMO), a 7-pass transmembrane domain protein of the Class F GPCR family, as the signal transducer [49]. In the absence of HH, PTCH1 inhibits downstream SMO (Figure 1.1a). The signaling cascade begins with the binding of HH to the PTCH1 receptor. This leads to the activation of SMO by releasing the PTCH1-mediated inhibition [42] and translocation of SMO to the primary cilia [42, 49, 52, 111, 125, 166]. However, the detail of this mechanism is poorly understood [128]. The active SMO contributes to the activation of the GLI transcription factors (two activators, GLI1 and GLI2; repressor GLI3) by inhibiting the SUFU-PKA complex via a poorly understood mechanism. These GLI transcription factors then move to the nucleus to bind their specific target genes, which result in the expression of GLI target genes (genes *ptch1* acts as a negative and *gli1* acts as a positive feedback loop) that are involved in cell proliferation and cell morphogenesis [3, 19, 69, 166] as described in Figure 1.1b.

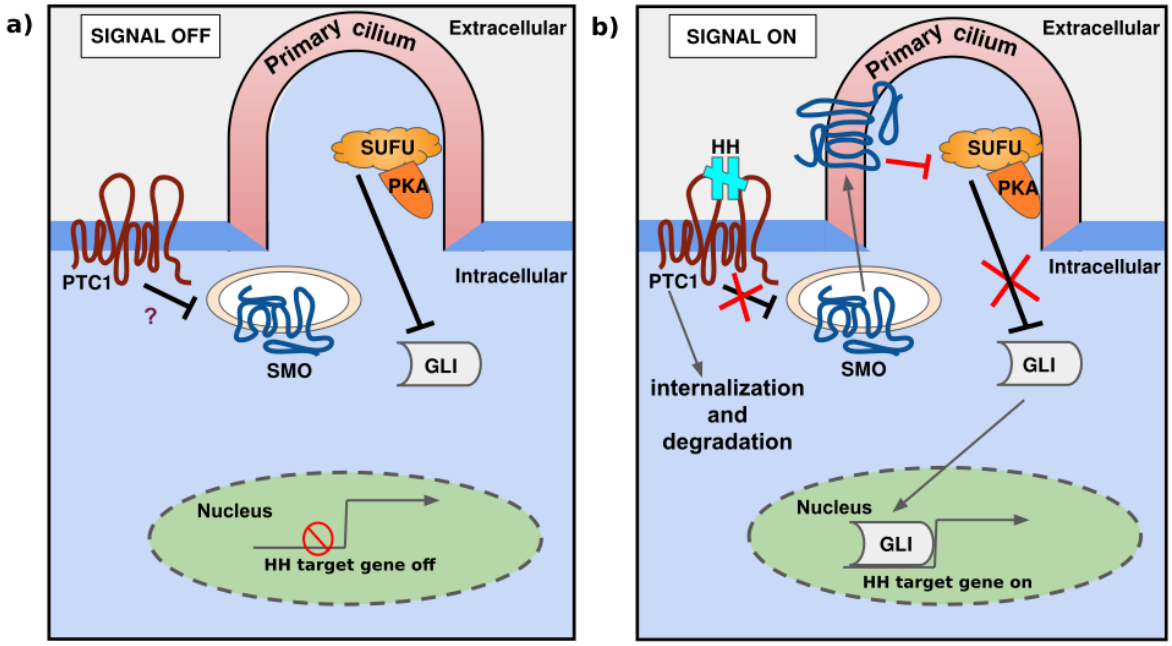


Figure 1.1: A schematic overview of the HH pathway. a) In the absence of HH, PTCH1 inhibits SMO, enabling SUFU and PKA to inhibit the activity of GLI transcription factors; and b) binding of HH to PTCH1 displaces it from the primary cilium and stops inhibition of SMO. This allows SMO to accumulate into the phosphatidylinositol-4-phosphate (PI4P) enriched primary cilium, resulting in downstream signal transmission and the transcription of HH target genes by the GLI transcription factors.

1.2 Role of the SMO and HH signaling pathway in cancer

SMO is referred to as a "proto-oncogene"; by definition, proto-oncogenes are genes that play a role in cell growth and division, and if they undergo mutations or other changes, they can become oncogenes, which promote uncontrolled cell growth and division [154]. Dysregulated activation of SMO in HH signaling pathway causes severe developmental abnormalities, serious diseases, and the development of various cancers, including basal cell carcinoma, leukemias, medulloblastoma, and cancers of the lung, colon, liver, pancreas, breast, and prostate [8, 50, 75, 96, 130, 137, 168]. In sporadic tumors, missense mutations in SMO have been shown to activate the HH pathway constitutively [92]. The role of SMO in various cancers is tabulated here in Table 1.1 (reviewed in [75]).

In the cardiovascular system, the HH pathway promotes both differentiation and proliferation. During early embryonic development, it is important for the morphogenesis of the heart to determine left/right asymmetry and control the correct looping of the heart tube [34]. HH

Table 1.1: SMO in cancer.

S.no	Cancer name	Cause	References
1.	Breast carcinoma	Expression of SMO (in ductal carcinoma)	[108, 165]
2.	Basal cell carcinoma	R562Q, W535L, R199W, S533N, V321M, L412F, F460L mutations in SMO	[8, 75, 92]
3.	Medulloblastoma	W535L, D473H, S533N, S278I mutations in SMO	[75, 112, 154, 185]
4.	Colorectal cancer	Overexpression of SMO	[96]
5	Intestinal tumorigenesis	Indirect role of SMO (by increasing Wnt signaling)	[75]
6.	Hepatocellular carcinoma	K575M mutation (missense) in SMO	[45, 145]
7.	Pancreatic cancer	D473H mutation in SMO, overexpression of SMO	[75, 90]
8.	Colon cancer	Overexpression of SMO, A324T, V404M, and T640A mutations in SMO	[46, 56, 96, 191]
9.	Lung carcinoma (non–small-cell lung cancer; NSCLC)	P641A mutation in SMO	[159]

signaling is involved in the maintenance and repair of adult cardiovascular tissues, as well as the regulation of angiogenesis [14, 149]. Table 1.2 provides an overview of the various tissues where the HH pathway plays a role in the development and its potential involvement in tumor formation [14]. Based on the role of HH signaling in the development of various cancer types which have been further classified into three categories. Category 1 represents cases where abnormal HH signaling initiates the formation of the tumor; Category 2 involves instances where HH signaling contributes towards the maintenance of the tumor; and Category 3 involves cases where the role of HH signaling is implicated, but its specific contribution is not yet fully understood [14]. According to a report on "carcinogenesis and tissue repair," HH, along with the Wnt signaling pathway, is activated in many cancers, and this activation is required for cancer growth [15, 153]. These cancers are responsible for roughly one-third of all cancer deaths [15, 153].

In this scinario, the HH signaling pathway has emerged as a promising therapeutic target for a range of human cancers. Consequently, SMO being a GPCR, key player of HH pathway, is highly druggable. SMO is easily accessible, has a suitable binding site, plays a important role in disease, its activity or function is modifiable by small molecules, and has available structural information (discusse in section 1.3.1 and 1.4). Therefore, SMO has become the most prominent target for cancer therapies, and various anti-cancer molecules that target SMO, including

cyclopamine, taladegib, and vismodegib, have been reported [51] (Figure 1.2). However, given the limited efficacy of these molecules against several cancers, there is a need for a rational drug design approach for a therapeutic intervention involving a molecular-level understanding of the mechanism of SMO activation.

Table 1.2: HH pathway involvement in tissue development and tumorigenesis. Based on the role of HH in their respective formation and/or proliferation, the tumors are placed in one of three categories. (Source: [14]).

S.no	**Role of HH in**		**Category of tumor**
	Normal tissue growth	**Cancer/Tumor growth**	
1	Eye, tooth, skin, and hair follicle	Basal cell carcinoma (BCC)	1
2	Central nervous system (CNS)	Medulloblastoma	1
3	Muscle, bone, epithelial, and soft tissue	Rhabdomyosarcoma	1
4	Muscle, bone, epithelial, and soft tissue	Breast carcinoma	3
5	Heart and vasculogenesis		
6	Pancreas	Pancreas	2
7	Gastrointestinal	Colorectal	2
8		Esophageal, gastric, and hepatic carcinoma	3
9	Thymocytes and hematopoiesis	Lymphoma and leukemia	3
10	Bronchial branching	Lung carcinoma	3
11	Ovary	Ovarian carcinoma	3
12	Testis and prostate	Prostate carcinoma	3
13		Hepatocellular carcinoma	

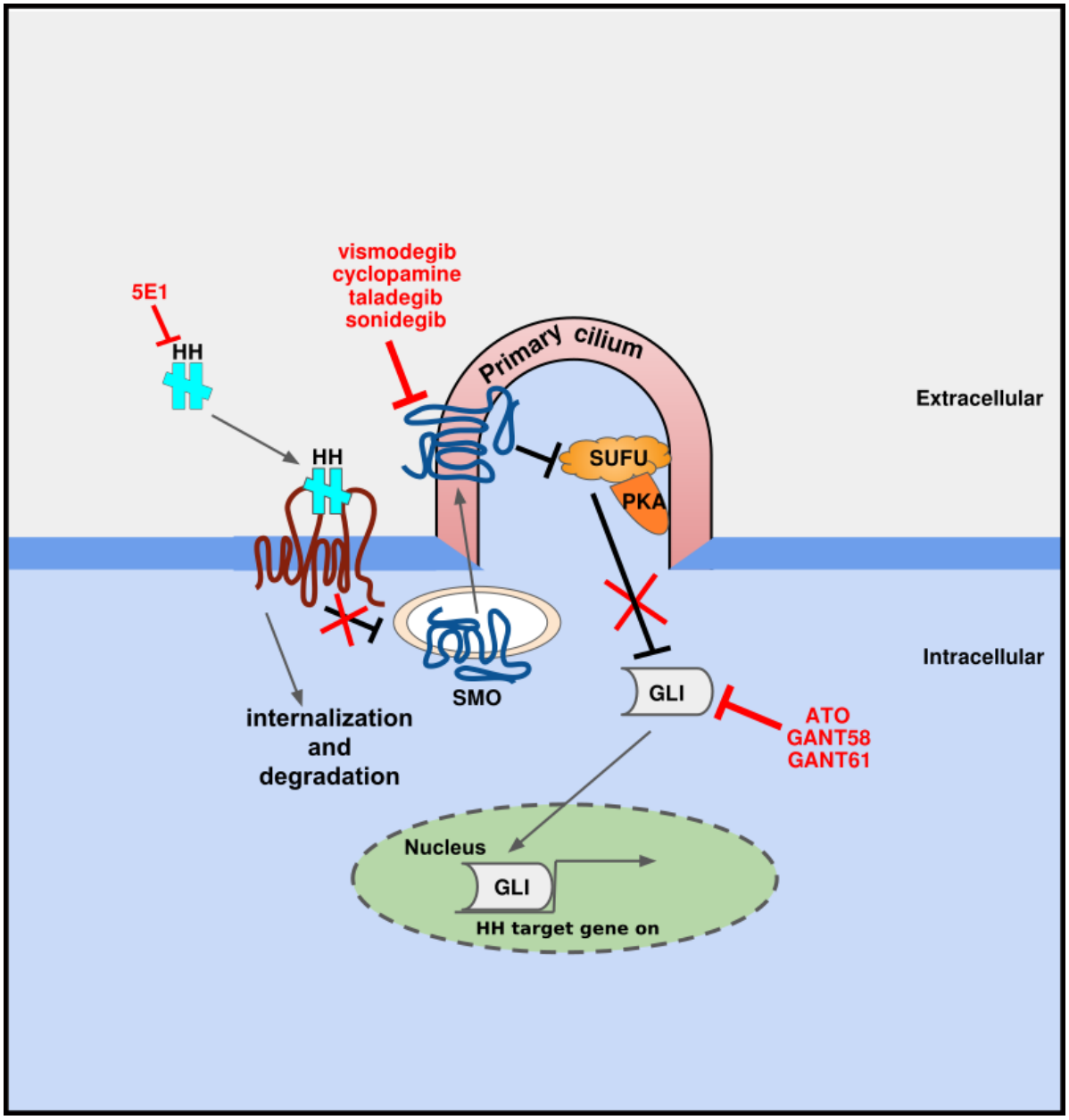


Figure 1.2: A schematic representation shows the reported antagonists of the HH signaling pathway at different stages.

1.3 The central role of SMO and its relationship with other key players in the HH signaling pathway

1.3.1 Smoothened (SMO) receptor- The queen of the HH pathway kingdom

SMO is a critical component of the HH signaling pathway. It transmits the signal across the cell membrane. SMO belongs to the Class F (Frizzled/FZD) family of GPCR proteins. It shows only 10% sequence similarity with class A GPCRs and high sequence similarity to the FZD receptor, which is involved in the WNT signaling pathway. The closely related FZD receptors are also seven-pass transmembrane proteins with cysteine-rich domain (CRD). However, the FZD receptors bind Wnt ligands, whereas SMO does not bind any morphogen. Instead, PTCH1 binds the extracellular signal, HH ligand, and releases the inhibition on SMO. The transition from active to inactive states of the HH pathway in mammalian cells includes a rapid translocation of SMO and other regulatory components to the primary cilium. In the absence of an HH ligand, the presence of ciliary SMO is minimal and limited to the base region surrounding the cilia. The structure of SMO as a seven-pass TM protein (described in Section 1.4) provides support to the belief that SMO can couple to heterotrimeric ($\alpha\beta\gamma$) G-proteins, a characteristic common to numerous examples of cell-surface seven-pass TM receptors for hormones, paracrine substances, and sensory stimuli [141].

As discussed in the previous section, SMO is highly druggable, and there are various small molecules that can modulate its activity. These may be agonists or antagonists. A few examples of agonists are the synthetic small molecules SAG [28] or tri-substituted purine derivative purmorphamine [182]. Examples of antagonists include cyclopamine- naturally occurring plant alkaloids [154], a few synthetic compounds such as SANT-1 [28], and vismodegib [132].

1.3.2 Other key players in the HH pathway

1.3.2.1 Hedgehog (HH) ligand

The HH signaling pathway is one of the important regulators of animal development and is found in all bilateral animals. It gets its name from its intercellular signaling molecule called HH, which is a polypeptide ligand and was first identified in Drosophila. HH is a segment polarity gene product of Drosophila that is important in forming its body plan. Mammals have three HH homologs, Desert hedgehog (DHH), Indian hedgehog (IHH), and Sonic hedgehog (SHH). SHH is the most well-studied vertebrate pathway ligand. The majority of what is known about HH signaling has come from research into SHH [32, 89, 120, 133]. A comparative study of three homologs of HH has been done by Pathi et al. to study the role of HH homologs

[119] summarized in Table 1.3.

Table 1.3: Role of HH homologs (source: [119]).

S.no	**HH homolog**	**Function**
1	SHH	• Establishment of the early left-right (L-R) axis in the chick embryo • Regulation of ventral cell fates in the central nervous system (CNS) • Specification of anteroposterior (A-P) axis in the limb
2	IHH	• Modulating chondrogenesis in the appendicular skeleton • Negative regulator of the differentiation of proliferating chondrocytes
3	DHH	• Germ-cell proliferation • Development of germ cells toward the later stages of spermatogenesis • Signaling peripheral nerve ensheathment • Nerve-Schwann cell interactions

SHH is initially translated as a $\sim$ 45kDa precursor, and it goes through an autocatalytic process to produce a $\sim$ 20kDa N-terminal domain or N-terminal signaling domain (SHH-N) and a $\sim$ 25kDa C-terminal domain (which does not have any currently known function in signaling) [32, 89, 120, 133]. During the cleavage, SHH-N is modified with a cholesterol moiety at its C-terminal and a palmitic acid group at its N-terminal (Figure 1.3a). Cholesteroylation helps SHH-N in the trafficking, secretion, and receptor interaction of the ligand [32, 89, 120, 133]. The crystal structure of human SHH-N (PDB ID- 3M1N) shows that this peptide has both α-helix and β-sheets secondary structure elements and is classified as in $\alpha + \beta$ protein category (Figure 1.3b) [120]. A cryo-EM structure of human PTCH1 in complex with SHH-N (PDB ID-6E1H) shows that SHH-N is bound near ECD1 and loop of ECD2 of PTCH1 (Figure 1.5) [53]. The interaction occurs mainly through polar and charged amino acid residues. The details of the interaction are discussed in the next section (1.3.2.2).

Unlike other signaling pathways, where the signaling molecules bind to their signal transducer protein, SHH does not bind to the HH signal transducer protein SMO.

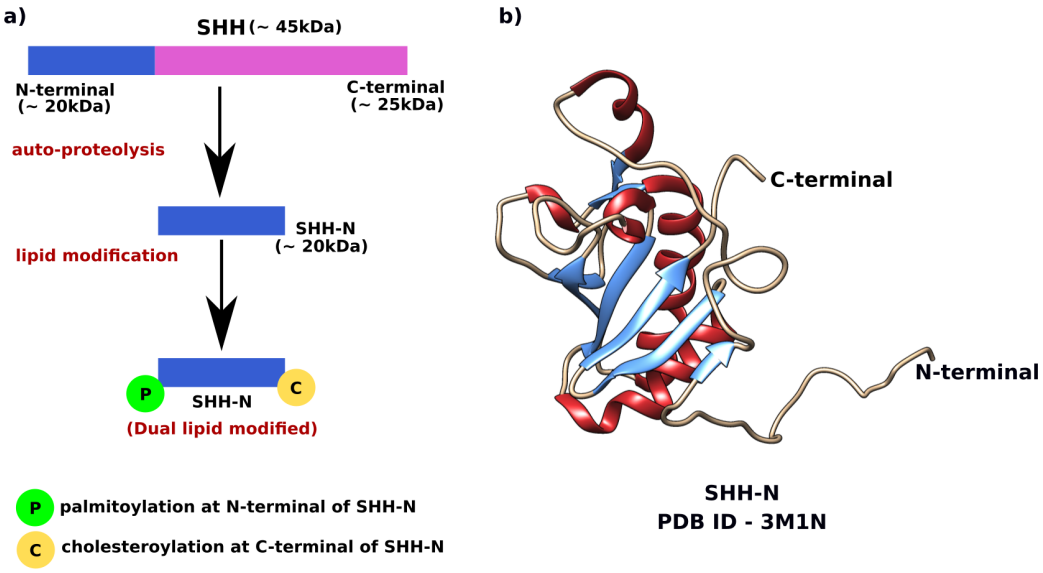


Figure 1.3: a) Overview of SHH post-translational modifications, and b) crystal structure of N-terminal domain of Human SHH.

1.3.2.2 Patched (PTCH) receptor

PTCH is the HH receptor that plays a crucial role in the HH pathway. Mammals have two PTCH genes, Patched1 (PTCH1; consists of 1447 residues in humans) and Patched2 (PTCH2). These two PTCH proteins are homologs of PTCH found in Drosophila. PTCH1 and PTCH2 have the expected 12 transmembrane topologies that consist of N- and C-terminal domains, both located inside the cell, along with six extracellular loops (ECLs) and five intracellular loops (ICLs). Among these loops, the central intracellular loop is the longest. Interestingly, PTCH1 and PTCH2 exhibit similarities with the resistance-nodulation-cell division (RND) family protein, a type of efflux pumps found in gram-negative bacteria, which are known to contribute significantly to multidrug resistance. They act as a cholesterol transporter [27, 53, 83]. It has the sterol-sensing domain (SSD) from TM2 to TM6 (Figure 1.4). ECL1 and ECL4 loops, two larger loops between TM2 and TM3, TM7 and TM8, respectively, are termed extracellular domains, ECD1 and ECD2. It has been reported that ECD1 and ECD2 play an important role in SHH binding to the receptor and are involved in forming the binding site for SHH-N (Figure 1.5a-b). Residue E212 of ECD1 interacts with K87, R123, and R153 on SHH-N (Figure 1.5c). E221 on helix 3 interacts with Y44 and K45 of SHH-N, while the preceding D217 forms a hydrogen bond with S177 (Figure 1.5c). ECD2 E947 and E958 on the H loop form a hydrogen bond with R155 and S151, respectively (Figure 1.5d). ECD1 and ECD2 move

toward each other upon SHH binding to accommodate it [27, 53].

PTCH1 was formerly thought to regulate SMO by directly interacting with it. The theory has evolved to show that there is no direct interaction between PTCH1 and SMO [155]. It is believed that PTCH1 controls SMO via some small molecules, such as cholesterol or an unknown sterol-like molecule, to exert its suppression on the HH pathway [16, 25, 83, 155].

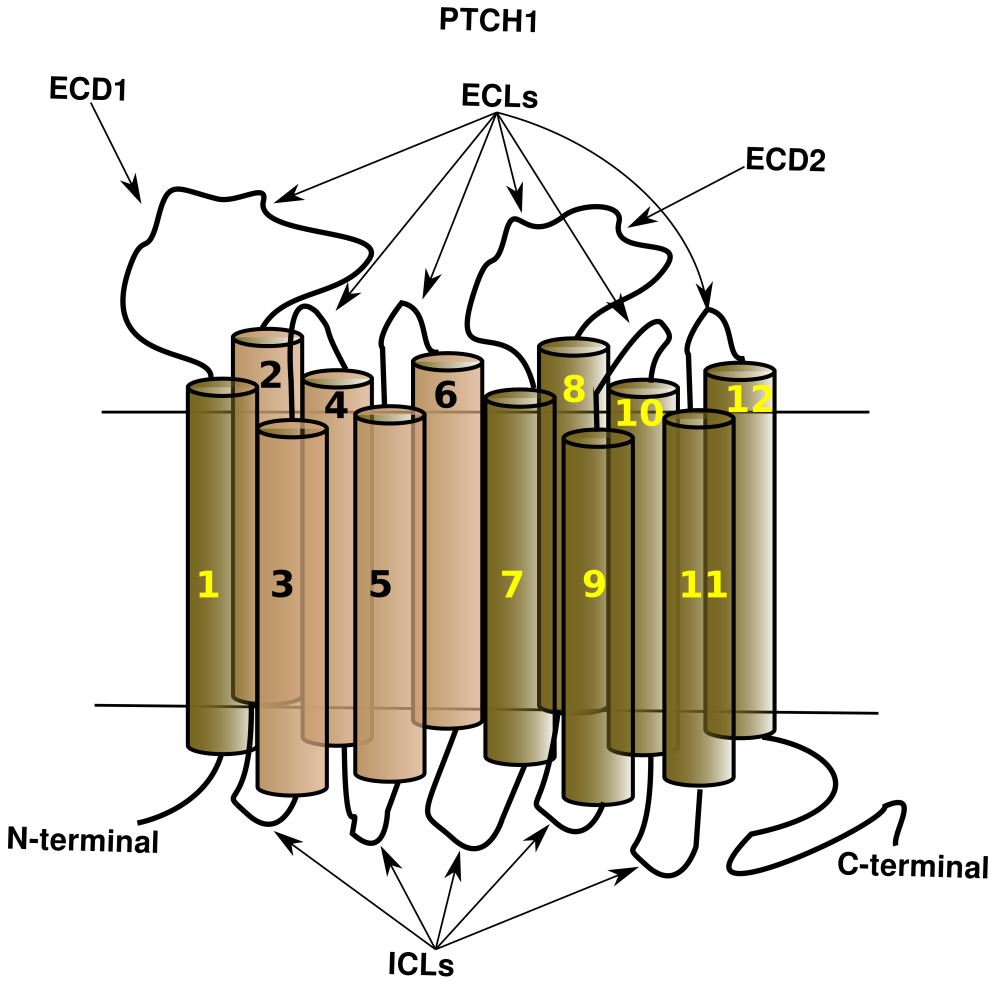


Figure 1.4: A schematic representation of PTCH1. TM helices are numbered, and TM2-TM6 contain the sterol-sensing domain (SSD).

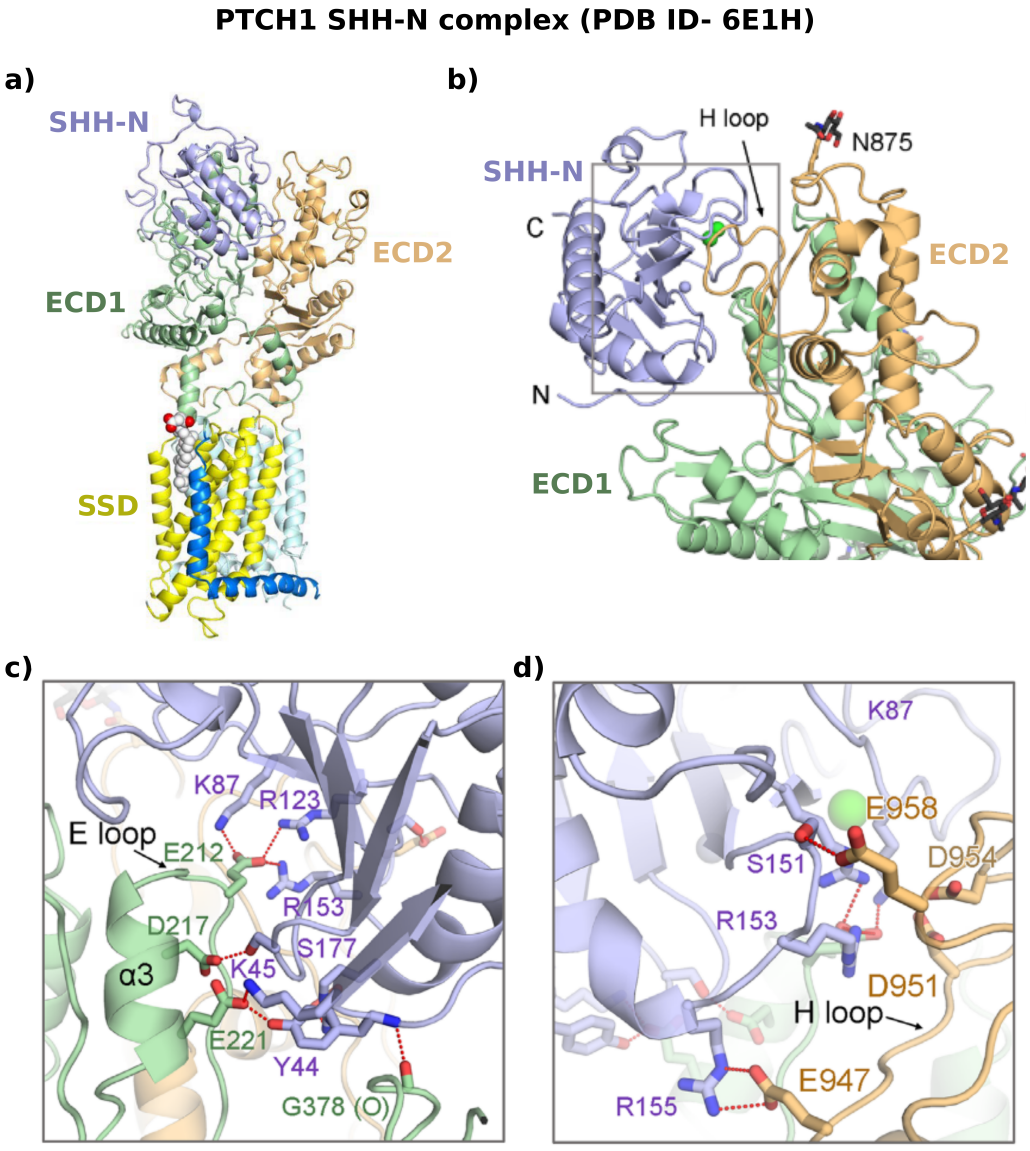


Figure 1.5: Structure of PTCH1 and SHH-N complex. a) Overall structure of PTCH1 with SHH-N. SHH-N is represented in ice blue color. The N-terminal intracellular helix and TM1 of PTCH1 are shown in blue. ECD1 and ECD2 of PTCH1 are colored green and tan, respectively. The SSD region containing TM2-6 of PTCH1 is represented in yellow, while TM7-12 of PTCH1 is shown in cyan. b) The interaction site between PTCH1 and SHH-N is marked with a rectangle. c) Shows the interacting residues between ECD1 of PTCH1 and SHH-N. d) Shows the interacting residues between ECD2 of PTCH1 and SHH-N (Source: [53]).

1.3.2.3 Heterotrimeric GTP-binding regulatory proteins (G-protein)

G-proteins are also important key players in any signaling pathway. They interact with various types of receptors (Wettschureck and Offermanns 2005). G-protein has three subunits: α, β, and γ subunits. The binding of GPCR to G-protein leads to the exchange of GDP (guanosine diphosphate) to GTP (guanosine triphosphate) attached to the Gα subunit. After which, this GTP-Gα is released from GPCR and also disassociated from the $\beta\gamma$ complex subunit (Figure 1.6).

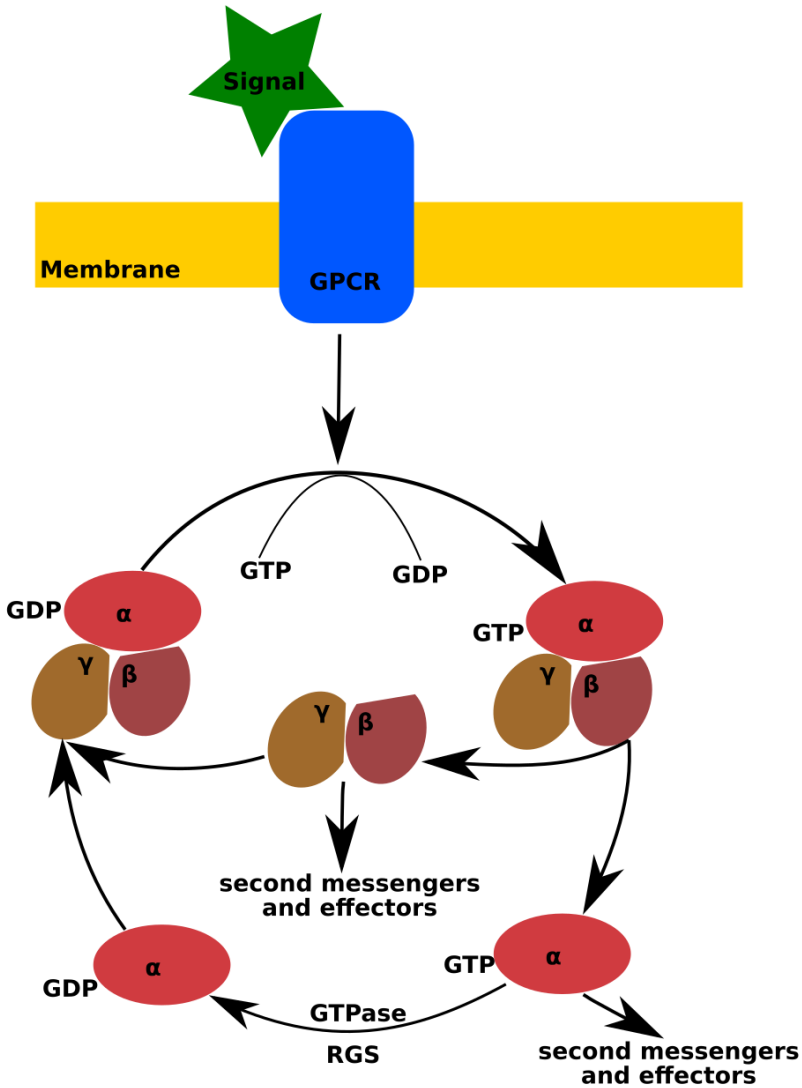


Figure 1.6: Schematic representation of G-protein and overview of their activation mechanism.

Vertebrate SMO has been reported to activate several members of the G_i family (G_{i1}, G_{i2}, G_{i3}, G_o, and G_z), which inhibit the production of cAMP [131, 163]; human SMO also appears to activate $G\alpha_i$-mediated pigment aggregation in Xenopus melanophores, although whether this interaction is direct remains unknown [41]. However, it does not bind to the member of G_s, G_q, and G_{12} families [131].

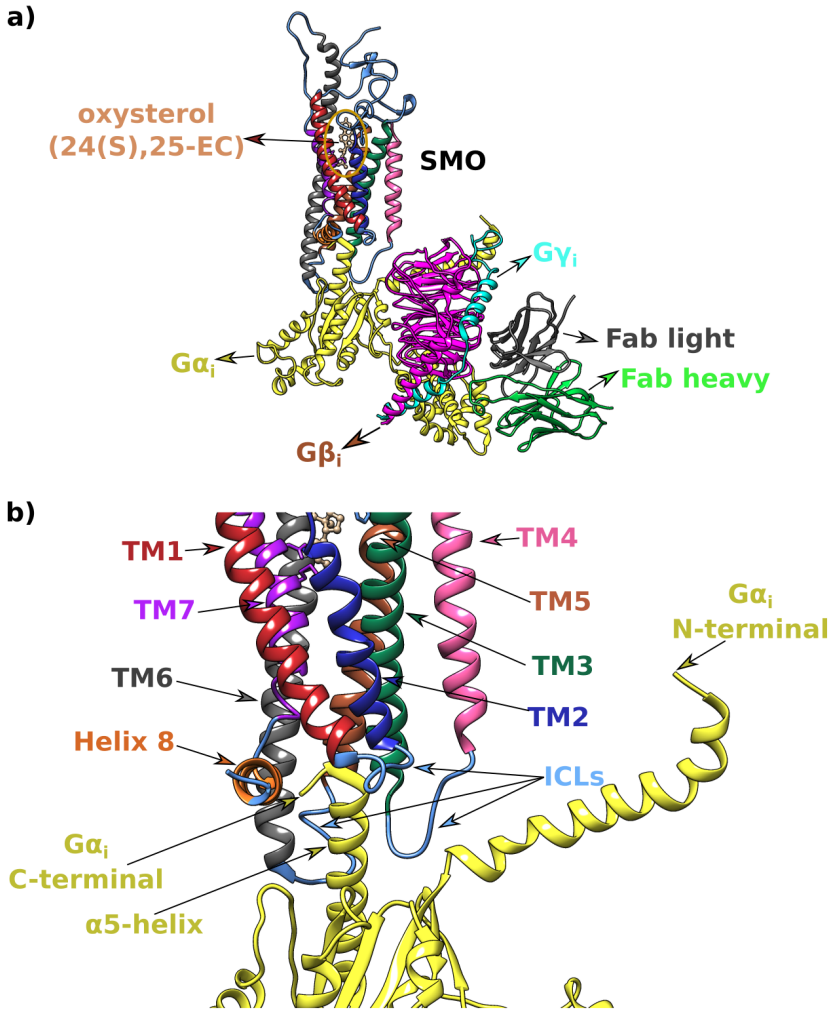


Figure 1.7: Figure 1.7: a) Shows the ribbon representation of the cryo-EM structure of the hSMO-G_i complex (PDB ID: 6OT0). The helices TM1-7 and helix 8, are colored red, blue, green, pink, brown, gray, purple, and orange. TMD bound oxysterol is shown in tan color. The three subunits $G\alpha_i$, $G\beta_i$, and $G\gamma_i$ are colored yellow, magenta, and cyan, respectively. Fab heavy and light are colored dark gray and lime green. b) Zoomed view of SMO-$G\alpha_i$ interaction.

The cryo-EM structure of the hSMO-G_i complex (PDB ID: 6OT0) shows the SMO and G_i-protein interactions (Figure 1.7) [126]. In this cryo-EM structure, α, β, and γ subunits of G_i are attached to SMO. Additionally, the transmembrane domain (TMD) of SMO contains an oxysterol molecule. When compared to the inactive hSMO (PDB ID: 5L7D) state, this structure shows an active conformation, with TM6 moving outward by $\sim$ 7Å and TM5 moving outward by $\sim$ 4Å at the cytoplasmic surface. The major contact between hSMO and G_i is formed by ICLs 1 to 3, TM3, and TM5-7 of hSMO and αN region, αN-β1 loop, and 5-helix of Gα_i (Figure 1.7b). As seen in Figure 1.7b, the C-terminal of the α5-helix of Gα_i inserts into the intracellular groove of hSMO.

1.3.2.4 Glioma-associated oncogene transcription factors (GLI) proteins

GLI is one of the important regulatory networks, which regulate animal development, and it is found in various species, ranging from flies to humans, indicating its conservation across different organisms [76]. Mammals have three transcription factors for HH pathway signal transduction: GLI1, GLI2, and GLI3 (GLI; glioma-associated oncogene family members 1, 2, and 3) [70, 72]. They are the first human transcription factors: zinc-finger proteins determined with a DNA-binding site and resolved with the 3D structure of their DNA-binding domain [70]. GLI1 exists only in the activator form, whereas GLI2 is processed in the active form and is suppressed by binding to SUFU in the cytosol [37, 80, 115, 116, 160]. Unlike GLI2, GLI3 exists in the repressor form and does not participate in signaling [170]. Hence, GLI2 acts as the principal transcriptional activator (GLI^A), while GLI3 acts as the principal transcriptional repressor (GLI^R) [10, 121, 167]. Although GLI1 does not contribute to HH signaling, it does create a positive feedback loop that increases the ability of the pathway to activate transcription [10, 116]SUFU is another negative regulator of the HH pathway. It regulates the activity of several GLI proteins. For example, SUFU suppressed the GLI^A function through the recruitment of a corepressor complex [31]. Additionally, it facilitates the formation of GLI^R by binding to GSK3 formation through binding GSK3 [85]. Active SMO promotes GLI activation by inhibiting SUFU at the primary cilium. SUFU is critical for mammalian HH signaling. It has been reported that SUFU inactivation in mice results in the activation of the ectopic pathway and death of the developing embryos [34, 151]. The study on drosophila SMO and mice SMO suggest that in the absence of HH, SUFU binds with the GLI proteins and retains them in the cytoplasm and hence stops the pathway activation [134].

1.3.2.5 Protein kinase A (PKA)

Protein kinase A (PKA) is a widely recognized kinase that is involved in a wide range of biological processes. In the HH pathway, the primary targets of PKA are the GLI transcription factors, which activate and repress HH target gene expression. PKA promotes the phosphory-

lation of GLI transcription factors, which results in the production of the GLI^R, the repressor forms of GLI discussed in the previous section, thereby suppressing the expression of HH target genes [88].

Table 1.4: Key players of HH signaling pathway, their property, and function.

Sl. no.	Name	Property	Function
1	HH (SHH/IHH/ DHH)	Signaling molecule	Acts as PTCH ligand
2	PTCH1	12-pass transmembrane	Inhibits SMO, receptor protein for HH, transport cholesterol
3	SMO	7-pass transmembrane, Class F GPCR family	Signal transducer
4	SUFU	Binding protein	Negative regulator, inhibits GLI
5	GLI	Transcriptional activators	Binds to target gene of HH pathway to switch on the HH pathway

1.4 Molecular architecture of SMO

The structure of SMO reveals that it has an extracellular domain (ECD), a TMD domain, and an intracellular C-terminal domain (ICD). The amino-(N-) and carboxy-(C-) termini of SMO are located at the ECD and ICD sides of SMO, respectively (Figure 1.8). The ECD consists of signal peptide (SP; 27 residues), cysteine-rich domain (CRD; 164 residues), and linker domain (LD; 30 residues). Like FZD receptors, the CRD of SMO contains four α-helixes and two short β-barrel motifs. The TMD (333 residues) consists of seven-transmembrane (7TM) α-helices, three extracellular loops (ECLs; ECL1, ECL2, and ECL3), three intracellular loops (ICLs; ICL1, ICL2, and ICL3) and a short helix 8. The connection between α-helices is facilitated by these ECLs and ICLs, and the helix 8 (connected to TM7) runs parallel to the membrane bilayer (Figure 1.8) [9, 24]. SMO has an "outside-in" topology that refers to the orientation where the N-terminal is at the extracellular region and the C-terminal is at the intracellular region (Figure 1.8). The long ICD is composed of 233 amino acid residues.

SMO receptors lack most of the class A GPCR motifs, such as D[E]RY (in TM 3), CWXP (in TM6), and NPXXY (in TM7) [169]. In SMO, TM6 is straighter than in class A GPCRs because of the absence of Pro residue in TM6. It has been reported that SMO has multiple Arg/Lys clusters in the ICD, keeping it inactive. The sequential phosphorylation of the ICD neutralizes the Arg/Lys cluster and converts SMO into an active form [190].

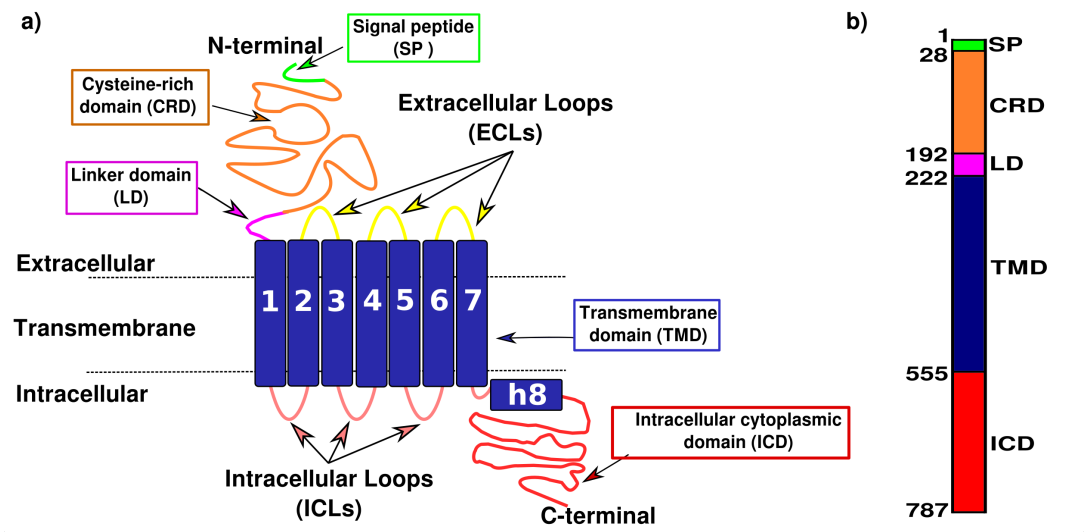


Figure 1.8: Molecular architecture of SMO. a) A cartoon representation of the full-length modeled SMO showing the extracellular (SP, CRD, and LD), transmembrane, and intracellular domains; and b) residue numbers indicating the domain boundaries and domains included in the simulations.

SMO activation has been a mystery for a long time. The first structure of hSMO came in 2013. Since then, several SMO structures have been solved by multiple laboratories (Figure 1.9 and Table 1.5. Currently, 18 structures (13 X-rays and five cryo-EM) of SMO (near full-length) have been made available in PDB. Some of these structures are complex with small molecules, which are agonists and/or antagonists. Most of these structures have thermostable protein (stretches) substituting some of the loops, which, given their flexibility, would have posed crystallization related challenges. Additionally, none of these structures are of full length, they are truncated from N- and C-terminal. Table 1.5 contains this information related to particular PDB IDs.

Table 1.5: SMO structures available in PDB with details.

S. no.	**PDB**	**Species (State)**	**Method (Resolution in Å)**	**Ligand (Type)**	**Other proteins**	**Truncated/ replaced residues**	**Reference**
1	7ZI0	Human (Inactive)	X-ray (3.0 Å)	Cholesterol (Agonist)	BRIL (replaced with ICL3)	SMO-BRIL(ICL3)-ΔICD	[84]

Continued on next page

Table 1.5 – continued from previous page

S. no.	PDB	Species (State)	Method (Resolution in Å)	Ligand (Type)	Other proteins	Truncated/ replaced residues	Reference
2	6XBK	Human (Active)	cryo-EM (3.2 Å)	Cholesterol (Agonist)	Gi/o + Antibodies	ΔCRD-ΔICD, 1–62, 538–787	[125]
3	6XBJ	Human (Active)	cryo-EM (3.9 Å)	Cholesterol (Agonist)	Gi/o	ΔCRD-ΔICD, 1–62, 538–787	[125]
4	6XBM	Human (Active)	cryo-EM (3.2 Å)	24,25-Epoxycholesterol (PAM Agonist)	Gi/o + scFv16	ΔCRD-ΔICD, 1–62, 538–787	[125]
5	6XBL	Human (Active)	cryo-EM (3.9 Å)	SAG, Cholesterol (PAM Agonist)	Gi/o + scFv16	ΔCRD-ΔICD, 1–62, 538–787	[125]
6	6O3C	Mouse (Active)	X-ray (2.8 Å)	SAG21k, Cholesterol (Agonist)	scFv16	ΔCRD-ΔICD, 1–64, 557-787	[43]
7	6OT0	Human (Active)	cryo-EM (3.9 Å)	24,25-Epoxycholesterol (Agonist)	Gi/o + scFv16	ΔCRD-ΔICD, 1-189, 557-787	[126]

Continued on next page

Table 1.5 – continued from previous page

S. no.	PDB	Species (State)	Method (Resolution in Å)	Ligand (Type)	Other proteins	Truncated/ replaced residues	Reference
8	6D32	African clawed frog (Active)	X-ray (3.8 Å)	Cyclopamine (Partial Agonist)	BRIL (replaced with ICL3 residues 402-415) + Fab light chain + Fab heavy chain	ΔCRD-SMO-BRIL(ICL3)-ΔICD, 1-35, 402-415, 526-787	[68]
9	6D35	African clawed frog (Active)	X-ray (3.9 Å)	Cholesterol (Agonist)	BRIL (replaced with ICL3 residues 402-415) + Fab light chain + Fab heavy chain	ΔCRD-SMO-BRIL(ICL3)-ΔICD, 1-35, 402-415, 526-787	[68]
10	5V56	Human (Inactive)	X-ray (2.9 Å)	TC114 (Antagonist)	Flavodoxin BRIL (replaced with ICL3 residues 434 to 443)	ΔCRD-SMO-FLA(ICL3)-ΔICD, 1–52, 559–787, 434-443	[188]

Continued on next page

Table 1.5 – continued from previous page

S. no.	PDB	Species (State)	Method (Resolution in Å)	Ligand (Type)	Other proteins	Truncated/ replaced residues	Reference
11	5V57	Human (Inactive)	X-ray (3.0 Å)	TC114 (Antagonist)	Flavodoxin BRIL (replaced with ICL3 residues 434 to 443)	ΔCRD-SMO-FLA(ICL3)-ΔICD, 1–52, 559–787, 434-443	[188]
12	5L7D	Human (Inactive)	X-ray (3.2 Å)	Cholesterol (Antagonist)	BRIL (replaced with ICL3 residues 429 to 445)	ΔCRD-SMO-BRIL(ICL3)-ΔICD, 1-57, 429-445, 554-787	[24]
13	5L7I	Human (Inactive)	X-ray (3.3 Å)	Vismodegib (Antagonist)	BRIL (replaced with ICL3 residues 429 to 445)	ΔCRD-SMO-BRIL(ICL3)-ΔICD, 1-57, 429-445, 554-787	[24]
14	4QIM	Human (Inactive)	X-ray (2.6 Å)	AntaXV (Antagonist)	BRIL (replaced with ICL3 residues 434 to 440)	ΔCRD-SMO-ΔICD, 1–189, 434-440, 555-787	[168]
15	4QIN	Human (Inactive)	X-ray (2.6 Å)	SAG1.5 (Agonist)	BRIL (replaced with ICL3 residues 434 to 440)	ΔCRD-SMO-ΔICD, 1-189, 434-440, 555-787	[168]

Continued on next page

Table 1.5 – continued from previous page

S. no.	PDB	Species (State)	Method (Resolution in Å)	Ligand (Type)	Other proteins	Truncated/ replaced residues	Reference
16	4N4W	Human (Inactive)	X-ray (2.8 Å)	Sant-1 (Antagonist)	BRIL (replaced with ICL3 residues 434 to 440)	ΔCRD-SMO-ΔICD, 1-189, 434-440, 555-787	[168]
17	4O9R	Human (Inactive)	X-ray (3.2 Å)	Cyclopamine (Antagonist)	BRIL (replaced with ICL3 residues 434 to 440)	ΔCRD-SMO-BRIL(ICL3)-ΔICD, 1-189, 434-440, 556-787	[174]
18	4JKV	Human (Inactive)	X-ray (2.5 Å)	Taladegib (Antagonist)	BRIL (fused to the truncated N terminus at S190)	BRIL–ΔCRD–SMO–ΔICD, 1-189, 555-787	[169]

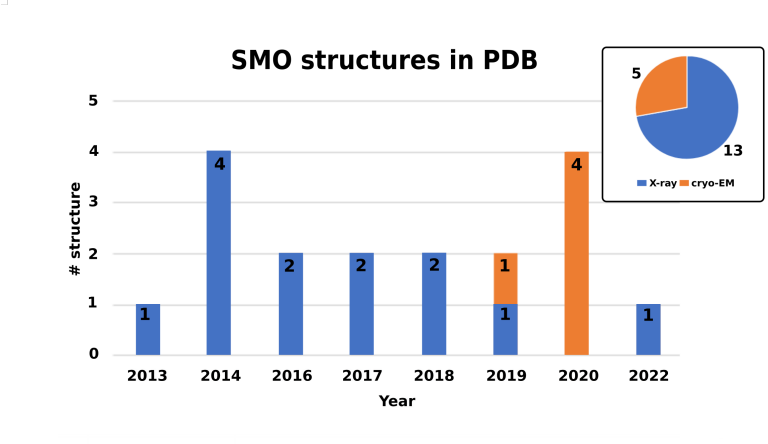


Figure 1.9: Statistics of SMO structures in the PDB databank.

1.5 Role of primary cilia in HH pathway

GPCRs (including SMO) are integral membrane proteins that reside in the plasma membrane of eukaryotic cells. It has been suggested that membrane lipids affect receptor conformation and dynamics, which are critical in transferring signals from extracellular to intracellular regions. The basic structure of the plasma membrane is provided by the heterogeneous phospholipid bilayer that acts as a stable barrier separating the inside from the outside of the cell. The phospholipid bilayer of the plasma membranes of animal cells harbors four major phospholipids: phosphatidylcholine (POPC/PC), phosphatidylethanolamine (POPE/PE), phosphatidylserine (POPS/PS), and sphingomyelin (SM), along with other lipids, including a fifth phospholipid, phosphatidylinositol (PIP), in varying concentrations. These phospholipids are distributed asymmetrically between the outer/upper and inner/lower leaflets of the membrane [73] (Figure 1.10a). The outer leaflet has a high content of PC, and SM, with ganglioside (GM). While the inner leaflet has POPE, POPS, PIP, and other charged lipids. In addition to the phospholipids, the plasma membranes also have glycolipids and cholesterol. The glycolipids (minor membrane components) are present on the cell surface, with their carbohydrate parts exposed. Cholesterol, on the other hand, is a key membrane component of mammalian cells and is present in both leaflets [73]. It is can be noted that both phosphatidylinositol 4-phosphate (PI4P) and phosphatidylinositol 4,5 bis-phosphate ($PI(4,5)P_2$) are present in the plasma membrane [110].

In vertebrates, the Hedgehog (HH) signaling pathway is intimately linked to primary cilia. Activation of the HH pathway triggers the translocation of the protein SMO to the primary cilium, where other signaling components are present, and where crucial signaling events take place. Primary cilia are specialized organelles that projects out from the surface of a cell (Figure 1.11) and functions as a signaling hub [128, 129]. Therefore, it is also important to examine the lipid composition of the primary cilium, which is difficult because of the technical issues involved in isolating and purifying membranes from mammalian primary cilia. However, some studies show that the main characteristic of primary cilia is the presence of inositol lipids. The experimental studies show that the PI4P is the major phosphoinositide of the primary ciliary membrane (Figure 1.11). Whereas another important PIP, ($PI(4,5)P_2$), is limited to the base of the primary cilia [110].

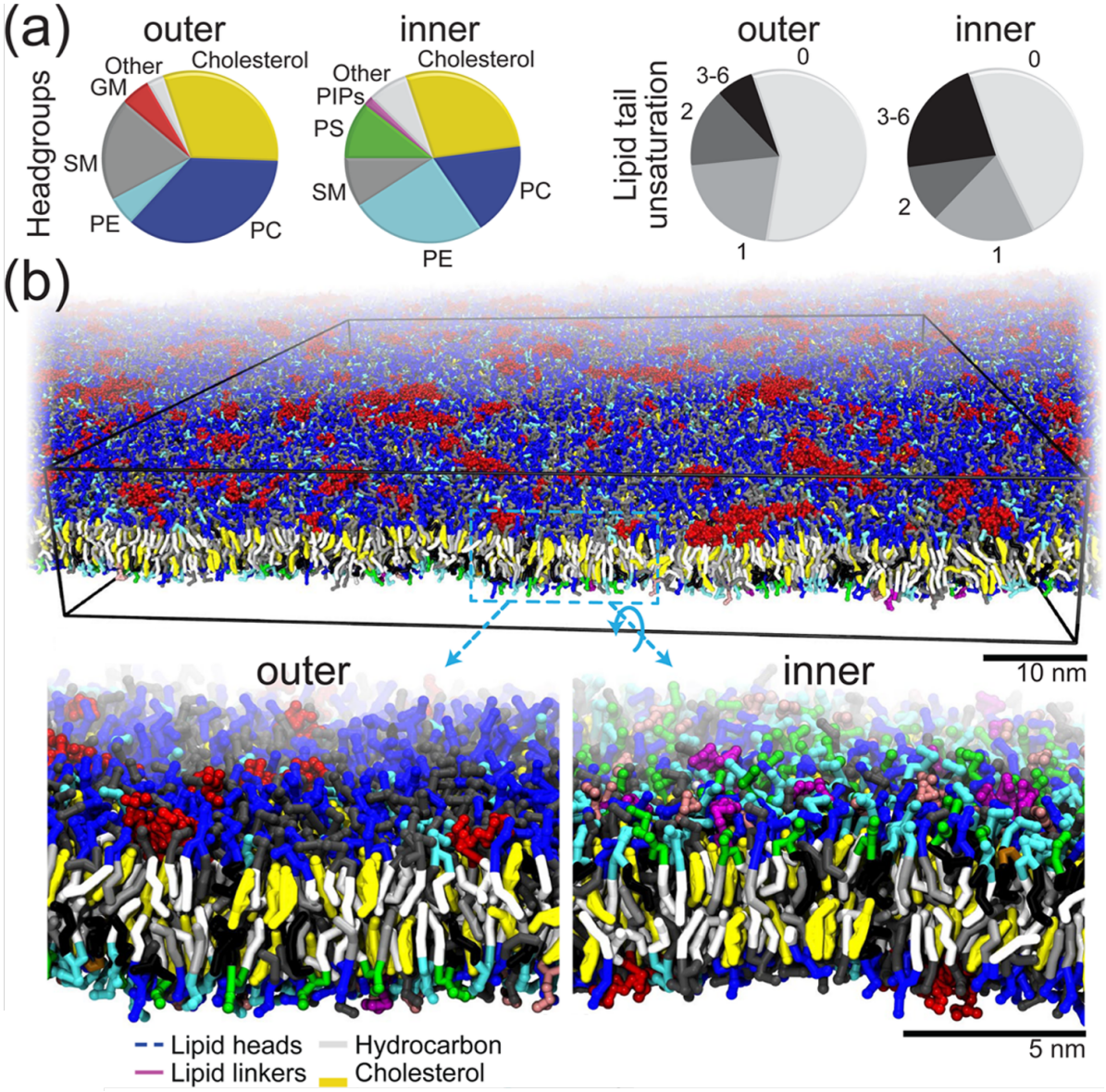


Figure 1.10: Distribution of phospho- and other lipids with cholesterol in the plasma membrane. (a) The first and second pie charts depict the distribution of the lipid headgroups in the outer and inner leaflets, respectively. The third and fourth pie charts depict the levels of lipid tail unsaturation in the inner and outer leaflets, respectively. (b) The plasma membrane and the cross section views with outer and inner leaflets. Cholesterols are shown as yellow, lipid headgroups are colored based on their type (PC - blue, SM - gray, PE - cyan, GM - red, PIPs - magenta, PS - green, PA - white), and the tails are colored according to the number of unsaturated bonds (0 - white, 1 - light gray, 2 - dark gray, 3-6 - black) (source: [73]).

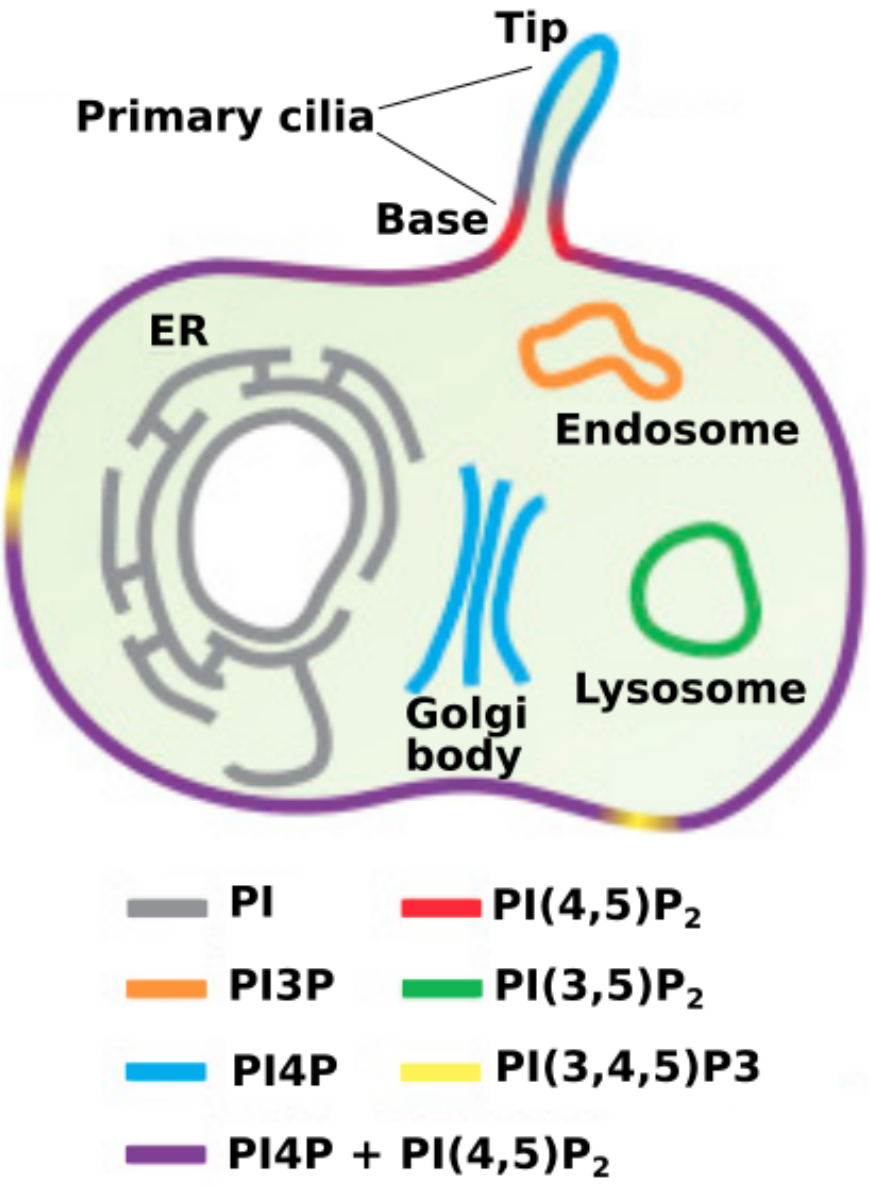


Figure 1.11: The distribution of phosphoinositides (PI) in various cellular membranes (Source: [110].

1.6 Understanding SMO dimerization

1.6.1 GPCR dimerization in general

SMO, as discussed in section 1.3.1 earlier, belongs to the class F family of GPCRs. One important and common aspect of GPCR functioning is GPCR dimerization [140]. Dimerization is basically the association of two identical or different GPCR subunits. It has been reported that GPCR dimerization is an important mechanism for regulating receptor signaling, trafficking, and pharmacology. Dimerization can happen in two ways a) homodimerization, where two identical subunits dimerize, or b) heterodimerization, where two distinct subunits dimerize (Figure 1.12). Many GPCRs, including adrenergic, dopamine, opioid, and serotonin receptors, have been reported to form homodimers that modulate receptor activity by altering ligand binding, downstream signaling, and trafficking.

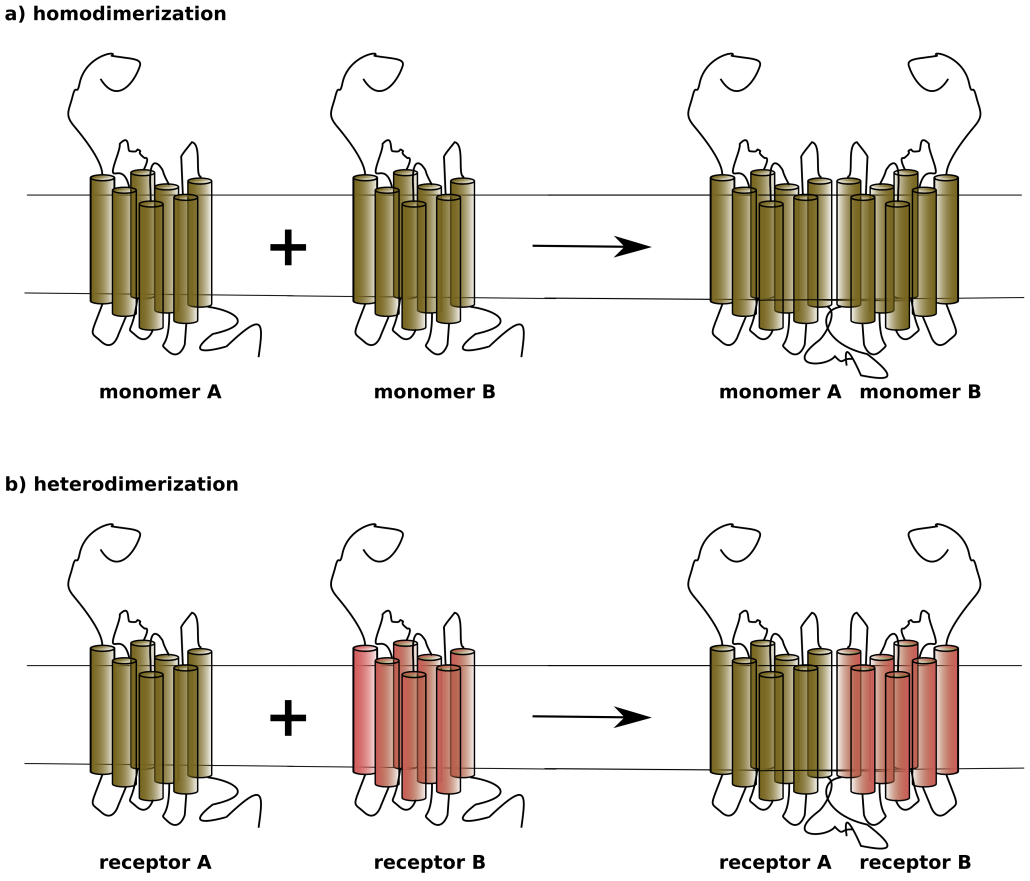


Figure 1.12: Schematic representation of GPCR illustrating the a) homodimerization and b) heterodimerization.

Earlier studies on receptor dimerization suggested that dimerization promotes receptor signaling activity [100, 105, 117, 123, 140]. An example is the dimerization of the β-adrenergic receptor, which has been shown to enhance receptor desensitization and internalization. Another is the dopamine D2 receptor, the homodimerization of this receptor increases its affinity for G-protein coupling [48]. On the other hand, the heterodimerization of GPCRs can result in novel signaling pathways and pharmacology, as well as alter the function of individual receptors. For example, the heterodimerization of dopamine D1 and D2 receptors results in synergistic effects on intracellular calcium signaling [59]. Similarly, the heterodimerization of the adenosine A2A and dopamine D2 receptors alters receptor trafficking and results in a novel pharmacology for the treatment of Parkinson's disease. Table 1.6 lists a few GPCRs that have been shown to homo dimerize.

Table 1.6: GPCRs known to dimerize are listed below.

S.no	Receptor name	Receptor class	Reference
1	Adenosine A1 receptor	Class A	[54]
2	Dopamine D1 receptor	Class A	[95]
3	Dopamine D2 receptor	Class A	[94]
4	M1 muscarinic receptor	Class A	[61, 122]
5	M2 muscarinic receptor	Class A	[122]
6	Metabotropic glutamate receptor 2 (mGluR2)	Class C	[7]
7	Neurotensin receptor 1 (NTS1)	Class A	[178]
8	Secretin receptor (SecR)	Class B	[7]
9	Serotonin 5HT2c	Class A	[95, 99, 139]
10	β2-adrenoreceptor	Class A	[123]
11	μ-opioid receptor (MOR)	Class A	[7]

1.6.2 SMO dimerization in Drosophila Melanogaster

There are multiple reports that suggest the dimerization of SMO in Drosophila Melanogaster (Table 1.7). It was reported that the D. melanogaster SMO (Drosophila SMO) can exist as homodimers in vivo. Additionally, dimerization has significance for the function. It promotes a high level of HH signaling activity (Zhao et al. 2007; Chen et al. 2010, 2011; Maier et al. 2014; Shi et al. 2011, 2013).

Zhao et al. [190] showed that the presence of HH leads to dimerization of the C-terminal cytoplasmic tails (C-tail), which is essential for pathway activation (Figure 1.13a). They also

showed that phosphorylation at multiple Ser/Thr residues in the Drosophila SMO C-tails leads to receptor cell surface accumulation and activation, which induces a conformational switch, ultimately leading to the dimerization of Drosophila SMO through its C-tails. In their fluorescence resonance energy transfer (FRET) experiment on Drosophila SMO tagged with CFP/YFP to its C- or N-terminal, they found significant increases in FRET C, which means the (dimerization through the C-terminal). They have also reported that in quiescent cells, Drosophila SMO adopts a closed conformation, which is an inactive state, where the C-terminal is in close proximity to ICL3. When HH binds to its receptor, PTCH, it triggers SMO to convert in an open conformation, which is an active state, where its C-terminus shifts away from ICL3 but near to the C-terminus of its interacting residues.

Table 1.7: Study of SMO dimerization in *Drosophila Melanogaster*

S.No	Species	Method	Summary	Reference
1	dSMO and mSMO	FRET	dSMO dimerzes by C-tail	[190]
2	dSMO	FRET	dSMO dimerzes by C-tail followed by Gprk2 mediated phosphorylation	[29]
3	dSMO	FRET	dimerization of dSMO-Cos2-Fu complexes	[143]
4	dSMO and mSMO	FRET	dSMO dimerzes by C-tail followed by Gprk2 mediated phosphorylation	[30]
5	dSMO	FRET	formation of oligomers/higher order clusters in lipid rafts of cell plasma membrane	[142]
6	dSMO	BRET	dSMO dimerization after phosphorylation by PKA/CK1 in SAID and and Gprk2	[98]
7	dSMO	FRET	Gprk2 promotes SMO dimerization	[77]

Another study [29, 30] (Chen et al. 2010, 2011) also showed that Drosophila SMO dimerization/ oligomerization takes place through its C-tail. Their study reveals that this process is dependent on HH concentration. According to their model, G protein-coupled receptor kinase 2 (Gprk2/GRK2) triggers Drosophila SMO activation and dimerization/ oligomerization by phosphorylating SMO C-tail at Ser741/Thr742, which happens after Protein kinase A (PKA) and Casein kinase I (CKI/CK1) phosphorylation at adjacent Ser residues (Figure 1.13b-c). In other words, HH triggers the phosphorylation of SMO C-tail by PKA and CK1, causing the C-tail to unfold. This unfolded C-tail then interacts with Gprk2, which phosphorylates it further. This prevents the C-tail from refolding and maintains the active form of Drosophila SMO. As a result, Gprk2 forms a dimer or oligomer that activates SMO by connecting their C-tails.

A study by Shi et al. [143] found that when there is no HH, Drosophila SMO phosphorylation is inhibited by the SMO-Cos2/Fu inactive complex (Figure 1.13d). When HH activates Drosophila SMO, it leads to the phosphorylation of SMO, Cos2, and Fu. As a result, SMO accumulates on the cell surface, accompanied by a conformational change in its C-tail. This change leads to the activation of signaling complexes involving SMO, Cos2, and Fu (Figure 1.13d). They hypothesized that the activation of signaling complexes by HH depends on the concentration of the HH signal.

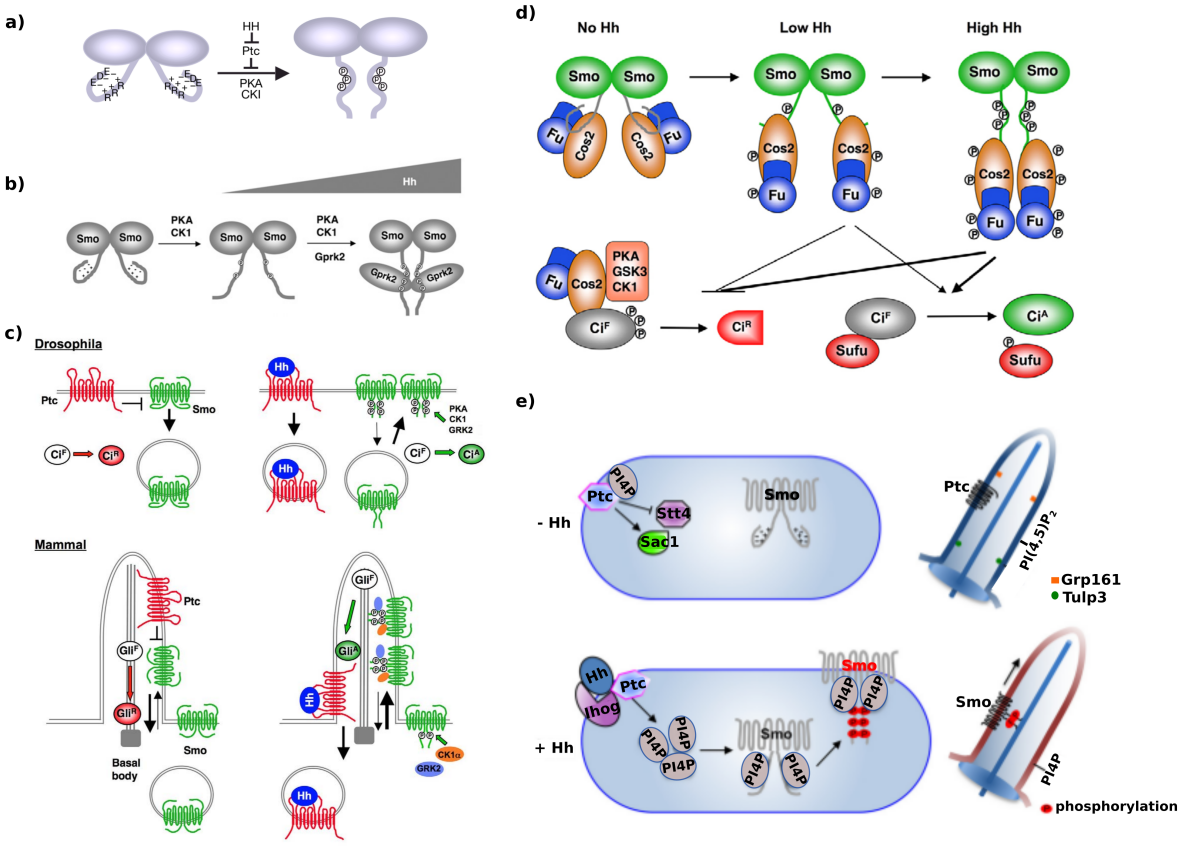


Figure 1.13: Different models of SMO dimerization in Drosophila (Source: [29, 77, 143, 190].

Another study [142] showed that lipid rafts of the cell plasma membrane are required for both Drosophila SMO oligomerization/higher order clustering and Drosophila SMO activation.From their experiments on, different construct of SMO (SMOΔN, SMOΔC, SMOΔNΔC, and SMO-CT), they suggest that N terminal and TM helices are important for oligomer (tetramer)/higher order cluster formation in Drosophila SMO. In SMOΔN, the residues 32–255 from N-terminal were truncated. In SMOΔC the residue 556-till end from C-terminal were truncated. In SMOΔNΔC has only residues 256–555 and both N- and C-terminal residues were removed. SMO-CT has only residue from C-terminal i.e, 556- till end. Their study also reports that

Drosophila SMO C-tail also forms a dimer.

Maier et al. [98] suggest that Drosophila SMO has SMO autoinhibitory domain (SAID), which contains four clusters for phosphorylation. Phosphorylation of three clusters at specific sites by PKA and CK1 is crucial and sufficient for the activation of SMO. Further, the phosphorylation initiates the dimerization of Drosophila SMO C-tails, which leads to high-level signaling activity. The phosphorylation of SAID has two important effects. Firstly, it prevents the ubiquitination and subsequent internalization and degradation of Drosophila SMO, allowing it to accumulate on the cell surface. Secondly, it facilitates the dimerization of Drosophila SMO, as shown by bioluminescence resonance energy transfer (BRET) experiments, resulting in a shift towards a more active conformation. These findings support previous research indicating that complete activation of Drosophila SMO requires phosphorylation by Gprk2. Gprk2 has been found to phosphorylate Drosophila SMO constitutively once it accumulates on the cell surface.

Later, Jiang et al. [77] studied the role of phosphatidylinositol 4-phosphate (PI4P) in SMO dimerization (Figure 1.13e). They also found that C-tail is important for receptor dimerization.

Based on the findings, it has been suggested that dimerization of the Drosophila SMO C-terminal region significantly increases signaling activity [30, 98].

1.6.3 SMO dimerization in vertebrates or mammals

In vertebrates, SMO has been shown to require translocation to primary cilia for its function, and the membrane lipids (membrane cholesterol and PI4P) play an important role in modulating the organization, dynamics, and function of the protein.

The study on mammalian SMO indicates that mammalian SMO may also exist as a constitutive dimer [30, 190]. In addition to this, a crystal structure of human SMO (PDB: 4JKV) available in the PDB database revealed the dimer form of SMO [169]. It should be noted that it is not clear whether the dimer observed in the crystal structure accurately represents the dimeric state of SMO in its natural cellular environment (in vivo), specifically the cell membrane [169, 190]. In this construct, SMO is observed to crystallize as a parallel homodimer (Figure 1.14) with the interface involving TM4 and TM5.

Duarte et al. [47] compiled a comprehensive dataset of validated transmembrane (TM) protein interfaces in order to study their features. They used the Evolutionary Protein-Protein Interface Classifier (EPPIC) (www.eppic-web.org), which they developed for evolutionary analysis. EPPIC effectively differentiates between biological interfaces and lattice contacts in crystal structures. EPPIC has 90% accuracy on soluble proteins, and they used it for their

TM protein dataset. They have suggested that human SMO and Drosophila SMO have 43% sequence identity, indicating a significant similarity in their protein sequences. Hence, the dimer interface in the PDB ID 4JKV (human SMO) is a biological interface [47].

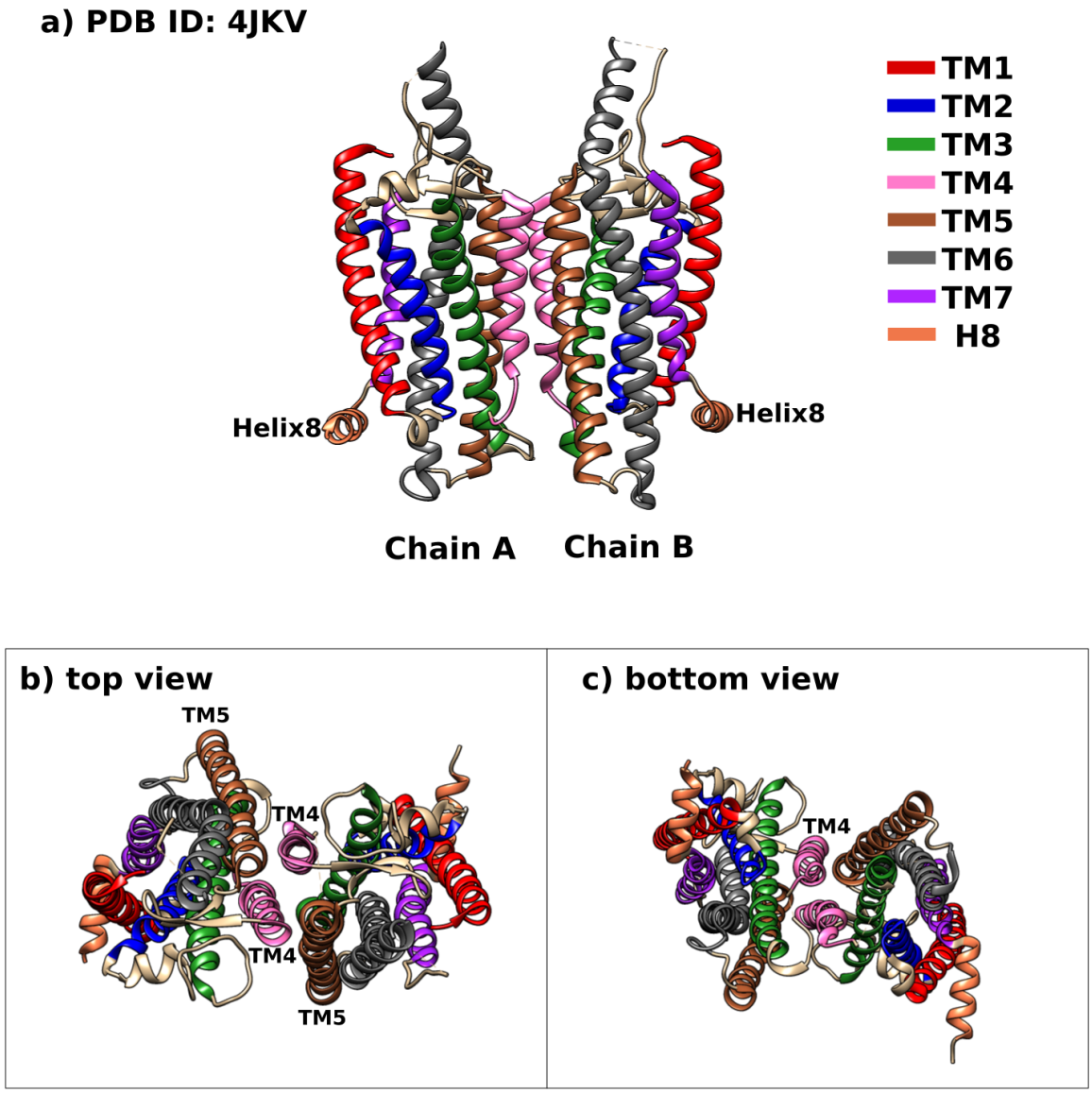


Figure 1.14: Crystal structure of SMO (PDB ID: 4JKV). a) SMO dimer interface showing the interaction between TM4-TM4 of chain A and chain B. b) and c) show the top and bottom views. The helices were displayed in ribbon representation with red, blue, green, pink, brown, gray, purple, and orange colors, respectively.

In summary, the dimerization of SMO in vertebrates is still unclear, while It is reasonably well understood in Drosophila SMO. Multiple studies have been done on Drosophila SMO dimerization, and there are many unanswered questions about vertebrate SMO dimerization, such as: How does vertebrate SMO dimerize? Do they also form oligomers? If yes, what is the size of the oligomers? What is the functional role of vertebrate SMO dimerization? Does dimerization/oligomerization have a relationship with membrane-lipid interaction? What are the controlling factors? Can we relate this to receptor druggability? Answers to these questions will help researchers understand the vertebrate SMO dimerization mechanism/process in great detail.

1.7 Significance of cholesterol in the HH pathway and related questions

All three proteins, HH, PTCH1, and SMO, contain cholesterol-binding domains/sites. The HH ligand has a cholesterol attachment at the C-terminal of SHH-N. The HH receptor PTCH1 also contains SSD, which binds to cholesterol. Further, the main focus of this thesis, SMO, also has multiple cholesterol-binding motifs/sites.

Despite multiple investigations, a thorough understanding of SMO activation by HH-bound PTCH1 remains elusive to the scientific community. However, there is enough evidence that these two TM proteins do not communicate directly [64, 86, 155]. In vertebrates, it has been observed that SMO accumulates in the primary cilia when the HH pathway is activated [42, 49, 52, 111, 125, 166]. In the absence of the HH ligand, PTCH1 is localized in the primary cilia. HH binding leads to the deactivation of PTCH1 molecules, which are then delocalized from primary cilia and are subsequently degraded [35, 135, 175, 186]. This possibly sets the stage for SMO to localize in the primary cilia, and initiate pathway activation. In conjunction with reports that small molecules, including cholesterol, have a regulatory effect on SMO [24, 67, 68, 97, 183], and given that bound cholesterol molecules have been reported in the 3D structures of SMO; a hypothesis involving cholesterol mediated translocation of SMO to the primary cilium, triggered by the binding of HH to PTCH1, has gained some currency [12, 35, 83, 128, 177].

1.8 Overview of the thesis

Chapter 1 introduces the HH pathway and significant details related to key players and briefly describes the significance of the problems addressed in this dissertation. This chapter also provides an overview of the dimerization mechanism of SMO in Drosophila and investigates the possibility of dimerization in humans, an area that is both emerging and of great significance. Chapter 2 gives an overview of the computational approaches used to carry out the investigations, methods/tools for structure generation, the principles of molecular dynamics simulations, details of atomistic and coarse-grained (CG) molecular dynamics simulations, and an explanation of potential energy functions. This also reports the CG parameterization of PI4P lipids, which is important for this study. Chapter 3 explores the sequence of SMO to get the maximum information from it. This includes the identification of known/new conserved motifs in helices, cholesterol, PI4P, and other lipid-binding motifs/sites. It is shown that the SMO has several cholesterol-binding motifs in its CRD, LD, and TMD, including the ICD. This chapter also discusses the re-defining of strict CCM. Chapter 4 describes the computational modeling of SMO, the dynamics of SMO in the primary cilium, and its interaction with different ciliary lipids (cholesterol and PI4P), which are in agreement with previous experimental reports. This also shows that SMO interacts with the strict CCM proposed in Chapter 3, and PI4P interacts with the Arg/Lys motifs, in agreement with previous experimental reports. Chapter 5 elucidates my attempt to study the structural changes in SMO in the various states; apo form (no activity), cholesterol in SMO-TMD (basal activity), cholesterol in SMO-CRD (medium activity), and cholesterol in SMO TMD+CRD (high activity). Chapter 6 summarizes the findings from this thesis and the possible outcomes of the performed work. This also discusses the impact of the findings reported here. This also discusses future directions for the work on this important pathway.

Chapter *2*

Methods

2.1 Computational methods

This chapter discusses the methods and algorithms used for the studies in the current thesis. Several computational approaches have been used to investigate the objectives discussed in Chapter 1. Following the paradigm of sequence -> structure -> dynamics and interaction -> function, first, the analysis of the sequence of SMO has been done to understand the inherent information coded therein, such as domains, motifs, etc. Section 2.2 discusses the approaches to decoding the maximum amount of information from the SMO sequence. The thesis includes some bioinformatic analysis to predict the motifs. After that, the complete structure of SMO was generated (discussed in Chapter 4) using an online server, a molecular modeling technique (discussed in the next section), and the workflow and algorithm behind the server are discussed in Section 2.3. The details will help to understand the selection of the best model based on the parameters discussed in Chapter 4. Lastly, long-time-scale molecular dynamics (MD) simulations have been carried out to understand the detailed mechanism of signal transduction based on SMO dynamics and its interaction with small molecules. Therefore, a brief introduction to the atomistic MD simulation and the coarse-grained (CG) MD simulation is provided in Section 2.4. A complex lipid membrane composed of five lipids (outer leaflet- POPC, POPE, and cholesterol; inner leaflet- POPC, POPE, POPS, PI4P, and cholesterol) has been used to mimic the primary cilia environment for SMO. The model ciliary membrane was prepared using the CHARMM-GUI web server, but the CG parameters of PI4P are not available on the CHARMM-GUI web server. Therefore, this chapter also contains the "CG parameterization of PI4P lipids" in Section 2.5, which was performed by me. This step is important as PI4P is an essential component for SMO systems.

2.2 Multiple sequence alignment

Multiple sequence alignment (MSA) is used to analyze and identify conserved motifs, domains, and functional elements within sequences. A consensus sequence was prepared that represents the most common or conserved amino acids at each position within a set of aligned sequences. The current work has used the GPCR database (GPCRdb; link- https://gpcrdb.org/) to gain information [87]. The SMO sequence was aligned with a few class A GPCRs to identify the conserved motifs in the TM helices, CRD, and ICD of the GPCR using the sequence alignment option provided in GPCRdb (Chapter 3). Further, MSA of SMO with all class F (frizzled) receptors has been performed to verify/identify the classical or alternative conserved motifs of class F. Those motifs for which a consensus sequence is known were searched in the SMO sequence by using the regular expression pattern search.

2.3 Molecular modeling

GPCR I-TASSER is an extension of the iterative threading assembly method (I-TASSER) server dedicated to predicting the 3D structure of GPCRs [187]. GPCR I-TASSER is a hybrid method that utilizes distant homology templates and low-resolution experimental data to generate a high-resolution structure of GPCR models. It uses multiple machine learning (ML) classifiers to predict the inter-TM contacts with an average accuracy of 62% in the top L/TM-5 predictions. It also uses the information from several mutagenesis experiments performed on GPCRs to identify the important residues and motifs and improves the quality of the model.

The GPCR-I-TASSER pipeline consists of three steps (Figure 2.1): 1) template identification (or ab initio construction of TM-helix) and experimental restraint collection, 2) Monte Carlo fragment assembly simulation, and 3) atomic-level structural refinement using fragment-guided MD (FD-MD). In the first step, the Local Meta-Threading Server (LOMETS) matches the GPCR sequence with the PDB library in order to identify potential templates and determine the super secondary structure. If there are no closely related templates available, then the 7TM bundle is created from scratch using an *ab initio* folding program. The resulting *ab initio* 7TM bundle, along with the LOMETS alignments and sparse restraints obtained from mutagenesis data in GPCR-RD, are used as input for structure assembly simulations in the subsequent step. It generated five models by using ten templates. The selected templates serve as a basis for constructing the 3D model. Templates are selected based on various parameters such as rank, Ident1, Ident2, coverage, normalized Z-score, and confidence score (C-score). The rank parameter in GPCR-I-TASSER indicates the position of the templates among the top ten used for threading. Ident1 represented the percentage sequence identity between the templates and the query sequence in the aligned region. Ident2 indicated the percentage sequence identity

between the whole template chain and the query sequence. The coverage parameter reflected the extent of the threading alignment, calculated by dividing the number of aligned residues by the length of the query protein. The normalized Z-score provided a measure of the quality of the threading alignments. A higher normalized Z-score indicated a better alignment. The C-score indicates the quality of the predicted models by taking into account the significance of the template alignments and the structure assembly simulation convergence parameters. Typically ranging from -5 to 2, a higher C-score indicates a model with high confidence and vice-versa. Overall, GPCR I-TASSER combines template-based modeling, loop modeling, and refinement techniques to predict the 3D structure of GPCRs. It makes use of available experimental data and computational tools to generate accurate models.

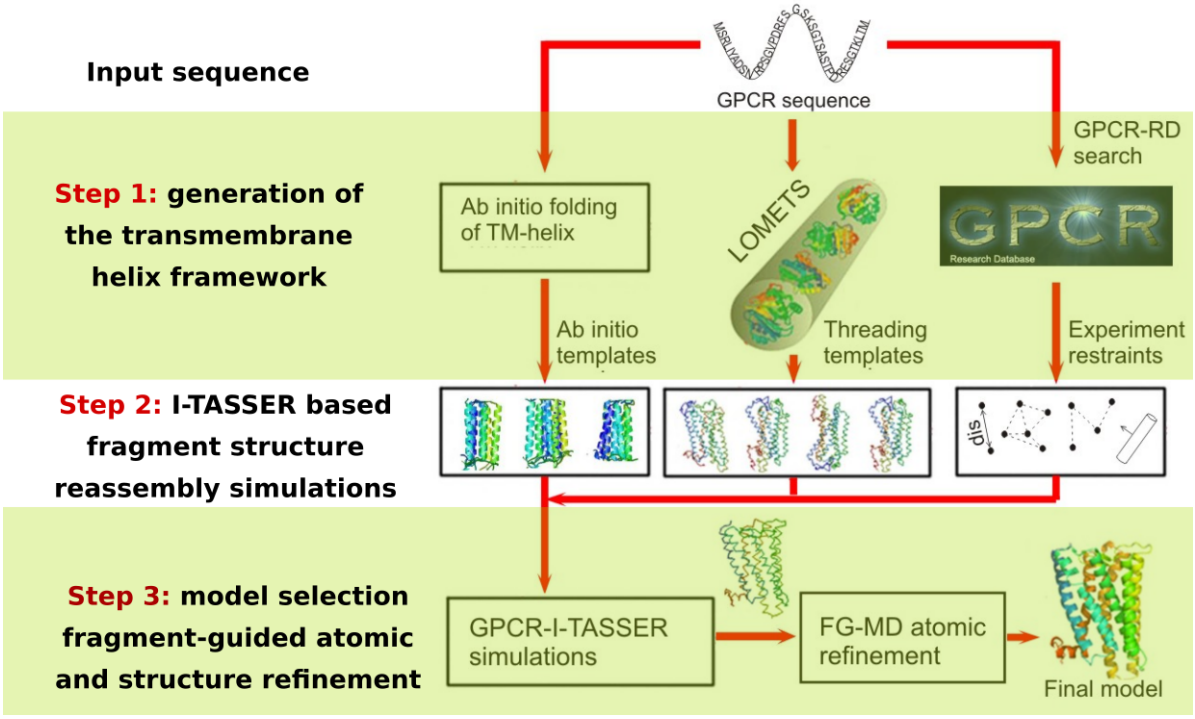


Figure 2.1: Workflow of the GPCR-I-TASSER for modeling GPCR structure (adapted from [187])

2.4 Molecular dynamics (MD) simulation: a computational approach to study the dynamics of biomolecules

Molecular dynamics (MD) simulation is the computational study of the motion of atoms within a molecular system using molecular mechanics by solving Newton's equations of motion (F=ma) numerically and using a force field to describe the interactions between the atoms or molecules. This allows the simulation to predict how the system will behave over time, such as

how the molecules will move and interact with each other. It is a powerful technique that computational biologists have that allows them to study the complex dynamics of macromolecules. It serves as a link between theoretical predictions and experimental observations, allowing for deeper insights into macroscopic phenomena that can be observed experimentally by examining the microscopic details of the systems of interest at the atomic level. Rapid growth in computational power and progress in algorithms continue to push the boundaries of MD simulation to study complex systems containing millions of atoms (i.e., larger length scales) and microscopic processes with longer timescales. Any biological activity is the result of time-dependent interactions between macromolecules or between macromolecules and micromolecules such as protein-protein, protein-nucleic acid, protein-ligand, etc. (Figure 2.2) [184].

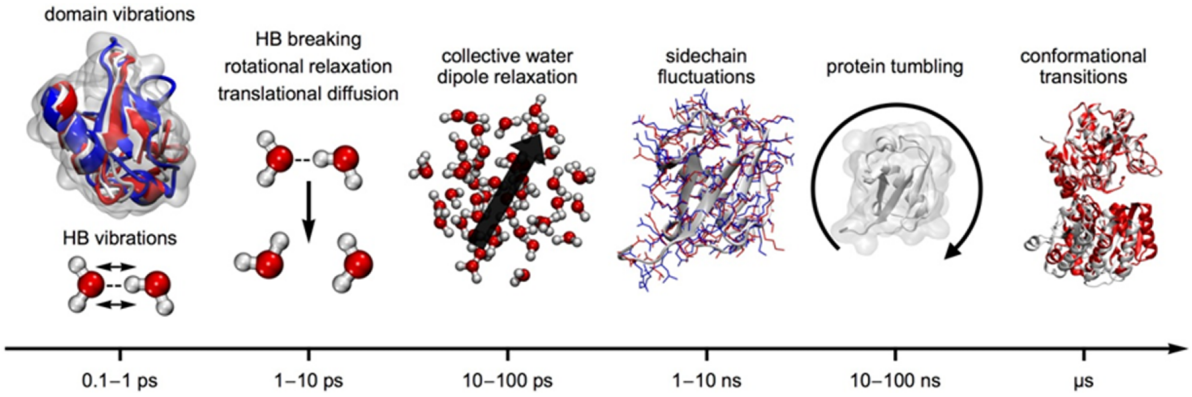


Figure 2.2: The hierarchy of time scales shows the varying rates at which motions occur in proteins. X-axis shows the timescale of the event (source: [184]).

Proteins exhibit a wide range of dynamic motions occurring at different timescales. Intramolecular vibrations (within the molecule) of secondary structure elements and domains of the protein, as well as intermolecular vibrations (between molecules) of the water hydrogen bond network, typically occur within the timescale of 0.1 to 1.0 picoseconds (ps). At slightly longer timescales, relaxation processes within the water hydrogen bond network, involving rearrangements of hydrogen bonds, rotation of single molecule, and translational diffusion, take place within the 1-10 ps range. On the other hand, collective dipole relaxation takes place in water between 10 and 100 ps. Moving to longer timescales, protein side chain fluctuations can be seen in the range of 1 to 10 nanoseconds (ns), emphasizing the dynamic nature of these functional groups. Protein rotational tumbling motions, which involve the overall rotation of the protein molecule, typically take place at a timescale of 10-100 ns. Furthermore, conformational transitions in proteins, such as hinge motions observed in glycogen synthase, can be studied at the microsecond (μs) timescale [184].

2.4.1 Atomistic simulations

Atomistic simulations provide insight into the behavior of biological systems at the atomic level. In MD simulations, the interactions between atoms are represented by empirical potential functions. These functions are used to solve the classical equations of motion. This results in the generation of a time-dependent trajectory of the system. From this trajectory, various structural and dynamic properties, including kinetic and thermodynamic properties, can be calculated by using the principles of statistical mechanics.

Potential energy functions The behavior of the system is determined by interatomic interactions, which are described by a potential energy function. The potential energy function, also known as a force field, defines the energy associated with the positions of the atoms in the system. In classical MD simulations, the potential energy of the system is calculated from the atomic coordinates, with the potential energy being given by the sum of the different bonded (bond stretching, angle bending, and dihedral or torsional rotation) and non-bonded (electrostatic and Van der Waals interactions) interactions occurring in the system [1]:

$$U(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[1 + cos(n\phi + \delta)] + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left[\frac{q_i q_j}{\varepsilon r_{ij}} + \varepsilon_{ij}\left(\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6}\right)\right]$$

In the above equation,
k_b = the spring constant corresponding to the stretching of the bond,
b = the instantaneous bond length for a given bond,
b_0 = the equilibrium bond length,
k_θ = spring constant corresponding to the bending of the bond angle,
θ = the instantaneous bond angle for a given bond angle,
θ_0 = the equilibrium bond angle,
k_ϕ = the height of the potential energy corresponding to the dihedral,
n = the number of maxima/minima that occur as ψ varies from $-\pi$ to π,
ϕ = the dihedral angle for a given dihedral,
δ gives the phase shift for ϕ,
q_i and q_j = the charge on particles i and j, respectively,
ε = the permittivity in a vacuum,
r_{ij} = the distance between the particles i and j,
$R_{min,ij}$ = the distance at which the Van der Waals energy between atoms i and j are minimum
ε_{ij} = the energy at this distance.

Bonded interactions are important for maintaining the connectivity and flexibility of the molecular structure. They influence the conformational changes, molecular vibrations, and overall geometry of the system. Whereas the non-bonded interactions are important for maintaining the overall structure and stability of the system and play an important role in determining molecular packing, solvation effects, and intermolecular interactions.

MD simulations allow for the exploration of GPCR conformations, the analysis of ligand-receptor interactions, and the characterization of receptor dynamics under different conditions. Several atomistic simulations of GPCRs have been reported. These simulations have identified important interactions of several GPCRs with lipids and ligands, e.g., mapping of the cholesterol interaction site. But the limitation of a longer time scale is an important issue. These simulations are limited to simulation times less than 100 ns and system sizes less than 10 nm. By employing a CG model (discussed in the next section), longer simulations for larger systems can be achieved.

2.4.2 Coarse-grained (CG) molecular dynamics and Martini mapping

CG simulations are a simplified representation of a system where groups of atoms are treated as a single interaction site or particle. In CG models, multiple atoms are combined into a single interaction unit (called CG beads), reducing the overall number of particles in the simulation and reducing (in comparison with an all-atom description) the number of degrees of freedom, which requires less resources and runs faster than an all-atom representation.

The MARTINI force field is one of the most widely used CG force fields, developed by Marrink and coworkers at the University of Groningen in 2004. Initially, this was for lipids but was extended to other molecules (proteins, nucleic acids, carbohydrates, etc.). The details and terms of MARTINI are described below:

Interaction Sites: It is based on four-to-one mapping, i.e., on average, four heavy atoms are represented by a single bead or interaction center [20, 101]. Interaction sites are classified into four main types: polar (P), nonpolar (N), apolar (C), and charged (Q) (Figure 2.3). Each particle type can be further classified into subtypes based on their hydrogen-bonding capabilities or degree of polarity. The hydrogen-bonding subtypes are represented by letters: d for donor, a for acceptor, da for both donor and acceptor and 0 for none. The degree of polarity subtypes ranges from 1 to 5 representing from low (1) to high (5). These subtypes provide a more detailed understanding of the chemical properties and atomic structure of the particles.

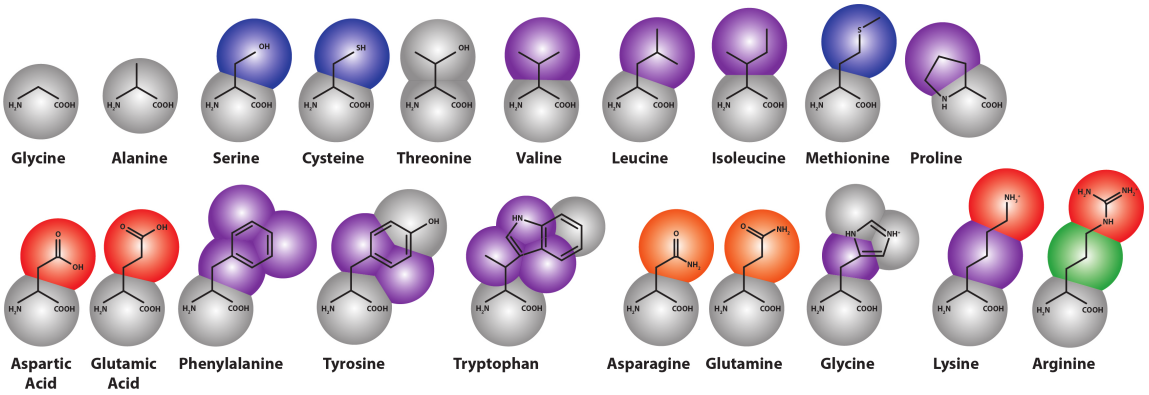


Figure 2.3: Representation of CG Martini amino acids. Martini beads are colored by type. Purple indicates apolar, blue and green represent intermediate, gray and orange denote polar, and red signify charged particles (Source: [20]).

The interactions between martini beads are represented as the addition of "bonded" and "nonbonded" terms.

$$V = V_{bonded} - V_{nonbonded} \tag{2.1}$$

The bonded term includes bonding, angle, dihedral, and improper dihedral potentials. And non-bonded includes Lennard-Jones and Coulombic potentials.

Bonded Interactions: A weak harmonic potential, V_{bond}(R), has been used to describe the bonds.

$$V_{bond}(R) = \frac{1}{2} k_{bond}(R - R_{bond})^2 \tag{2.2}$$

Here, R_{bond} is equilibrium distance.
σ is 0.47 nm.
K_{bond} is a force constant with 1250 $kJmol^{-1}nm^{-2}$.
The LJ interaction is not considered between particles that are bonded to each other. Bonded particles are, on average, slightly closer to each other than non-bonded particles (for which the equilibrium distance is $21/6\sigma$).

A weak harmonic potential, $V_{angle}(\theta)$, of the cosine type is employed for the angles to describe the chain stiffness

$$V_{angle}(\theta) = \frac{1}{2} k_{angle}(cos(\theta) - cos(\theta_0))^2 \tag{2.3}$$

LJ interactions between the second nearest neighbors are not ruled out. The force constant K_{angle} is 25 $kJmol^{-1}$ and the equilibrium bond angle θ_0 is 180° for aliphatic chains.

The force constant K_{angle} is 45 $kJmol^{-1}$ (the original value K_{angle} = 35 $kJmol^{-1}$) and the equilibrium angle θ_0 is 120° for the angles involving the cis double bond. The force constant K_{angle}, is 45 $kJmol^{-1}$ and the equilibrium angle θ_0 is 180° for trans-unsaturated bonds.

The proper dihedrals ($V_{dihedral}(\phi)$) are utilized to enforce the secondary structure of the peptide backbone.

$$V_{dihedral}(\phi) = k_{dihedral}[1 + cos(n\phi - \phi_d))] \tag{2.4}$$

The improper dihedral angle potential ($V_{imp\_dihedral}(\psi)$) is used to prevent out-of-plane distortions of planar groups.

$$V_{imp\_dihedral}(\psi) = k_{imp\_dihedral}(\psi - \psi_d)^2 \tag{2.5}$$

Nonbonded Interactions: The nonbonded interactions are described by a shifted Lennard-Jones (LJ) 12-6 potential energy function.

$$U_{LJ}(r) = 4\varepsilon_{ij}\left[\left(\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6}\right)\right] \tag{2.6}$$

Here, σ_{ij} represents the closest distance of approach between two particles.
ε_{ij} represents the strength of their interaction.
σ is 0.47 nm for each interaction pair (except for the two special classes of rings and antifreeze particles).

Charged groups (type Q) with a charge q, interact through a Coulombic energy function. The interactions incorporate explicit screening with a relative dielectric constant ε_{rel} set to 15.

$$V_{el} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_{rel}r_{ij}} \tag{2.7}$$

Ring Particles: To maintain the geometry of small ring compounds, 2 or 3 to 1 mapping has been employed, where multiple atoms in the ring have assigned to a single representation. A special particle set called "S" represents these ring structures. In this set, the interactions between ring particles are adjusted to be smaller and weaker compared to regular particles. Specifically, the size parameter (σ) of the LJ potential is reduced to 0.43 nm instead of 0.47 nm, and the energy parameter (ε) is scaled down to 75% of its original value. For example, a hydrophobic particle of type SC1 (representing a carbon atom in the ring; C1 of set S) belonging

to a cyclohexane ring interacts with other SC1 particles using the modified LJ potential with $\varepsilon = 0.75 \times 3.5\ kJmol^{-1}$ and $\sigma = 0.43$ nm. However, its interaction with normal C1 particles remains unaffected, with $\varepsilon = 3.5\ kJmol^{-1}$ and $\sigma = 0.47$ nm.

CG simulations are particularly useful for studying systems with long-range interactions or large assemblies of molecules. However, they lost some atomic-level detail, which may limit their accuracy in capturing fine-grained structural features and dynamics. CG simulations have been extensively employed to study the structural dynamics, receptor conformation, activation mechanism, allosteric modulation, effects of mutations, and ligand binding of GPCRs [63, 93, 140, 156].

Both atomistic and CG simulations have advantages and disadvantages, and the choice is determined by the specific research question, the system of interest, and the computational resources available. Atomistic simulations are best suited for studying detailed molecular interactions and processes, whereas CG simulations are better suited for exploring larger systems and longer time scales. The combination of these methods has also been used to develop a thorough understanding of intricate biological systems.

The current thesis uses both atomistic and CG simulations to answer the specific research questions related to SMO that have been asked in Chapters 4, 5, and 6. In Chapter 4, CG simulations of SMO in a modeled ciliary membrane have been performed to study the larger structural movement in SMO (larger conformational changes such as domain movement). Whereas in Chapter 5, atomistic simulations of SMO bound to cholesterol in the CRD and TMD in the modeled ciliary membrane have been performed. The objective here is to see the structural changes in SMO upon cholesterol binding at various positions.

2.5 Parameterization of a CG model of phosphatidylinositol 4-phosphate (PI4P) lipid

For the preparation of the CG system, SMO in the modeled ciliary membrane, CG lipids, CG water beads, and CG SMO were used. The composition of primary cilia is heterogeneous, and the current thesis uses a five lipid component (phosphatidylcholine; POPC, phosphatidylethanolamine; POPE, phosphatidylserine; POPS, PI4P, and cholesterol) model of the ciliary membrane (discussed in Chapter 1, Section Role of Primary Cilia). The system was prepared using the input generator martini maker option available in the CHARMM-GUI server [78, 127, 181]. PI4P, being a signature lipid for the composition of the tip of primary cilia

and an important signaling lipid in the HH pathway, needs to be introduced in the simulation system. The CG PIP2 and PI4P models (old Martini model) have been used by Sansom's group [4, 148] to study the binding mechanism of the PIP5K1A kinase to the membrane (Figure 2.4). However, the CG parameters for PI4P lipids are not available in the CHARMM-GUI server. Therefore, I have parameterized the PI4P lipid CG model.

Martini lipidome (`http://www.cgmartini.nl/index.php/force-field-parameters/lipids`) has several itp(s) of several lipids with phosphatidylinositol [PIP], phosphatidylinositol 3-phosphate (PI3P, named POP1 on server) and phosphatidylinositol 3-4 bi-phosphate (PIP2(3,4), named POP2 on server). In the CG representation, the POPC headgroup has two hydrophilic groups: the choline (type Q0) and the phosphate group (Qa). For POPE lipids, Qd is the positively charged group, representing the hydrogen-bonding capacities of the amine moiety. The glycerol ester moiety is represented by two intermediate hydrophilicity (Na) sites. Each lipid tail is represented by four hydrophobic particles (C1), each of which contains 16 methylene/methyl units. Standard bonded interactions have been used. In the phospholipid topology, the bond length of the glycerol backbone is 0.37 nm (mimics the underlying chemical structure better). Oleoyl tails consist of five particles, four of which are of the C1 type, and the central particle (C3) is slightly more polar to represent the polarizable characteristic of the double bond.

Wassenaar et al. described the CG beads of POPC [172] using the Martini topology with the bead order: NC3-PO4-GL1-GL2-C1A-D2A-C3A-C4A-C1B-C2B-C3B-C4B.

In my work, I have utilized the POPC, PIP, PI3P, and PIP2(3,4) Martini lipid parameters as a reference and extended them to generate the CG parameter for PI4P (Figure 2.5). Upon examining the topology files (itp) of the Martini/CG lipids from the martini lipidome, I utilized the existing CG lipid model as a reference to develop the parameterization for PIP2. Specifically, I focused on the 1'-phosphate, glycerol, and acyl tails of PIP2, following the definition provided in the POPC CG model. For the ring particles, C1, C2, and C3 were designated as P1, P1, and P4, respectively, aligning with the particle definitions in PIP, PI3P, and PIP2 (3,4) models. The 4'-phosphate particles were assigned as Qa type beads with a charge of -2, while the 1'-phosphate particles were assigned as Qa type beads with a charge of -1, consistent with the POPC CG model. Non-bonded and bonded interactions were defined in a similar manner to the POPC CG model and other CG models for PIP.

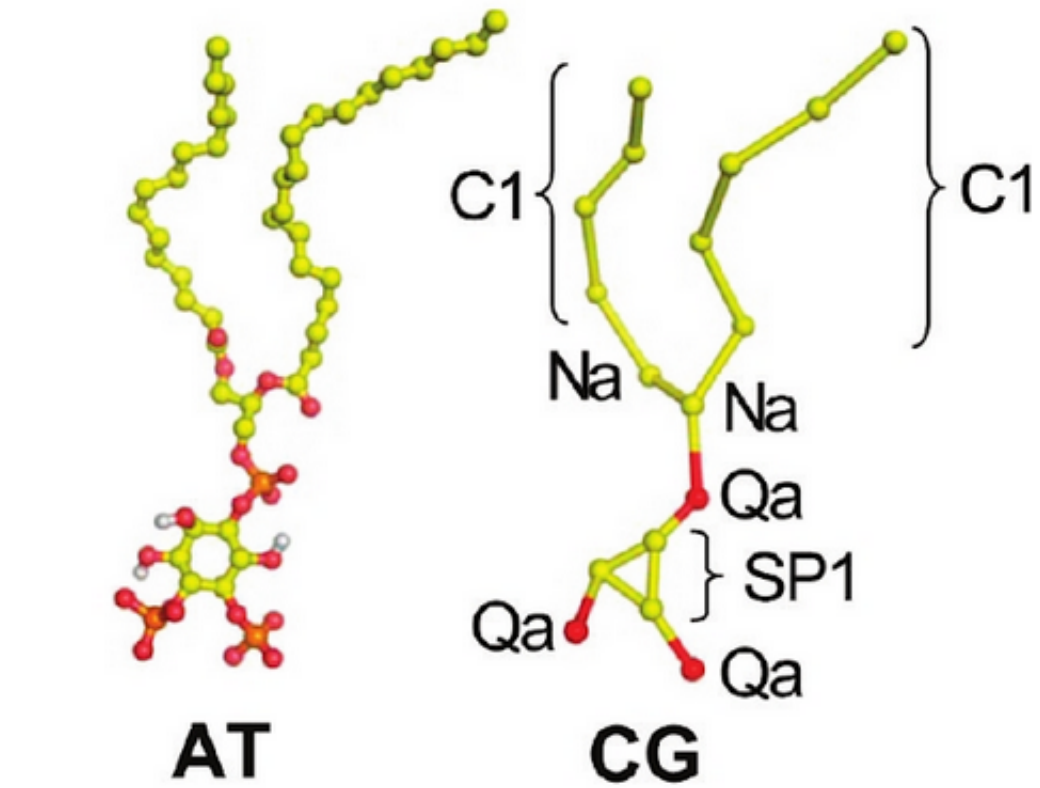


Figure 2.4: Comparison of atomistic (AT) and coarse-grained (CG) models of phosphatidylinositol 4,5-bisphosphate (PIP2). The MARTINI particle types used in the CG model are shown for reference (source: [148]).

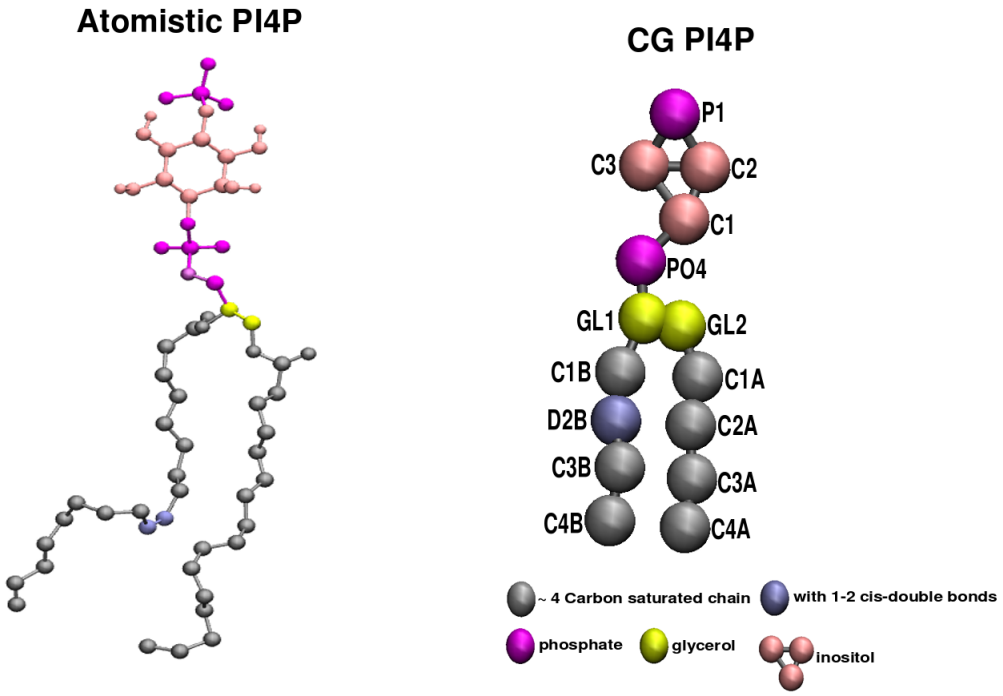


Figure 2.5: CG representation of PI4P based on POPC, PIP, PI3P, and PIP2(3,4) CG models.

Chapter *3*

Investigating the sequence and structure of Smoothened (SMO): Highlighting Lipid Binding and Other Functionally Conserved Motifs of GPCRs

3.1 Introduction

Smoothened (SMO), which is the primary focus of this thesis, is a G-protein coupled receptor (GPCR). GPCRs constitute the largest and the most diverse group of membrane receptors involved in signal transduction in eukaryotes. There are several classes of GPCRs that show variations in length and other features. However, the basic architecture of all GPCRs reveal three major domains: the extracellular domain (ED), the transmembrane domain (TMD), and the intracellular cytoplasmic domain (ICD). They all show structural similarities. They exhibit homology with other proteins in the specific class of GPCRs. Although the percentage of similarity may vary, this sequence similarity extends to other classes as well. Additionally, they also share a common mechanism of function [106]. The typical pathway for signal transduction is from ED to ICD via TMD. Being TM receptors, GPCRs are known to interact with lipids through their lipid-binding motifs, which are essential for their function and signaling. Cholesterol, a lipid, has been shown to play an important role in the dynamics and function of GPCRs. Receptors that interact with cholesterol have a characteristic amino acid sequence that can bind to cholesterol molecules [138]. A few important cholesterol-binding motifs in GPCRs are the CRAC (Cholesterol Recognition/interaction Amino acid Consensus), CARC (inverted CRAC), CRAC-like motifs, and *strict* cholesterol consensus motif (CCM). Not only membrane cholesterol but also other membrane components, *viz.*, sphingolipids and phospholipids, have been reported as modulators of ligand binding in GPCRs. A few other lipid-binding motifs are sphingolipid-binding motifs (SBM) that interact with sphingolipids and Arg/Lys clusters that interact with charged lipids (phosphatidylinositol; PI and phosphatidylserine; PS). The details of these motifs are described later.

Smoothened (SMO), a class F GPCR (protein of interest), also contains a few of these motifs mentioned above. Class A GPCRs are the most studied GPCRs. It is believed that all GPCRs have evolved from a common ancestral receptor. Despite their sequential diversity, they share common structural patterns and similar mechanisms of activation. In addition, the availability of data such as structures, more experiments, and evolutionary conservation provide a useful reference point for comparing the structures of other GPCR classes. Notably, SMO and other class F receptors show very little (less than 10% sequence identity) sequence similarity with class A GPCRs [169], and lack most of the conserved class A motifs, including D[E]R3.50Y in TM3, CWXP6.50 in TM5, and NP7.50XXY in TM7. The numbers above are based on the GPCR Ballesteros-Weinstein (B & W) [11] numbering scheme, where the first number denotes the helix number, and the digits after the decimal represent the position of a specific residue with respect to the most conserved residue of that TM helix. The D[E]R3.50Y motif is located in TM3, and R is the most conserved residue in TM3 and is represented by the number 3.50. The first position of this motif is usually D, but it can also be E (in square brackets). The first two amino acids in the motif are C and W, and the third position can vary and is denoted by "X" to indicate any amino acid can occupy that position. Similarly, NP7.50XXY is located in TM7, with P as the most conserved residence. N, P, and Y are in the first, second, and fifth positions, respectively. The third and fourth positions can include any amino acid.

However, there are a few residues or sets of residues that are conserved in class F, which could possibly be an important alternative to classical motifs (involved in molecular switches and receptor activation) of other classes of GPCRs. SMO contains lipid-binding motifs/sites that interact with sterol/cholesterol, PI4P, POPS, etc.

3.1.1 Comparison of SMO sequence with other class F GPCRs

An earlier study showed that a conserved KTXXXW motif is present in helix 8 of class F GPCRs. Instead of this, in SMO, there is a "KATXXXW" motif (residues 539–545) with an extra residue "A" between K and T (Figure 3.8b) [169]. A cation- interaction, between R451 in TM6 and W535 in TM7, similar to the 3-7 lock switch [158], reported in some of the crystal and cryo-EM structures of SMO, has been hypothesized to be the ionic lock (6-7 lock instead of 3-7 lock) in SMO (Figure 3.1). It has been suggested that it breaks during its cholesterol-binding-induced transition to the active state [13, 68, 158]. Furthermore, two salt-bridge networks in SMO: DRE (or D-R-E) and WGM (or W-G-M), which are conserved in class F receptors (Figure 3.1) [13], have recently been identified. The salt-bridge interactions between D473 (TM6), R400 (TM5), and E518 (TM7) constitute a D-R-E network [13] situated at the extracellular end of TMD. While the WGM network, between W339 (TM3), G422 (TM5), and M449 (TM6), is present in the lower TMD region [13].

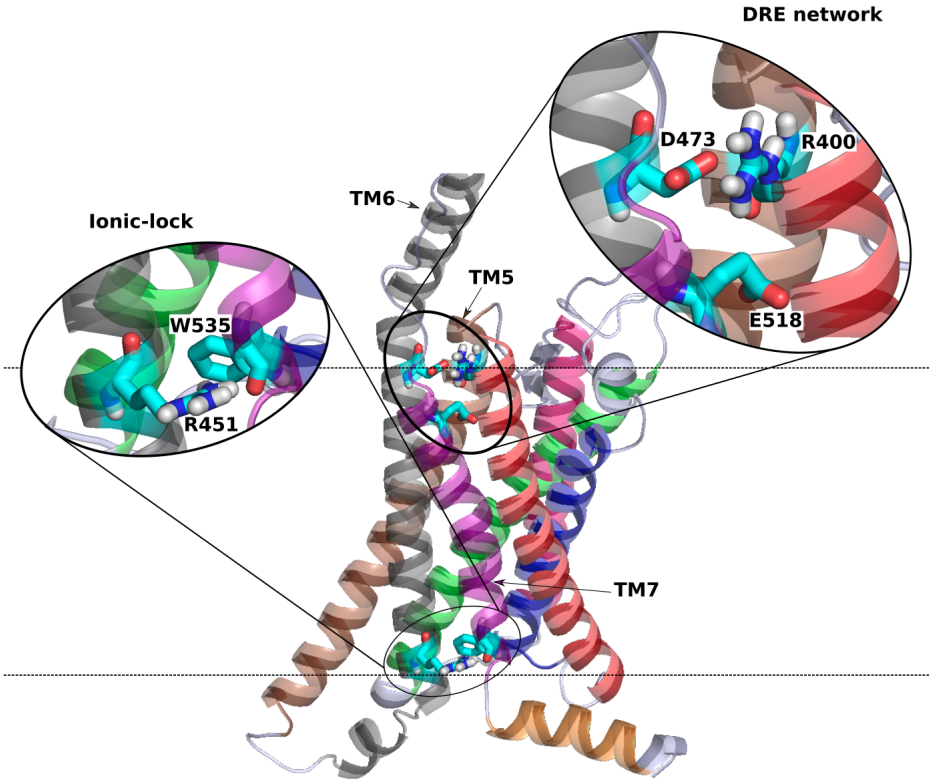


Figure 3.1: A schematic representation of SMO illustrating DRE, WGM, and ionic-lock interaction networks.

3.1.2 Cholesterol binding sites/motifs in GPCRs

A recent review on the binding of cholesterol to GPCRs summarized that proteins that interact with cholesterol molecules have characteristic amino acid sequence patterns, which are called cholesterol-binding motifs [138]. Several GPCRs have been found to possess these cholesterol-binding motifs. These include cytochrome P450scc, mouse apolipoprotein A-I, mouse caveolin 1, Streptomyces cholesterol oxidase, serotonin 1A receptor, β2-adrenergic receptor, and rhodopsin [138].

Depending upon their constitution, these are referred to as structural (*strict* CCM) or as sequence motifs, namely, cholesterol recognition amino acid consensus (CRAC), inverted CRAC

(CARC), and CRAC-like motifs [138].

1) CRAC (Cholesterol recognition/interaction amino acid consensus) motif-
The CRAC motif consists of a specific linear sequence of amino acids that are arranged in a particular order from the N-terminal to the C-terminal. This motif is characterized by a branched apolar residue, either LEU or VAL, followed by a segment of one to five amino acids of any type, then an aromatic TYR residue as central residue, another segment of one to five amino acids of any type, and finally a basic LYS or ARG residue.

The pattern of the motif can be represented as -(L/V)-$(X)_{1-5}$-Y-$(X)_{1-5}$-(R/K)-, where $(X)_{1-5}$ represents one to five residues of any amino acid.

2) CARC motif-
The CARC motif is very similar to the CRAC motif. Here the orientation of the specific linear sequence of amino acids is in the opposite direction than that of the CRAC motif. Hence, it is also known as "inverted CRAC". Unlike the CRAC motif, where the central residue is TYR, the CARC motif could have any aromatic residue (TYR, PHE, or TRP) as the central residue.

The pattern of the motif can be represented as -(R/K)-$(X)_{1-5}$-(W/Y/F)-$(X)_{1-5}$-(V/L)-, where $(X)_{1-5}$ represents one to five residues of any amino acid.

3) CRAC-like motif-
This is another important cholesterol interaction motif, similar to the CRAC motif, is CRAC-like motif. The linear sequence of this motif is similar to the CRAC motif except for the central residue. Here, the central aromatic residue must be the PHE.

The pattern of the motif can be represented as -(L/V)-$(X)_{1-5}$-F-$(X)_{1-5}$-(R/K)-, where $(X)_{1-5}$ represents one to five residues of any amino acid.

4) Cholesterol-binding structural motifs (*strict* CCM) in SMO-
CCM is a structural motif that was first reported in the crystal structure of β2-adrenergic receptor (β2AR) [58] and later identified in other GPCRs [58, 114, 140, 152]. Spanning both TM4 and TM2, they are separated in the sequence space; unlike in the sequence motifs, residues in this motif are in close spatial proximity. Based on the GPCR sequence homology, the *strict* CCM for GPCRs has been defined by four residues: (i) a positively charged residue (R, K) between positions 4.39–4.43, (ii) an aromatic residue (W, Y) at 4.50, (iii) a β-branched hydrophobic residue (I, L, V) at 4.46, and (iv) an additional aromatic residue (W, Y, F) at 2.41 [138]. The positions here are according to the Ballesteros-Weinstein (BW) numbering scheme

for GPCRs [11, 58]. In the BW numbering scheme, the first number denotes the helix identifier (1–7). The second number denotes the residue position relative to the most conserved position, which is assigned the number 50 [11].

3.1.3 Experimentally and computationally identified cholesterol binding sites/ motifs

Till date, 13 X-ray crystal structures and five cryo-electron microscopy structures of SMO have been reported (see Table 1.5 Chapter 1 for more details). Notably, most of these structures have a truncated CRD or truncated ICD. Analysis of these structures, and computational investigations, searching for sterol binding sites reveal essentially two distinct clusters of sites where cholesterol and/or other sterols can potentially interact or bind. One of these is in its extracellular CRD, and another is in the TMD [24, 43, 68, 84, 125]. Interestingly, a crystal structure of human SMO (hSMO) revealed cholesterol bound to a hydrophobic groove in the CRD (PDB ID: 5L7D; hereafter called the `CRD_5L7D` [24]. A more recent structure of hSMO reported by the same group (PDB ID: 7ZI0) reveals cholesterol in another CRD binding site; hereafter called the `CRD_7ZI0`) [84]. Two other crystal structures of African clawed frog and mouse SMO reveal the cholesterol-binding site in CRD (PDB IDs: 6D35 and 6O3C; hereafter called the `CRD_6D35` and `CRD_6O3C`) [43, 68]. Deshpande et al. also identified a sterol bound deep within the 7TM core of mouse SMO (PDB ID: 6O3C; here-after called the `TMD_6O3C` site) [43]. In the most recent electron microscopy structure of the hSMO-G_i complex with SAG (an SMO agonist), Qi et al. [125] reported four sterol-binding sites in SMO. Two of these four sites (sites 1-4) are situated in the TMD, one in the CRD, and one in the extracellular extension of TM6 to connect other sites in the TMD and CRD (Figure 3.2). Site 1 is situated in the central part of TMD and involves residues from TM2, TM3, TM5, TM6, and TM7. In contrast, site 2 is an extension of site 1 towards the upper TMD region and involves additional residues from LD and ECL2. Site 3 is further away from the membrane than any other binding site in the TMD and involves residues of TM6 and LD. The fourth sterol-binding site (Site 4) is defined by the hydrophobic pocket created by the residues in the CRD. The details of residues involved in the above-mentioned sterol-binding sites are summarized in Table 3.2 [60], have also observed a distinct cholesterol interaction site that constitutes extracellular portions of TM2, TM3 helices, and ECL1 and is hence referred to as the TM2/3e site (Figure 3.2) [60]. Another cholesterol-binding site near the cytoplasmic region (referred to as a cytoplasmic binding pocket (CBP)) has been identified by a molecular docking study [129]. This site includes residues from the cytoplasm-facing portion of TM1, TM2, TM6, and Helix 8.

Table 3.1: List of cholesterol-bound SMO structures available in the PDB.

S. no.	PDB ID	Species	Method	State	Ligand	Position	Reference
1.	7ZI0	Human	X-ray	Inactive	Cholesterol	CRD	[84]
2.	6XBK	Human	cryo-EM	Active	Cholesterol	TMD	[125]
3.	6XBJ	Human	cryo-EM	Active	Cholesterol	TMD	[125]
4.	6XBM	Human	cryo-EM	Active	24,25-Epoxy-cholesterol	TMD and LD	[125]
5.	6XBL	Human	cryo-EM	Active	cholesterol	TMD	[125]
6.	6O3C	Mouse	X-ray	Active	SAG21k, cholesterol	both CRD and TMD	[43]
7.	6OT0	Human	cryo-EM	Active	24,25-Epoxy-cholesterol	TMD	[126]
8.	6D35	African clawed frog	X-ray	Active	Cholesterol	CRD	[68]
9.	5L7D	Human	X-ray	Inactive	Cholesterol	CRD	[24]

Table 3.2: Cholesterol-binding sites identified by experimental and computational methods.

S. no.	Binding sites	Residues	Position	Method	Reference
1.	CRD_7ZI0	95, 105, 108, 109, 111, 112, 114, 156, 157, 161, 210, 596, 600	CRD region (CRD, LD, ECL3)	X-ray	[84]
2.	Sterol-binding site1 (Site1) (PDBs: 6XBJ, 6XBL)	281, 325, 329, 332, 394, 404, 408, 459, 470, 521, 525	Central TM region	Cryo-EM	[125]
3.	Sterol-binding site2 (Site2) (PDB:6XBK)	219, 281, 325, 329, 332, 386, 404, 408, 459, 463, 470, 521, 525	Central-upper TM region	Cryo-EM	[125]
4.	Sterol-binding site3 (Site3) (PDB: 6XBM)	196, 215, 220, 484, 485, 488	Extracellular extension of TM6 and the LD	Cryo-EM	[125]

Continued on next page

Table 3.2 – continued from previous page

S. no.	Binding sites	Residues	Position	Method	Reference
5.	Sterol-binding site4 (Site4)	95, 108, 109, 111, 112, 130, 156, 157	CRD	Cryo-EM	[125]
6.	`CRD_5L7D`	95, 108, 109, 112, 130, 156, 157, 164, 210, 491, 492, 496	CRD, LD, TM6 (Cholesterol binding groove in CRD)	X-ray (5L7D)	[24]
7.	Extracellular portions of TM2 and TM3 helices with ECL1 (TM2/3e)	276, 279, 283, 286, 292, 312, 313, 316, 317, 320	Upper TM region (inside TMD)	MD simulation	[60]
8.	`CRD_6O3C`	99, 109, 112, 113, 115, 116, 160, 164, 165	CRD	X-ray (6O3C)	[43]
9.	`TMD_6O3C`	329, 395, 398, 408, 412, 415, 467, 470, 474, 525, 526, 528, 529, 532	Inside TMD	X-ray (6O3C)	[43]
10.	`CRD_6D35`	68, 81, 82, 129, 133, 134	CRD	X-ray (6D35)	[68]
11.	Cytoplasmic binding pocket (CBP)	251-266, 339-346, 446-454, 535-553	Lower TM region	Molecular docking	[129]

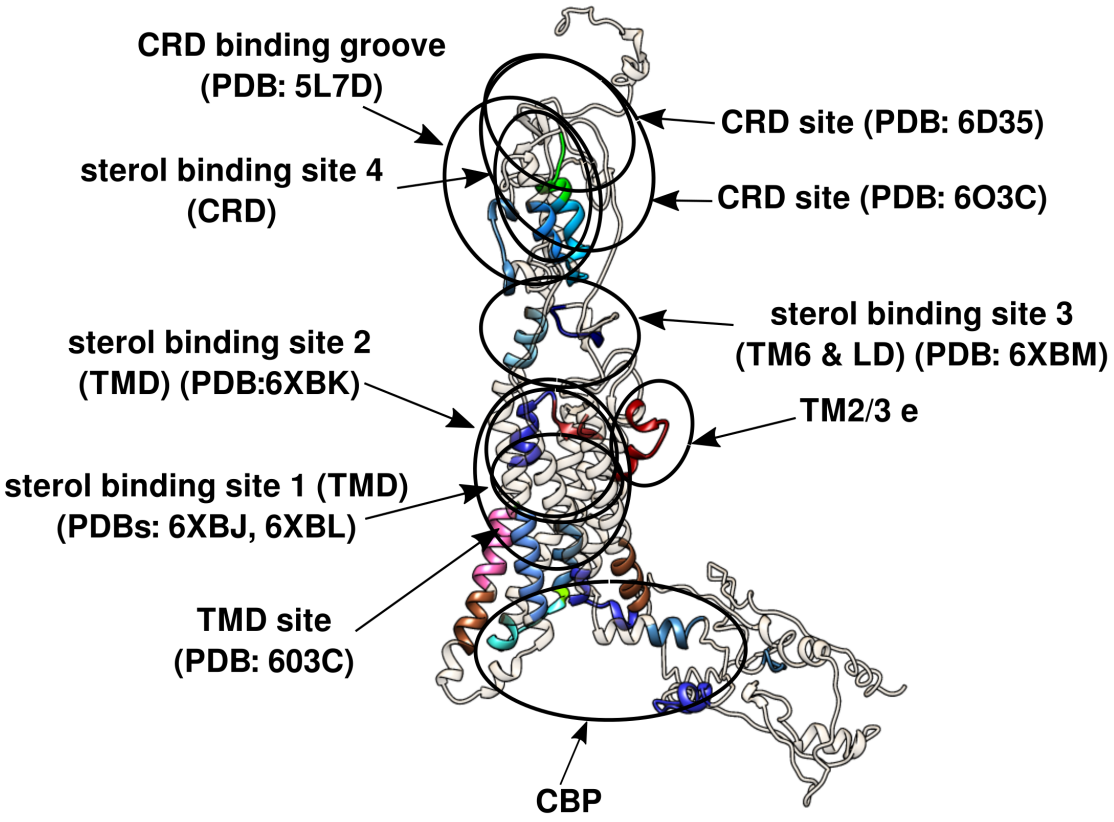


Figure 3.2: A schematic representation of SMO displaying experimentally and computationally identified cholesterol-binding sites (for more information, see Tables 3.1 and 3.2).

3.2 Method details

Most motifs have been discovered using multiple-sequence alignments (MSA) that give a consensus sequence. This helps in defining the pattern of a particular motif. The amino acid sequence of SMO has been taken from the NCBI database. The specific pattern of the lipid-binding motif has been well-defined in previous studies (discussed in the result section). A Python script using regular expressions has been written to locate and predict these lipid-binding motifs/sites. This script uses the well-known motif pattern to detect the presence of lipid-binding motifs in SMO. Subsequently, the conserved motifs involved in receptor activation alternatives to class A (or other GPCRs) have been identified by doing an MSA of SMO with a few representatives of other classes of GPCRs and an MSA of SMO with all known frizzled receptors of class F GPCRs.

3.3 Result and Discussion

This thesis focuses on the interaction of SMO with cholesterol and PI4P. Therefore, only the cholesterol-binding sites/motifs, the PI4P binding motif, and the sphingolipid-binding motif (SBM) are discussed.

3.3.1 Cholesterol-binding sites/motifs in SMO

3.3.1.1 Cholesterol-binding sequence motifs: CRAC, CARC and CRAC-like motifs in SMO

The primary sequence of SMO is characterized by 4 CRAC, 13 CARC, and 4 CRAC-like motifs distributed in its CRD, LD, TMD, and ICD regions. These motifs are depicted in Figure 3.3 in a 3D schematic representation of SMO and in Figure 3.4 in a snake plot diagram. The details are listed in Tables 3.3, 3.4, and 3.5. Among 21 cholesterol-binding motifs, SMO has some overlapping sequences between (i) CARC6 and CRAC2 and (ii) CRAC-like3, CRAC3, and CARC8 motifs. CRD contains four motifs (1 CRAC and 3 CARC motifs) within the residue range 105-167. One CARC motif is present in residues 204-210 in the LD. A total of 13 motifs (3 CRAC, 6 CARC, and 4 CRAC-like) are present in the TM helices. Most of these motifs (8 motifs: 2 CRAC, 4 CARC, and 2 CRAC-like) are located in the lower regions of the TM helices (TM3, TM4, TM5, TM6) and ICL1 and ICL2. Five motifs (1 CRAC, 2 CRAC-like, and 2 CARC motifs) are present in the upper regions of the TM helices (TM2, TM5, and TM6) and ECL1 and ECL2. Helix 8 contains one CARC motif, and ICD has two CARC motifs.

Table 3.3: Details of cholesterol-binding CRAC sequence motifs in SMO, including residue positions, sequence, length, and presence in SMO region.

CRAC motif				
S. no	**Residues**	**Sequence**	**Length**	**Domain**
1.	125-133	**L**LCA**V**YMP**K**	8	CRD
2.	346-356	**L**GTT**Y**QPLSG**K**	11	ICL2
3.	392-400	**V**GYKN**Y**RY**R**	9	ECL2 and TM5
4.	411-421	**V**LIVGG**Y**FLI**R**	11	TM5

Table 3.4: Details of cholesterol-binding CARC sequence motifs in SMO, including residue positions, sequence, length, and presence in SMO region.

CARC motif				
S. no	**Residues**	**Sequence**	**Length**	**Domain**
1.	105-112	**K**LVL**W**SG**L**	8	CRD
2.	114-126	**R**NAPRC**W**AVIQP**L**	12	CRD
3.	159-167	**R**ERG**W**PDF**L**	9	CRD
4.	204-210	**K**SW**Y**ED**V**	7	LD
5.	257-267	**R**NSNR**Y**PAVI**L**	11	ICL1 and TM2
6.	344-353	**K**ALGTT**Y**QP**L**	10	TM3 and ICL2
7.	356-363	**K**TS**Y**FHL**L**	8	TM4
8.	395-405	**K**NYR**Y**RAGFV**L**	11	ECL2 and TM5
9.	451-463	**R**LGIFG**F**LAFGF**V**	13	TM6
10.	482-489	**R**SFRD**Y**V**L**	8	TM6
11.	546-552	**R**RTWCR**L**	7	Helix 8 and ICD
12.	567-579	**K**MIAKA**F**SKRHE**L**	13	ICD
13.	726-733	**R**QGA**W**TL**V**	8	ICD

Table 3.5: Details of cholesterol-binding CRAC-like sequence motifs in SMO, including residue positions, sequence, length, and presence in SMO region.

CRAC-like motif				
S. no	**Residues**	**Sequence**	**Length**	**Domain**
1.	246-257	**L**FTLAT**F**VADW**R**	12	TM1 and ICL1
2.	282-291	**L**AQ**F**MDGAR**R**	10	ECL1
3.	386-395	**V**SGIC**F**VGY**K**	10	ECL2
4.	423-430	**V**MTL**F**SI**K**	8	TM5

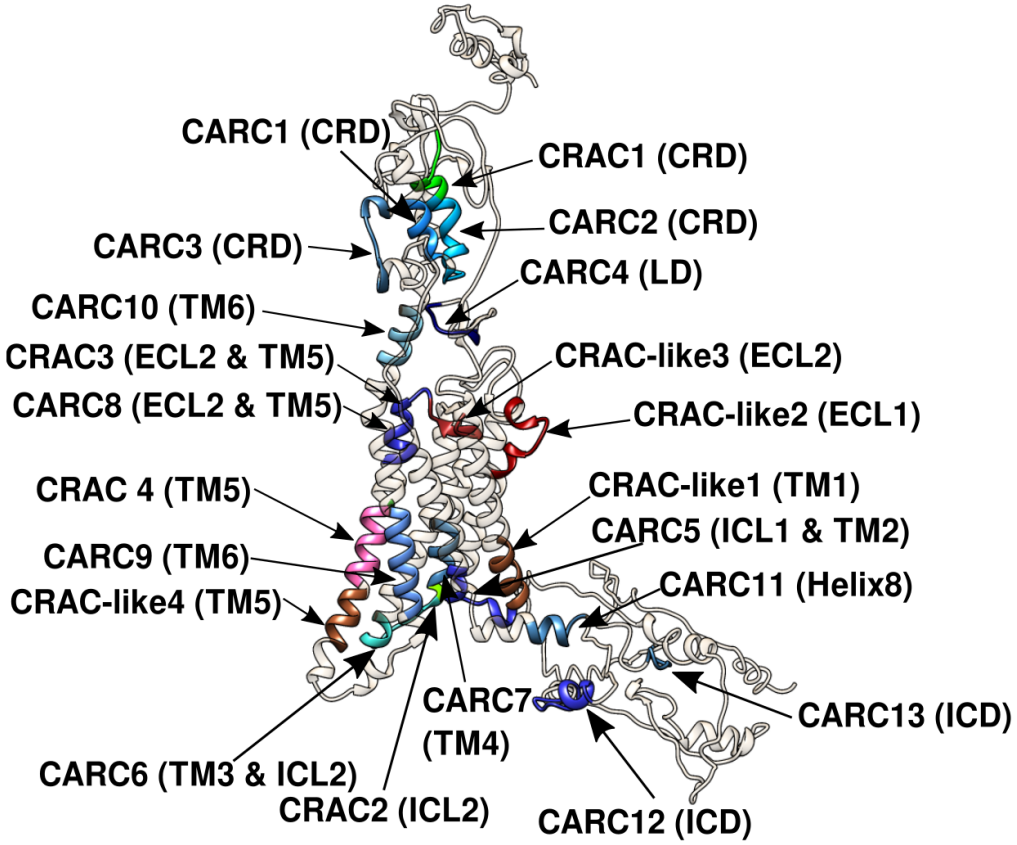


Figure 3.3: A schematic representation of SMO displaying computationally identified CRAC, CARC, and CRAC-like cholesterol-binding sequence motifs.

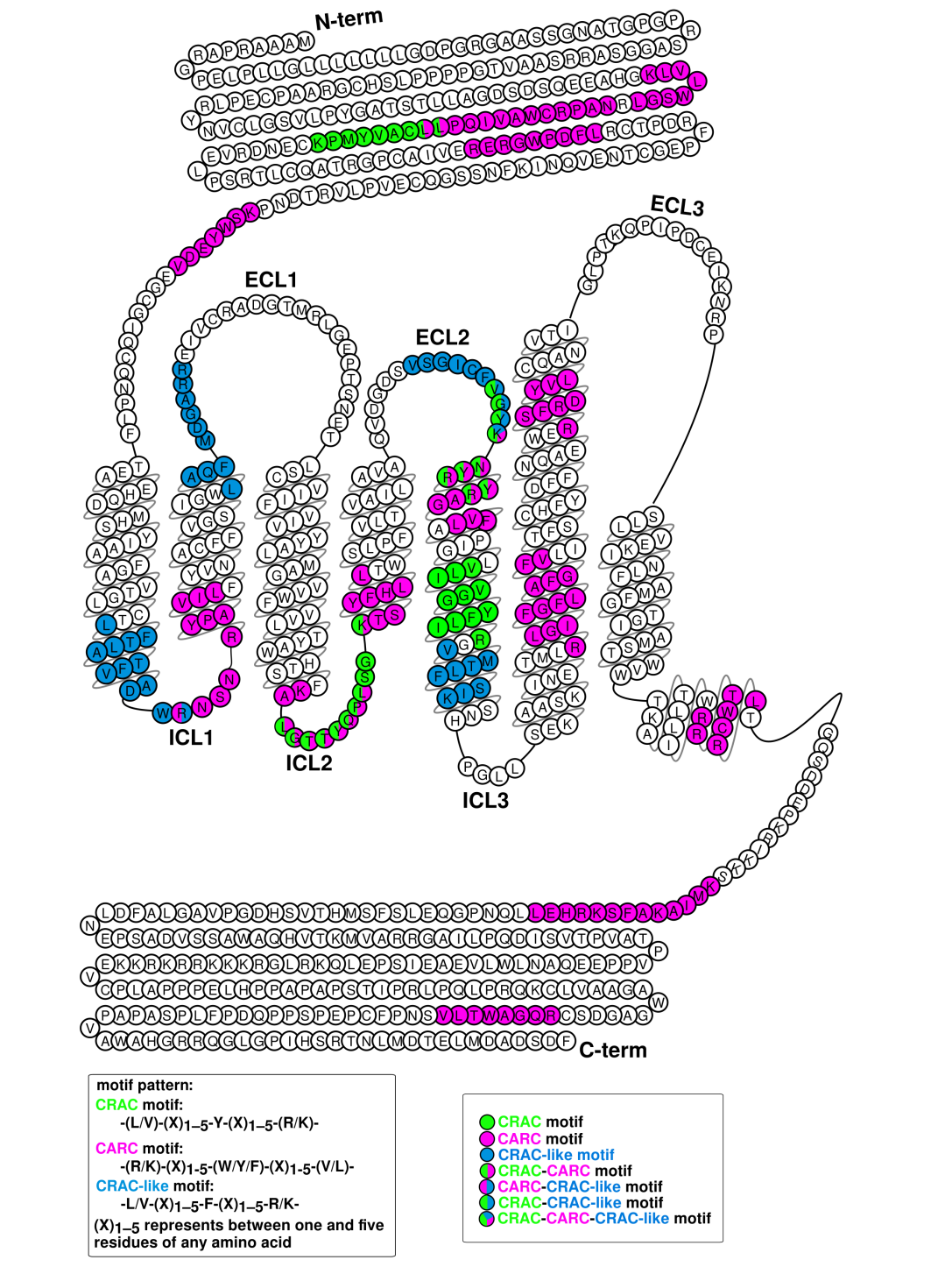


Figure 3.4: A schematic representation of SMO highlighting cholesterol-binding motifs. The CRAC, CARC, and CRAC-like motifs in the sequence are highlighted in green, pink, and blue.

3.3.1.2 Cholesterol-binding structure motif: *strict* CCM in SMO

Building upon an earlier attempt to partially identify the CCM [103], the component residues of a '*strict* CCM' in SMO were mapped. The fit could be achieved using some minor adjustments (Figures 3.5a and 3.5b): (i) Considering a positively charged residue K356 at position 4.41 that comes between the position range from 4.39 to 4.43, (ii) keeping W365 at position 4.50, which overlays well with the component residue at position 4.50, (iii) considering L362 at position 4.47 (instead of H361 at 4.46), which fulfills the requirement of a β-branched hydrophobic residue, and (iv) considering Y262 at position 2.39, which is conserved in class F GPCRs (the most conserved residue in TM2 is C273 at 2.50). Therefore, residues K356 (4.41)-W365 (4.50)-L362 (4.47) of SMO correspond to the CCM, and together with, Y262 (2.39) correspond to the *strict* CCM found in other GPCRs (Figures 3.6a and 3.6b).

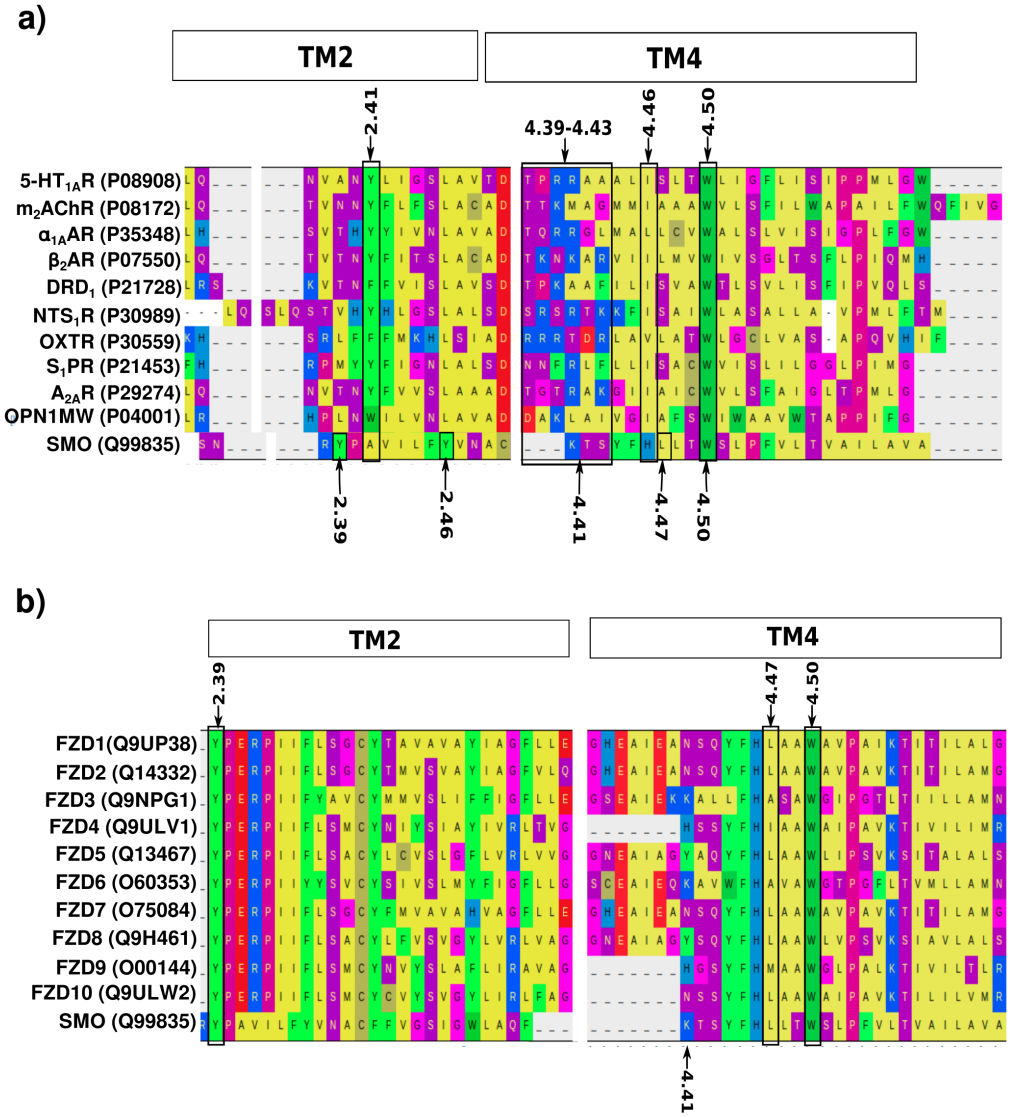


Figure 3.5: *strict* CCM in SMO. a) Sequence alignment of SMO with class A GPCRs to map the *strict* CCM in SMO; b) Sequence alignment of SMO with class F GPCRs with *strict* CCM residues of SMO. The corresponding protein accession numbers are indicated in parentheses.

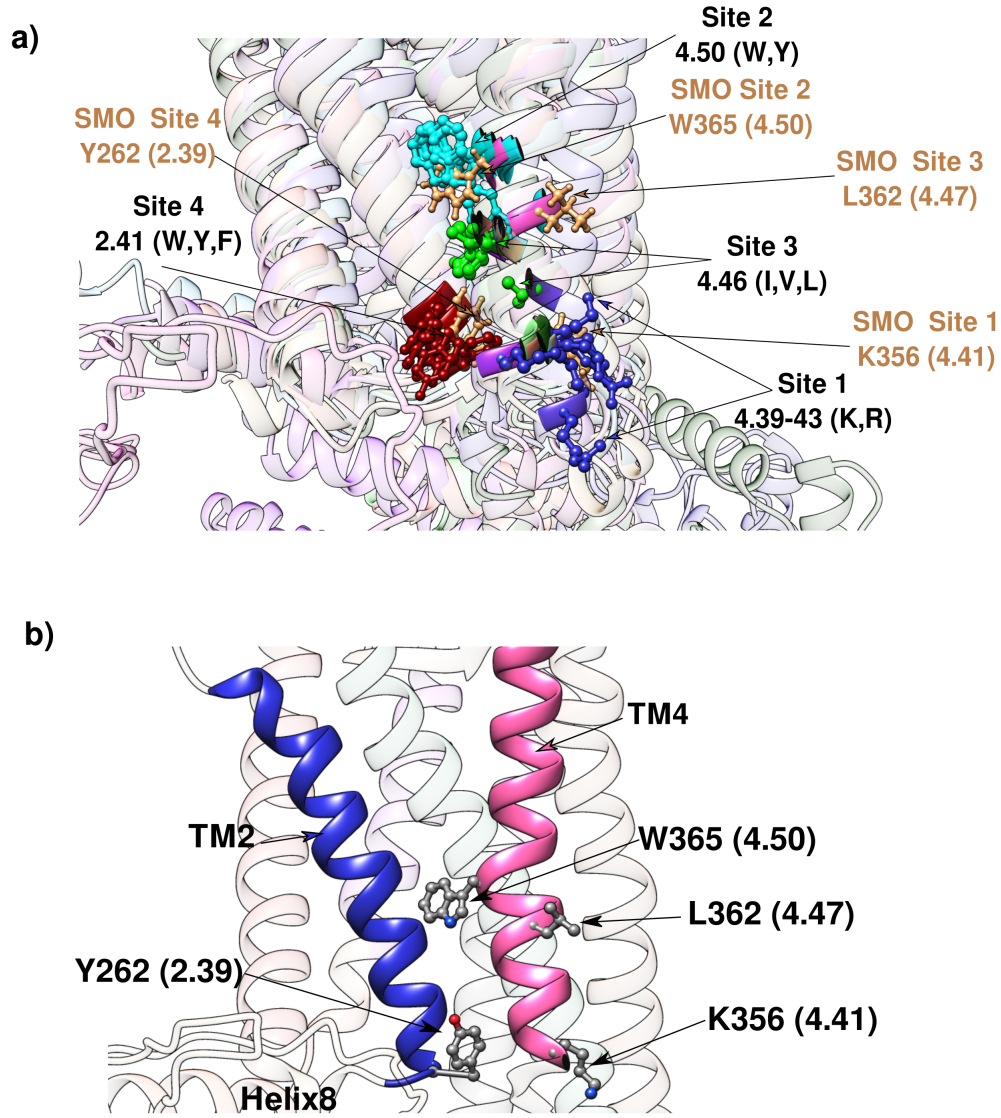


Figure 3.6: *strict* CCM in SMO. a) Structure alignment of SMO with class A GPCRs in ribbon representation and the residues of *strict* CCM (in the ball and stick representation). Key residue 1 (blue) at position 4.39-4.43 is arginine or lysine. Key residue 2 (cyan) at position 4.50 is the most conserved tryptophan. Another allowed amino acid in this position is tyrosine. Key residue 3 (green) at position 4.46 is isoleucine or leucine or valine. Key residue 4 (maroon) at position 2.39 is tryptophan or phenylalanine or tyrosine. The four-key residues of SMO corresponding to the *strict* CCM are highlighted in tan color; b) shows all four *strict* CCM key residues only in the SMO TMD.

3.3.2 Sphingolipid-binding motif in SMO:

Sphingolipids are an important class of bioactive phospholipids without a glycerol backbone that accounts for 10–20% of total membrane lipids [74]. They play an important role in signal transduction and cell recognition. Sphingomyelin (SM) is a type of sphingolipid [147]. Several experimental and computational studies have suggested that SM interacts with membrane cholesterol [75, 128].

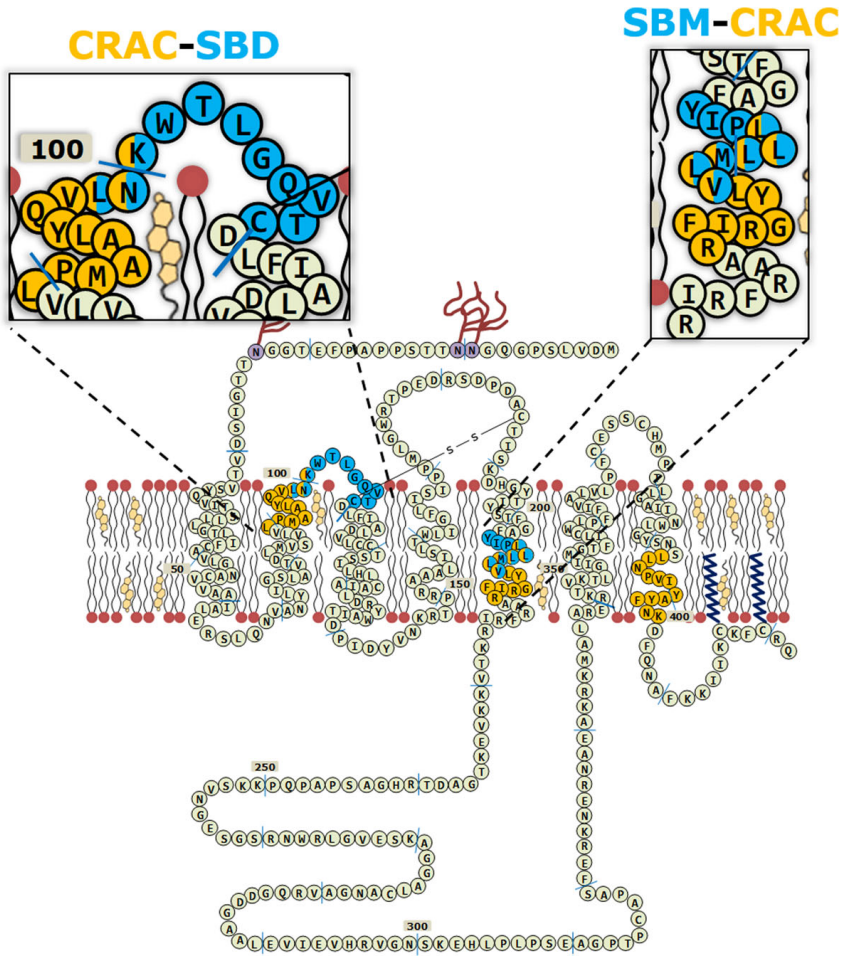


Figure 3.7: A schematic representation of the human serotonin1A receptor embedded in the membrane. The overlapping lipid binding/interaction motifs are highlighted. The sequence of human serotonin1A is shown as circles. The sphingolipid binding domain (SBD) present in TM2 and the sphingolipid binding motif (SBM) present in TM5 are highlighted in cyan. The two boxes show the overlapping between a) CRAC motif I (yellow) and SBD (cyan) in TM2; and b) CRAC motif II (yellow) and SBM (cyan). The common residues between CRAC motif I-SBM and CRAC motif II-SBD are depicted in both cyan and yellow (Source: Sarkar and Chattopadhyay, 2020 [138]).

A recent study has highlighted the role of SM lipids in the accessibility of cholesterol to SMO in the primary cilium [128]. It has been suggested that SM sequesters membrane cholesterol and restricts cholesterol from reaching the SMO. Several GPCRs, including cholecystokinin, oxytocin, and secretin receptors, as well as subtypes of human serotonin receptors, have been found to contain a sphingolipid-binding motif (SBM) and/or a sphingolipid binding domain (SBD) [17, 26, 74, 124, 144].

The pattern of SBM is -(I/L/T/V)XX(I/L/T/V)(I/L/T/V)XX(I/L/T/V)(F/W/Y)-, where X can be any of the 20 amino acids (Figure 3.7) [144].

An SBM pattern search suggested that SMO also has SBM "IQPLLCAVY" in its CRD from 122 to 130 in the CRD. SMO has a CRAC motif from "LLCAVYMPK" from 125 to 133 (Figure 3.8). However, there is no evidence that suggests SM interactions with SMO. The cholesterol and sphingolipid interaction, the presence of cholesterol in the CRD, and the CRAC motif with an overlapping SBM give the idea that cholesterol and sphingolipid could interact simultaneously. The identification of SBM in SMO may lead to result in the construction of new relationships between SMO and SM.

Figure 3.8: Snake plot representation of SMO CRD illustrating the overlap of SBM (122-130) and CRAC (125-133). The common residues between the SBM and CRAC motifs are colored pink.

3.3.3 Arginine/Lysine (Arg/Lys; R/K) clusters in SMO

Inositol plays an important role in the signaling pathway (Tan and Brill 2014). PI4P is the signature lipid for the tip of primary cilia [35, 52, 77, 110, 135]. They are anionic lipids and interact with the positively charged amino acid of the receptor. Arg/Lys cluster is a group of positively charged amino acid residues (arginine and lysine) that interact with anionic phos-

pholipids (phosphatidylinositol; PI and phosphatidylserine; PS). A previous study on SMO and PI4P interaction shows that Drosophila SMO has four arginine clusters (R1-R4) in its C-terminal [77]. They have shown that mouse SMO (mSMO) also has these arginine clusters in its C-terminal. Their experiments on mSMO suggest that the R3 and R4 clusters are responsible for PI4P-associated binding and activation.

After that, the sequence of hSMO ICD has been observed in detail. It was found that there are five Arg/Lys clusters of positively charged residues (arginine and lysine) in hSMO (Figure 3.9), similar to the arginine clusters in Drosophila SMO. Arg/Lys 1 (residue 561-567), Arg/Lys 2 (residue 571-576), Arg/Lys 3 (residue 624-629), Arg/Lys 4 (residue 667-680), and Arg/Lys 5 (residue 703-711).

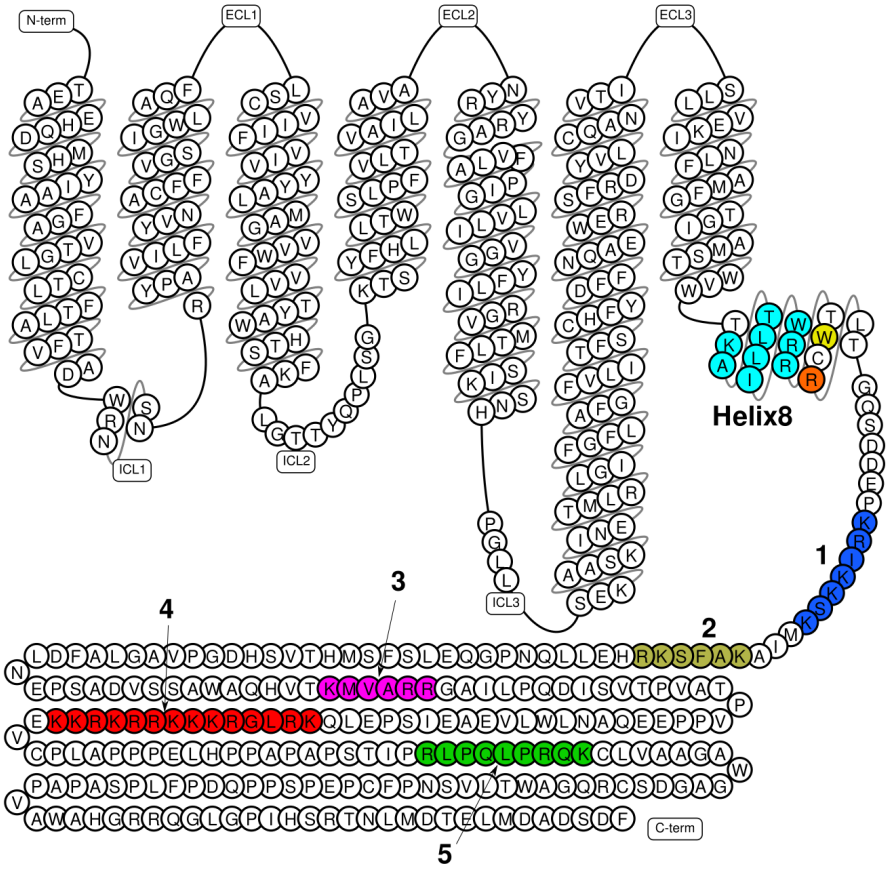


Figure 3.9: A snake plot representation of SMO highlights helix 8 (cyan), W549 (yellow), and R551 (orange) next to helix 8. The multiple Arg/Lys clusters are highlighted in blue, olive green, magenta, red, and green, respectively.

3.3.4 Conformational switches or conserved GPCR motifs in SMO

As discussed earlier, SMO shows less than 10% sequence similarity with class A GPCRs. There is some pattern that is common from class A to class F GPCR. The sequence alignment of the class F GPCRs helps in the identification of conserved motifs in TM helices.

3.3.4.1 The presence of the conserved Class F motifs in SMO

MSA of the TM6 in Class F receptors with SMO shows the presence of conserved residues. The consensus sequence of these residues in TM6 is CYFYE (Figure 3.10a). Here, SMO has H at the second position instead of Y and D at the last position in place of E. Both D and E are negatively charged amino acids. Here, C is the most conserved residue and is numbered 6.50. Class A GPCRs have CWXPY motif [158], where P is the most conserved residue 6.50 that imposes the bend in TM6. These conserved CYFYE residues could be an alternative to the CWXPY motif in class A GPCRs.

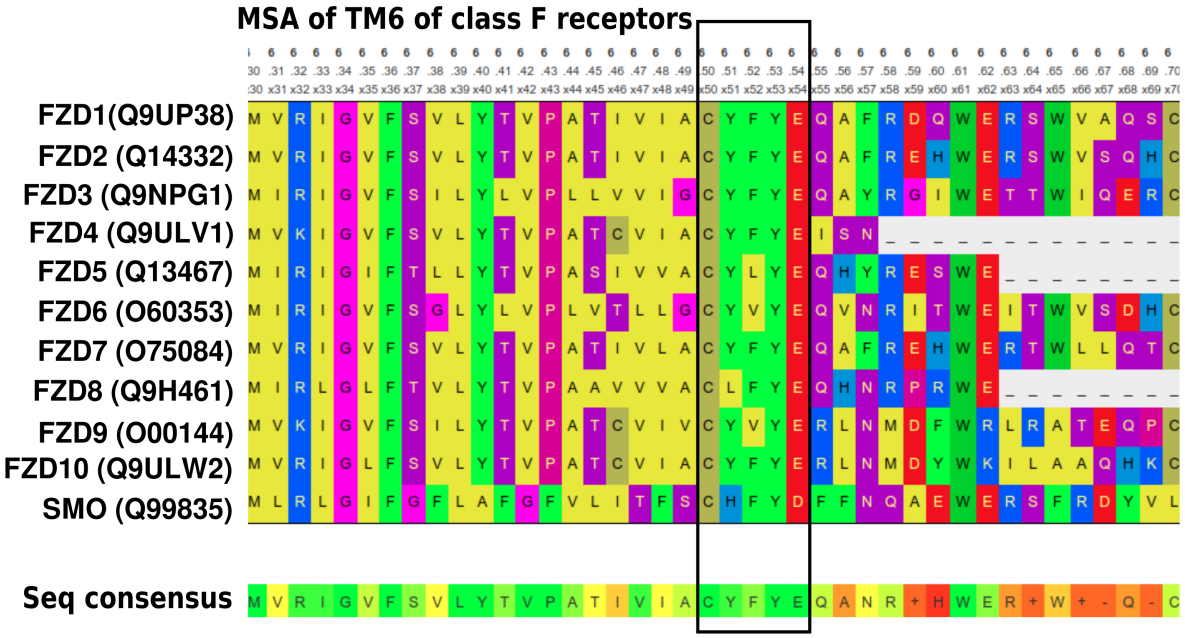


Figure 3.10: Sequence alignment of SMO TM6 (for better visualization, only some part of TM6 sequence MSA is shown) with class F GPCRs. The set of conserved residues in a TM6 is marked in rectangles.

Similarly, MSA of the TM7 in Class F receptors with SMO shows the presence of the consensus motif GIAMSTWVW (Figure 3.11).

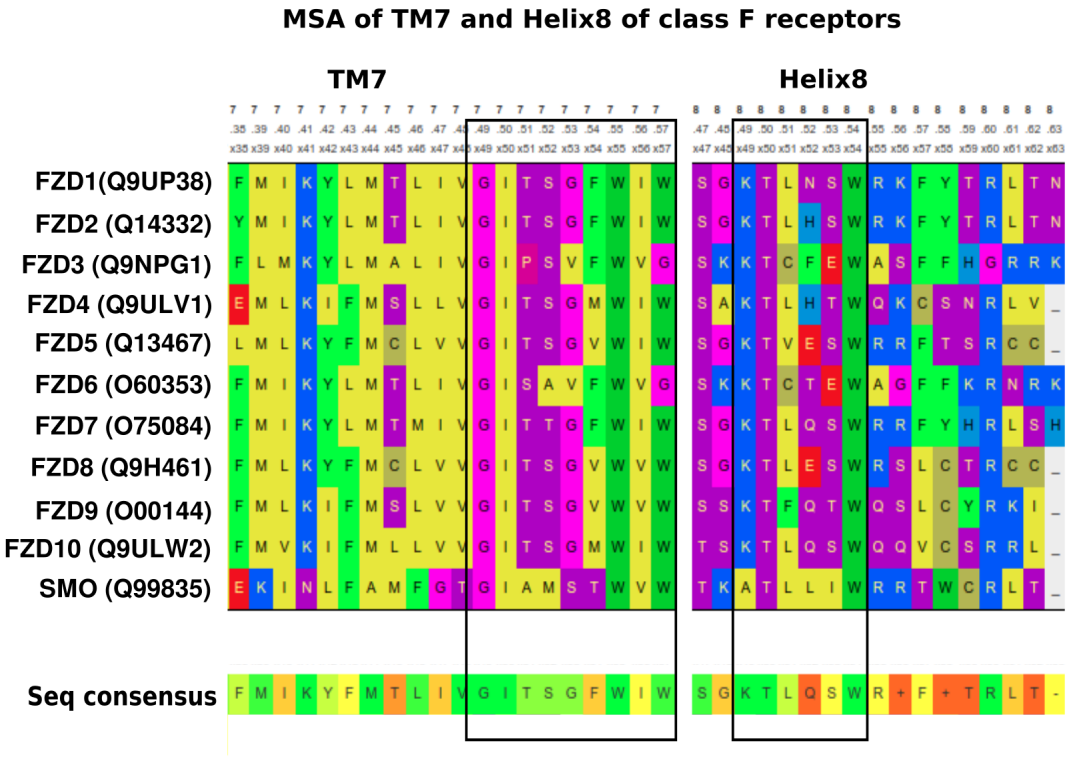


Figure 3.11: Sequence alignment of SMO TM7 and helix 8 (for better visualization, only some part of TM7 and helix 8 sequence MSA are shown) with class F GPCRs. The set of conserved residues are marked in rectangles.

The functional significance of these two motifs can not be explained by only the MSA of SMO with class A and/or other GPCRs. Therefore, further investigations using experimental or computational methods are necessary to gain a deeper understanding of the functional implications and potential roles of these motifs.

3.4 Conclusion

The SMO sequence has been analyzed to identify conserved domains and motifs, which can provide clues about the function and evolutionary history of the class F receptor or other GPCRs. Most of these domains and motifs are shared across different classes of GPCRs, and a few are specific to class F subfamilies. This chapter summarizes the known lipid-binding motifs/sites and other important conserved motifs of SMO, as well as the identification of conserved GPCR motifs of the class F receptor in SMO. The next two chapters (Chapters 4 and 5) explored the interaction of membrane cholesterol with these known cholesterol-binding motifs/sites along with the identified sequence and the redefined strict CCM. The importance of Arg/Lys clusters in the stabilization of SMO in primary cilia has been highlighted in this study. The sequence analysis and new motif/domain identification of class F conserved motifs aid in our understanding of some key features of class F receptors. The SMO sequence analysis provides valuable information that can be used to understand its structure, function, and dynamics.

Chapter 4

Structure and dynamics of SMO in ciliary membrane and its interactions with membrane lipids

4.1 Introduction

Smoothened (SMO) is essentially involved in three sequentially related events in the Hedgehog (HH) pathway: 1) HH binding to Patched1 (PTCH1), 2) activation of SMO, and 3) activation of the GLI family of transcription factors. This work focuses on the second component, which is currently understood to involve the mediation of HH bound PTCH1 in initiating some conformational changes in SMO [64, 102], which is important for its activation [5]. It has been established that in response to HH signaling SMO localizes to the primary cilium, and this localization is essential for SMO to function [52, 128]. In addition to cholesterol, ciliary membrane are rich in phosphatidylinositol-4-phosphate (PI4P) and other phospholipids that likely play a important role in activating SMO. Further, recent structures (PDB ID(s): 6D35, 6XBK, and 6XBJ) reveal bound cholesterol and a sterol channel in SMO [68, 125], as discussed in chapter 3, and there is evidence that small molecules, including cholesterol (and PI4P), regulate SMO activity [77, 128]. Given this scenario, there are several yet unanswered questions related to the stepwise molecular mechanism of this event.

The first step towards SMO activation is its translocation to the primary cilia (Figure 1.1; Chapter 1), as a result of HH binding to PTCH1, followed by the PTCH1 dislocation. However, it is still unclear how exactly this happens. What is the driving force behind the translocation of inactive SMO to the primary cilia? What mediates the anchoring of SMO? Does this also provide the driving force for the translocation to take place spontaneously when PTCH1 vacates the cilia? There is an even greater lack of clarity regarding the subsequent steps. They are related to how cholesterol, PI4P, and other lipids in the primary cilium coordinate to control SMO activity [77, 128]. What could be the sequence of events? Are there other small molecules involved? What is the role of cholesterol-binding motifs in the interactions of cholesterol with

cilia bound SMO? How are the conformational changes in different SMO domains related to cholesterol binding to SMO? Where does the cholesterol bind initially? Does it bind to the CRD to transmit the signal to TMD, or does direct cholesterol binding to TMD help in SMO activation? It has been reported that SMO mutants, including those lacking the CRD, are functionally viable in a membrane environment [18, 21]. What does this imply in this context?

The 3D structure of SMO helps to comprehend the SMO activation mechanism. There are 18 structures (crystal and cryo-EM) of SMO Table 3.1; Chapter3 in the PDB (Protein Data Bank) [150]. All the available structures have the transmembrane (TM) region, but the loop region has been replaced by another protein, such as apocytochrome b562RIL (BRIL), or flavodoxin (since the protein is highly dynamic because of loops and difficult to crystallize) with truncated N and C terminals. The complete N to C terminal structure of SMO is not available (including all loop regions). Therefore, modeling a complete SMO is imperative. In this chapter, the modeling of the full-length 3D structure of SMO was first described. Then, the structural dynamics of SMO in the primary cilium were investigated in the presence of cholesterol and other membrane lipids (PI4P) using CGMD simulations of appropriately designed membrane protein models [as described in Chapter1 section 1.5 and Chapter 2 section 2.5]. The interaction of ciliary lipids with SMO and the motifs/sites present in SMO, discussed in chapter 3, were also studied. This study sheds light on some of the questions mentioned above. In addition to advancing the understanding of the molecular level mechanism of the HH pathway, the results widen the scope for applying rational drug design approaches targeting the HH pathway.

4.2 Methods Details

4.2.1 3D structure of full-length SMO

The SMO linear sequence was obtained from UniProtKB. The full-length SMO structure was generated by the GPCR I-TASSER [187] web server. The generated models were validated using MolProbity [39]. Furher, the best model structure was energy minimized using the steepest descent algorithm available in GROMACS and used for further studies.

4.2.2 Coarse-grained (CG) system setup

The functional SMO structure was simulated using coarse-grained molecular dynamics (CGMD) in two different membrane models: (a) a heterogeneous bilayer made up of phosphatidylcholine (POPC), phosphatidylethanolamine (POPE), phosphatidylserine (POPS), PI4P, and cholesterol; and (b) a homogeneous phosphatidylcholine (POPC) bilayer. The composition of the outer and inner leaflets of the heterogeneous bilayer (modeled ciliary membrane) was chosen to realistically represent the physiological ciliary membrane [33, 60, 73]. Table 4.1 and

Figure 4.1a give the composition details. It may be noted here that in the initial setup, the distribution of the lipids was random. However, during the course of the simulation, it changes due to the presence of SMO. Control simulations of the POPC membrane with SMO and the modeled ciliary membrane without SMO are performed. The molecular systems for simulations were prepared using the CHARMM-GUI input generator [78, 127, 181]. The energy minimized SMO structure (without SP) was converted to a CG representation using the MARTINI 2.2 force field [40, 107] and embedded into the heterogeneous ciliary and homogeneous POPC lipid bilayers (Figure 4.1b and 4.1c). A heterogeneous lipid bilayer with phosphatidylinositol 3-phosphate (PI3P) lipids was generated using the replacement method as the CG parameter of PI4P is not available at CRAMM-GUI web server. Both homogeneous and heterogeneous bilayer systems were solvated using the standard MARTINI water model [101] and neutralized with 0.15 M NaCl. Then PI3P lipids from the heterogeneous bilayer system were replaced with PI4P lipids to mimic the ciliary membrane. PI4P parameters were generated using Martini mapping version 2.0 [148], as discussed in the Methods chapter (see Chapter 2, Section 2.5 for more details). A ciliary membrane control system was prepared by simply removing SMO from the SMO in the modeled ciliary membrane system setup, as described above (Figure 4.1d).

Five independent simulations of SMO in the modeled ciliary membrane, each of 15 μs, were performed with different random distributions of lipids around the SMO and different initial velocities. Similarly, three independent simulations for both control simulations, SMO in the POPC membrane and ciliary membrane without SMO (membrane only system), each of 15 μs were performed. The total simulation time, including the control simulations, was 165 μs. The details of the simulations performed, along with their system composition, were summarized in Table 4.1.

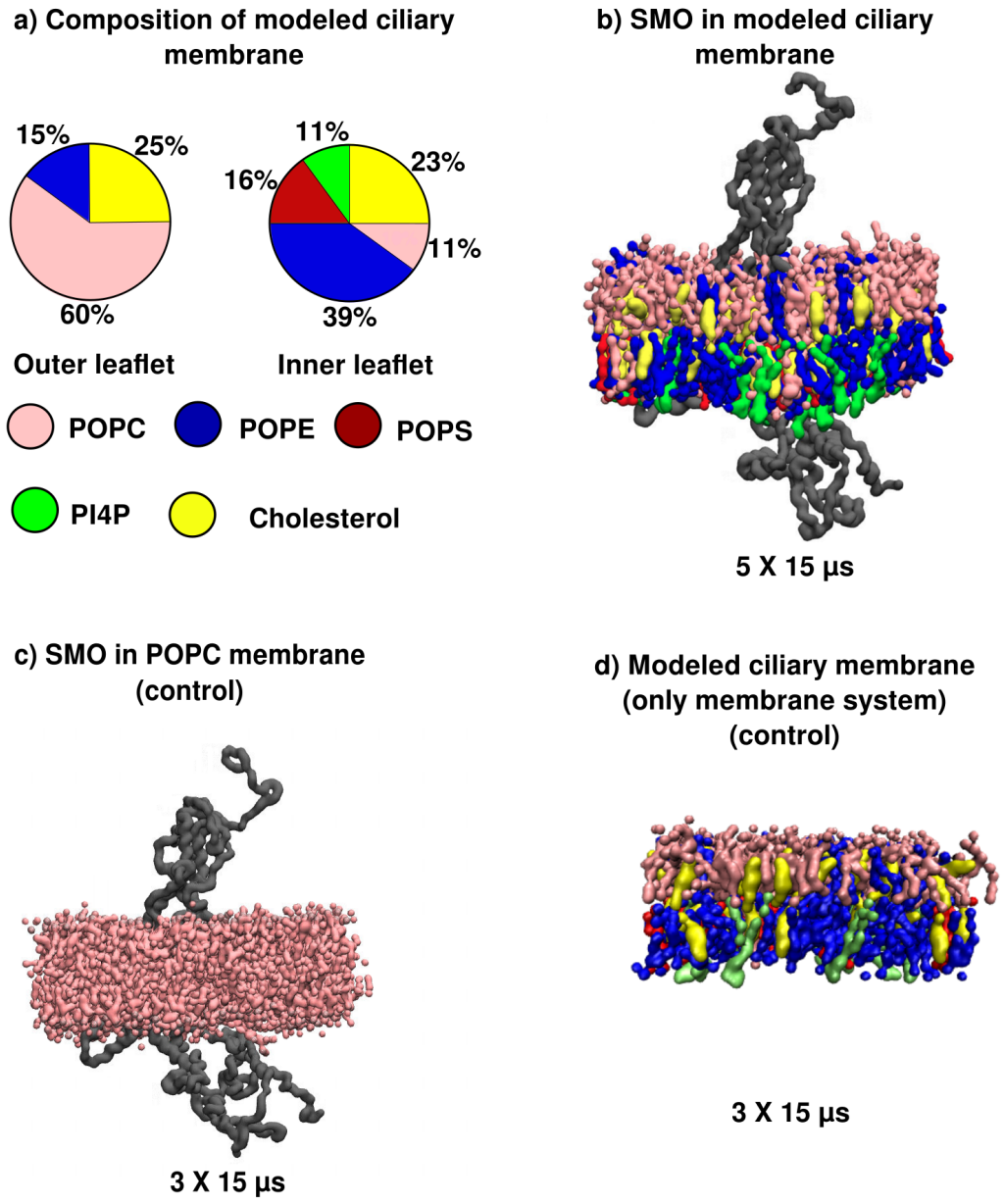


Figure 4.1: CG simulation systems overview a) composition of the modeled ciliary membrane; the initial structure of b) SMO in the modeled ciliary membrane; c) SMO in the POPC membrane, and d) membrane only system (modeled ciliary membrane without SMO). SMO is shown in gray; water and ions are not shown for the clarity of the images.

Table 4.1: Details of studied simulation systems.

S. No.	Description	Type	Membrane Composition (in numbers)		Number and duration of simulations (μs)
			Outer leaflet	Inner leaflet	
1	SMO in modeled ciliary membrane (heterogeneous)	CG	POPC: 120 POPE: 30 Cholesterol: 50	POPC: 24 POPE: 84 POPS: 34 PI4P: 24 Cholesterol: 50	5 x 15 μs
2	SMO in POPC membrane (homogeneous)	CG (control)	POPC: 200	POPC: 202	3 x 15 μs
3	Modeled ciliary membrane (Membrane only system)	CG (control)	POPC: 120 POPE: 30 Cholesterol: 50	POPC: 24 POPE: 84 POPS: 34 PI4P: 24 Cholesterol: 50	3 x 15 μs

4.2.3 Simulation parameters

All simulations were carried out using GROMACS version 2019 [162], with MARTINI forcefield version 2.2 [40, 107]. Energy minimization was carried out using the steepest descent algorithm and subjected to six successive equilibrations of a total of 500 ns using the leap-frog algorithm [146]. After the equilibration, the production run was carried out for 15 μs. The temperature of each molecular group in the system was weakly coupled to a thermostat at 310 K using the v-rescale algorithm [23]. Semiisotropic pressure was controlled at 1 bar using the Parrinello-Rahman algorithm [23]. Initial velocities for the simulations were chosen randomly from a Maxwell distribution [136]. The reaction field approach [157] was used for non-bonded interactions with a Coulomb cut off of 1.1 nm and a potential shift Verlet modifier. Bond lengths were kept constant using the LINCS algorithm [62]. Periodic boundary conditions were maintained along the x, y, and z directions. A time step of 20 fs was used, and snapshots were taken every 100 ps for analysis.

4.2.4 Backmapping of CG structures

The CG structures were converted back to atomistic models using the backward.py script [173] provided by the MARTINI website and the CHARMM-GUI martini to all-atom converter (https://charmm-gui.org/?doc=input/converter.martini2all) with the CHARMM 36 force field for both protein and lipids.

4.2.5 Analysis

Trajectories were analyzed using tools implemented in GROMACS, and VMD utilities [71] together with in-house scripts. VMD was used for visualization, and figures were produced using VMD, Bendix VMD plugin [36], UCSF Chimera [104], PyMOL (https://pymol.org), and Inkscape (www.inkscape.org). Plots were generated using Xmgrace [161], Gnuplot [179], and in-house Python scripts.

4.2.5.1 Structural analysis

The first 1.5 μs of the production run was considered as the unrestrained equilibration phase (required to allow the systems to become stable), and analysis was performed from 1.5 μs onwards. Analyses such as root-mean-square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), helical tilt, and solvent accessible surface area (SASA) were performed using tools within GROMACS [162]. The RMSD was calculated with reference to the starting structure of each simulation. The helical tilt of an individual TM helix was determined by measuring the angle between a cylindrical model fitted to the helix axis and the normal direction of the bilayer, using the *bundle* function of GROMACS.

4.2.5.2 Dynamic cross correlation matrix (DCCM) and Principal component analysis (PCA)

PCA was performed on backbone atoms by using the Normal Mode Wizard plugin of VMD-1.9.1 version. DCCM [6] were calculated using the in-house script as a time average along the simulation trajectory to study the correlated motions between different regions of SMO. PCA and DCCM analyses were carried out on the concatenated trajectory of 67.5 μs and 40.5 μs of SMO in the ciliary membrane and SMO in the POPC membrane systems, respectively.

4.2.5.3 Contact analysis

In order to quantitatively investigate the interaction dynamics, the *mindist* function available in GROMACS was used to calculate the minimum of all distances between any two CG beads, belonging to the two interacting entities (lipids and SMO residues), in every frame of the trajectory. The interacting entities (cholesterol, PI4P, and other lipids) were appropriately number-tagged. The contacts between the cholesterol-binding motifs of SMO and individual cholesterol molecules in the membrane were also calculated by *mindist* function, and heatmaps were generated by a Python script.

4.2.5.4 Occupancy of Cholesterol, PI4P, and POPS lipids around SMO

The occupancies of cholesterol, PI4P, and POPS lipids around SMO were calculated using in-house protocols. First, the distance between each residue and cholesterol/PI4P/POPS lipid has been calculated using the *mindist* function of GROMACS for the last 13.5 μs of each simulation system. Then the frames, where the distance value is within the cut off distance of 0.6 nm, are masked as 1. Occupancy is calculated by this formula- $(\frac{no.of frames masked 1}{total no.of frames} \times 100)$. The final occupancy was calculated by taking the mean across all replicates.

4.2.5.5 Tilt angles of Cysteine-rich domain (CRD) and intracellular cytoplasmic domain (ICD)

The tilt angles of CRD and ICD relative to TMD were calculated by the *angle* function of GROMACS. The backbone beads of P69, V210, and W535 residues defined the CRD tilt. Similarly, ICD tilt was defined by the backbone beads of V210, V536, and E650.

4.2.5.6 Membrane dynamics

The dynamical characteristics of cholesterol molecules in the membrane for each replicate simulation, such as cholesterol concentration in the bilayer and movement of cholesterols within the bilayer, were analyzed by identifying the position of individual cholesterol in the outer and inner leaflets for each frame with reference to the starting structure using the FATSLiM analysis tool [22].

4.3 Results

4.4 Modeling of SMO

The linear sequence of SMO (UniProt ID- Q99835) was retrieved from UniProtKB and submitted to the GPCR I-TASSER [187] web server for full-length SMO structure. The inactive SMO structures 5L7D and 4JKV were used as templates to generate five models that satisfy all of the topological requirements for GPCRs [24, 169]. A model with a C-score of -1.37 and a lower RMSD with both templates was chosen as the SMO structure (Table 4.2). The selected model showed 0.322 Å and 0.730 Å RMSD with the 5L7D and 4JKV templates, respectively. The model was also validated using MolProbity [39] to estimate the stereo-chemical correctness of the structure. It showed that 71.46% (561/785) of all residues were in favored (98%) regions, and 86.37% (678/785) of all residues were in allowed (> 99.8%) regions with 107 outliers (Figure 4.2a). The model was energy minimized after that. Steepest descent algorithm available in GROMACS was used for this energy minimization. After that, 82.9% (651/785) of all residues

were in favored (98%) regions, and 93.6% (735/785) of all residues were in allowed (>99.8%) regions with 50 outliers (Figure 4.2b).

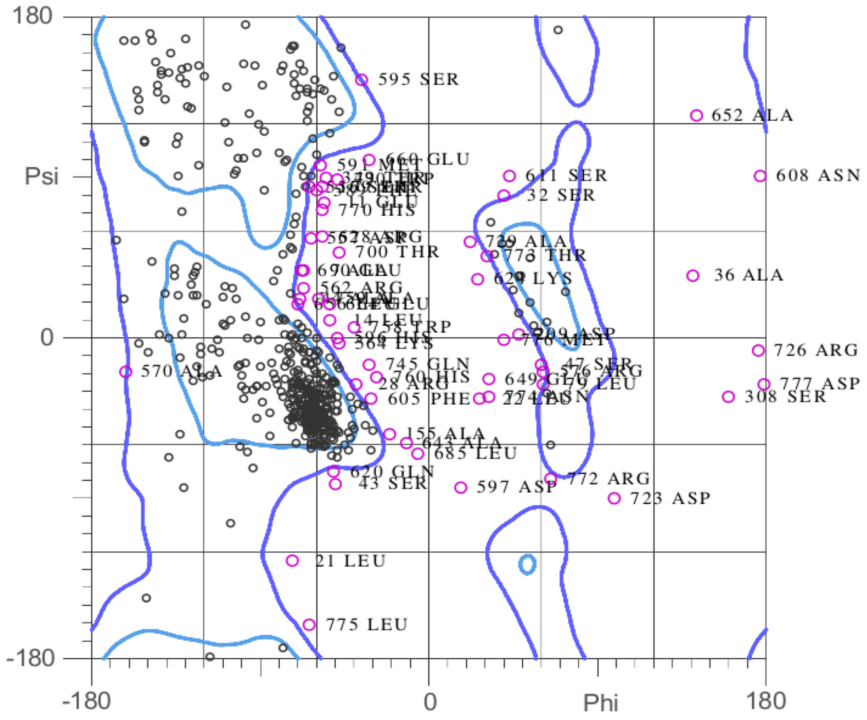
Table 4.2: Five models generated by GPCR I-TASSER with their C-score.

S. no.	**Model**	**C-score**
1	Model 1	-1.37
2	Model 2	-1.44
3	Model 3	-3.86
4	Model 4	-3.78
5	Model 5	-3.93

The distinct domains of the modeled SMO are shown in Figure 4.3a. The extracellular domain (ECD) spans from amino acid residues 1 to 221 and is composed of three subdomains: the signal peptide (SP) from residues 1 to 27, the CRD from residues 28 to 191, and the linker domain (LD) from residues 192 to 221. The transmembrane domain (TMD) is composed of seven TM helices, labeled TM1 through TM7, and spans from residues 222 to 554. These helices are linked by three intracellular loops (ICLs) labeled as ICL1, ICL2, and ICL3. The ICD spans from residues 555 to 787. The N-terminus of SMO is located in the ECD, while the C-terminus is located in the ICD. SP was removed from the modeled structure to obtain the functional SMO for simulation and further study (Figure 4.3b).

The results reported herein refer to the simulation of the dynamics of SMO, in the modeled ciliary membrane system. Other simulations involving SMO in POPC membranes or membrane systems without SMO were only used for comparison purposes. A visual investigation of the simulation trajectories using VMD showed that SMO and the different components of the ciliary membrane interact with each other dynamically, leading to changes in the distribution of the membrane components on the one hand, and in the conformation of SMO, on the other. Cholesterol molecules were seen to move around within and across the membrane leaflets, spending a lot of time in the TM region, while taking turns to interact with different SMO domains and with other membrane lipids. In contrast, the other lipids move around within their respective leaflets, while interacting with specific membrane proximal regions of SMO.

a) before minimization



b) after minimization

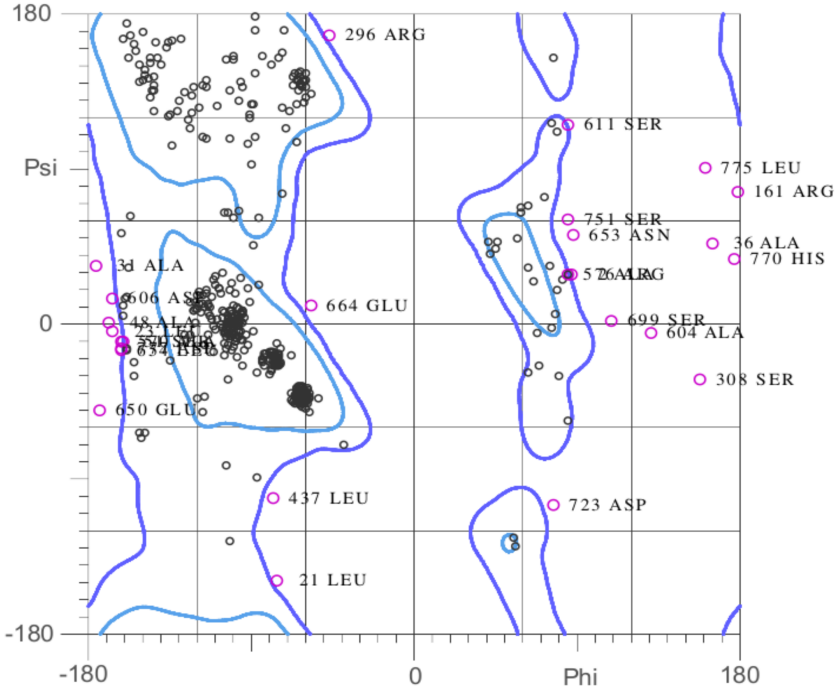


Figure 4.2: Ramachandran plot of modeled SMO before and after minimization.

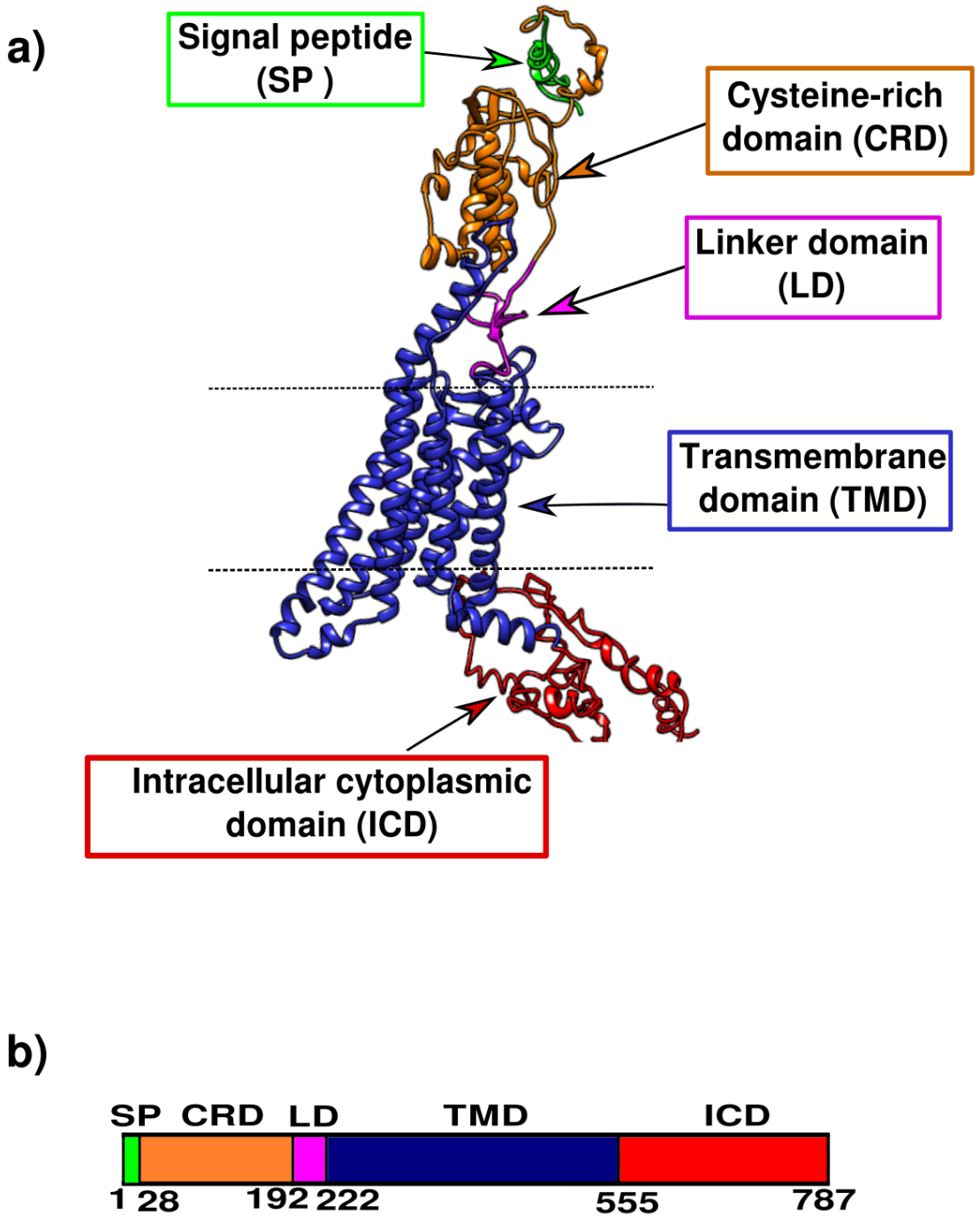


Figure 4.3: Molecular architecture of SMO. a) A cartoon representation of the full-length modeled SMO showing the extracellular domain (SP, CRD, and LD), transmembrane domain (TMD), and intracellular cytoplasmic domain (ICD); and b) residue numbers indicating the domain boundaries and domains included in the simulations.

4.4.1 Cholesterol molecules in the membrane interact with different domains of SMO

Figure 4.4 shows the domain-wise minimum distance plots for cholesterol molecules with SMO during the simulations of all five replicates in modeled ciliary membranes. The TMD and ICD interact with cholesterol molecules with distances less than 0.6 nm throughout the simulation trajectory in all replicates. However, the interactions of other SMO regions, namely the CRD and LD domains, were intermittent, and the interaction patterns differ slightly in the five replicates. For example, for a few frames, the distance between any residue of CRD - any cholesterol is 0.6 nm in replicates 1, 2, 3, and 5. Similarly, the LD – cholesterol distance value was also ∼0.6 nm in a few frames in replicates 1, 4, and 5. Two representative structures of SMO along with cholesterol molecules within a cut-off distance of 0.6 nm, at 9.15 μs and 2.28 μs time points from replicate 1 (Figure 4.5a and Figure 4.5b) selected based on the proximity of cholesterol to CRD and LD, respectively, provide a glimpse of these interactions. Overall, it may be concluded that the interaction of cholesterols with CRD, and LD regions were intermittent, while the cholesterol interactions with TMD were persistent. ICD was seen to interact with cholesterol throughout the simulation time (Figure 4.4).

Detailed VMD-assisted visual analysis of the trajectories, in addition to the distance plots, showed that the CRD bends towards the outer leaflet of the membrane and interacts with membrane cholesterols (Figure 4.4a, 4.5a, and Video 4.1). One of the frames selected based on the proximity of cholesterol to CRD (t= 9.15 μs) (Figure 4.5a) shows the interaction between CRD and two cholesterols (cholesterol referred to as chol; chol 5 and chol 52) through three residues P39, P41, and R42. The backmapped atomistic structure of this shows that one cholesterol (chol 52 makes a hydrogen bond with the side chain of R42 (Figure 4.5a). While in replicate 5, cholesterol interaction with CRD was observed multiple times between 5 and 8 μs and again from 13.38 to 14.99 μs. Overall, from all replicates, cholesterol interacts with R28, G29, A30, N35, G38, P39, G40, P41, R42, S43, A44, R49, D95, S96, Q99, E100, E101, R151, R159, E160, R161, D165, F166, L167, R168, and R173 residues. The interaction of LD with membrane cholesterol was also observed (Figure 4.5b and Video 4.2), and residue V210 was found to interact with membrane cholesterol (chol 48) (Figure 4.5b). The backmapped atomistic structure shows additional interactions of this cholesterol with CRD residues R159 and TM5 residues R398 and Y399 (Figure 4.5b).

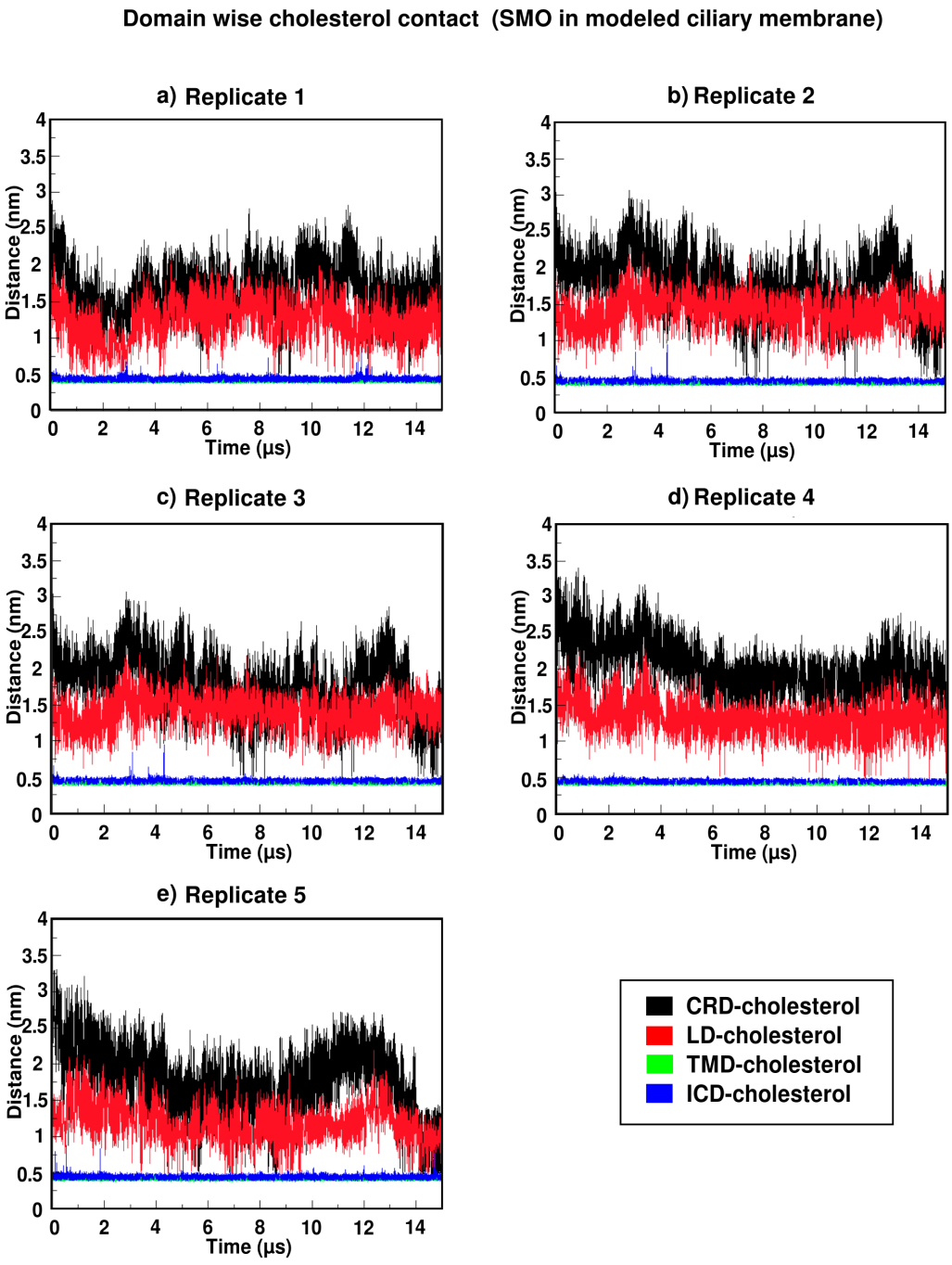


Figure 4.4: Domain-wise SMO and membrane cholesterol contacts. (a-e) The distance map for all five replicates of the simulation system shows the minimum distance between any membrane cholesterol and the corresponding SMO domain.

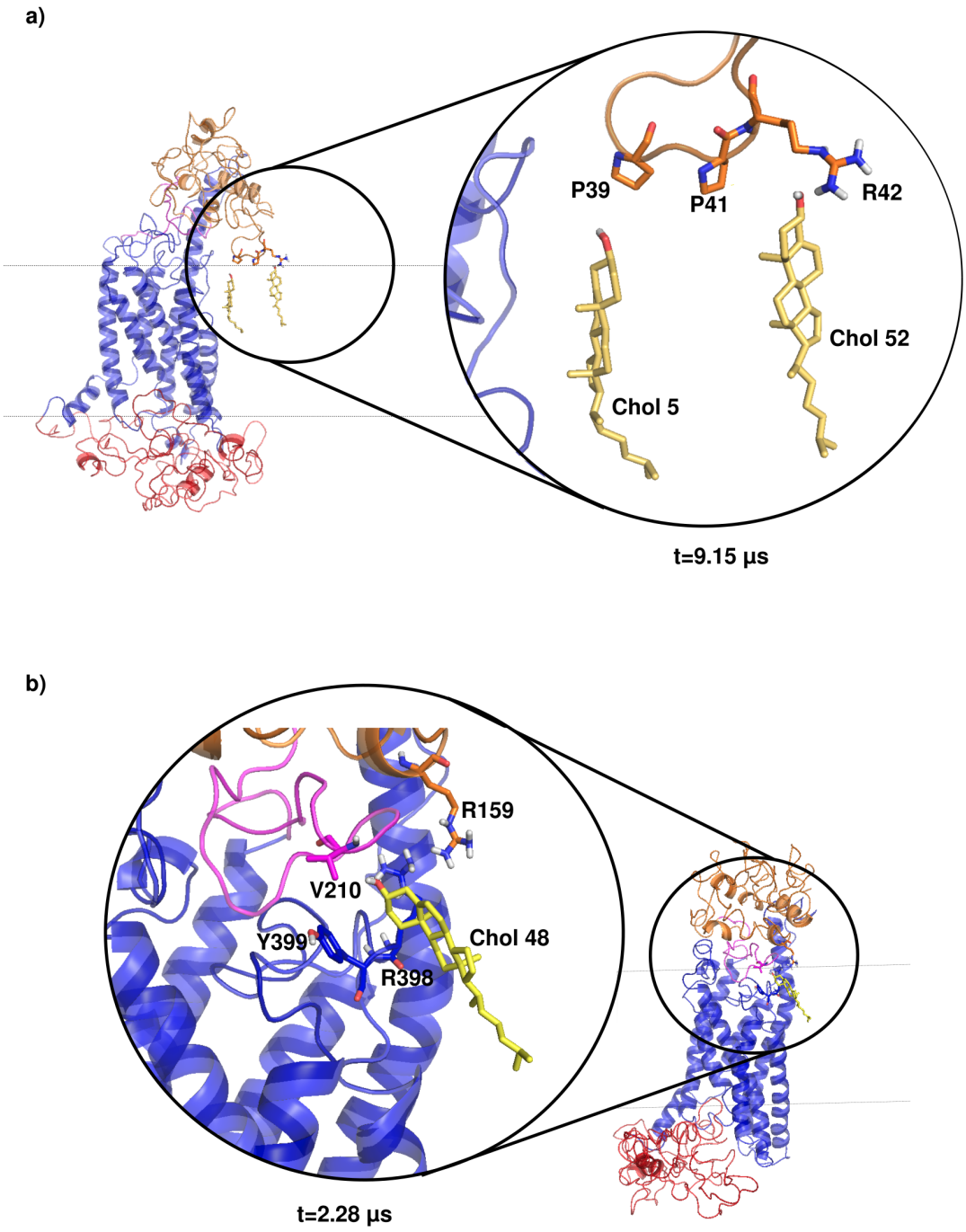


Figure 4.5: Interaction of membrane cholesterols with CRD and LD of SMO, shown as a backmapped atomistic structure. a) Interaction between CRD (orange) and two cholesterol molecules (yellow sticks) at t = 9.15 μs and b) Interaction between LD (magenta) and cholesterol (yellow sticks) at t = 2.28 μs time point.

Any membrane cholesterol is thought to interact with SMO if it is within 0.6 nm of any SMO residue. And based on this criterion, the regions of higher cholesterol occupancy in SMO TMD during the simulations (Figure 4.6) were identified. High cholesterol occupancy was observed near TM1, TM3, TM4, TM5, and TM6 (Figure 4.4). The distance plot of any residues of ICD with any cholesterol shows persistent interaction of ICD with cholesterol (Figure 4.4). The maximum occupancy for any cholesterol with any residue of ICD was ∼43%. In ICD, interactions with cholesterols were observed through the LEU-PRO rich region (685 to 713), particularly with the residues L685, P687, P688, P689, P694, A695, A697, P702, R703, P705, Q706, L707, P708, R709, Q710, and L713.

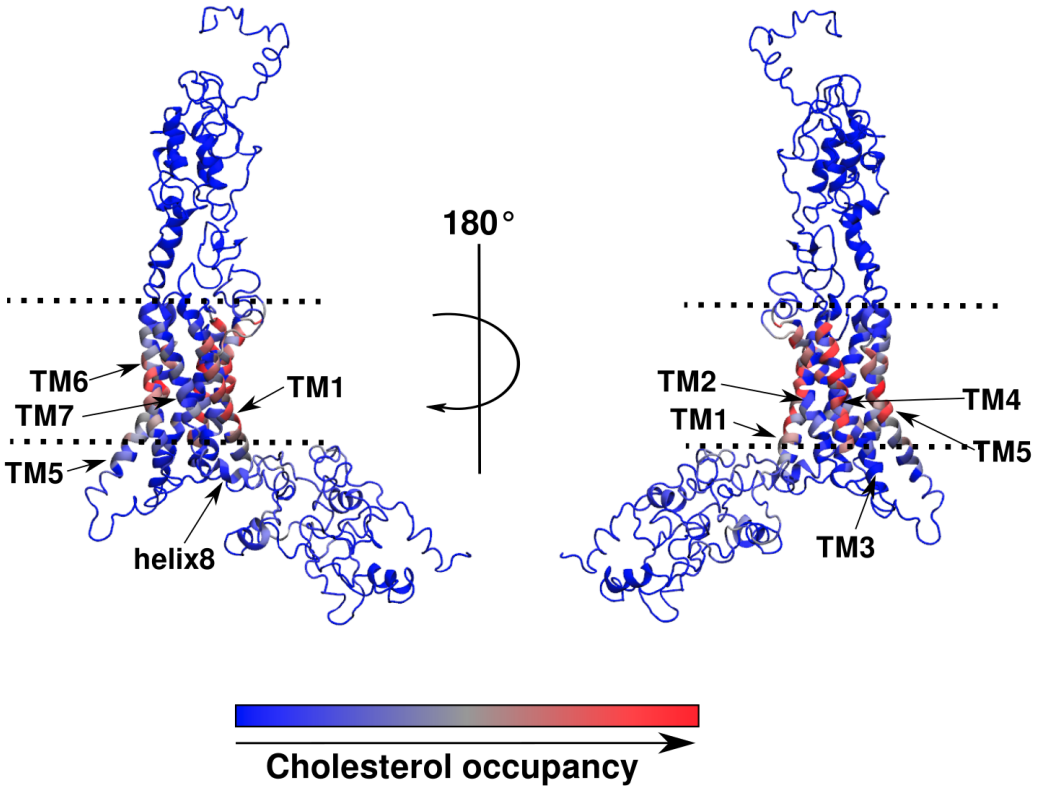


Figure 4.6: Cholesterol occupancy around SMO. Cholesterol occupancy around SMO was mapped onto the structure, with each residue colored from blue (less occupancy) to red (high occupancy). The dots represent the phosphate head group beads of ciliary lipids. The mean occupancy is calculated across all replicates, using a 0.6 nm distance cut off.

4.4.2 Cholesterol-binding motifs in TM4, TM5, and TM6 persistently interact with cholesterols

Thereafter, the persistence of interactions of individual cholesterol molecules with each of their proximal cholesterol-binding motifs for all five replicate simulations was estimated. Most of these motifs showed significant interactions with cholesterol in the membrane. The interaction patterns, however, showed some variations across replicates. For example, the CARC7 motif shows different extents of cholesterol interactions in different replicate systems. A cholesterol molecule (chol 78) was found near CARC7 in replicate 1 with greater than 21% occupancy, interacting with residues L362 and L363 in TM4 (Figure 4.7).

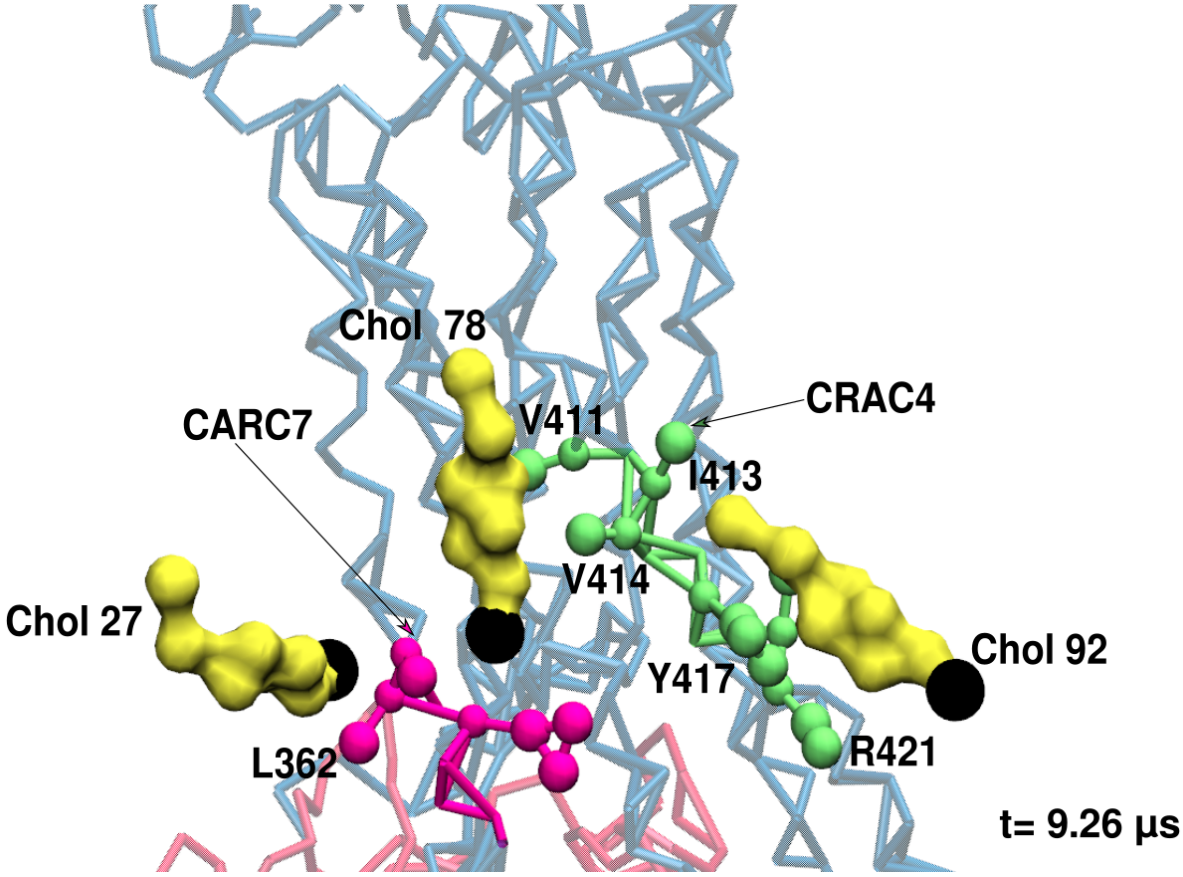


Figure 4.7: Interaction of cholesterol with CRAC4 and CARC7 motifs. The structure shows the interaction of three cholesterols with cholesterol-binding CRAC4 and CARC7 motifs at 9.26 μs from replicate 1. The color scheme is as follows: SMO TMD (blue) and ICD (red), cholesterol (yellow) and ROH bead (black), CRAC4 motif (green) from TM5, and CARC7 motif (pink) from TM4.

While in replicate 2, CARC7 showed 55% occupancy with cholesterol (chol 81). Two of the seven residues in the CARC7 motif, viz., Y359 and F360, were seen to interact with cholesterol molecules frequently, but transiently. Such transient interaction of CARC7 with cholesterol was also observed in the other four replicates (Figure 4.8b-e).

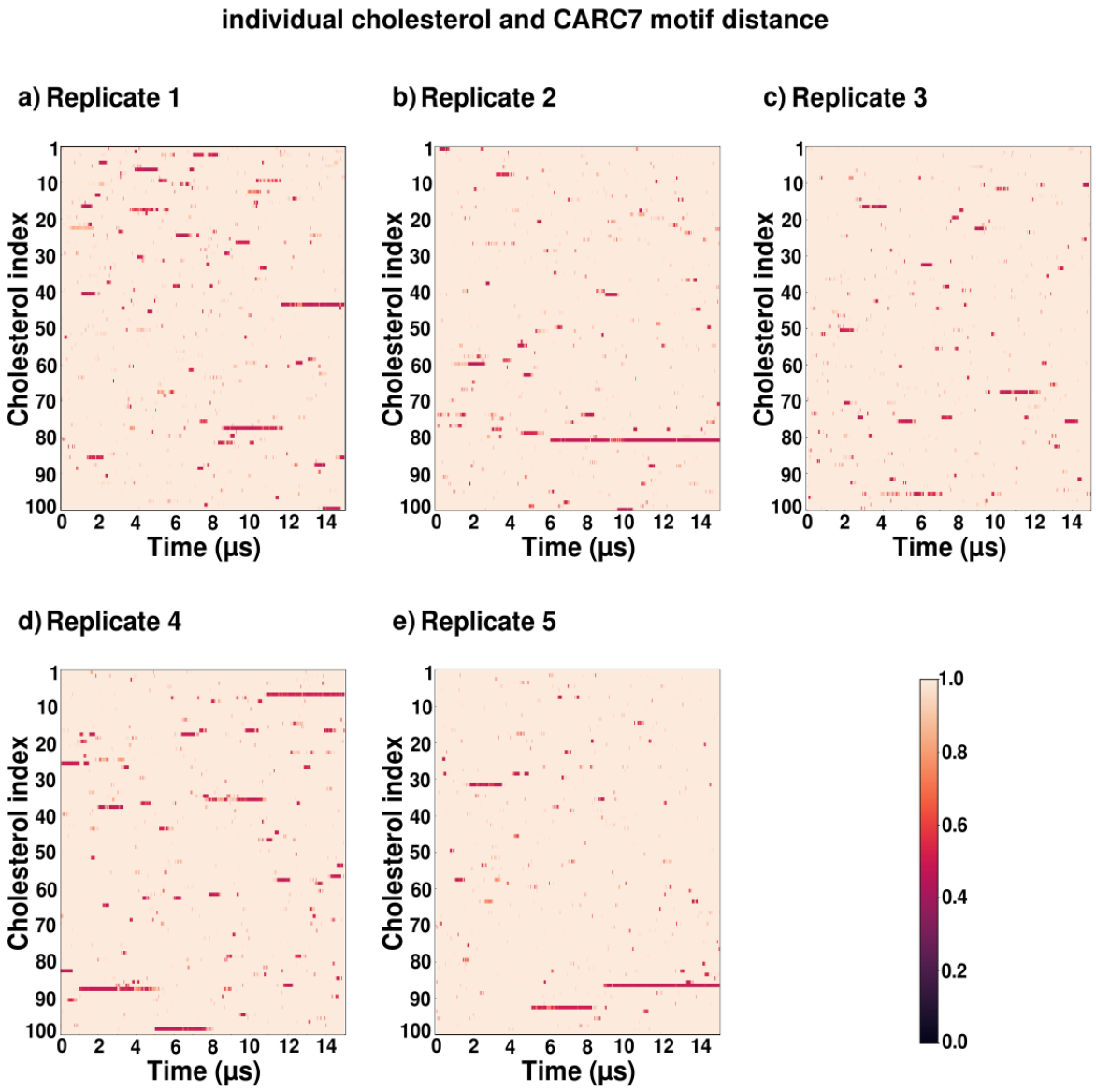


Figure 4.8: Distance between individual membrane cholesterol and the CARC7 motif for all five replicates of SMO in modeled ciliary membrane.

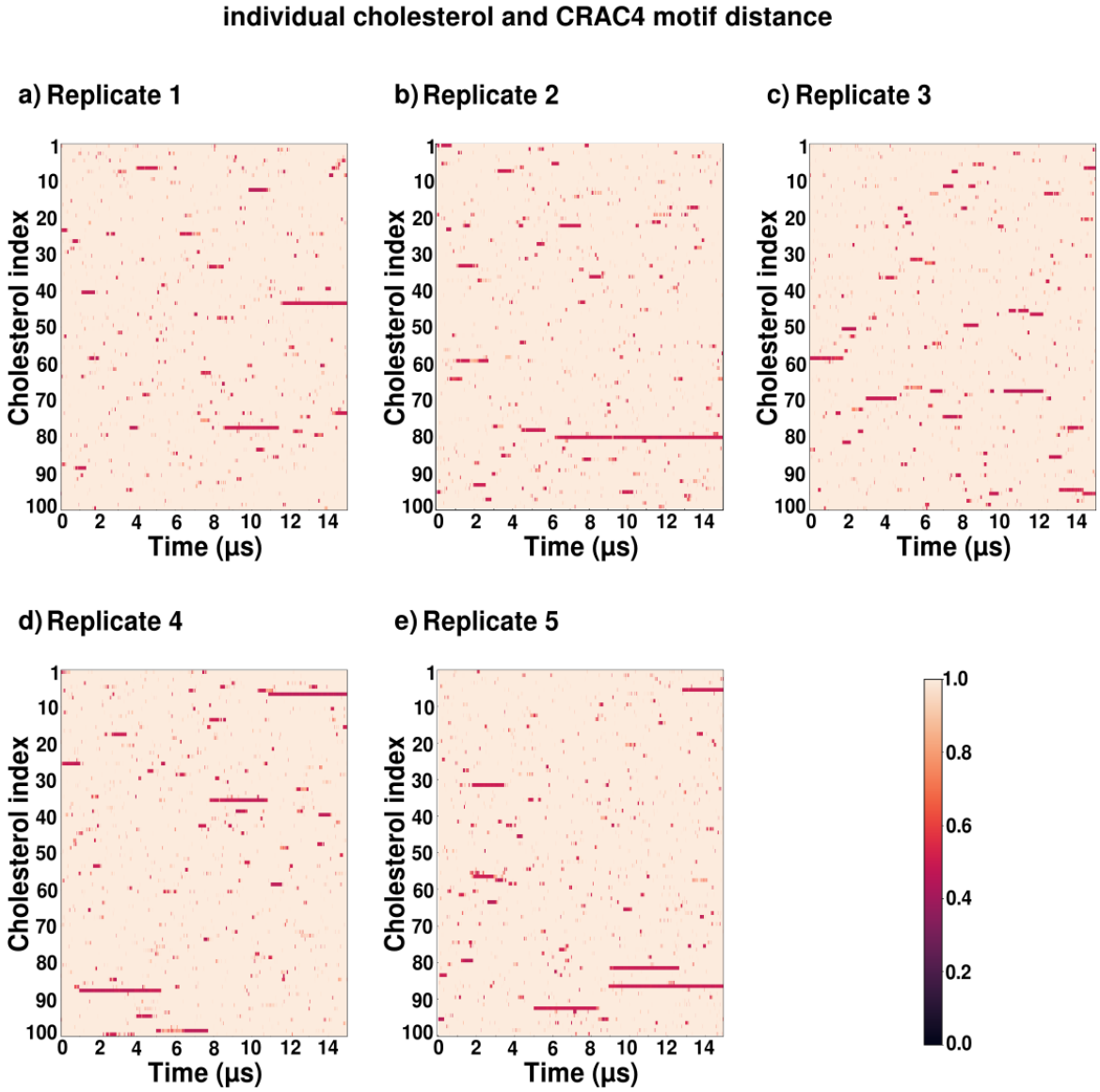


Figure 4.9: Distance between individual membrane cholesterol and the CRAC4 motif for all five replicates of SMO in modeled ciliary membrane.

The CRAC2 motif present in ICL2, preceding CARC7, also showed different extents of cholesterol interactions in different replicate systems. The CRAC4 motif present in TM5, and proximal to CARC7, showed stable interaction with cholesterol molecules, especially with those that interact with CARC7, in all five replicate systems (Figures 4.8, 4.9, and Video 4.3). There were multiple cholesterol interactions with this motif. Another motif, CARC9, present in TM6 and near CRAC4, also interacted with more than one cholesterol molecule simultaneously or at different times. This was observed in two replicates, for reasonable durations, though less than

those of CARC7 interactions. Similarly, CARC8, CARC9, CRAC-like2, and CRAC-like1 also had $> 80\%$ total cholesterol occupancy in all replicates (Figure 4.10). Motif CARC8, present in ECL2 and TM5, showed intermittent interaction with cholesterol molecules in all replicates. However, the interaction distances suggest relatively loose binding.

Taken together, these results suggest that only a few cholesterol-binding motifs are involved in any significant interactions with cholesterol molecules. Based on persistence and proximity considerations, the motifs present in the lower TM region (CARC7-TM4, CRAC4-TM5, CARC9-TM6) have more interaction with cholesterol than other motifs, except CARC8 (ECL2 & TM5), which were present in the upper TM region (Figures 4.9, 4.10 and Video 4.3).

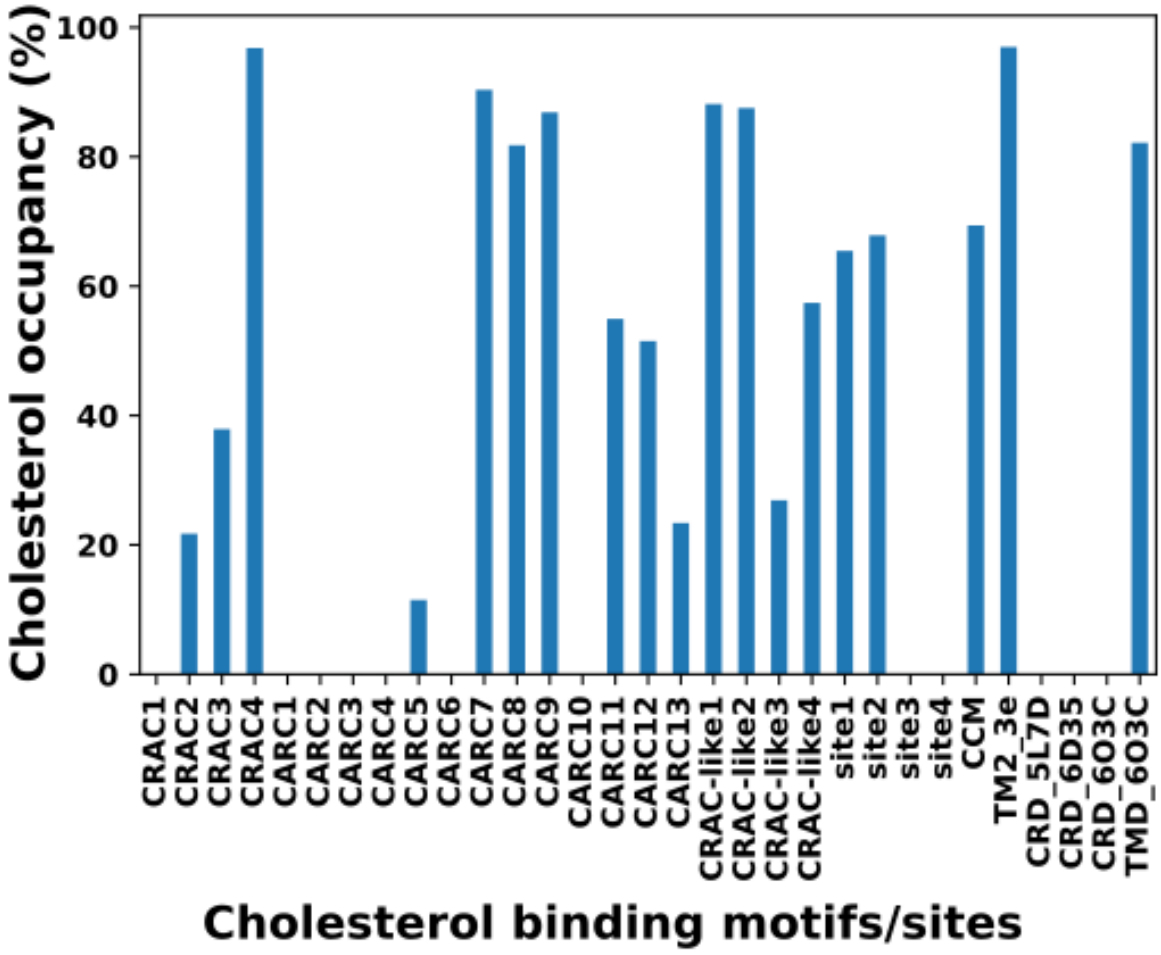


Figure 4.10: Cholesterol occupancy at cholesterol and sterol binding sites and motifs of SMO. The values shown are averaged over five simulations.

Interestingly, due to the inter-motif proximity of some of these motifs present in the TM bundle (Figure 3.3b), it was possible for the same cholesterol molecule to interact with more than one motif simultaneously. This may be one of the reasons for the concentration of cholesterol molecules in these regions. Thus, interactions of cholesterol with CRAC2-ICL2 were not persistent, possibly because of the intracellular location of the motif. Notably, Figures 4.8-4.10 confirmed that the respective cholesterol molecules make stable contacts with motifs present in TM4, TM5, and TM6. For example, cholesterol molecules were found to associate with these respective sites for more than 2 μs. Particularly notable in this context was the 65% occupancy of cholesterol (chol 81) near the CARC7 motif present in TM4 (Figure 4.8b). Given that cholesterol molecules were rarely found to interact persistently, some of these persistent interactions need some special attention, and will be discussed below.

4.4.2.1 The residue-wise interaction pattern of the *strict* CCM in SMO with membrane cholesterol shows variation across replicates

Several cholesterols (chol 78, 51, 67) were observed in the vicinity of the *strict* CCM (discussed in chapter 3) in the simulation trajectories of all replicates, albeit with large variations in their respective interaction patterns. All cholesterols within 0.6 nm of the *strict* CCM were tracked in the simulation trajectories of all five replicates. Figure 4.11 shows the minimum distance plots for any cholesterol molecule with any of the *strict* CCM residues for all five replicates. Figure 4.16 show the interaction of cholesterol with each of the four residues of the *strict* CCM and Y269. While persistent cholesterol– *strict* CCM interactions were observed in replicate 1 and replicate 2, the interactions were intermittent in replicates 3, 4, and 5 (Figure 4.11). Again, Figure 4.12 also highlights the role of K356 in mediating the persistent interaction with cholesterol molecules in replicate 1 and replicate 2. However, the interactions of cholesterol molecules with K356 were intermittent in replicates 3, 4, and 5. Both W365 (Figure 4.13) and the hydrophobic residue of CCM, L362 (Figure 4.14) did show persistent interaction with cholesterol molecules in replicates 1 and 2. However, there were some intermittent interactions with cholesterol in other replicates. These interactions were mostly unstable and showed persistent contact for only a few nanoseconds to a maximum of less than a microsecond during the entire duration of the simulations. Notably, in replicate 1, the interaction of cholesterols with K356, W365, and L362 were persistent for around 1-microsecond duration (Figure 4.12-4.14).

Figure 4.17a shows the details of the interaction of two cholesterols (chol 61 and 78) at time 9.15 μs in the replicate 1 system. The ROH bead of the cholesterols (chol 61 and 78) interacts with the side-chain of K356 and W365. The fourth residue from TM2, Y262, does not seem to contribute much, as the Y262-cholesterol distance value was always greater than 0.6 nm in all replicates (Figure 4.12d). Although Y269 (2.46), another aromatic residue, is also conserved in Class F GPCRs (Figure 3.5b), it is located next to W365 (Figure 4.17b). The cholesterol

interaction with Y269 was also reported, and it was discovered that this residue interacted with cholesterol frequently in all replicates (Figures 4.12e, 4.13b, and Video 4.4).

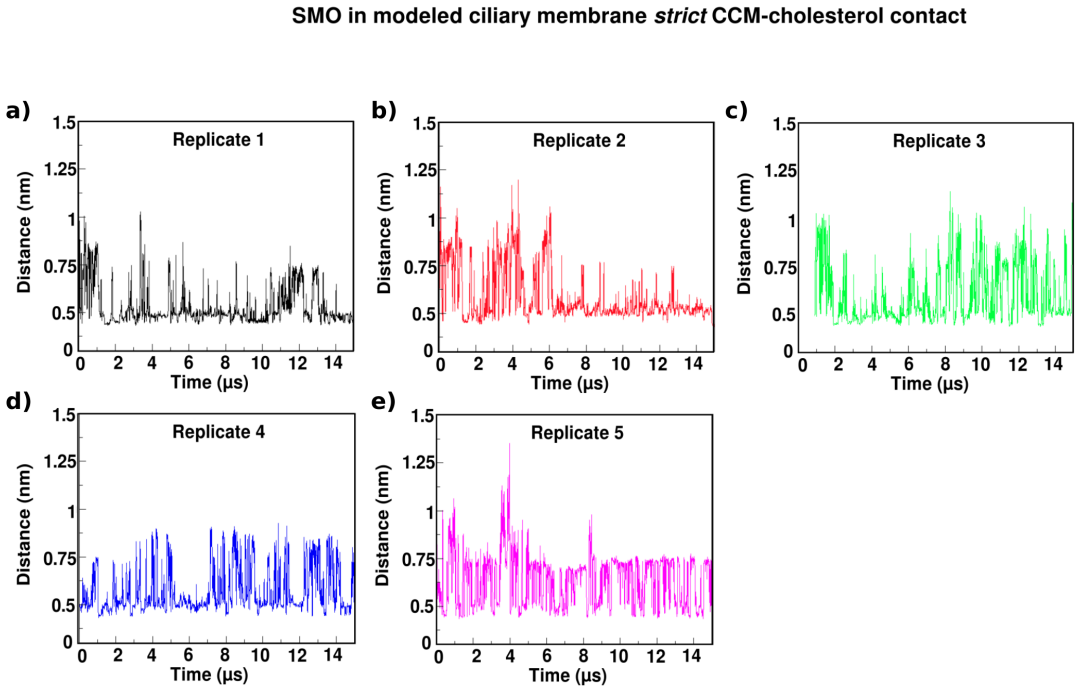


Figure 4.11: The interaction between *strict* CCM and cholesterol. Distance map for all five replicates along the simulation trajectory showing the minimum distance of any membrane cholesterol with any residues of *strict* CCM.

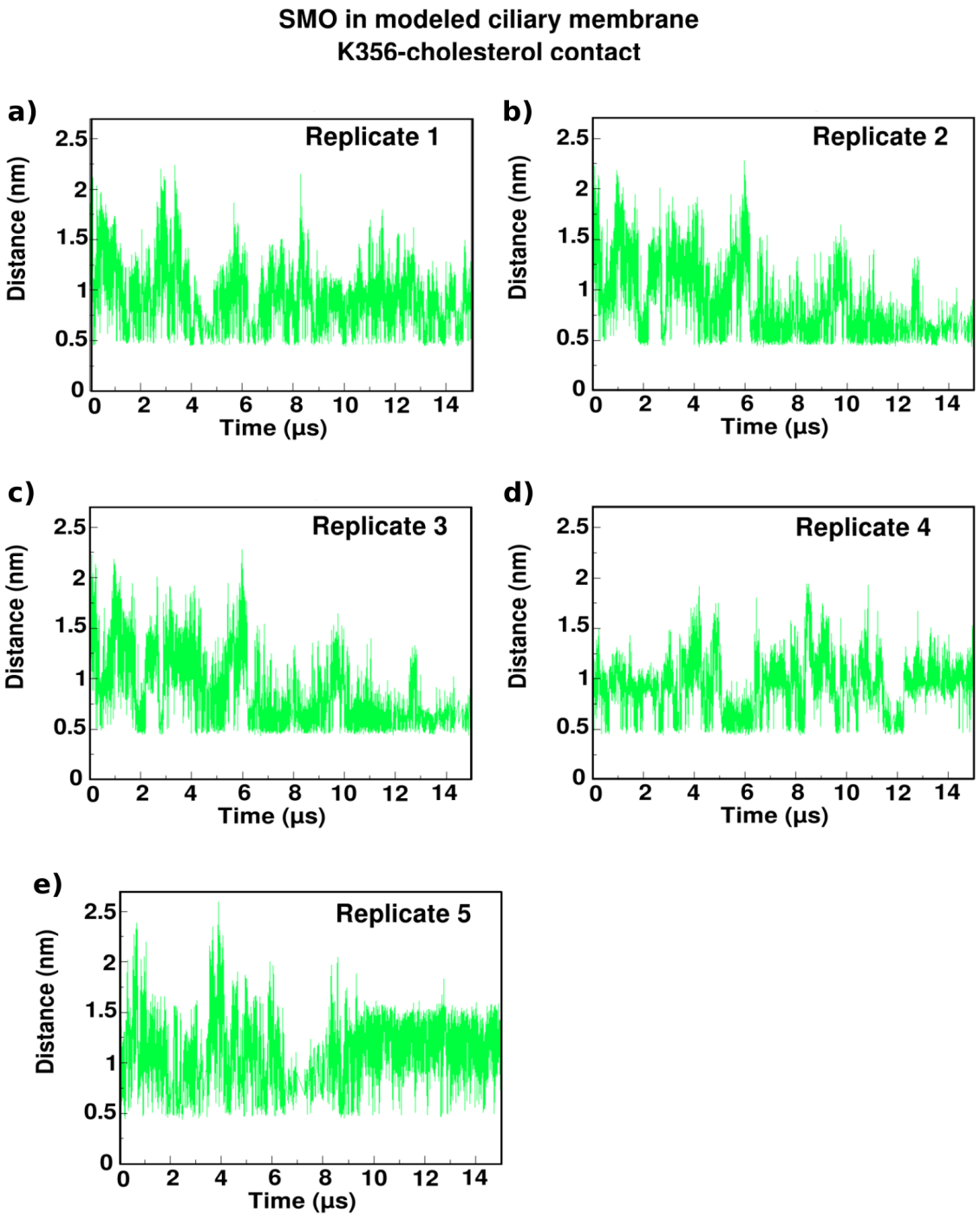


Figure 4.12: The interaction between K356 of *strict* CCM and cholesterol. Distance map for all five replicates along the simulation trajectory showing the minimum distance of any membrane cholesterol with K356.

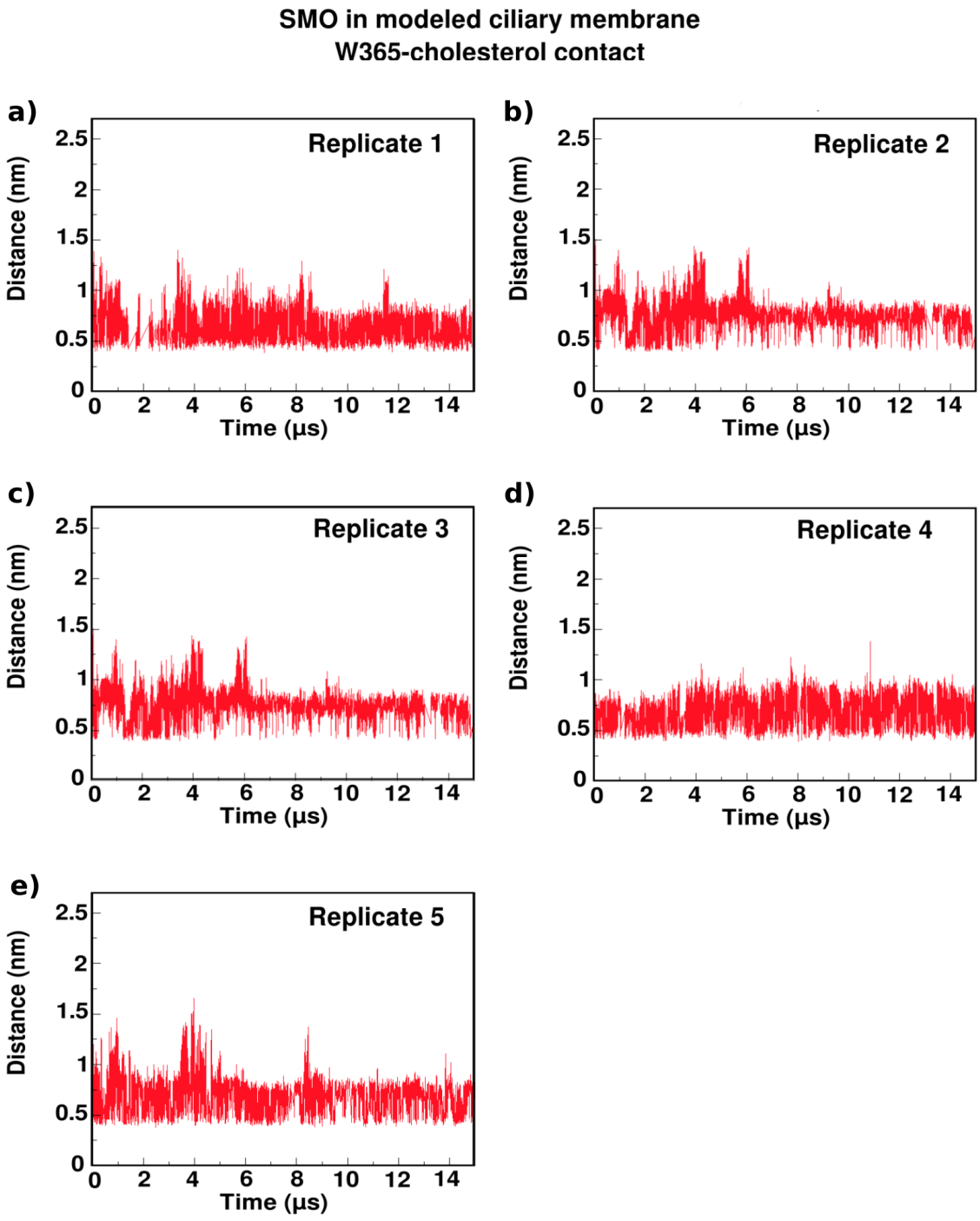


Figure 4.13: The interaction between W365 of *strict* CCM and cholesterol. Distance map for all five replicates along the simulation trajectory showing the minimum distance of any membrane cholesterol with W365.

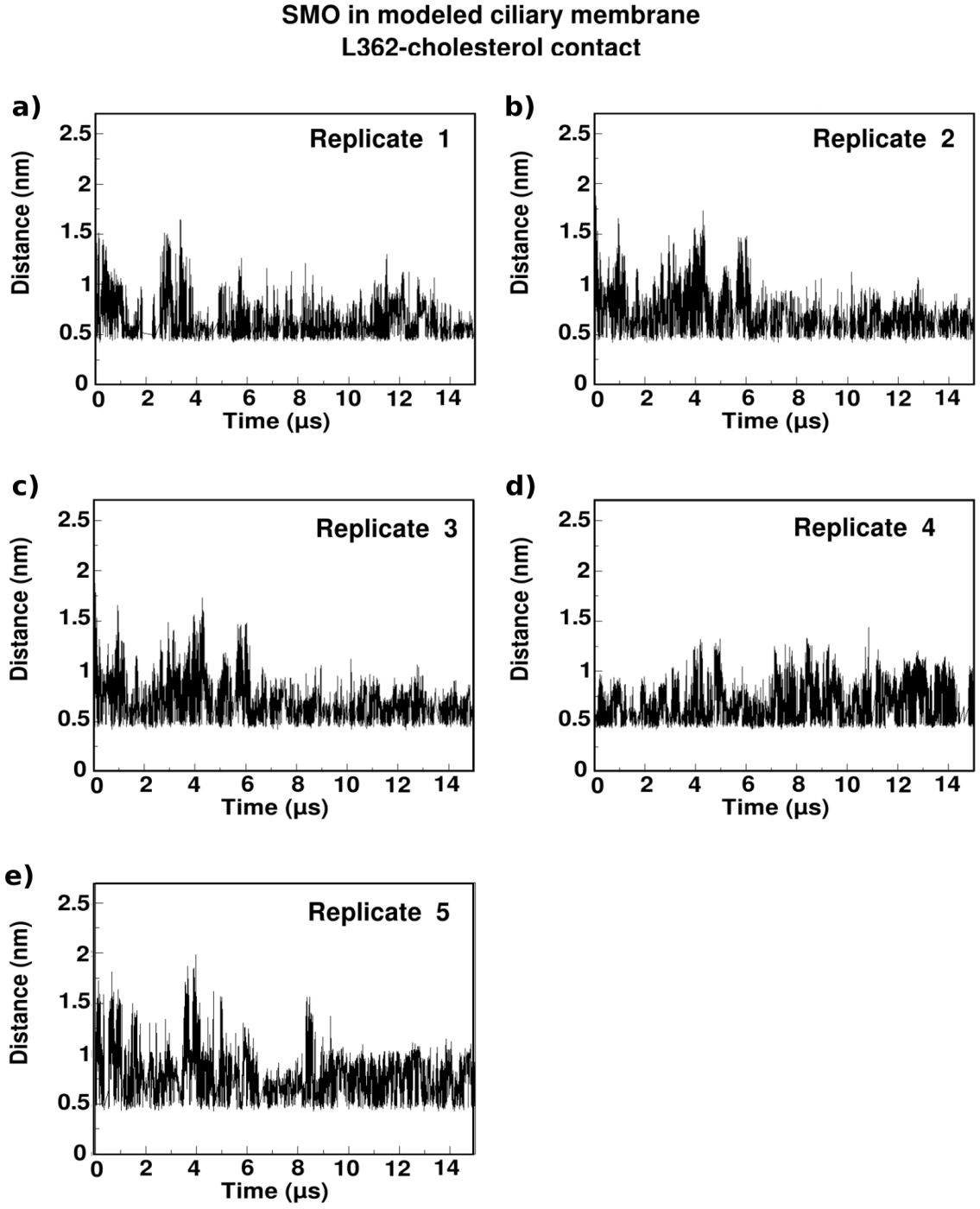


Figure 4.14: The interaction between L362 of *strict* CCM and cholesterol. Distance map for all five replicates along the simulation trajectory showing the minimum distance of any membrane cholesterol with L362.

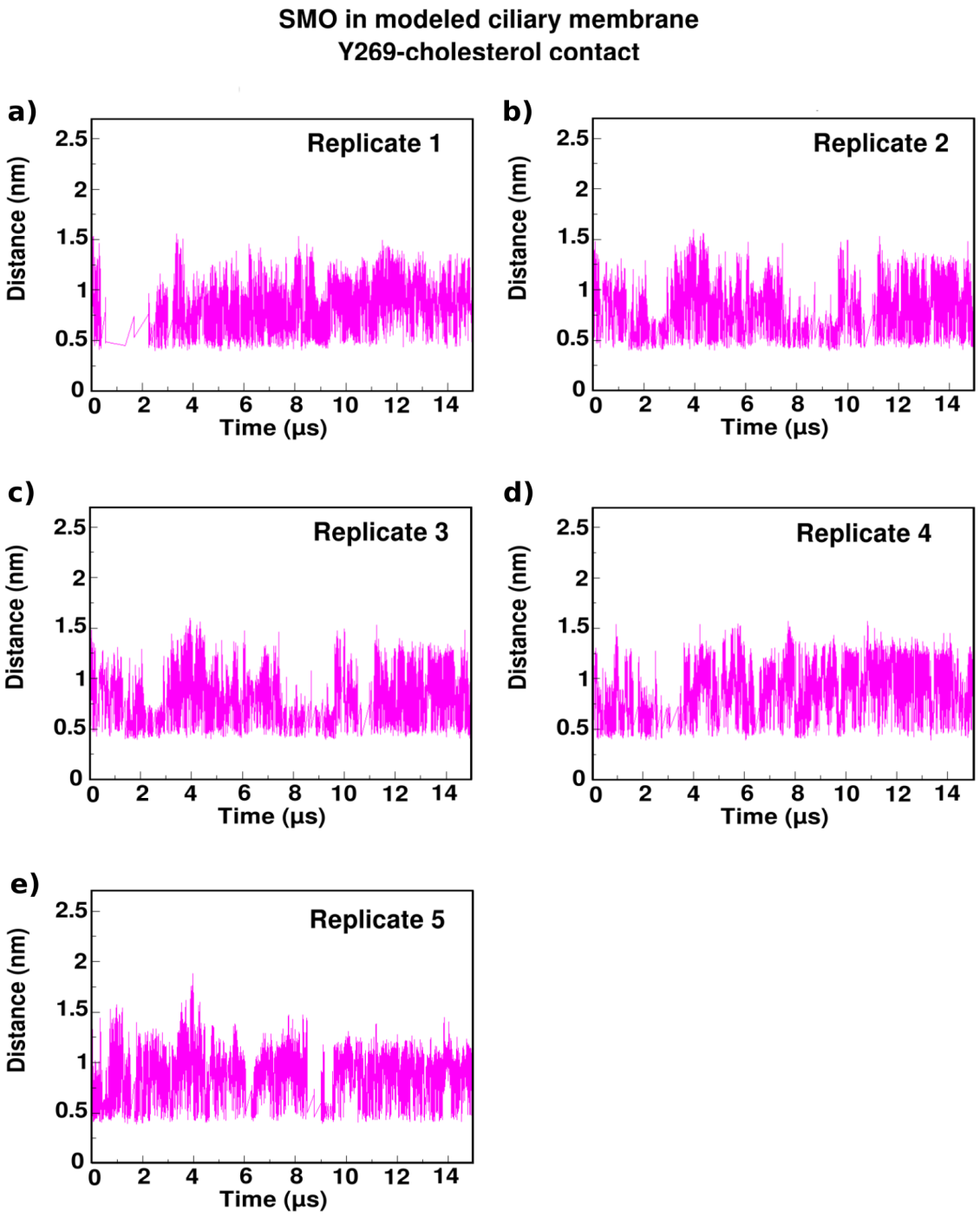


Figure 4.15: The interaction between Y269 of *strict* CCM and cholesterol. Distance map for all five replicates along the simulation trajectory showing the minimum distance of any membrane cholesterol with Y269.

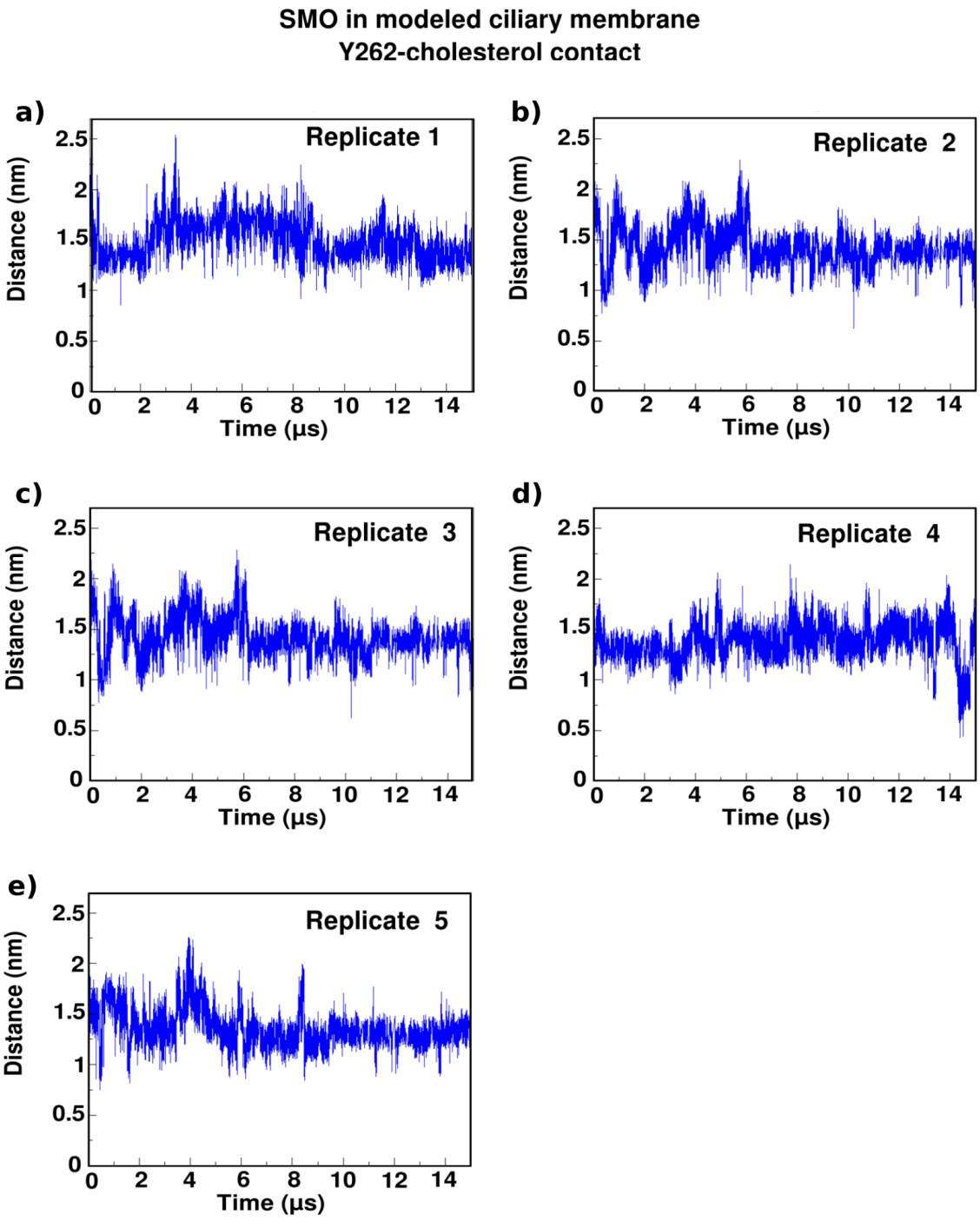


Figure 4.16: The interaction between Y269 of *strict* CCM and cholesterol. Distance map for all five replicates along the simulation trajectory showing the minimum distance of any membrane cholesterol with Y269.

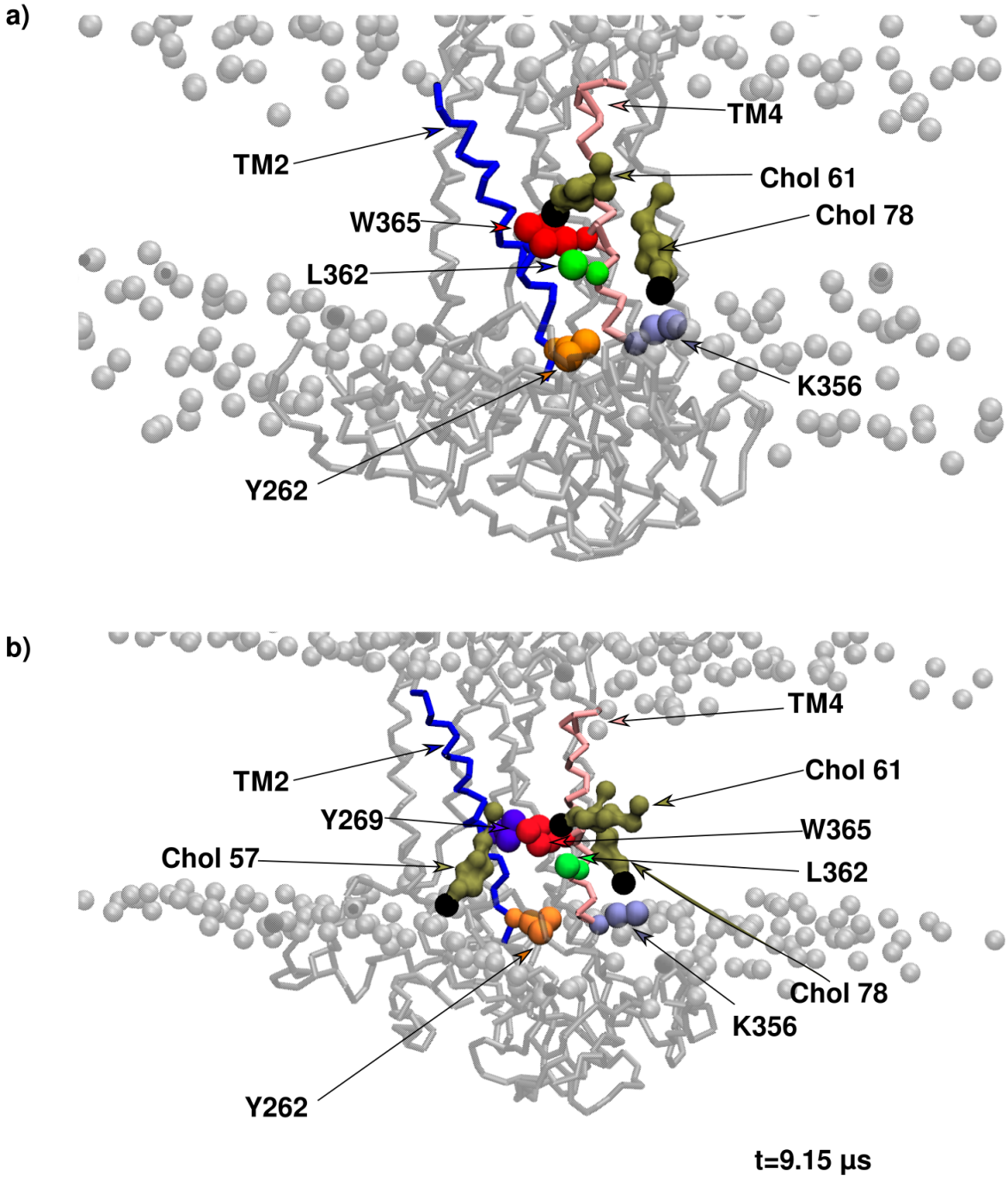


Figure 4.17: A representative snapshot at t=9.15 μs from replicate 1 shows the *strict* CCM-cholesterol interaction. The CG structure shows two cholesterols (chol 78 and 61; tan and the ROH group in black) interacting with *strict* CCM residues K356, L362, and W356. a) The key residues Y262, K356, L362, and W356, colored orange, ice-blue, green, and red. b) Residue Y269 (dark blue), near residue W365, interaction with cholesterol (chol 57).

4.4.2.2 Analysis of PI4P interactions with SMO provides insights into the role of localization of PI4P to primary cilia

Apart from cholesterol, another important lipid class participating in the HH pathway, particularly for SMO activation, is the anionic PI4P lipids that are present in the inner leaflet of the primary cilium. PI4P interaction hotspots were observed on the surface of SMO, where PI4P molecules interact with SMO at the intracellular region of TM helices (TM1, TM3, TM4, and TM5), ICL1, helix 8, and ICD (Figure 4.18a). Further, the conserved motif of SMO "KATXXXW" (residues 539-545) present in helix 8 interacts with PI4P lipids. PI4P lipids were also seen to form strong contacts with R546 and R547 with high occupancy. Furthermore, SMO has five Arg/Lys clusters in the ICD: 561-567 (cluster 1), 571-576 (cluster 2), 624-629 (cluster 3), 667-680 (cluster 4), and 703-711 (cluster 5) (Figure 3.9; chapter 3). All five clusters showed higher occupancy for PI4P lipids. Cluster 4 has a contiguous stretch of ten basic residues. C1, C2, C3, and headgroup phosphate beads of PI4P interact with the sidechain beads of Arg and Lys residues. Most of the PI4P lipids accumulate near the ICD region. Interestingly, in keeping with an earlier report [35], that W549 and R550 of mouse SMO are essential for its localization in the primary cilia, these two residues, W549 and R551 of human SMO, were found to interact with PI4P in the present system (Figures 4.19 and 4.20). PI4P lipids interacted with W549 transiently in replicates 1 and 2 (after 4.3 μs), but persistently in replicates 3, 4, and 5 (Figure 4.19). At the same time, R551 showed persistent interactions in all replicates (Figure 4.20). A representative structure at time 13.56 μs showed that the side chains of W549 and R551 interact with the glycerol backbone beads of two PI4P lipids (index 9 and 15) (Figure 4.18b). These observations suggest that the accumulation of anionic PI4P molecules near Arg/Lys clusters holds an important key to its ciliary role. For example, the high occupancy of PI4P near the fourth and fifth Arg/Lys clusters in the ICD helps ICD to interact with the membrane surface. This also helps in mediating strong cholesterol-TMD interactions near the lower leaflet of the membrane (will be discussed later in the discussion section).

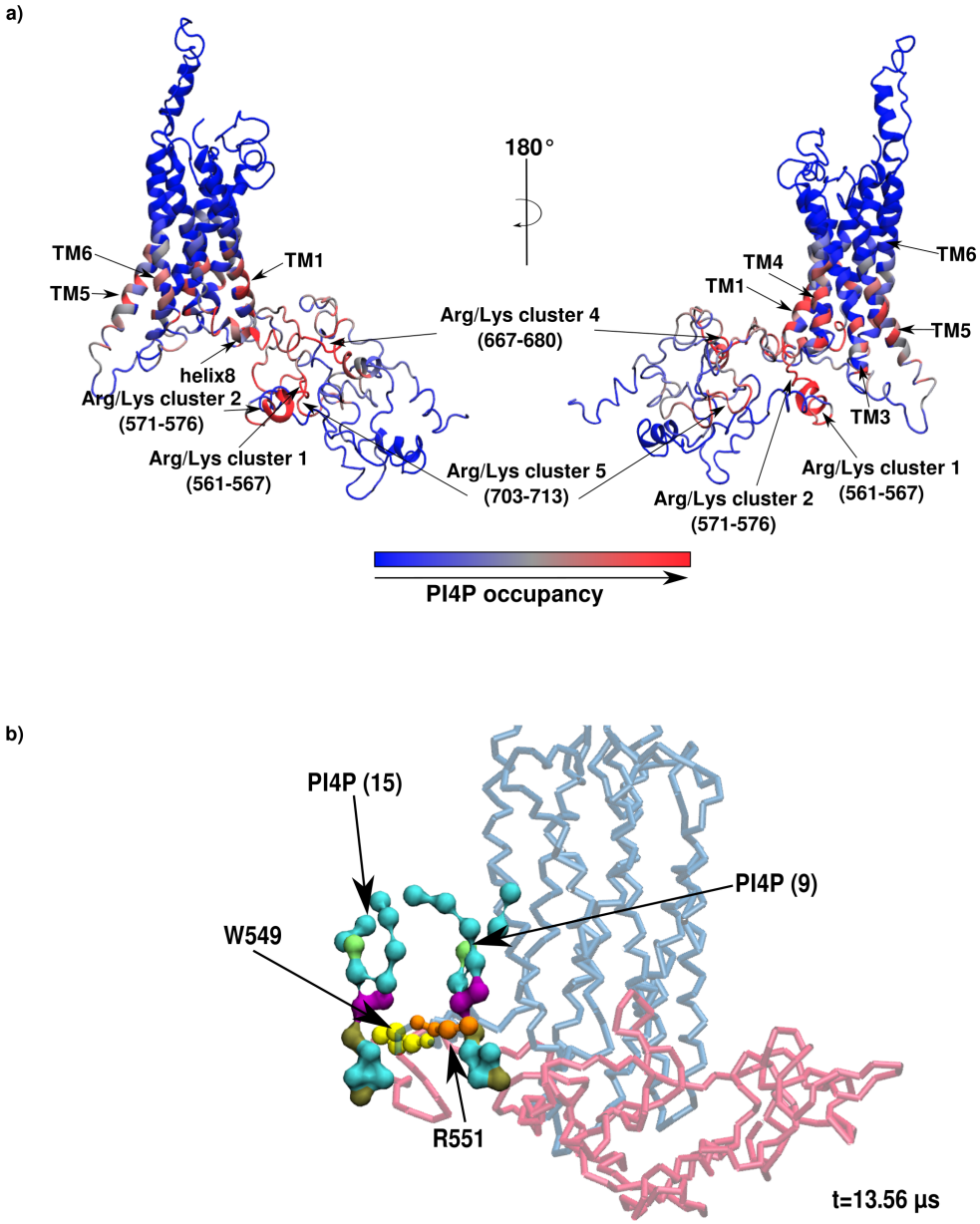


Figure 4.18: a) PI4P occupancy around SMO TMD was mapped onto the structure, with each residue colored from green (less occupancy) to red (high occupancy). TM helices are marked as TM1, TM2, TM4, TM5, TM6, TM7, and helix 8. The mean occupancy was calculated across all replicates, using a 0.6 nm distance cutoff. b) The interaction of PI4P with W549 and R551. The CG structure of 13.56 μs from replicate 1 shows the interaction of 9, 11, and 15 PI4P lipids with W549 and R551. The color scheme is as follows: SMO TMD (blue), ICD (red), W549 (yellow), and R551 (orange), hydrocarbon chains of PI4P lipids (cyan), glycerol backbone (purple), and headgroup phosphate (PO4) beads (tan).

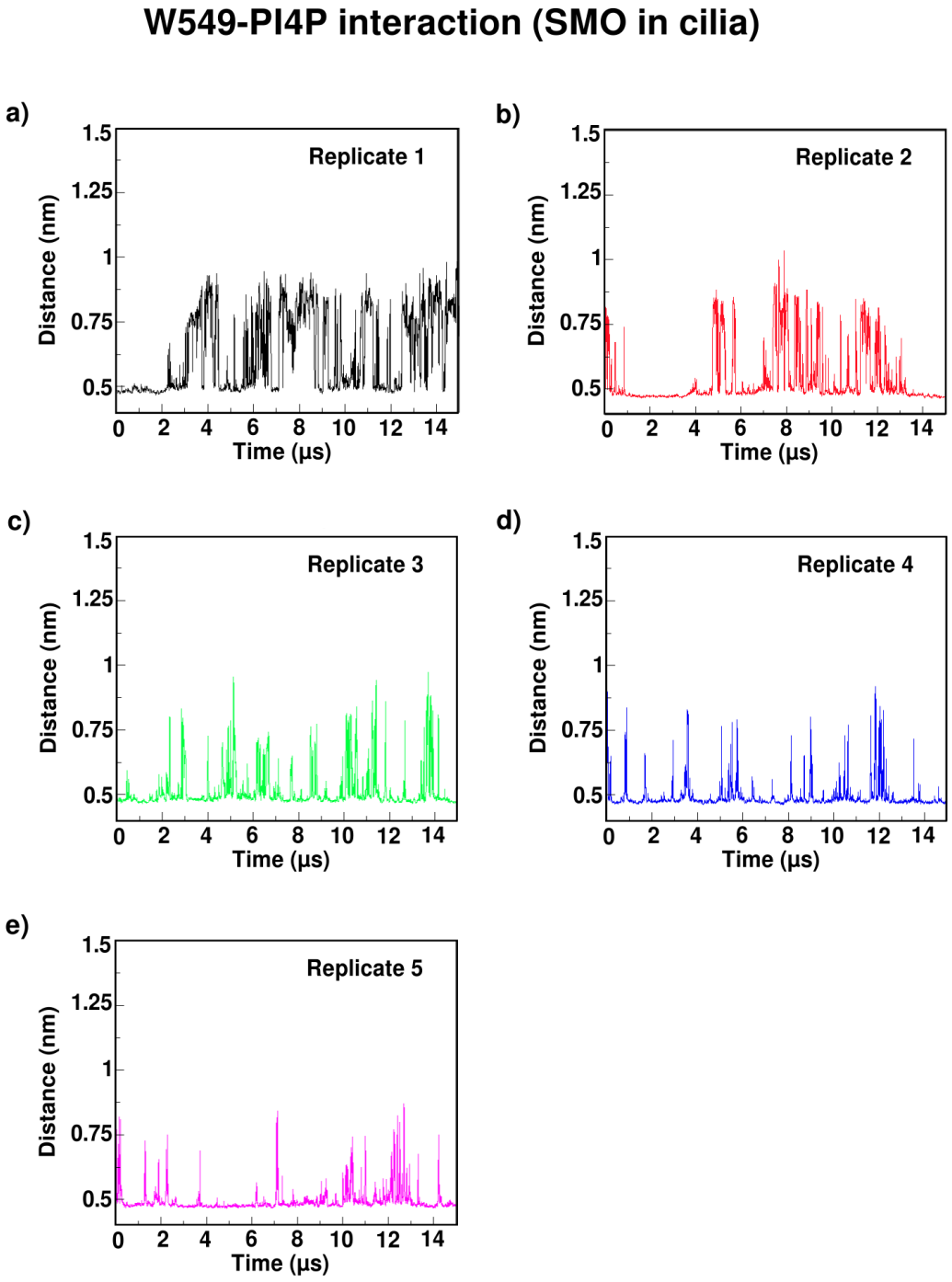


Figure 4.19: W549 and PI4P interactions. Distance map for all five replicates (a-e) along the simulation trajectory showing the minimum distance of any PI4P lipid with W549.

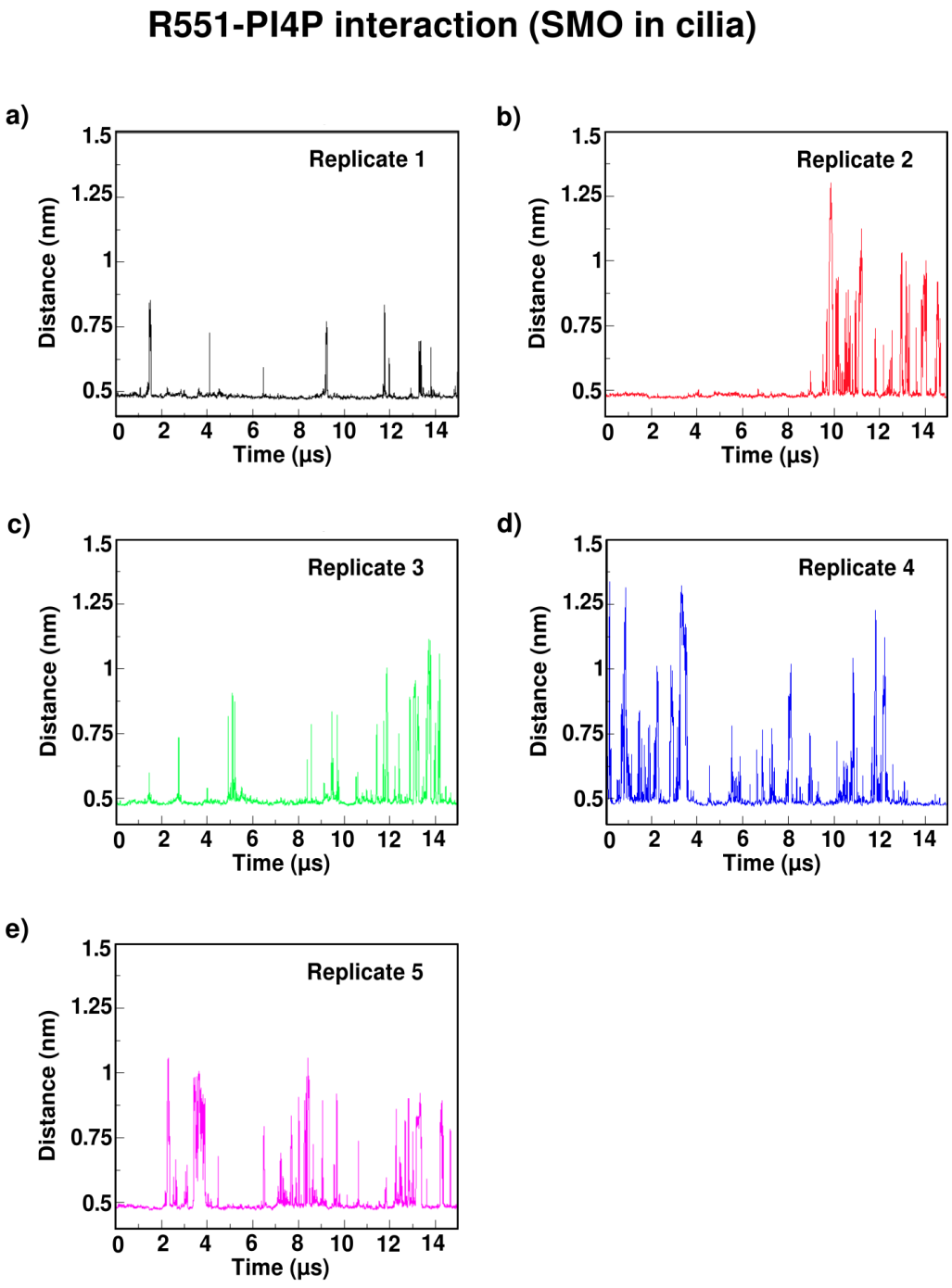


Figure 4.20: R551 and PI4P interactions. Distance map for all five replicates (a-e) along the simulation trajectory showing the minimum distance of any PI4P lipid with R551.

4.4.2.3 Translational dynamics of lipids in the ciliary membrane

The translational dynamics of the lipids were also examined, especially cholesterol, in the ciliary membrane. There are 50 cholesterol molecules to begin with, in each of the two leaflets of the ciliary membranes (with or without SMO; see Table 4.1 for more details) respectively. The general observation was that as the simulation progressed, some cholesterol molecules moved from the inner to the outer leaflet. Eventually, the cholesterol concentration in the outer leaflet became higher than that of the inner leaflet during the simulation of the ciliary membrane, both in the presence and absence of embedded SMO. Interestingly, though only very occasionally, the cholesterol concentration goes higher in the inner leaflet during the simulation in the presence of SMO. In the absence of SMO, the ciliary membrane cholesterol concentration went close to 50-50 (number) in some frames, but the outer leaflet always showed a high cholesterol concentration (Table 4.3 and Figure 4.21). This could be a result of the cholesterol-binding motifs found in SMO TMD, which draw in cholesterol molecules. PI4P lipids were seen to accumulate near the ICD and lower TM regions of SMO. A few POPS lipids were seen to interact with the lower region of TM5. However, SMO selectivity was higher for PI4P than POPS. Some cholesterol molecules were sandwiched between PI4P lipids in the cholesterol-binding motifs present in TM4, TM5, and TM6, which mediate persistent interactions with the cholesterol molecules. These PI4P lipids interacted with the cholesterol molecules through their phospholipid tails (Figure 4.22).

Table 4.3: Average cholesterol concentration (in numbers) during simulation.

Leaflet position	**SMO in modeled ciliary membrane**				
	Replicate 1	Replicate 2	Replicate 3	Replicate 4	Replicate 5
Outer	56	56	57	56	57
Inner	45	45	44	44	44
Leaflet position	**Modeled ciliary membrane (only membrane system)**				
	Replicate 1	Replicate 2	Replicate 3	-	-
Outer	56	56	56	-	-
Inner	44	44	44	-	-

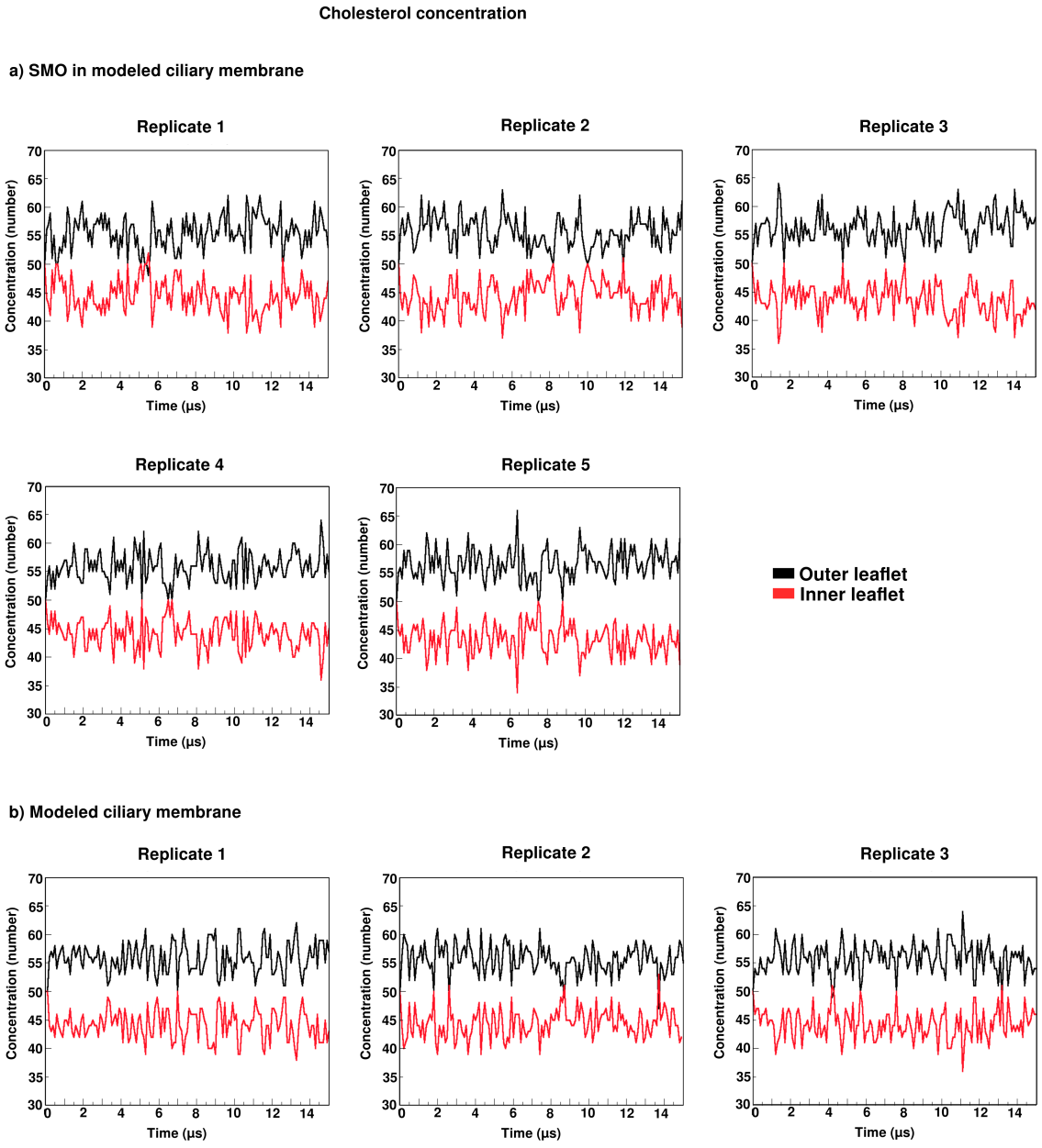


Figure 4.21: Dynamics of cholesterol in ciliary membrane. Cholesterol concentration in the outer and inner leaflets within the ciliary bilayer a) in the presence of SMO for all five replicate systems and b) in the absence of SMO for all three replicate systems.

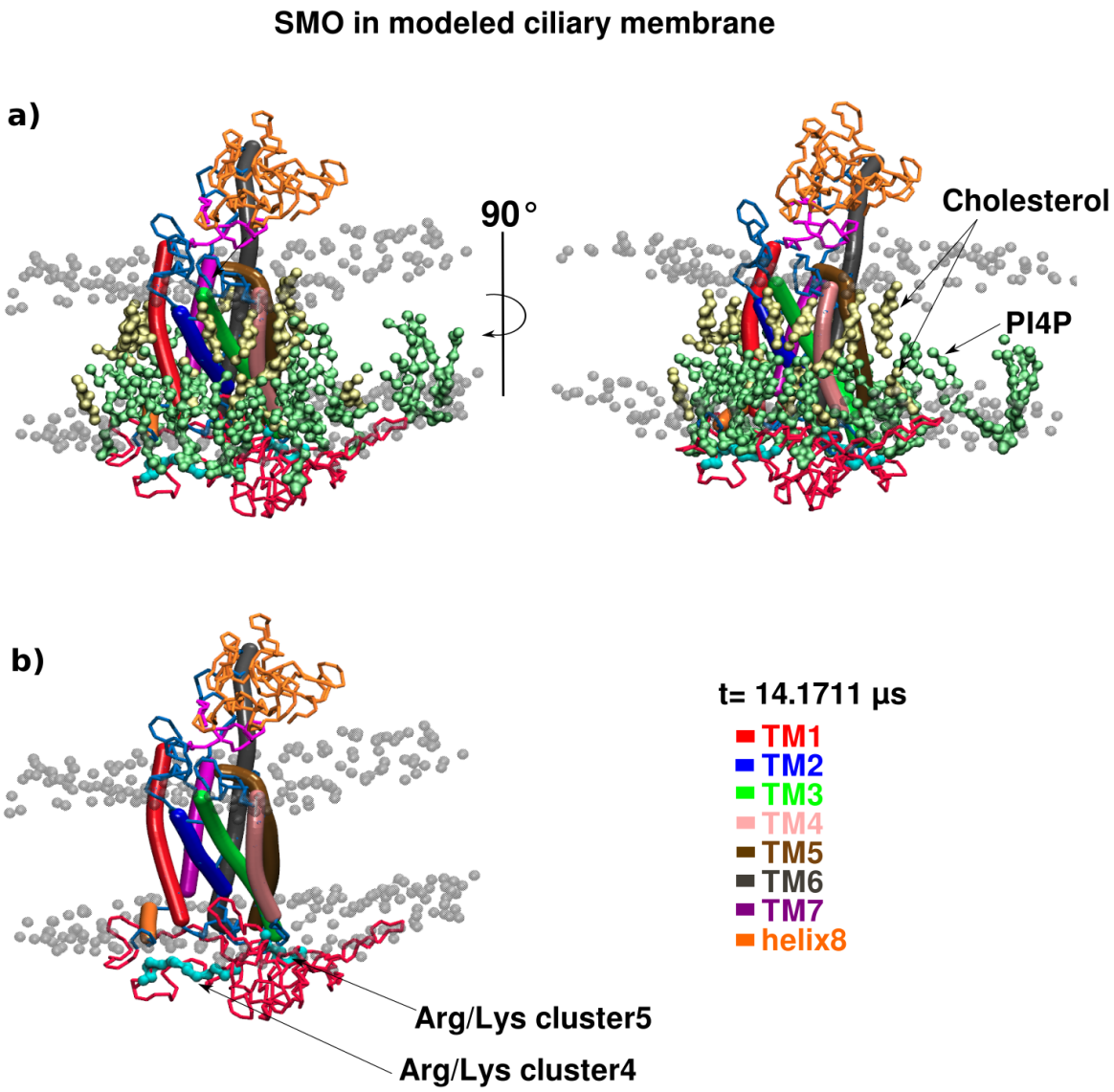


Figure 4.22: Interaction of PI4P-cholesterol together with SMO and sequestering of cholesterol molecules. a) Shows the interaction of cholesterol and PI4P with SMO TMD and ICD at t = 14.17 μs; b) Same representation without cholesterols and PI4P lipids. The color scheme is as follows: CRD (orange), LD (magenta), ICD (red), Arg/Lys cluster 4 and Arg/Lys cluster 5 (cyan), all ICLs and ECLs (ice-blue).

4.4.2.4 Apart from the increase in compactness, no major structural changes were observed during the simulation of SMO

Domain-wise SMO dynamics were also explored after the lipid dynamics. The CRD and ICD were dynamic regions (RMSD and RMSF plots; Figures 4.23 and 4.24, respectively), whose movements lead to increased compactness of SMO during simulation. LD also showed some movements, but less as compared to CRD and ICD. TM helices were relatively less dynamic than the non-helical (CRD, LD and ICD) regions of SMO. Though the overall RMSD for the TMD does not change significantly, some of the replicates showed some fluctuations in the helices TM1, TM2, TM4, TM6, and TM7 RMSD values with respect to their starting structures (Figure 4.25). This was also supported by the helix tilt analysis (Figure 4.26). In all replicates, greater fluctuations were observed in the loops ECL1, ECL3, and ICL3 (relatively less than ECLs but higher than other ICLs).

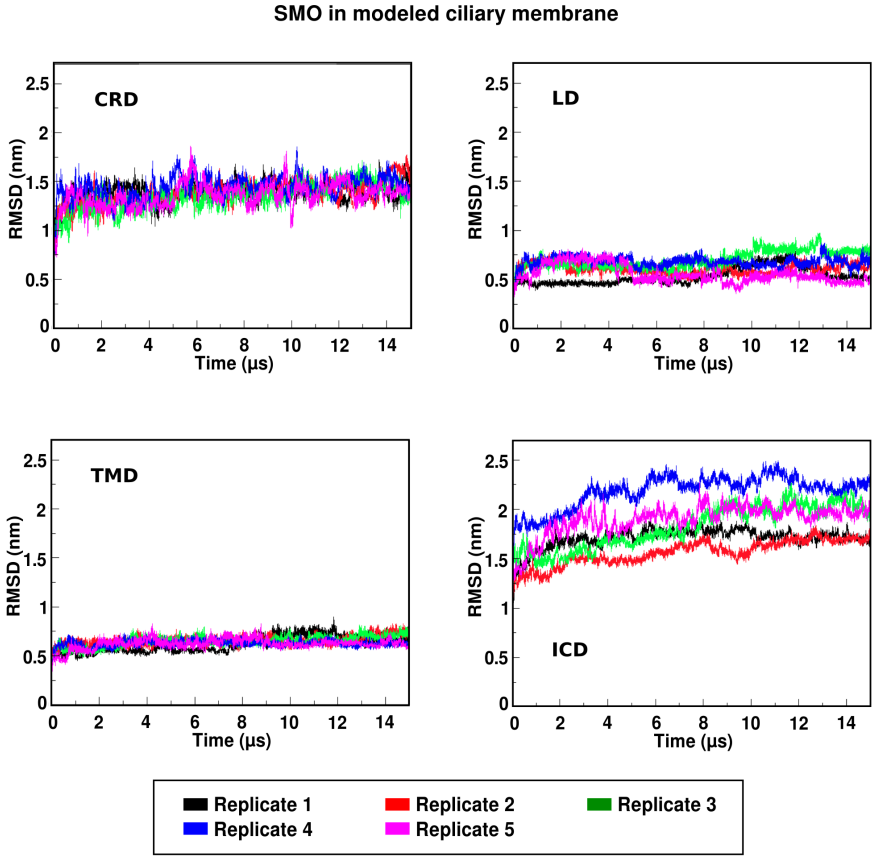


Figure 4.23: Domain-wise RMSD plots of SMO (CRD, LD, TMD, and ICD) with respect to their starting structures in the modeled ciliary membrane.

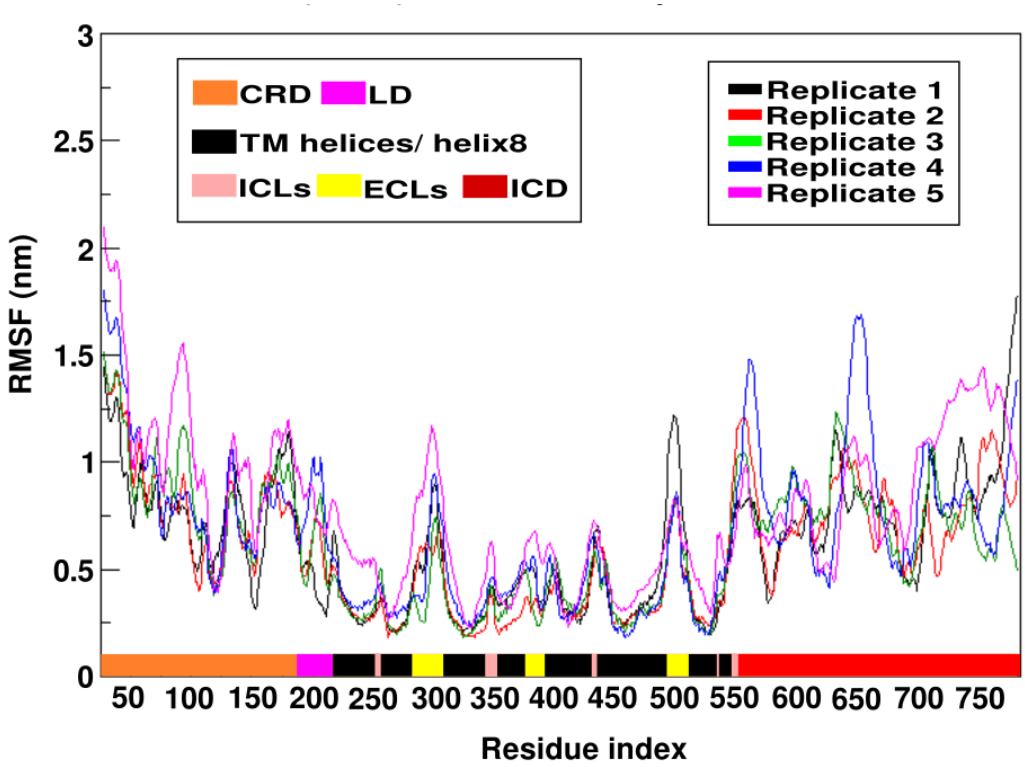


Figure 4.24: RMSF for the backbone beads of the SMO in modeled ciliary membrane. Structural domains are indicated by the color bar as per the scheme CRD in orange, LD in magenta, TM helices in black, ECLs in yellow, ICLs in pink and ICD in red.

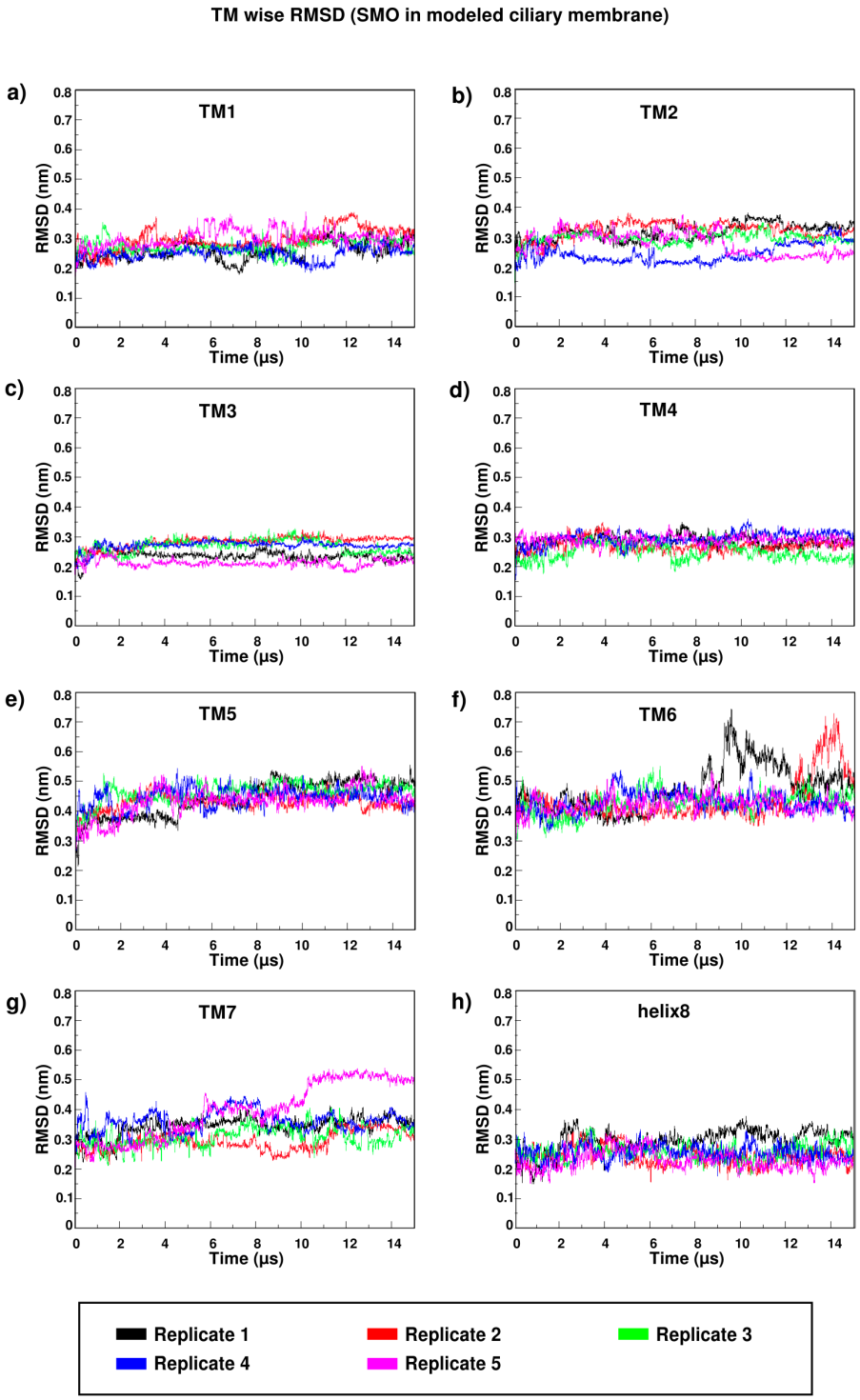


Figure 4.25: a-h) TM-wise RMSD plots of SMO (TM1, TM2, TM3, TM4, TM5, TM6, TM7, and helix 8) with respect to their starting structure in modeled ciliary membrane.

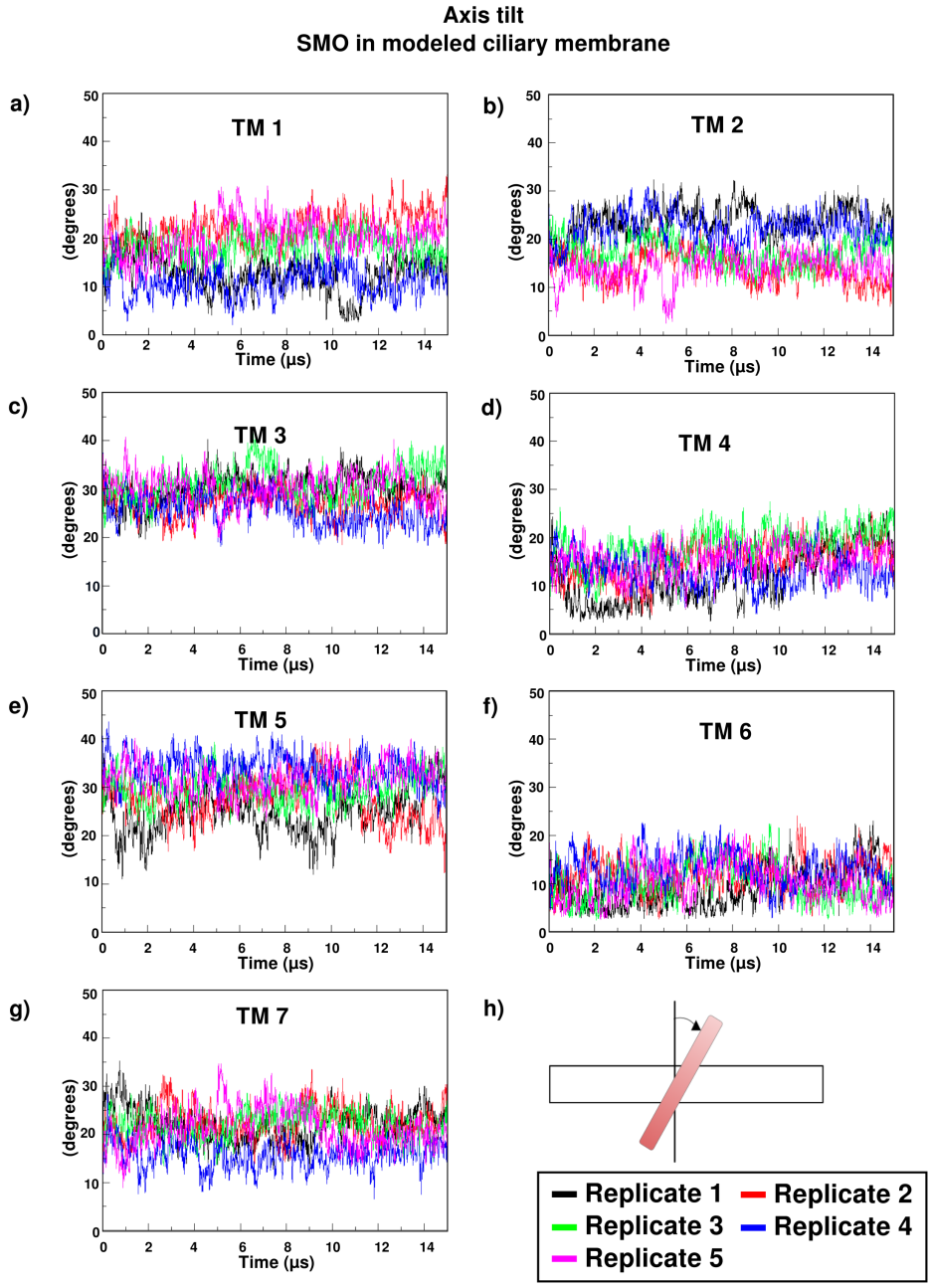


Figure 4.26: Axis tilts of 7 TM helices. a-g) Tilts of the individual 7 TM helices with respect to the z-axis were plotted for the entire simulation length for all replicates of SMO in modeled ciliary membrane, and h) a schematic representation showing axis tilt calculation for TM helices.

4.4.2.5 Conformational changes in SMO involve changes in the tilt angles of the CRD and ICD with respect to the TMD

Analysis of RMSD and RMSF, and visual examination using VMD, revealed the movements in CRD and ICD regions. In conjunction with earlier reports, the observations support the assumption that the membrane lipids (cholesterol and PI4P) strongly influence the conformational dynamics of both CRD and ICD. The CRD showed a far greater conformational diversity in the presence of these lipids, compared to that in the simulation in a simple POPC membrane. The tilt angles of CRD and ICD relative to TMD were calculated to measure conformational changes. CRD tilt is defined by the angle between the backbone beads of P69, V210, and W535 (Figure 4.27a). ICD tilt is defined by the angle between the backbone beads of V210, V536, and E650 (Figure 4.27b). Figures 4.27c-d showed the CRD and ICD tilts, respectively, in all replicates of SMO in the modeled ciliary membrane. It was observed that while the tilt angle of ICD varied with respect to TMD in different replicates, the CRD tilt values indicated the movement of CRD towards the TM bundle during the simulation in all replicates.

4.4.2.6 DCCM and PCA reveal several distinctive features of overall SMO dynamics in the ciliary membrane

In order to understand the correlated movements in SMO, and to identify the roles of the membrane lipids cholesterol and PI4P, the DCCM analysis (Figure 4.28) and PCA (Figures 4.29 and 4.30) were performed on the concatenated trajectories of both SMO in modeled ciliary membrane (67.5 μs) and SMO in POPC (40.5 μs) systems, respectively. The DCCM of SMO in the two-membrane condition, representing correlation coefficients calculated as a time average over the duration of the simulation, is shown in Figure 4.28. The entire spectrum of correlation ranges from +1 to -1. Blue represents a positive correlation, white represents no correlation, and red represents a negative correlation. The significant changes in the ciliary membrane (Figure 4.28a) and POPC membrane (Figure 4.28b) are marked with boxed regions 1 to 16. Figure 4.28 clearly shows different patterns of blue islands, indicating positively correlated fluctuations of SMO domains implicated in the presence of the ciliary lipids. A closer examination, however, suggests several notable features that indicate the changes associated with the presence of ciliary lipids and consequent stabilization of the SMO.

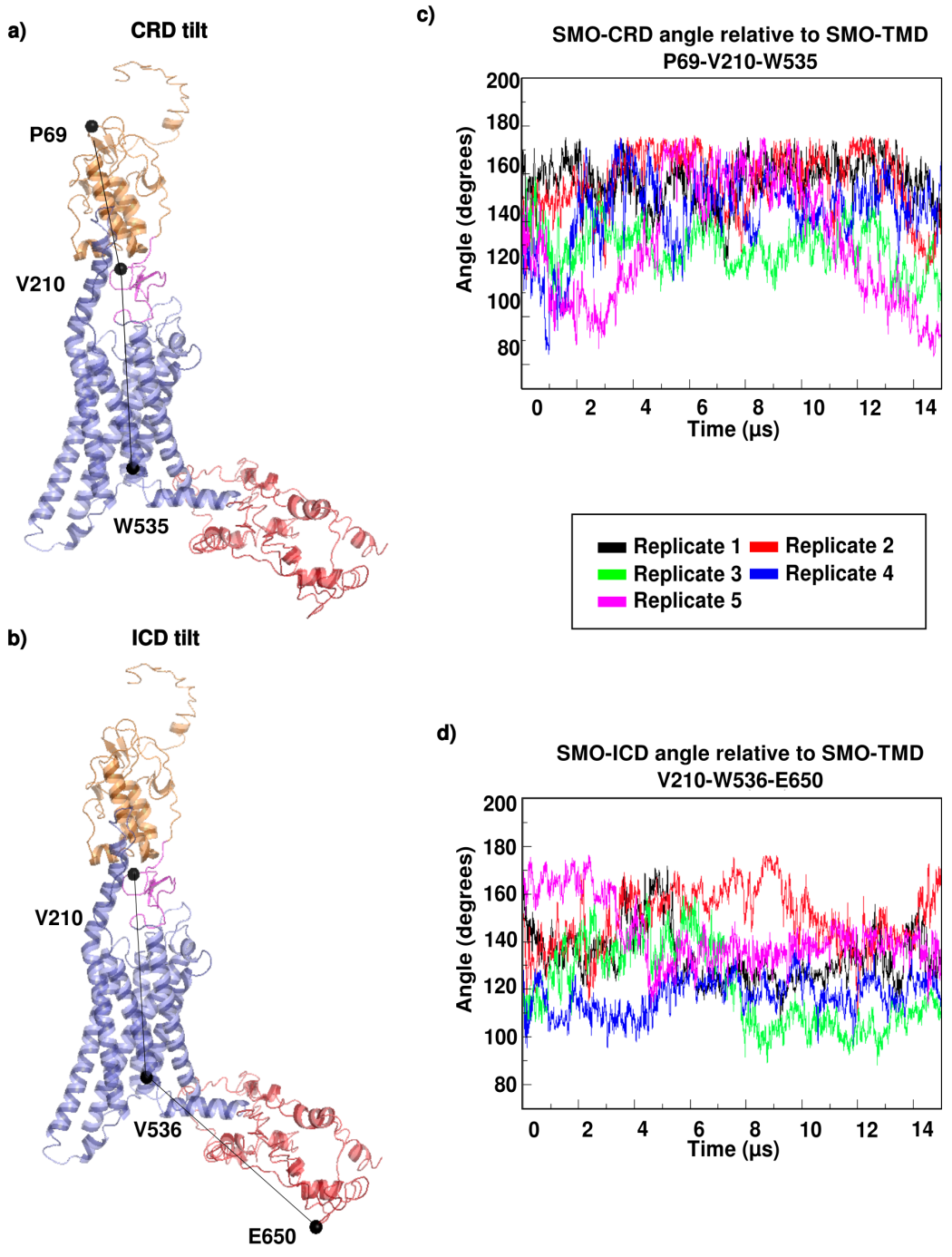


Figure 4.27: CRD and ICD tilt with respect to TMD. a) CRD angle defined by the backbone beads of P69, V210, and W535 residues, and b) ICD angle defined by the backbone beads of V210, V536, and E650 residues; c) CRD tilt and d) ICD tilt in all replicate systems of SMO in modeled ciliary membrane.

The following features resulting from the interactions of SMO with cholesterol and PI4P in the modeled ciliary membrane were revealed by the maps. Box1 showed an increased in self-correlation of CRD in the presence of cholesterol and other ciliary lipids in the ciliary membrane as compared to the POPC membrane. Boxes 2 and 3 demonstrated an increased correlation of LD with CRD and TMD in the ciliary membrane compared to the POPC membrane. Box 4 showed that the negative correlation between Arg/Lys cluster 4 and a part of C-terminal residues of CRD in the ciliary membrane was greater than that in the POPC membrane. Box 5 showed an increase in negative correlation between ICD C-terminal residues (684 to 787) and CRD. Box 6 highlighted an increase in negative correlation of TM1 with Arg/Lys cluster 3 in the presence of the ciliary lipids. Box 7 showed that the negative correlation between TM4 and ECL1 in the ciliary membrane was greater than in the POPC membrane. Box 8 showed an increase in negative correlation of TM4 with TM2-TM3 including ICL1 in ciliary membrane. Box 9 showed that the TM4 correlation with TM4-TM6 was more positive in the ciliary membrane, likely due to the interaction of cholesterols with these TM helices. Box 10 showed that the TM2 correlation became positive in the presence of cholesterol and ciliary lipids. Box 11 depicted an increase in negative correlation between ECL1 and Arg/Lys clusters 1 and 2 in the ciliary membrane. Box 12 showed that the part of ICD (residues ∼576 to 629) became more correlated in the presence of cholesterol and other ciliary lipids. Box 13 depicted an increase in negative correlation of Arg/Lys cluster with TM4 and TM5. Box 14 showed an increase in positive correlation of ECL2 with C-terminal residues of ICD (∼711 to 787). Box 15 depicted a weak positive correlation of Arg/Lys 5 with TM6 and TM7 in the POPC membrane but a stronger positive correlation in the ciliary membrane. Box 16 showed that residues 672 to 711 and 561 to 576 in ICD became positively correlated in the presence of cholesterol and PI4P in the ciliary membrane.

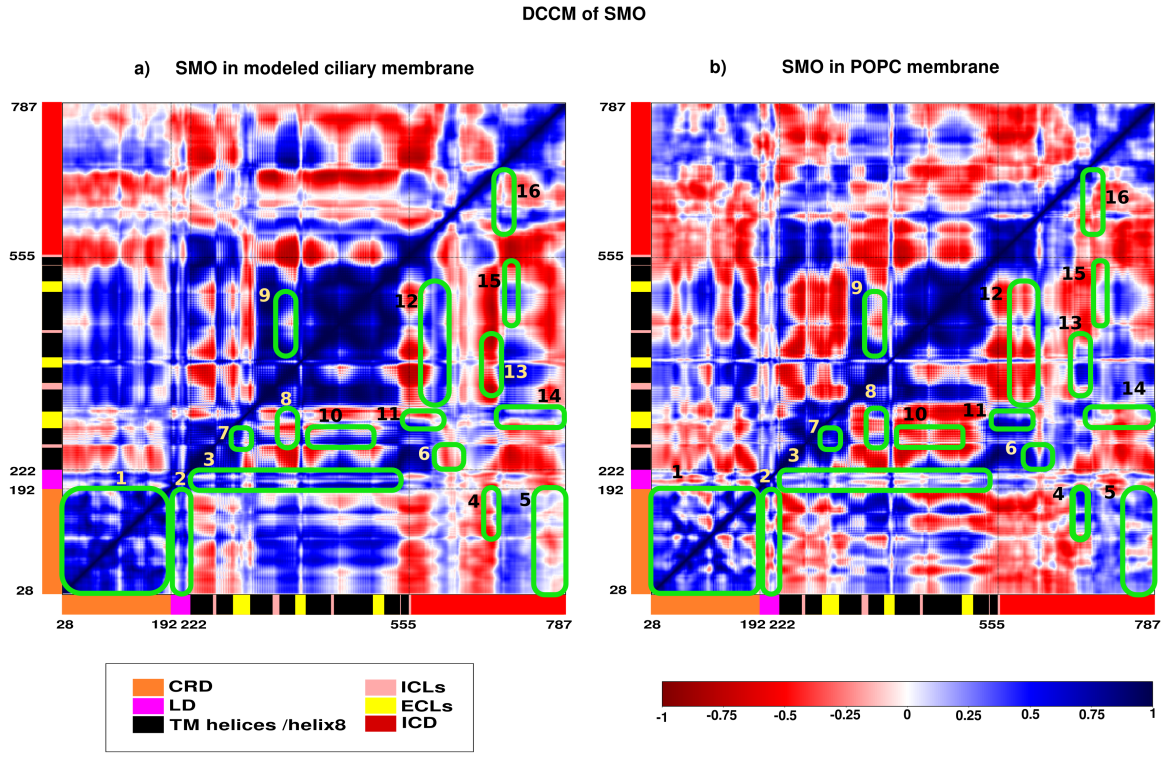


Figure 4.28: Time-averaged DCCM for the SMO backbone bead pairs during the simulation. DCCM of the concatenated trajectory of a) SMO in modeled ciliary membrane and b) SMO in POPC membrane showing the correlation between the domain movements. Domain demarcations are shown along the axes. Domain demarcations are shown along the axes. Box 1 indicates CRD self-correlations. Boxes 2 and 3 indicate the correlations of LD with CRD and TMD, respectively. Boxes 4 and 5 show the correlations of CRD with Arg/Lys cluster 4 and residues 684-787. Boxes 6 and 7 show the correlations of TM1 with Arg/Lys cluster 3 and ECL1 with TM2, respectively. Boxes 8 and 9 show the correlations of TM4 with TM1-ICL1-TM2-ECL1-TM3 and TM5-ICL3-TM6. Box 10 indicates the correlations of TM2 with TM5-ICL3-TM6. Box 11 shows the correlation of ECL1 with ICD residues (∼555 to 576, including Arg/Lys clusters 1 and 2). Box 12 indicates the correlations of ICD residue (∼576-629) with TM3 to TM7. Box 13 indicates the correlations of Arg/Lys cluster 4 with TM4-TM5. Box 14 indicates the correlations of ECL2 with ICD residues ∼711 to 787. Box 15 indicates the correlation of Arg/Lys cluster 5 with TM6-TM7. Box 16 indicates the correlation of ICD residues ∼675 to 711 and with the rest of the ICD. Domain boundaries are represented in dotted lines. Domain boundaries are represented in dotted lines.

To complement the DCCM analysis, PCA was performed on the simulation trajectories (Figures 4.29 and 4.30). The first three principal components, in the modeled ciliary membrane dynamics, captured around 58.81% (67.19% in the POPC dynamics) of all protein motions.

The first mode (PC1), accounting for 28.91% (37.91% in the POPC dynamics) of total motion, depicts the motions of the ICD and CRD towards the membrane plane. These motions were noticeably lesser in the POPC membrane compared to that in modeled ciliary membrane. However, in the ciliary bilayer system, high fluctuations were observed in the ICD, including in Arg/Lys clusters 1 and 2, but other Arg/Lys clusters showed less RMSF. Further, it showed high fluctuations in TM5, TM6, and ECL3. This mode also captured the Arg/Lys cluster 5 movements towards the membrane near the TM6-TM7 region, in conjunction with the DCCM result (Figure 4.28) in the ciliary membrane. In contrast, in the POPC system, CRD and Arg/Lys cluster 5 showed higher fluctuations as compared to the ciliary membrane. A clear comparison of mode 1 in two different membrane systems shows ICD residues, which showed higher fluctuations in the ciliary membrane and lower fluctuations in POPC. This suggests that the presence of ciliary lipids had a significant impact on the ICD residues. The second mode (PC2), in the modeled ciliary system dynamics, accounting for 18.92% of the total motion (24.28% in the POPC dynamics), captured the movement of CRD and ICD towards the membrane plane. This mode showed the correlated motion of the CRD and ICD while the TMD remained rigid. This mode showed a high RMSF value for C-terminal residues of CRD ($\sim$33 to 45 and 90 to 100). In contrast, the second mode (PC2) in the POPC dynamics showed that a part of the CRD moves away from the membrane, while the other part of CRD ($\sim$130 to 140) moves toward the membrane. The ECLs were relatively more flexible than the ICLs, except for ECL2. The third mode (PC3) of the modeled ciliary system, accounting for 10.98% of the total motion, captures the simultaneous movement of part of CRD (residues $\sim$90 to 114; CARC1 motif- 105 to 112) and ICD towards the membrane. Here, the fluctuations of ICD were dampened compared to that in PC1. In contrast, PC3 in the POPC system, (which comprises only 5% of total motion), shows only very little movement of CRD, no major movements in the ICD, and a rigid TMD.

The PCA analysis clearly showed that SMO showed major structural movements (CRD and ICD, especially Arg/Lys clusters) in the presence of cholesterol and other membrane lipids. The TMD dynamics were restricted in the absence of ciliary lipids.

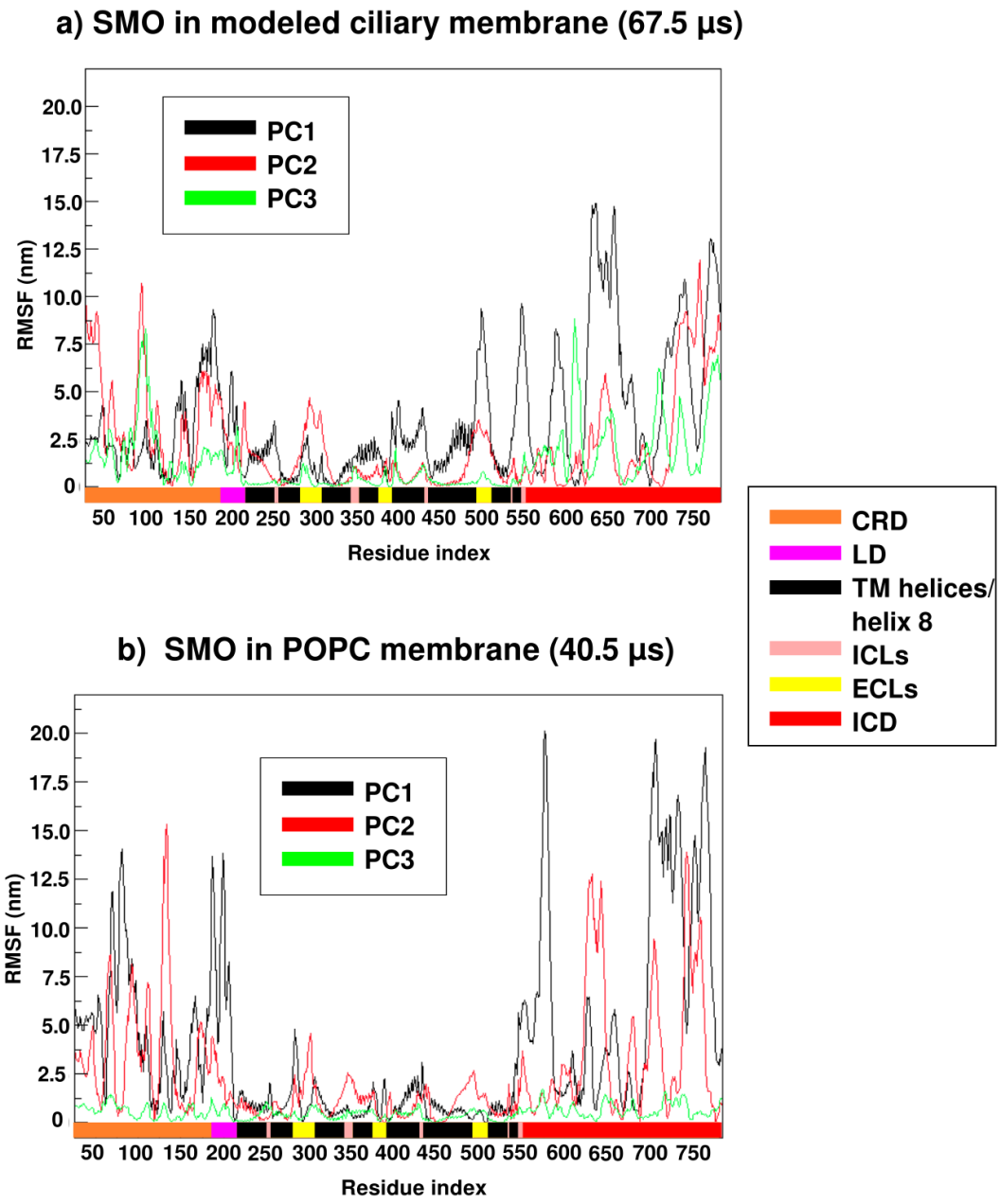


Figure 4.29: Essential dynamics of SMO. RMSF plots corresponding to the PC1, PC2, and PC3 for SMO in a) ciliary membrane and b) POPC membrane with domain demarcation along the abscissa.

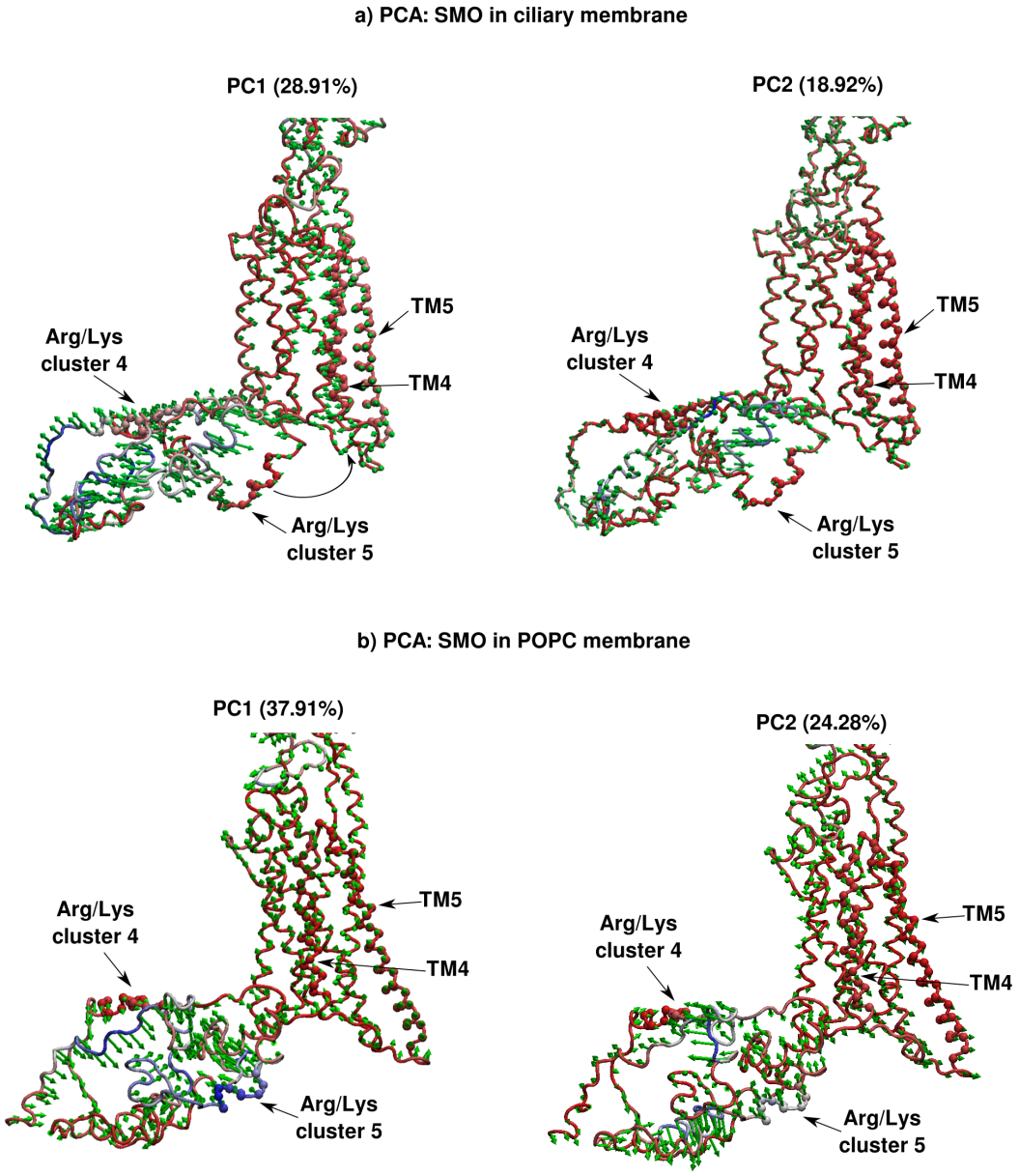


Figure 4.30: Essential dynamics of SMO. PC1 and PC2 of SMO ICD region in a) ciliary membrane and b) POPC membrane, shows the movement of Arg/Lys cluster 4 and Arg/Lys cluster 5. Only TMD and ICD of SMO are shown here. The green arrows show the direction and magnitude of the motion corresponding to the principal components.

4.5 Discussion

Results of the CG-MD simulations of SMO in the modeled ciliary membrane showed that SMO and the different constituents of ciliary membrane interact with each other, leading to dynamic changes in the distribution of the membrane components along with conformational changes in SMO. Cholesterol molecules were seen to move around within and across the membrane leaflets, spending a lot of time in the TM region, while taking turns to interact with different SMO domains and with other membrane lipids. In contrast, the other lipids (especially PI4P) moved around within their respective leaflets, while interacting with specific membrane proximal regions of SMO.

That cholesterol is a key modulator of SMO activation, and therefore of signal transduction in the HH pathway, has been demonstrated by several researchers [128, 186]. In this context, analysis of the MD simulations of SMO in the ciliary membrane model, discussed in this chapter, helped to reiterate the role of cholesterol-binding motifs and sites within SMO [24, 43, 125]. These analyses also helped to identify and define a *strict* CCM, which may well steer cholesterol into the hitherto identified binding sites [43, 138] within the TMD of SMO. The results showed that membrane cholesterols selectively interact with specific sites on the CRD and TMD of SMO, consistent with previous experimental and computational studies [24, 43, 125]. These interactions play an important role in the localization of SMO to the primary cilia and its subsequent activation [24, 43, 52, 109, 125, 128].

4.5.1 SMO cholesterol interaction

The simulations revealed two types of cholesterol-SMO interactions. The first type of interaction was essentially transient interactions with residues that are not part of any motifs, such as interactions with different CRD, TM1, TM2, TM3, and TM4 upper region residues. The second type of interaction, such as cholesterol interactions near the lower region of TM2, TM3, TM4, TM5, and TM6, was more persistent. It was shown that the binding of cholesterol with CRD changed its conformation and orientation with respect to TMD [24, 75, 109]. The study suggested that the CRD was highly dynamic and bent towards the ciliary membrane (Figures 4.20, 4.31, and Video 4.1). Interestingly, the forays of the CRD towards the membrane appeared to be in concert with membrane cholesterol molecules moving upwards from the lower leaflet to the upper leaflet to make contacts. For example, the CRD-cholesterol interaction in replicate 1 was for a very short duration (9.14 to 9.16 μs) (Video 4.1). However, in replicate 5, the cholesterol-CRD interaction lasted for a longer duration (13.38 to 14.99 μs). Some of these interactions were with any cholesterol-binding motif (CARC3; residues 159 to 167) present in CRD and with other residues. Membrane cholesterols showed the interaction with R28, G29, A30, N35, G38, P39, G40, P41, R42, S43, A44, R49, D95, S96, Q99, E100, E101, R151, R159,

E160, R161, D165, F166, L167, R168, and R173 residues. Out of these residues, D95 was reported to be important for cholesterylation [65, 183]. D95 and a few other residues, Q99, E160, and D165, were reported to interact with cholesterol in crystal and cryo-EM structures [24, 43, 125]. Although the residues of CARC3 were interacting with cholesterol fleetingly, this was observed in three independent replicate simulations. It was difficult to say if the movement of CRD was due to the movement of cholesterol in the membrane or the cholesterol movement was due to that of CRD. Additionally, in some frames, residue V210, which was a part of the CARC4 motif present in LD, also interacted with cholesterols.

The DCCM of SMO in ciliary and POPC membranes was compared to learn more about how cholesterol and PI4P influence SMO conformational dynamics. Particularly, the correlated motion of TM4, TM5, and TM6 with ICD residues (including Arg/Lys clusters) in the ciliary membrane suggested that the ciliary lipids, such as, cholesterol, and PI4P, play an important role. During the simulation, PCA showed that there were essential movements of CRD and ICD towards the membrane, including movements of Arg/Lys clusters 4 and 5 towards TM6-TM7. Interestingly, these clusters were previously reported to interact with PI4P [77].

The displacement of helix 8 was observed at three different time points, where the mode of SMO-cholesterol interaction varied (Figure 4.30). In the first structure (initial structure), the distance between the extracellular end of TM5 and TM6 was more than in the structures at 2.28 μs and 9.15 μs. It was because in the initial structure the interaction of SMO with membrane cholesterols is mediated only through the TMD and ICD domains. The LD - cholesterol interactions at $\sim$2.28 μs led to changes in the conformation of LD. At 9.15 μs, a change in the alignment of CRD with respect to TM6 towards the membrane was observed, and CRD bent towards the membrane to interact with membrane cholesterols. At this point, the distance between TM5 (extracellular tip) and TM6 decreased with a slight change in the intracellular part of TM6. The movement of TM6 ultimately also caused the displacement of helix 8, which changed the conformation of ICD (Figure 4.31). Further, the alignment of the 7.5 μs structure with the last frame (15 μs) structure, of SMO in modeled ciliary membrane, also clearly showed the shift in the extracellular region of TM1, TM6, and TM7 with a slight shift in helix 8 (Figure 4.32). To investigate the impact of cholesterol and PI4P on TM helices further, the last frame (15 μs) from both systems was superimposed. The structural comparison of SMO in the last frame (15 μs) between the POPC membrane and the modeled ciliary membrane showed a major shift in helix 8 with tilting of all TM helices except TM6 and bending of TM5 at the extracellular region (Figure 4.33). Therefore, it was predicted that the conformational changes in SMO were mediated from CRD to LD and reorganized the TM bundle with a conformational disposition in ICD (Figure 4.34).

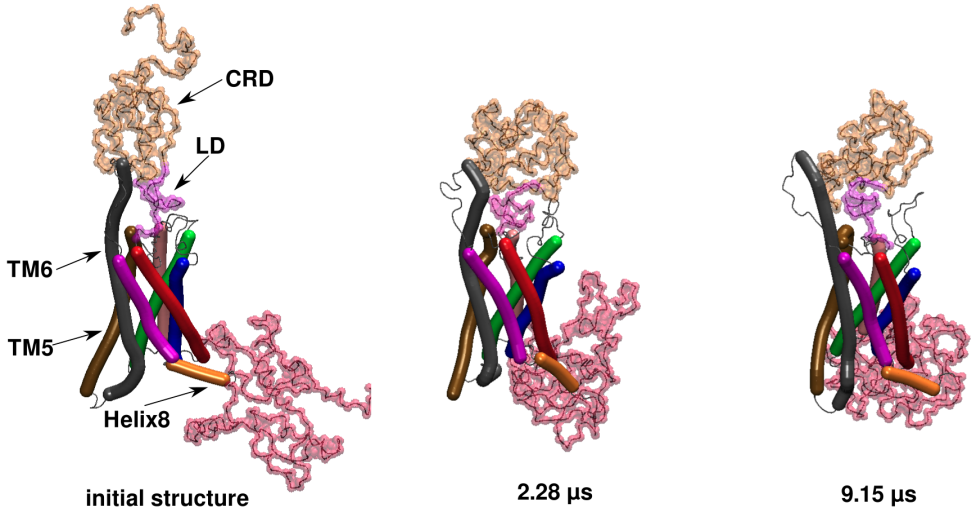


Figure 4.31: Different conformations of SMO (in modeled ciliary membrane) domains and TM helices. Three representative snapshots (initial structure, t= 2.28 μs, and 9.15 μs) from replicate 1 simulation showing different conformation of CRD (orange), LD (magenta), ICD (red), helix 8 (orange tube) and changes in TM helices during simulation in modeled ciliary membrane.

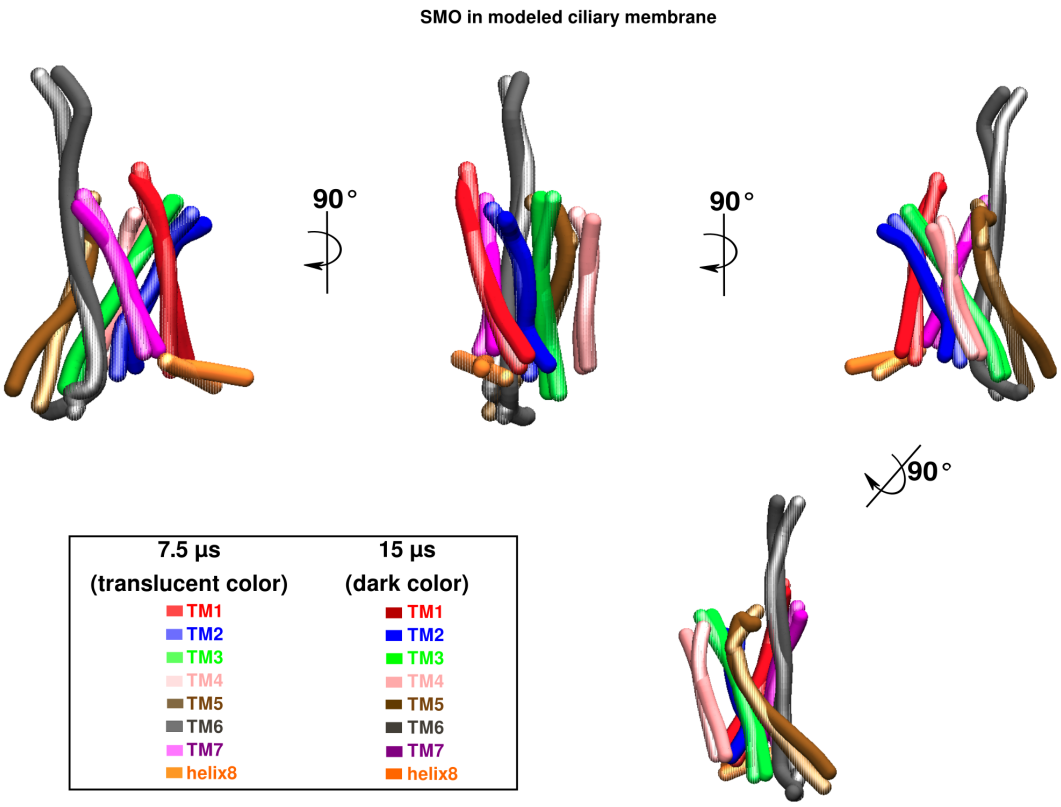


Figure 4.32: Superposition of 7.5 μs and 15 μs structures of SMO shows structural changes in TM helices. The TM helices of structure correspond to the 7.5 μs are shown in translucent color than 15 μs in ciliary membrane.

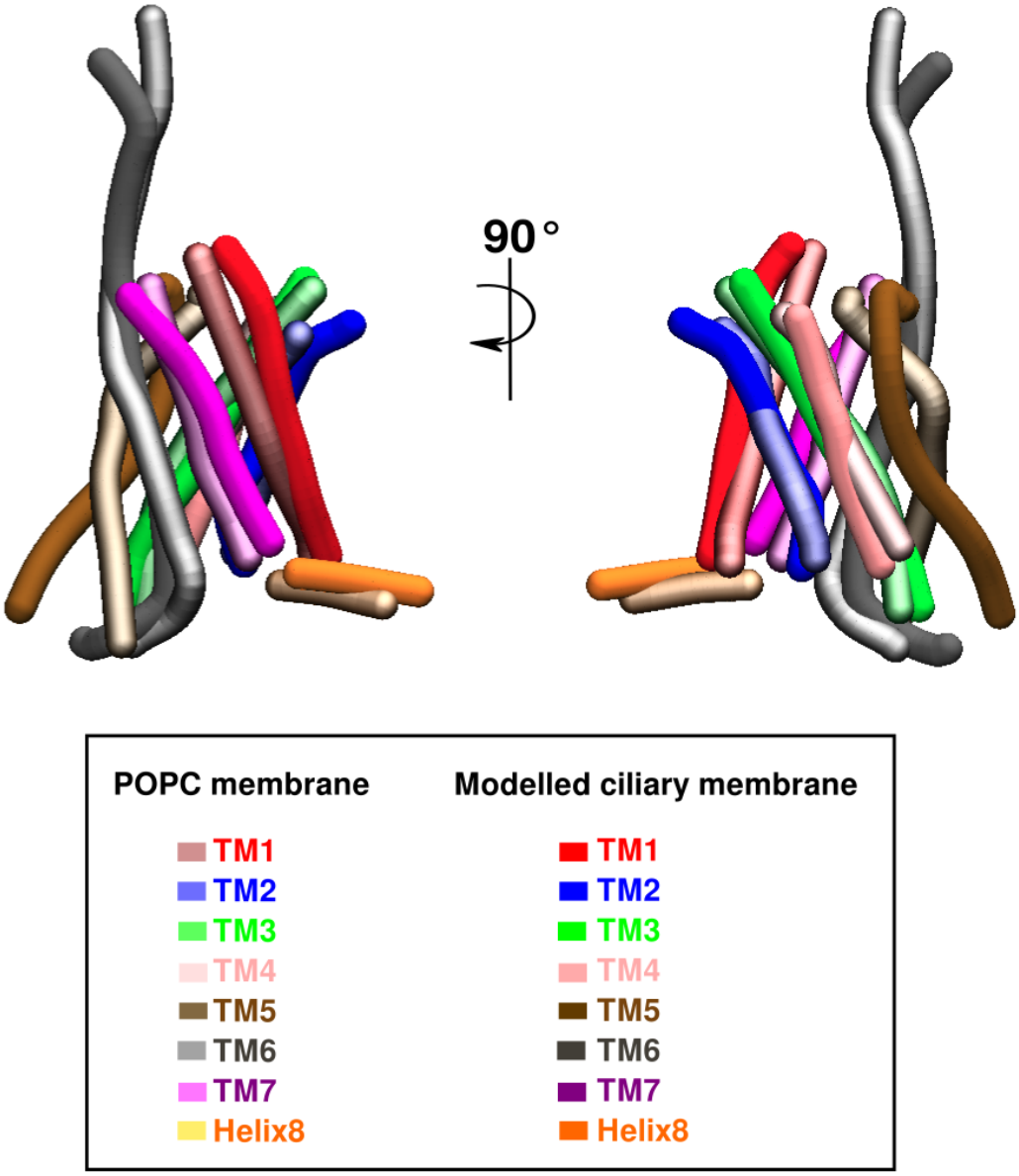


Figure 4.33: Superposition of the final structures (15 μs) of SMO in POPC and ciliary membrane. The TM helices of structure correspond to the 15 μs in POPC membrane are shown in metallic color than 15 μs in ciliary membrane.

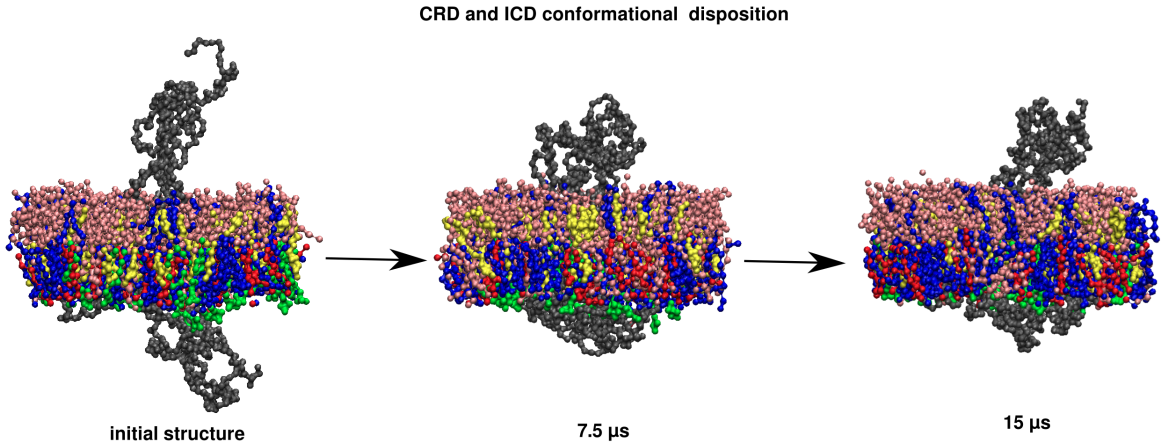


Figure 4.34: The conformational disposition of CRD and ICD of SMO in modeled ciliary membrane (replicate 1) shown by structures at three different time frames (initial structure, t = 7.5 μs, and 15 μs. The color scheme is as follows: SMO (gray), POPC (pink), POPE (blue), POPS (red), PI4P (green) and cholesterol (yellow).

4.5.1.1 The role of cholesterol-binding motifs and the interesting scenario of an unusually persistent SMO-cholesterol interaction in the TMD region

The results showed that, though some cholesterol-SMO interactions do not involve cholesterol-binding motifs, such motifs do have some role in the interactions of cholesterol with SMO. Cholesterol-binding motifs such as CARC7 and CRAC4 in the lower TM region interact with cholesterol, along with other motifs in the lower TMD. The same cholesterol molecule can interact with multiple cholesterol-binding motifs due to their arrangement in the TMD (Figure 3.3; chapter 3). Video 4.3 showed that CRAC4 (TM5) and CARC7 (TM4) motifs participated in forming cholesterol interaction sites with cholesterol located between the two TM helices. These cholesterol interactions with TM4 and TM5 were from the inner leaflet, and in most replicates, were persistent with one cholesterol molecule, supplemented by several additional transient interactions with different cholesterol molecules. Comparable observations showing very high cholesterol occupancy at CRAC motif on TM5 in the human serotonin 1A, were reported earlier [139]. Recently, a study on human serotonin1A receptor and cholesterol found that the CRAC motif's lysine residue (K101) in TM2 is critical in cholesterol sensitivity. Other than in the lower region of TM4 and TM5, the cholesterol occupancy was also observed near CARC8 (TM6) and CARC9 (TM5). This result is in close agreement with two recently identified sterol binding sites present in TMD [125], as few of the identified residues are part of

cholesterol-binding motifs CARC8 and CARC9.

A computational study on a truncated SMO (only LD and TMD) was performed in POPC/ cholesterol and modeled ciliary membrane with Phosphatidylinositol 4,5-bisphosphate (PIP2; to mimic the base of primary cilia) by Hedger et al., [60]. They identified higher occupancy of cholesterol near TM2/3e (residues from the upper region of TM2, TM3, and ECL1; Figure 3.2 and Table 3.2; chapter3) than that of any cholesterol-binding site (CRAC/CARC motifs) present in TMD. They found the TM4 and TM5 intracellular ends have very less cholesterol occupancy as compared to the TM2/3e [60]. In contrast, a higher cholesterol occupancy was observed at CARC7, CARC8, and CRAC4, along with TM2/3e, in the studied simulations. It is possible that the movement of ICD towards the membrane was caused by the interaction of ICD with cholesterol and PI4P. Thus, persistent interactions between cholesterol and the cholesterol-binding motifs located in the lower TMD were formed.

4.5.1.2 Redefining the *strict* CCM in SMO

The *strict* CCM has been suggested to bind with cholesterol in GPCRs. The CCM is a well-known cholesterol-binding site found in GPCRs originally identified in β2AR. This CCM comprises of three residues from TM4 (i) a positively charged residue (R, K) between positions 4.39-4.43, (ii) an aromatic residue (W, Y) at 4.50, and (iii) a β-branched hydrophobic residue (I, L, V) at 4.46. To be a *strict* CCM a fourth aromatic residue (W, Y, F) from TM2, usually at 2.41 , should be present. Multiple sequence alignments of GPCRs identified the CCM in 44% of human GPCRs [58]. McCabe and Leahy (2015) also attempted to define a CCM in SMO with only two residues W365 (4.50) and H361 (4.46) [103]. Hedger et al., (2019) did not see any significant interaction of cholesterol with the CCM of SMO identified by McCabe and Leahy, which corresponds to the CCM of the β2AR [60]. Here, I have redefined the CCM motif by doing the multiple sequence alignment of SMO with several GPCRs (Figures 3.5b-d). Residue K356 (4.41)-W365 (4.50)-L362 (4.47) of SMO were defined as CCM, and the addition of Y269 (2.46) correspond to the *strict* CCM. L362 (4.47) was a more suitable candidate to fulfill the requirement of a β-branched hydrophobic residue. K356 (4.41) was the most suitable positively charged residue to be considered in the range from 4.39 to 4.43. Initially, two residues Y262 (2.39) and Y269 (2.46) were identified suitable for the CCM to become a *strict* CCM (considering the fourth residue from TM2). Cholesterols were observed to be present near L362, W365, and K356, but no significant interaction of cholesterol with Y262 was found. However, it was identified that Y269 was proximal to W365 and had more interactions with cholesterol than Y262 (Figures 4.12-4.16 and Video 4.4). Therefore, it was proposed that K356 (4.41)-W365 (4.50)-L362 (4.47) and Y269 (2.46) could be the *strict* CCM in SMO.

4.5.1.3 Cholesterol is stable around and inside TMD

Crystal and cryo-EM structures, and computational studies showed that SMO has another sterol binding site 'inside' its TMD that can activate SMO [43, 125]. Cholesterol molecules were not observed to enter the TM bundle in the simulations performed in this study and in previous study [60]. Instead, higher cholesterol occupancy was seen near residues of site 1, site 2 [125], and the TM site [43]. But these interactions were through the involved residues of TMD from the outer surface of the TM helices. Interestingly, a persistent interaction of cholesterol with the lower regions of TM4 and TM5 from the 'outer side' of the bundle was observed in the performed simulations. Is this unusual persistence only due to the proximity of the two cholesterol-binding motifs, namely, CARC7 and CRAC4?

A simulation was also performed with cholesterol positioned inside the TMD (Figure 4.35a) [43, 125, 128]. The observation revealed that cholesterol remained stable within the TM site (Figure 4.35b and Video 4.5) and did not exit. This could be attributed to an activation energy barrier that prevents a smooth entry of cholesterol molecules into the interior of TMD within the limited duration of the simulations.

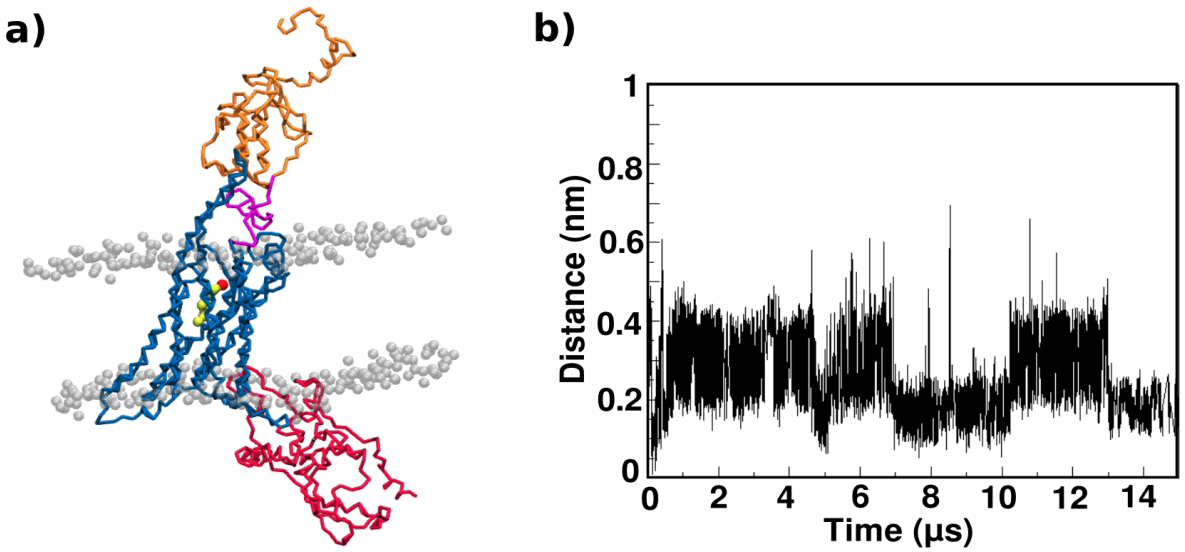


Figure 4.35: A cholesterol molecule positioned inside TMD. a) Initial structure of cholesterol placed inside TMD near site 1 sterol binding site. b) Average distance between the center of mass of site 1 residues and center of mass of cholesterol along the simulation trajectory.

4.5.1.4 Cholesterol-PI4P interplay with SMO

A closer look revealed that the interaction dynamics of SMO involved other lipids, and there was a combined interplay of Arg/Lys clusters, PI4P lipids, several cholesterols, and CARC7 (TM4) and CRAC4 (TM5). PI4P lipids, which were confined in the inner leaflet of primary cilia, were observed to interact with ICD, helix 8, lower TM regions (TM1, TM3, TM4, TM5, and TM6), and ICL1. The presence of basic residues in these lower TM regions allowed them to interact with the PIP4 head group. High occupancy of PI4P (71.63%, 95.32%, and 64.46%, respectively) was observed in W545, along with R546 and R547 residues of helix 8. Further, it was reported that W549 and R551 were particularly important for ciliary localization [35, 52]. The observations indicated that W549 and R551 of helix 8 had PI4P occupancies of 80.29% and 67.28%, respectively, during the total simulation time. It was suggested from the data that helix 8 interact with the ciliary membrane. Additionally, previous reports have emphasized the critical role of higher PI4P occupancy near the basic residues in the fourth Arg/Lys cluster for SMO transport and enrichment in cilia [35]. In this context, the interaction of membrane cholesterol with PI4P seemed particularly important for the persistent interaction of cholesterol with CARC7 and CRAC4 motifs. PI4P lipids constrained the movement of inner leaflet cholesterol molecules, which would otherwise participate in frequent movements between the leaflets. Note that cholesterol movement was rapid in the upper leaflet, where PI4P lipids were not present (Figure 4.21). Here, the cholesterol molecules, which had persistent contact with CARC7, CARC8, and CRAC4 present in the lower TM region of TM4, TM6 and TM5, respectively, were seen trapped by PI4P lipids interacting with lower TM4, TM5, helix 8, Arg/Lys cluster 4, and Arg/Lys cluster 5 of ICD (Figure 4.30). So, such a scenario of the combined interplay of Arg/Lys cluster, PI4P lipids, several cholesterols, and CARC7 (TM4) and CRAC4 (TM5) together explained the long residence time of cholesterol, and this helps in the anchoring of SMO to the ciliary membrane.

4.6 Conclusion

SMO activation is associated with the translocation of the SMO to the primary cilium. The simulations revealed a notable change in the CRD, together with a rearrangement of the TM bundle and ICD, when cholesterol and other ciliary lipids were present in the membrane. The interaction of membrane cholesterols with cholesterol-binding motifs at TM4 and TM5, and PI4P lipids with the intracellular proximal region of TM helices led to rearrangement of the TM bundle. While the interactions of CRD and LD with cholesterol molecules (for a few ns to μs) helped them adopt multiple conformations, the interaction of helix 8 and ICD with cholesterol and PI4P lipids leds to the reorientation of ICD. This study showed that PI4P interaction with W549 (helix 8) and R551, along with the Arg/Lys clusters 4 and 5 in the central region of SMO ICD, helped PI4P accumulation. These accumulated PI4P lipids sequestered cholesterol

molecules near the lower TMD, which helped the cholesterol to make stable interaction with SMO near the cluster of cholesterol-binding motifs at the lower TM region, including the *strict* CCM site. The anchoring of SMO to primary cilia happened through PI4P, and cholesterol moves within outer or inner leaflets and between the leaflets. The interaction of cholesterols with SMO was sometimes persistent and sometimes transient. Overall, the study provided some insights into how localization of SMO in the primary cilia, leading to its stability and conformational dynamics being modulated by the ciliary environment, may be critical for its activation (Figure 4.36). Further experiments, including those leading to the identification of endogenous ligands of SMO, are needed to fully understand the downstream processes leading to morphogenesis, at the molecular level. The study will facilitate the designing of new experiments.

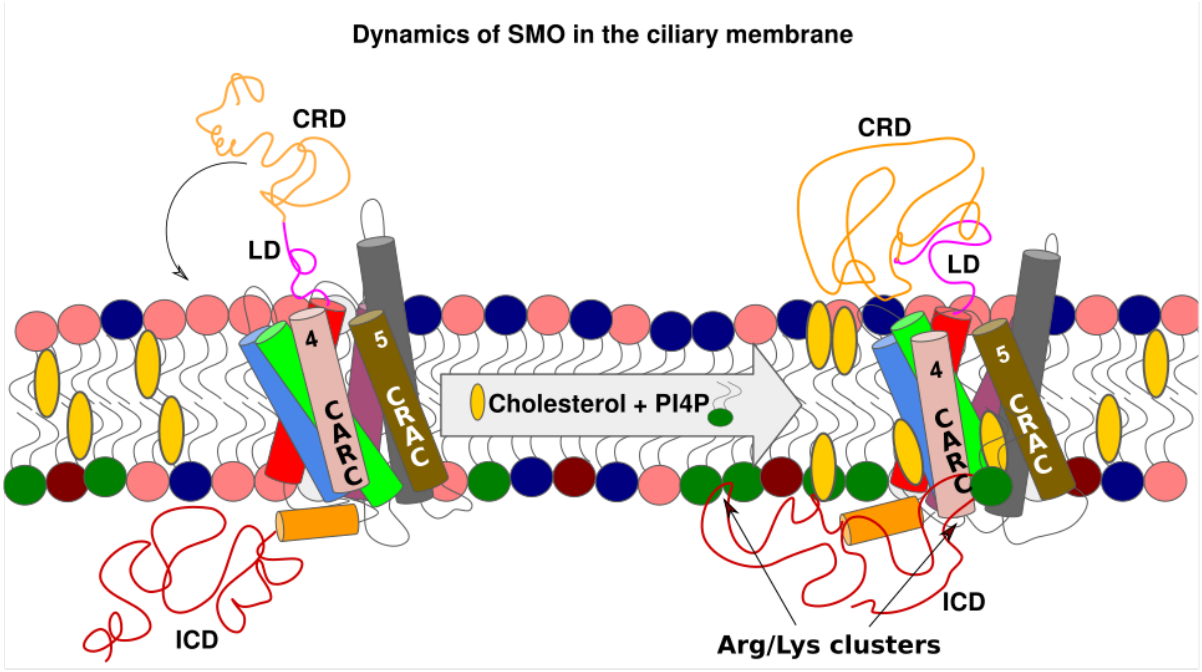


Figure 4.36: A schematic diagram of SMO in modeled ciliary membrane showing the dynamics of SMO domains and the interaction of membrane cholesterol with CRAC7 (TM4) and CRAC4 (TM5) motifs, PI4P interaction of ICD and cholesterol and PI4P interaction.

Chapter *5*

Putative role of cholesterol in shaping the structural and functional dynamics of Smoothened (SMO)

5.1 Introduction

The presence of cholesterols in the cryo-EM structure of PTCH1 (PDB ID: 6MG8) supports the idea of PTCH1 being a cholesterol transporter [189]. Other studies have also reported that in vertebrates, PTCH1 uses its cholesterol transporter function to control SMO via depleting the cholesterol concentration in the ciliary membrane, and thereby reducing the accessible cholesterol pool for SMO [82, 83]. In this context, it is important to note that cholesterol has been invoked in the activation of SMO, by several groups [24, 97, 109]. Inhibition of SMO allows the Suppressor of Fused (SUFU) and Protein Kinase A (PKA) to inhibit the Glioma-associated (GLI) transcription factors. The first important step of HH signaling, namely the onward transmission of the signal from PTCH1 to SMO [24], involves the binding of HH ligand to PTCH1, leading to its displacement from the primary cilium to release the repression on SMO. This leads to the movement of SMO to the tip of the primary cilium, where it binds to the SUFU-PKA complex to inhibit its repression of GLI transcription factors. The activated GLI transcription factors then translocate to the nucleus and promote the transcription of target genes [3, 8, 12, 19, 69, 128, 130, 166].

With the translocation of SMO to the primary cilium being a necessary prerequisite, the current understanding of the activation of the downstream processes envisages an additional role of cholesterol as an agonist (ligand) inducing the binding of the translocated SMO to the SUFU complex. Several studies have also reported that SMO has more than one ligand-binding site, *viz.,* (a) in the cysteine-rich domain (CRD) and (b) in the transmembrane domain (TMD) [24, 43, 60, 68, 84, 125, 126, 169, 174, 188]. The identification of CRD and TMD ligand-binding sites and the presence of cholesterol in both ligand-binding sites make the activation of SMO upon cholesterol binding even more puzzling and interesting [24, 43, 68, 84, 125]. The detailed

mechanism underlying the cholesterol-induced effect on the structure and dynamics of SMO is not very clear. Though there are multiple structures of SMO available in the PDB in complex with cholesterol and epoxy-cholesterol at both TMD and/or CRD binding sites (Table 3.1), they do not provide any clues regarding whether cholesterol binds to TMD or CRD first.

In this context, a recent publication [84] provides significant insights. The detailed study by Kinnebrew et al. revealed that sterols function as HH-regulated orthosteric ligands at the CRD and as allosteric ligands in the TMD to regulate SMO activity and, thus, HH signaling. They suggested that PTCH1 can control SMO in two ways: (a) it depletes the accessible cholesterol from the outer leaflet of the membrane and indirectly inhibits SMO activity, and (b) it removes cholesterol directly from the SMO CRD by an unknown mechanism. They also found that the sterol binding to (inside) the TMD site increases SMO activity regardless of HH exposure and noted that the presence of cholesterol in both the TMD and CRD binding sites is not required for SMO activation. Thus, SMO has four states based on the presence of cholesterol in neither, either, or both of the CRD and TMD binding sites related to its activity (Figure 5.1) [84]. State 1 is the apo form of SMO, where there is no bound cholesterol in either of the binding sites, and hence there is no activity. State 2 is where cholesterol is present inside the TM bundle, which initiates basal activity. In State 3, cholesterol is present at the binding groove of CRD and shows medium activity (better than State 2). Lastly, state 4 is the holo form of SMO, where cholesterol is bound to both TMD and CRD binding sites, showing the highest activity. Since the primary cilium is rich in cholesterol, SMO mostly exists in state 2, with cholesterol bound to the TMD, and PTCH1 inhibits the transition from state 2 to the fully active state 4 by reducing accessible cholesterol in the outer leaflet of the membrane. In this hypothesis, Kinnebrew et al. suggested that cholesterol binding in both the TMD and CRD binding sites is not necessary for SMO activation. When there is no HH signal, the SMO TMD site is occupied with membrane cholesterol, as there is an abundance of cholesterol in primary cilia. In this situation, overexpressed PTCH1 transports the ciliary cholesterol and reduces accessible cholesterol in the outer leaflet [83], which prevents cholesterol from binding to the CRD site. When PTCH1 is inactivated by HH, the accessible cholesterol concentration increases, which allows cholesterol to bind to the CRD binding site [84].

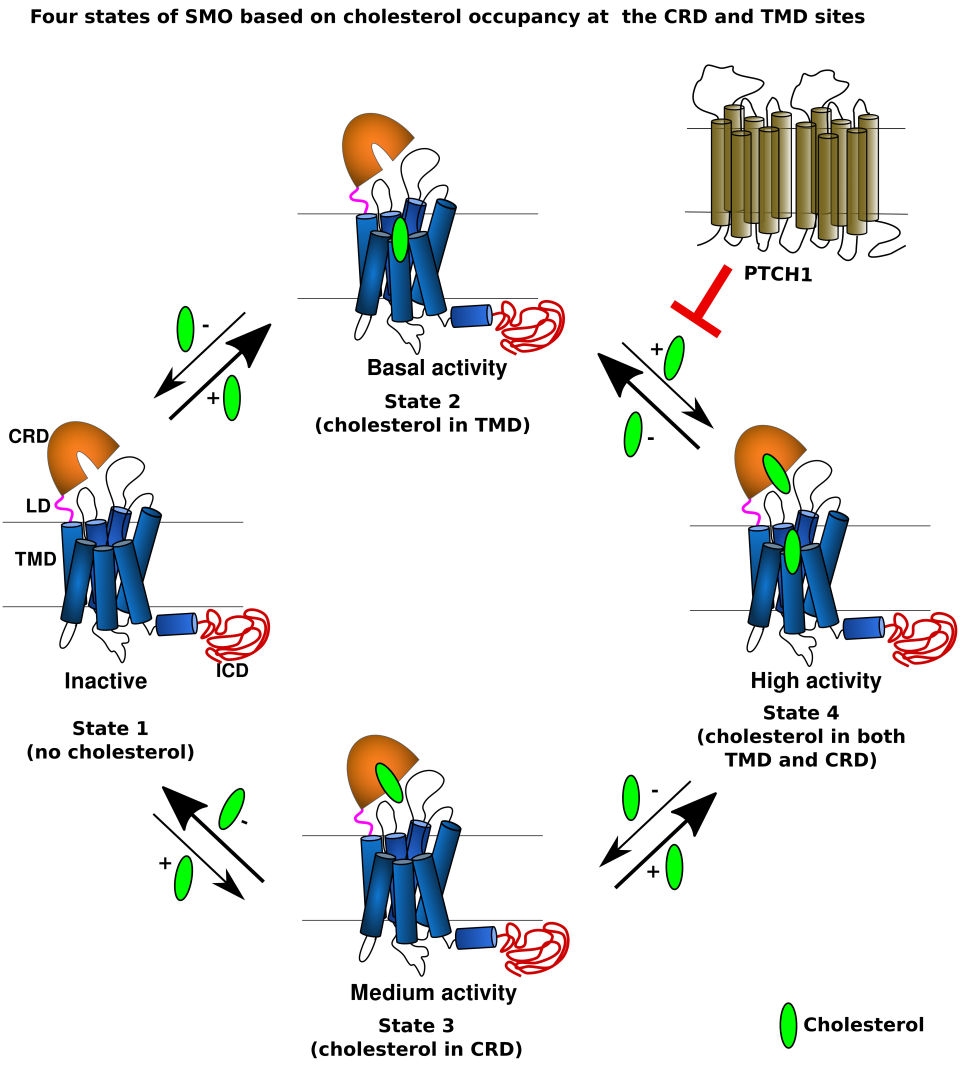


Figure 5.1: Four states of SMO defined by sterol occupancy of the CRD and TMD sites with varying signaling activities (adapted from: [84]). The width of the arrow shows the strength of SMO state conversion. State 1 represents a cholesterol-free state and signifies an inactive SMO. State 4 represents fully cholesterol-occupied as cholesterol is present in both TMD and CRD binding sites and signifies fully active SMO. State 2 and state 3 represent two additional states with cholesterol present in any of the TMD or CRD binding sites. State 2 signifies the basal (no SHH) conditions where SMO-TMD is occupied with cholesterol, and SMO-CRD is free because of less concentration of accessible cholesterol in the outer leaflet of the membrane. Overexpressed PTCH1 uses its transporter activity to reduce the cholesterol from the outer leaflet. Hence, this hinders SMO from going to state 4 from state 2. State 3 signifies the medium activity with cholesterol present in SMO-CRD.

It may be noted that the above hypothesis apparently does not consider earlier reports suggesting that the localization of SMO to the primary cilia is linked with the displacement of PTCH1 from the primary cilia upon HH binding to PTCH1 [42, 52, 171, 176]. This raises the question of how PTCH1 can deplete cholesterol levels in the outer leaflet of the SMO bound primary cilia, if the localization of SMO is linked to its displacement [171, 176]. In this context, it must be mentioned that the entry and exit of SMO into the primary cilia is a complex process that is possibly dependent on more than one factor, including ciliary dynamics during the cell cycle. Thus, it is in principle possible for SMO to translocate to the primary cilia even in the absence of HH signaling [42, 52, 81, 171, 176, 180]. Further, it has also been reported that SMO can show some activity even in the absence of HH signaling [44, 125]. It is also possible to visualize a modification of Kinnebrew's hypothesis by combining it with this process of SMO localization in the primary cilia in response to HH binding driven displacement of PTCH1 [171, 176]. Be that as it may, depending on the strength of HH signaling, it is possible to imagine situations where PTCH1 and SMO molecules colocalize within the primary cilia, while PTCH1 concentration decreases with the increase of HH signaling. In that situation, a scenario can be imagined where there will be a growing population of SMO molecules, with cholesterol bound to its TMD, in the primary cilia, coupled with a progressively decreasing population of PTCH1, and hence, a progressively increasing population of cholesterol molecules in its outer leaflet. The transition from state 1 to state 2 and on to state 4 may then be mapped with the progressive increase in HH signaling.

In the context of the role of ciliary lipids stabilizing SMO in primary cilia, based on a coarse-grained (CG) simulation study described in the previous chapter, it was suggested that cholesterol from the modeled ciliary membrane is bound at multiple sites of SMO, including cholesterol binding sequence and structure motifs [91]. Apart from providing several insights regarding the role of ciliary lipids, and their interactions with SMO binding sites, this study also provided the basis for addressing larger questions. For example, those related to the effect of cholesterol, as an agonist ligand, binding to SMO and hence on the transmission of the signal from the extracellular CRD to intracellular ICD, particularly in the context of the four-state hypothesis proposed by Kinnebrew et al. [84]. In order to obtain a detailed understanding of the structure and dynamics of the relevant systems at the molecular level, necessary for pursuing this goal, atomistic molecular dynamics simulations of four SMO systems in cholesterol-containing modeled ciliary membranes have been carried out. These systems correspond to the four SMO activity related states described by Kinnebrew et al. [84]. The results reported here provide a detailed understanding of the differential dynamics of all three cholesterol-bound states relative to that of the cholesterol-free state. Not only do they corroborate most of the earlier experimental findings, they also provide important pointers toward the possible mechanism of signal transduction, consequent to cholesterol binding, leading to different activity levels.

5.2 Methods

5.2.1 System setup

Four simulation systems were prepared corresponding to the four states of SMO in the ciliary membrane: (a) apo SMO, (b) holo SMO-TMD (cholesterol bound to SMO-TMD), (c) holo SMO-CRD (cholesterol bound to SMO-CRD), and (d) holo SMO with two cholesterol each in TMD and CRD (Figure 5.2). In system 1, only SMO was embedded in the ciliary membrane. While in system 2, one cholesterol was placed near the sterol binding site 1 in SMO-TMD, identified by Qi et al. [125] (Figure 3.2 and Tables 3.1 and 3.2; Chapter 3). The hydroxyl group was oriented towards the extracellular side. Further, in the third system, one cholesterol was placed in the hydrophobic groove of SMO-CRD identified by Byrne et al. [24] (Figure 3.2 and Tables 3.1 and 3.2; Chapter 3). Lastly, the fourth system contains two cholesterols; one cholesterol was placed in SMO-TMD and another in SMO-CRD in the same position as in the second and third systems.

All the systems were generated using CHARMM-GUI [78, 127, 181]. The minimized SMO structure, with or without cholesterol, was embedded into the modeled heterogeneous ciliary membrane. The composition of the outer and inner leaflets of the bilayer was chosen to mimic the ciliary membrane [60] used in the previous study [91]. Systems were solvated using the standard TIP3P [79] water model and neutralized with 0.15 M NaCl. The distributions of lipids around the SMO were random and had different initial velocities. Three independent simulations of each system were performed for 300 ns. The total simulation time is 3.6 μs. The details of the simulations performed, along with their system composition, were summarized in Table 5.1.

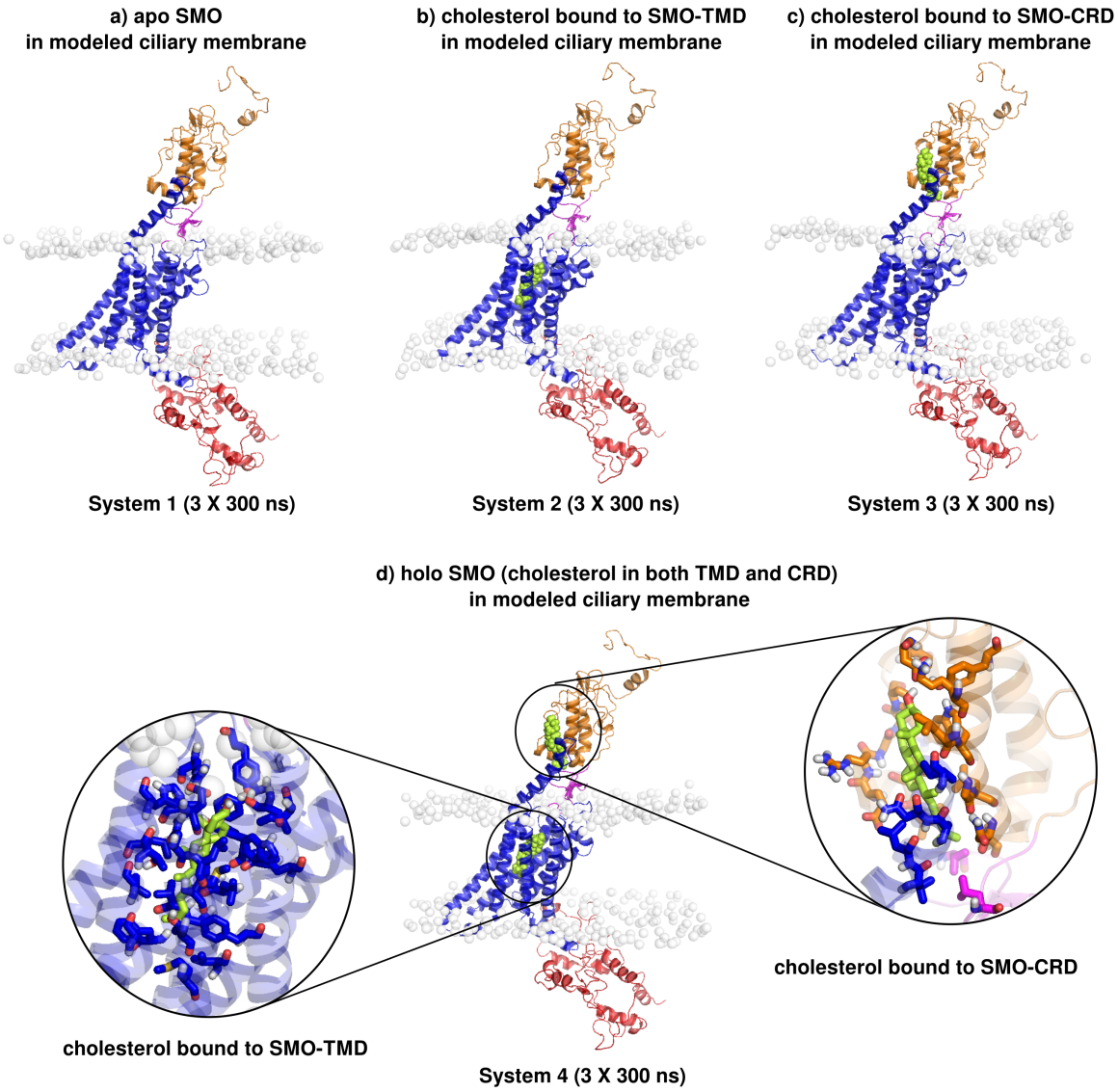


Figure 5.2: The initial structures of simulation systems a) system 1: apo SMO, b) system 2: holo SMO-TMD, c) system 3: holo SMO-CRD, and d) system 4: holo SMO (cholesterol in both CRD and TMD binding sites). SMO is shown in cartoon representation and colored domain-wise; CRD, LD, TMD, and ICD are colored in orange, magenta, blue, and red, respectively. White spheres represent the membrane boundary. Bound cholesterols are colored green and are shown in the sphere representation for systems 2, 3, and 4. The inset used to show the position of the bound cholesterol in TMD and CRD in system 4 also shows the position of cholesterol in TMD and CRD in systems 2 and 3, respectively. The interacting residues are shown in the stick model, and water and ions are not shown here for clarity.

Table 5.1: Details of atomistic molecular dynamics simulation systems corresponding to the four states of SMO.

S. No.	**Description**	**State**	**Cholesterol bound to SMO domain**	**Membrane Composition (in numbers)**		**Number and duration of simulations (ns)**
				Outer leaflet	**Inner leaflet**	
System 1	SMO in modeled ciliary membrane (apo SMO)	Inactive	No bound Cholesterol	POPC: 120 POPE: 30 Cholesterol: 50	POPC: 24 POPE: 84 POPS: 34 PI4P: 24 Cholesterol: 50	3 x 300 ns
System 2	Cholesterol in SMO-TMD in modeled ciliary membrane (holo SMO-TMD)	Basal activity	One Cholesterol in TMD	POPC: 120 POPE: 30 Cholesterol: 50	POPC: 24 POPE: 84 POPS: 34 PI4P: 24 Cholesterol: 50	3 x 300 ns
System 3	Cholesterol in SMO-CRD in modeled ciliary membrane (holo SMO-CRD)	Medium activity	One Cholesterol in CRD	POPC: 120 POPE: 30 Cholesterol: 50	POPC: 24 POPE: 84 POPS: 34 PI4P: 24 Cholesterol: 50	3 x 300 ns

System 4	Cholesterol in both TMD and CRD of SMO in modeled ciliary membrane (holo SMO)	High activity	Two Cholesterols, one in TMD and another in CRD	POPC: 120 POPE: 30 Cholesterol: 50	POPC: 24 POPE: 84 POPS: 34 PI4P: 24 Cholesterol: 50	3 x 300 ns

5.2.2 Simulation parameters

All simulations were carried out using GROMACS version 2019 [162], with CHARMM36 force-field [66]. Energy minimization was carried out using the steepest descent algorithm and subjected to six successive constrained equilibrations of 2 ns each (a total of 12 ns). After the equilibration, the production run was carried out for 300 ns. The temperature was maintained at 310 K using the Nosé-Hoover thermostat [113]. Semiisotropic pressure is controlled at 1 bar using the Parrinello-Rahman algorithm [118]. Initial velocities for the simulations were chosen randomly from a Maxwell distribution [136]. Electrostatics was described using the particle-mesh Ewald method with a 1.2-nm cutoff [38]. Van der Waals interactions were modeled with a 1.2-nm cutoff using the Verlet method [164]. Long-range dispersion corrections were applied for energy and pressure. All bonds were constrained to the equilibrium lengths using the LINCS algorithm [62]. Periodic boundary conditions were maintained along the x, y, and z directions. A time step of 2 fs was used, and snapshots were taken every 100 ps for analysis.

5.2.3 Analysis

Trajectories were analyzed using tools implemented in GROMACS and visual molecular dynamics (VMD) utilities [71], along with in-house scripts. VMD was used for visualization. Figures were produced using VMD, Bendix plugin [36], UCSF Chimera [104], PyMOL (`https://pymol.org`), and Inkscape (`www.inkscape.org`). Plots were generated using in-house Python scripts.

5.2.3.1 Trajectory concatenation

The first 10 ns of the production run were considered as unrestrained equilibration phase (required to allow the systems to become stable). Therefore, the individual replicate trajectories from 10 to 300 ns were combined to create the concatenated 870 ns trajectories for each system.

5.2.3.2 TMD volume analysis

The volume of SMO-TMD was calculated by the *traj_cavity* tool implemented within GROMACS. It analyzes the changes in the accessible volume of a given set of residues involved in forming cavities.

5.2.3.3 Dynamic cross correlation matrix (DCCM) and Principal component analysis (PCA)

DCCM was calculated using the Bio3D package [55] of R as a time average along the 870 ns simulation trajectory to study the correlated motions between different regions of SMO. PCA was performed on TMD backbone atoms for all systems concatenated trajectories using the Normal Mode Wizard plugin of the VMD-1.9.1 version.

5.2.3.4 Clustering analysis

Clustering was performed on the 870 ns concatenated trajectory of each system using the *gmx cluster* function available in GROMACS. The coordinates of the backbone atoms of SMO TMD were selected as input with the κ-means algorithm and the RMSD cutoff of 0.2 nm. Each cluster has a representative structure of cluster members.

5.2.3.5 Real average structure calculation

The average structure for each system was calculated separately using the *gmx rmsf* function in GROMACS with the *-ox* option. Next, the RMSD between each frame of the respective system and the corresponding average structure (used as the reference) was calculated. The structure with the lowest RMSD was then chosen as the real average structure.

5.3 Results

There are two loosely defined binding sites in SMO, one located in the transmembrane domain (TMD) and the other in the cysteine-rich domain (CRD). System 1 represents the cholesterol-free SMO state (apo SMO), i.e., no cholesterol in either of the two binding sites, whereas systems 2, 3, and 4 represent different cholesterol-bound SMO states (Figure 5.3). System 2 is holo SMO-TMD, where cholesterol is bound to the TMD; system 3 is holo SMO-CRD, where cholesterol is bound to the CRD; and system 4 is holo SMO, where cholesterol is bound to both the TMD and the CRD.

5.3.1 Binding of cholesterol to either or both the binding sites leads to an increase in the TMD volume

The volume of the SMO TMD was calculated for all four simulated systems (Figure 5.3). The TMD volumes in systems 1, 2, 3, and 4 are 2796.136 Å^3, 3473.904 Å^3, 3226.944 Å^3, and 3438.232 Å^3, respectively, in the initial structure. The analysis shows that the average TMD volume increases from 3494 Å^3 (SD: 498) in system 1 (apo state) to 4116 Å^3 (SD: 384) in system 2, 3662 Å^3 (SD: 350) in system 3, and 4162 Å^3 (SD: 380) in system 4. This suggests that in systems 2 and 4, the presence of cholesterol in TMD significantly increases the volume of the TMD. In contrast, in system 3, where cholesterol is only present in the CRD, the increase in TMD volume (also confirmed by the inter-helical distances at the upper and lower regions of the TMD discussed below, (Table 5.2) while noticeable, does not appear to be significant.

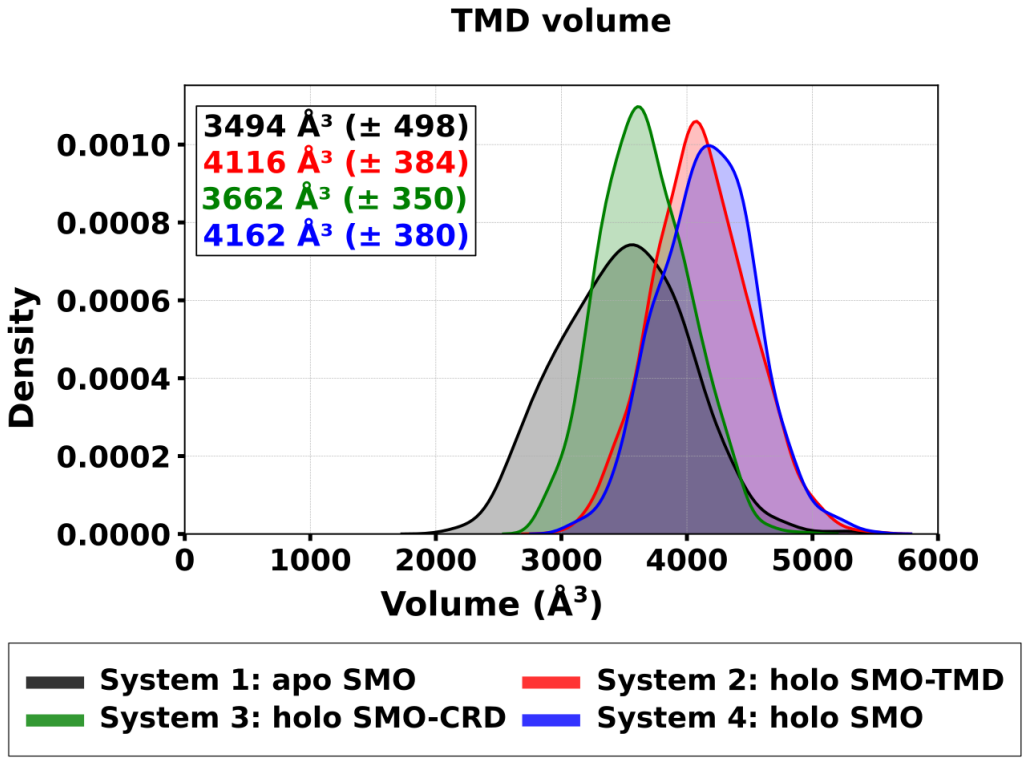


Figure 5.3: Volume of SMO-TMD. The histogram of SMO-TMD volume was calculated by the *traj_cavity* tool for system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). The TMD volume increased by the binding of cholesterol to either or both of the binding sites.

In addition, the inter-helical distances at the upper and lower regions of the TMD were calculated to investigate the structural details underlying these observations on changes in TMD volume. The TM residues selected to calculate inter helical upper and lower distance (D1 and D2) are described in Figure 5.4.

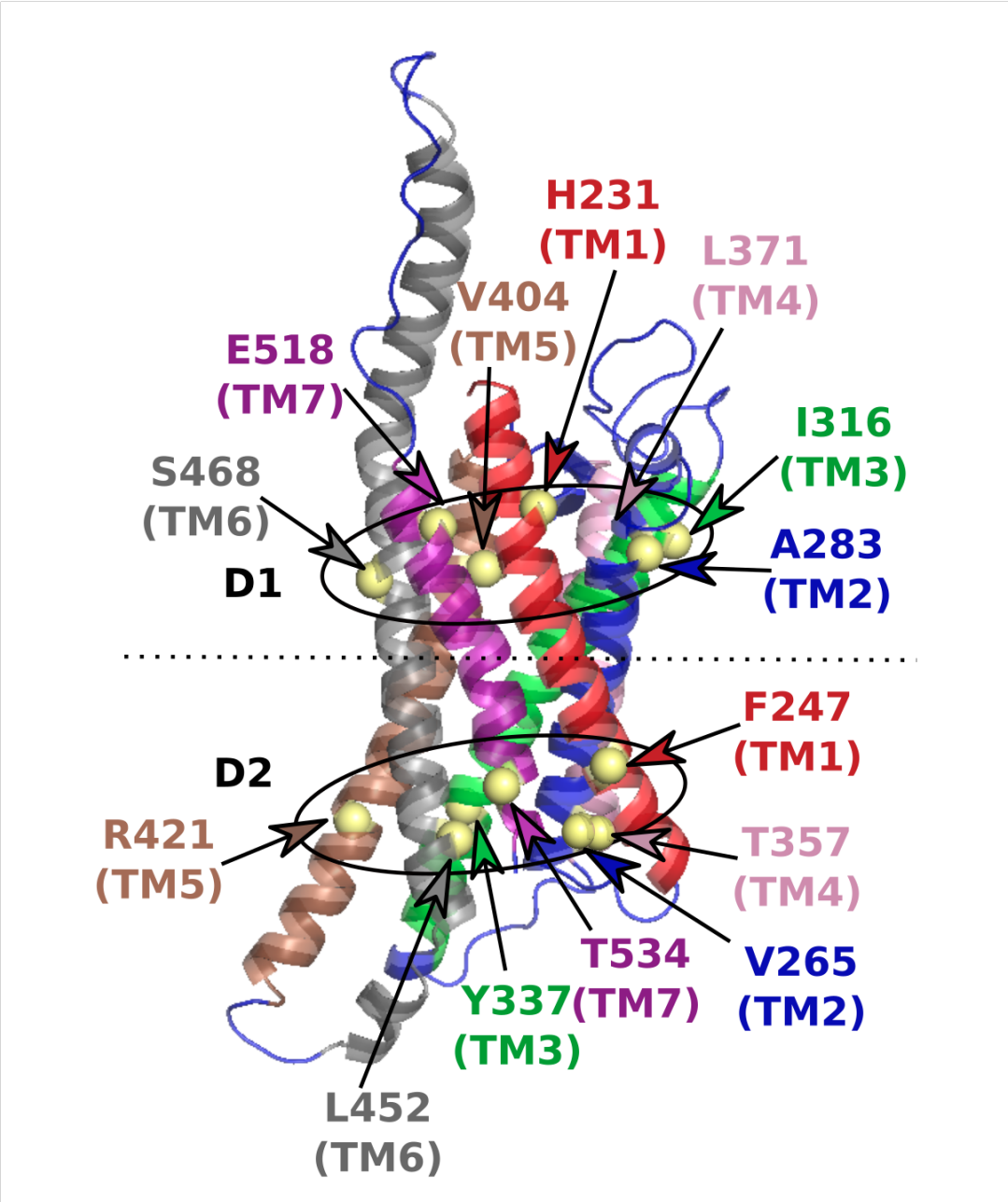


Figure 5.4: A schematic representation of SMO-TMD. TM helices are colored red, blue, green, pink, brown, gray, and purple, respectively. The residues selected for upper TM-TM distance (D1) and lower TM-TM distance (D2) calculations are shown in yellow spheres.

It is observed that in both systems 2 and 4, respectively, where cholesterol is bound to TMD, the helices TM1 and TM5 move away from TM6. This is evidenced by noticeable increase in both the upper and lower TM -TM distances. It may be noted that these distances actually decrease in system 3. Along with this, the lower region of TM2-TM7, TM3-TM7, TM5-TM7, and TM6-TM7 distances have increased (Figure 5.4 and Table 5.2 with a decrease in TM3-TM4 and TM5-TM6 distances in system 4 (Table 5.2).

Table 5.2: Inter-helical average distances (nm) of TM helices between D1 and D2 residues from upper and lower TMD regions, respectively.

Upper TM-TM residue average distance				
	System 1	**System 2**	**System 3**	**System 4**
TM1 (H231) -TM2 (A283)	1.25 (±0.07)	1.19 (±0.05)	1.17 (±0.04)	1.23 (±0.05)
TM1 (H231) -TM3 (I316)	2.04 (±0.08)	2.09 (±0.08)	1.93 (±0.1)	2.07 (±0.09)
TM1 (H231) -TM4 (L371)	2.83 (±0.07)	2.82 (±0.06)	2.73 (±0.07)	2.87 (±0.12)
TM1 (H231) -TM5 (V404)	2.26 (±0.07)	2.27 (±0.06)	2.21 (±0.06)	2.25 (±0.06)
TM1 (H231) -TM6 (S468)	2.27 (±0.05)	2.29 (±0.05)	2.23 (±0.04)	2.37 (±0.08)
TM1 (H231) -TM7 (E518)	1.17 (±0.04)	1.19 (±0.04)	1.17 (±0.04)	1.19 (±0.04)
TM2 (A283) -TM3 (I316)	1.1 (±0.06)	1.19 (±0.08)	1.07 (±0.08)	1.13 (±0.08)
TM2 (A283) -TM4 (L371)	2.3 (±0.06)	2.36 (±0.06)	2.25 (±0.07)	2.33 (±0.07)
TM2 (A283) -TM5 (V404)	2.24 (±0.06)	2.25 (±0.08)	2.19 (±0.08)	2.17 (±0.1)
TM2 (A283) -TM6 (S468)	2.73 (±0.06)	2.79 (±0.06)	2.68 (±0.06)	2.85 (±0.06)
TM2 (A283) -TM7 (E518)	1.97 (±0.07)	1.97 (±0.08)	1.92 (±0.06)	1.93 (±0.11)
TM3 (I316) -TM4 (L371)	1.51 (±0.04)	1.5 (±0.03)	1.5 (±0.04)	1.52 (±0.04)
TM3 (I316) -TM5 (V404)	1.84 (±0.04)	1.84 (±0.04)	1.81 (±0.05)	1.78 (±0.07)

TM3 (I316) -TM6 (S468)	2.75 (±0.04)	2.85 (±0.06)	2.72 (±0.06)	2.9 (±0.06)
TM3 (I316) -TM7 (E518)	2.33 (±0.05)	2.39 (±0.09)	2.25 (±0.07)	2.29 (±0.12)
TM4 (L371) -TM5 (V404)	1.06 (±0.05)	0.99 (±0.05)	1.0 (±0.05)	1.02 (±0.07)
TM4 (L371) -TM6 (S468)	2.16 (±0.05)	2.21 (±0.05)	2.12 (±0.04)	2.3 (±0.11)
TM4 (L371) -TM7 (E518)	2.43 (±0.06)	2.42 (±0.09)	2.34 (±0.05)	2.39 (±0.18)
TM5 (V404) -TM6 (S468)	1.17 (±0.04)	1.26 (±0.05)	1.17 (±0.04)	1.32 (±0.08)
TM5 (V404) -TM7 (E518)	1.53 (±0.05)	1.56 (±0.08)	1.48 (±0.05)	1.48 (±0.12)
TM6 (S468) -TM7 (E518)	1.15 (±0.03)	1.16 (±0.05)	1.13 (±0.03)	1.22 (±0.06)
Lower TM-TM residue average distance				
	System 1	**System 2**	**System 3**	**System 4**
TM1 (F247) -TM2 (V265)	1.53 (±0.03)	1.53 (±0.04)	1.51 (±0.03)	1.53 (±0.03)
TM1 (F247) -TM3 (Y337)	2.54 (±0.05)	2.59 (±0.05)	2.52 (±0.04)	2.58 (±0.07)
TM1 (F247) -TM4 (T357)	2.65 (±0.08)	2.65 (±0.07)	2.58 (±0.06)	2.66 (±0.07)
TM1 (F247) -TM5 (R421)	3.17 (±0.06)	3.23 (±0.05)	3.16 (±0.06)	3.22 (±0.11)
TM1 (F247) -TM6 (L452)	1.83 (±0.06)	1.88 (±0.07)	1.79 (±0.05)	1.9 (±0.13)
TM1 (F247) -TM7 (T534)	1.04 (±0.05)	1.04 (±0.05)	0.98 (±0.06)	1.06 (±0.04)
TM2 (V265) -TM3 (Y337)	1.37 (±0.04)	1.39 (±0.05)	1.37 (±0.05)	1.37 (±0.06)
TM2 (V265) -TM4 (T357)	1.17 (±0.09)	1.16 (±0.08)	1.11 (±0.06)	1.18 (±0.07)
TM2 (V265) -TM5 (R421)	2.21 (±0.07)	2.23 (±0.06)	2.24 (±0.07)	2.22 (±0.11)

TM2 (V265) -TM6 (L452)	1.37 (±0.05)	1.36 (±0.05)	1.33 (±0.05)	1.39 (±0.06)
TM2 (V265) -TM7 (T534)	1.57 (±0.06)	1.54 (±0.04)	1.55 (±0.04)	1.58 (±0.13)
TM3 (Y337) -TM4 (T357)	1.31 (±0.1)	1.26 (±0.12)	1.25 (±0.09)	1.19 (±0.13)
TM3 (Y337) -TM5 (R421)	0.93 (±0.05)	0.94 (±0.06)	0.97 (±0.05)	0.95 (±0.09)
TM3 (Y337) -TM6 (L452)	1.24 (±0.05)	1.22 (±0.05)	1.2 (±0.05)	1.22 (±0.04)
TM3 (Y337) -TM7 (T534)	2.0 (±0.08)	2.04 (±0.06)	2.0 (±0.05)	2.09 (±0.15)
TM4 (T357) -TM5 (R421)	2.17 (±0.08)	2.12 (±0.15)	2.16 (±0.1)	2.07 (±0.15)
TM4 (T357) -TM6 (L452)	2.12 (±0.11)	2.05 (±0.11)	2.0 (±0.1)	2.05 (±0.13)
TM4 (T357) -TM7 (T534)	2.62 (±0.11)	2.58 (±0.09)	2.52 (±0.08)	2.62 (±0.14)
TM5 (R421) -TM6 (L452)	1.49 (±0.06)	1.48 (±0.05)	1.47 (±0.04)	1.45 (±0.06)
TM5 (R421) -TM7 (T534)	2.38 (±0.09)	2.44 (±0.08)	2.4 (±0.07)	2.47 (±0.16)
TM6 (L452) -TM7 (T534)	1.02 (±0.08)	1.05 (±0.08)	1.04 (±0.07)	1.1 (±0.18)

5.3.2 Dynamics of TMD-bound cholesterol

The TMD has multiple binding sites for cholesterol within the TMD region and hence indicates the presence of a channel in different reported structures. Apart from the sterol binding site 1 spanning the middle TMD region, where cholesterol was initially placed in both systems 2 and 4, respectively, other reported binding sites of note are sterol-binding site 2 spanning from the middle TMD region to the upper TMD region and the TMD site in PDB ID 603C spanning the middle and lower middle TMD regions. The dynamics of the TMD-bound cholesterol was examined to understand the possible role of these multiple binding sites. This was carried out by tracking its movements within these binding regions, in both systems 2 and 4, respectively. Two residues from each of the TM helices, one from the upper TM region and one from the lower TM region, were selected for this analysis. The set of the seven upper TM region residues

was labeled as D1, and that consisting of the seven lower TM region residues was labeled as D2.

In both systems, the distances of the COM (Center of Mass) of TMD-bound cholesterol from those of D1 (upper TM) residues (Figure 5.5a) and D2 (lower TM) residues (Figure 5.5b), respectively, reveal upward and downward movements of, the cholesterol moiety, within the TMD channel. Multiple snapshots from systems 2 and 4 with minimum and maximum distance values (between COMs of TMD-bound cholesterol and those of D1 or D2, respectively) were chosen to investigate this movement (Figures 5.5c and 5.5d). It was observed that in system 2, cholesterol primarily visits the upper TM region, accompanied by some movements towards the middle TM region. In contrast, the cholesterol moiety in system 4 spends more time in the upper middle to lower middle TM region. A comparative analysis of relevant interactions was carried out to rationalize these up-and-down oscillations of TMD-bound cholesterol, and the observed variations in the two systems, in terms of the interaction of bound cholesterol with multiple available binding sites in TMD.

The TMD residues with occupancy greater than 30% within a radius of 0.4 nm of TMD-bound cholesterol were identified (Table 5.3). Some of these interactions of cholesterol with TMD residues are shown in a representative structure of system 4 (Figure 5.6a). The hydrophobic and hydrogen bond interactions observed during simulation may be summarized as follows:

- **Interacting residues common to both systems 2 and 4:** W281, L325, V329, F332, **Y394**^{*4}, V404, I408, A459, **H470***, **N521***, M525 (initial residues corresponding to sterol binding site 1) and V270, F274, M326, G328, **S387**^{*4}, I389, F391, **R400**^{*4}, F462, V463, T466, **E518**^{*4}, L522, A524, T528, and M532 (additional residues)
- **Residues interacting only in system 2:** C273 and **D384**^{*2}
- **Residues interacting only in system 4:** L335, L405 and L412

NB. Residues making hydrogen bond with cholesterol OH are marked * (in both systems 2 and 4), ^{*2} (in system 2 only) and ^{*4} (in system 4 only).

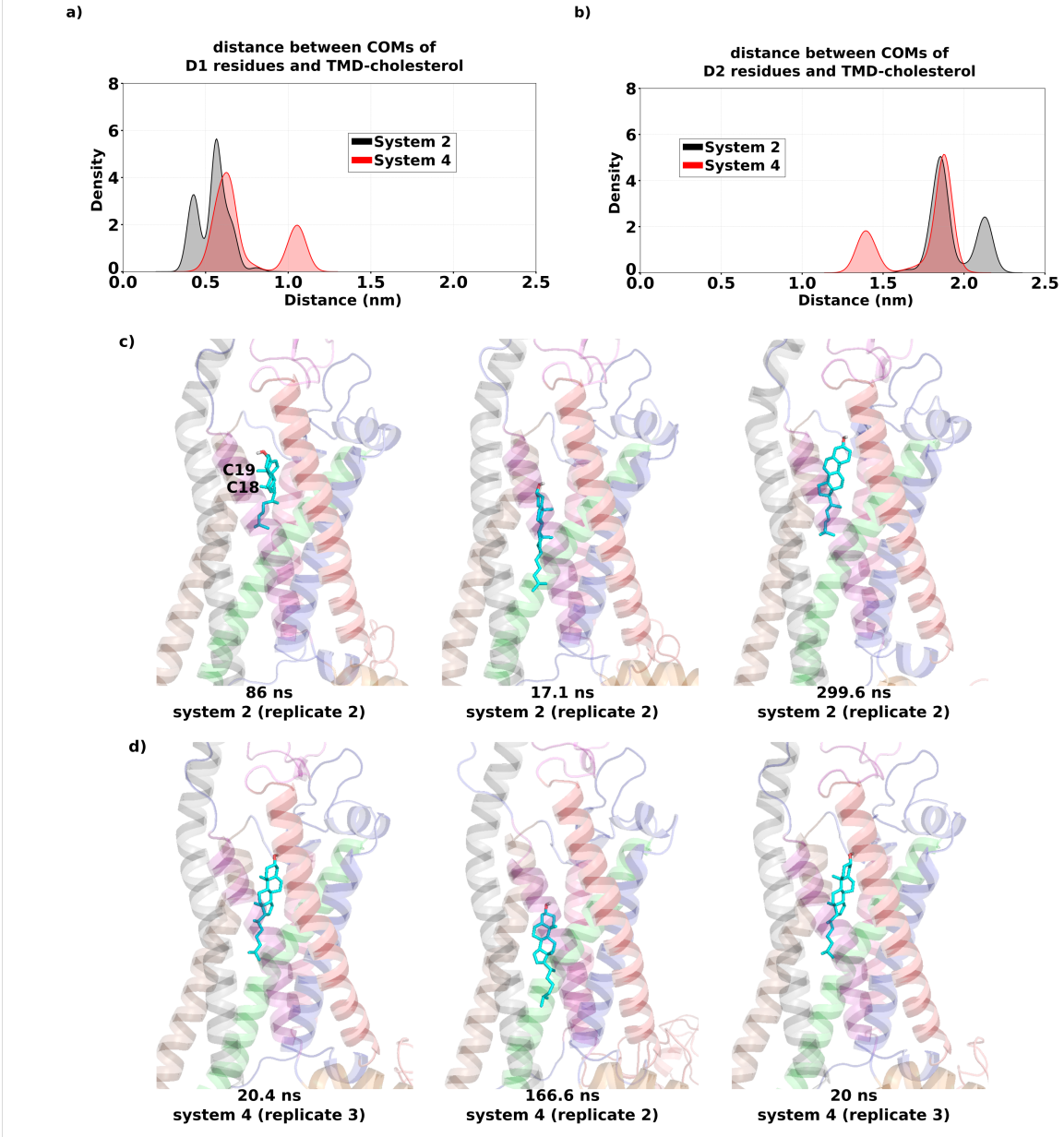


Figure 5.5: Movement of TMD-bound cholesterol. a) Shows the distance between COMs of D1 residues (from upper TMD) and bound cholesterol, and b) shows the distance between COMs of D2 residues (from lower TMD) and bound cholesterol (For details of D1 and D2 residues see Figure 5.4). Representative structures of SMO-TMD in c) system 2 and d) system 4 showing the upward and downward movements of bound cholesterol in the different positions. The rotation of bound cholesterol along the long axis is shown by the C18 and C19 (β face) positions near different TM helices. The helices were displayed in transparent ribbon representation with red, blue, green, pink, brown, gray, purple, and orange colors, respectively. In contrast, the TMD-bound cholesterol was shown using stick representation and colored cyan, and oxygen atoms were depicted in red.

Table 5.3: SMO residues with positions and occupancy percentages within 0.4 nm of TMD and CRD-bound cholesterol. * indicates H-bond with bound cholesterol.

Residues within 0.4 nm (4 Å) of CRD-bound cholesterol with > 30.00%					
Residue position	**CRD binding groove**	**System 2**	**Occupancy (%)**	**System 4**	**Occupancy (%)**
CRD	D95	D95*	48	D95*	39
CRD		K105*	80	K105*	98
CRD	L108	L108*	100	L108	100
CRD	W109	W109*	100	W109	100
CRD		G111	100	G111	100
CRD	L112	L112	100	L112	100
CRD		N114	88	N114	95
CRD		A115	83	A115	99
CRD	Y130	Y130*	62	Y130*	87
CRD	I156	I156	100	I156	100
CRD	V157	V157	100	V157	100
CRD		E160	94	E160	74
CRD		R161*	89	R161	99
CRD		G162*	78	G162	79
CRD	P164	P164	95	P164	99
CRD		F166	37	-	-
LD		L197	98	L197	99
LD		Y207	78	Y207	82
LD	V210	V210	94	V210	100
LD		C213	34	C213	65
TM6		V488	85	V488	98
TM6		L489	78	L489	100
TM6	Q491	-	-	-	-
TM6	A492	A492	85	A492	100
TM6		N493	37	N493	95
TM6	I496	I496	35	-	-
Residues within 0.4 nm (4 Å) of TMD-bound cholesterol with > 30.00%					
Residue position in TMD	**Site 1**	**System 2**	**Occupancy (%)**	**System 4**	**Occupancy (%)**
Lower (TM2)		V270	68	V270	85

Table 5.3 – continued from previous page

Middle lower (TM2)		C273	30		
Middle (TM2)		F274	100	F274	100
Upper (TM2)	W281	W281	84	W281	77
Middle (TM3)	L325	L325	100	L325	100
Middle upper (TM3)		M326	83	M326	55
Middle (TM3)		G328	87	G328	98
Middle (TM3)	V329	V329	100	V329	100
Middle lower (TM3)	F332	F332	62	F332	99
Lower (TM3)		-	-	L335	32
ECL2		D384*	43	-	-
ECL2		S387*	85	S387*	56
ECL2		I389	83	I389	59
ECL2		F391	100	F391	100
ECL2	Y394*	Y394*	73	Y394*	71
Upper (TM5)		R400*	32	R400*	92
Upper (TM5)	V404	V404	92	V404	99
Upper (TM5)		-	-	L405	30
Middle (TM5)	I408	I408	100	I408	100
Middle (TM5)		-	-	L412	30
Middle lower	A459	A459	40	A459	90
Middle lower (TM6)		F462	38	F462	52
Middle (TM6)		V463	100	V463	100
Middle (TM6)		T466	100	T466	100
Upper (TM6)	H470	H470	91	H470*	47
Upper (TM7)		E518*	43	E518*	31
Upper (TM7)	N521	N521	100	N521*	100
Upper (TM7)		L522	86	L522	80
Middle (TM7)		A524	56	A524	88
Middle (TM7)	M525	M525	100	M525	100
Middle lower (TM7)		T528	100	T528	100
Lower (TM7)		M532	63	M532	95

Apart from the TM3 residue F332 (62% in system 2 and 99% in system 4) and the TM6 residues A459 (40% in system 2 and 90% in system 4) and H470 (91% in system 2 and 47% in system 4), all the residues belonging to the sterol binding site 1 have greater than 75% occupancy with TMD-bound cholesterol in both systems 2 and 4, respectively. It may be noted that F332 and A459, both situated in the lower middle TMD region, having much higher occupancy in system 4 compared to that in system 2 and H470, situated in the upper TMD region, having much higher occupancy in system 2 compared to that in system 4, reinforces the differential cholesterol dynamics observed above from the snapshots (Figures 5.5c and 5.5d). The additional residues that came within the 0.4 nm cut off range in both the systems also reveal the same trend. That is, most of the upper TMD region residues have a higher percentage occupancy in system 2, while the majority of the residues from the lower TMD region have a higher percentage occupancy in system 4 (Table 5.3). Of the additional interactions observed in only one of the systems, D384 (only in system 2) belonging to ECL2 and L335 belonging to the lower middle region of TM3 are noteworthy. They are in tune with cholesterol binding to the upper and lower middle TMD regions in systems 2 and 4, respectively.

Apart from its up-down movement, the snapshots in Figures 5.5c and 5.5d also suggest that the cholesterol molecule rotated about its long axis during the simulation. Examination of TMD residues within a 0.4 nm cut off from C19 of TMD bound cholesterol shows that during the simulation, the two methyl groups C18 and C19 on its rough β face frequently orient towards TM1, TM3, TM5, TM6, and TM7, while the smooth α-face gets oriented towards the other two TM helices (TM2 and TM4). Interestingly, a detailed interaction analysis shows differences in the interaction pattern in the two systems, respectively (Figures 5.5c and 5.5d).

In summary, TMD-bound cholesterol is highly dynamic. Within the TMD binding site, it moves up and down, and also rotates about its long axis. The dynamics is however different in the two systems. The dynamics in system 4 appears to be restricted compared to that in system 2. This is possibly due to the presence of additional cholesterol in the CRD binding site, which could be pushing the TMD-bound cholesterol from the upper TMD region towards the middle or lower middle TMD region. The resultant shift explains the differences in the interaction environment of cholesterol in the two systems. This restriction in dynamics can also be inferred from RMSF plots (Figure 5.7a) of TMD-bound cholesterol in the two systems, where the RMSF in system 2 is significantly higher than that in system 4. The RMSF plot also reveals the high flexibility of the isooctyl chain of cholesterol.

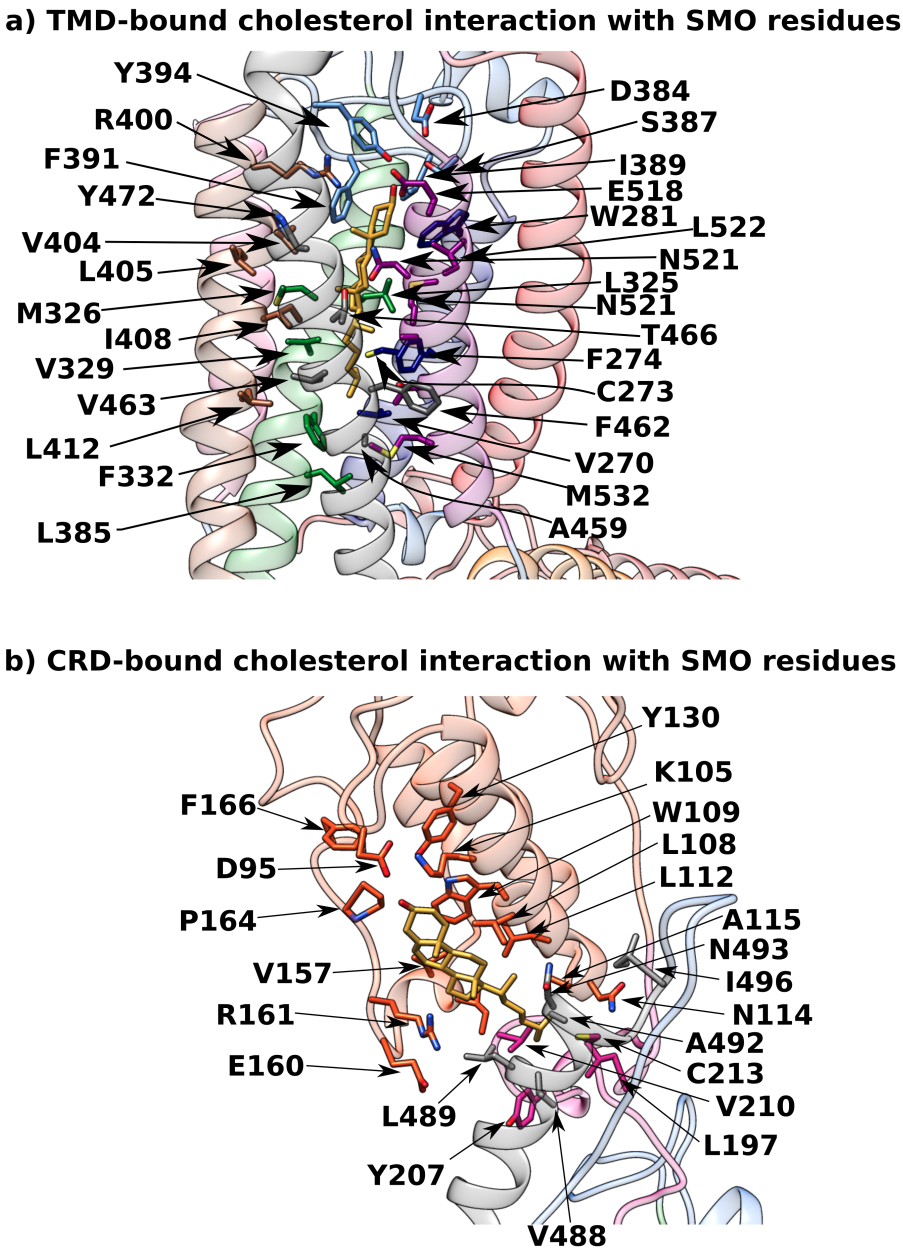


Figure 5.6: A representative structure from system 4 shows SMO residues within a 0.4 nm cutoff around a) TMD-bound and b) CRD-bound cholesterol.

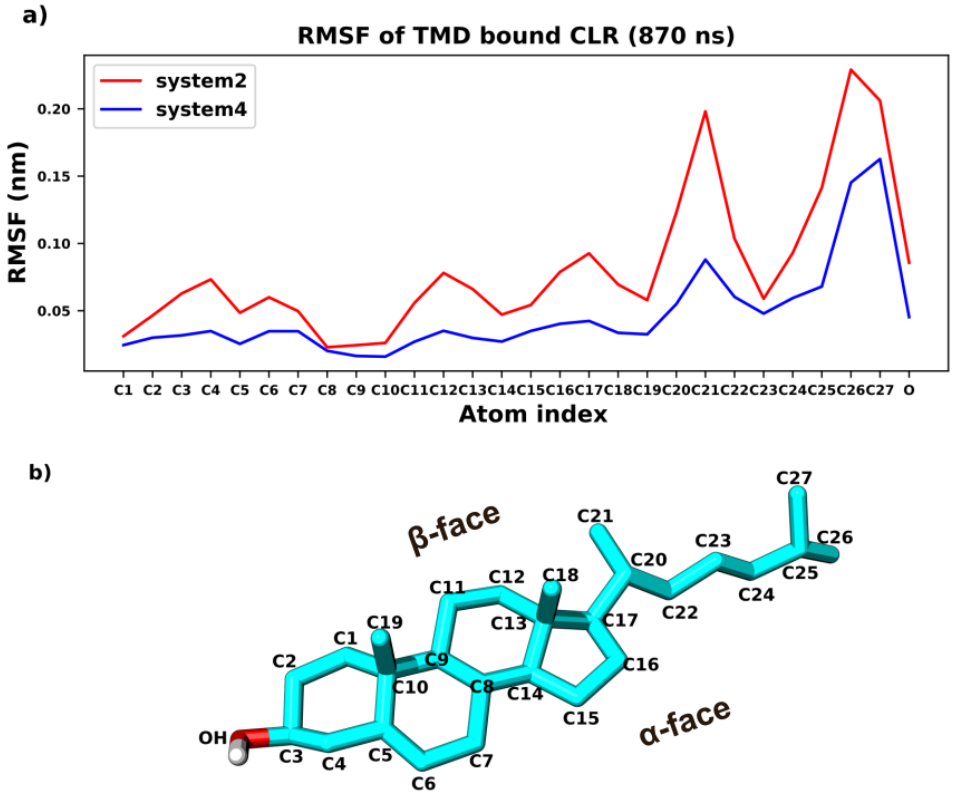


Figure 5.7: a) RMSF of TMD-bound cholesterol in systems 2 (in red) and 4 (in blue); b) structure of cholesterol with atom names and numbers.

5.3.3 Dynamics of CRD-bound cholesterol

In the initial structures of systems 3 and 4, cholesterol was placed in the CRD binding groove (residues: D95**, L108*, W109*, L112, Y130**, I156, V157, P164, V210, Q491, A492, I496) as in the structure 5L7D (Figures 5.2c and 5.2d) (residues marked ** are involved in both systems; residues marked * are involved in only one of the 2 systems). The listing of SMO residues, with greater than 30%, occupancy within 0.4 nm radius of cholesterol bound to CRD (Table 5.4), revealed that in addition to all the above residues, with the exceptions of Q491 and I496, the residues K105**, G111, N114, A115, E160, R161*, G162*, L197, Y207, V488, L489, N493, and C213, interacted with CRD-bound cholesterol in the MD trajectories of both systems 3 and 4. Interestingly, with the initial binding groove residue Q491 missing from both the lists, the trajectory of system 3 showed contacts with two additional residues compared to those with system 4: the initial binding groove residue I496, and an additional residue F166 (Figure 5.6b and Table 5.4). It may also be noted that out of all the residues

listed above, the list of CRD residues involved in H-bonds with CRD-bound cholesterols (Table 5.4), implicated the involvement of the residues D95, K105, and Y130 in both systems 3 and 4, while the residues L108, W109, R161, and G162 were implicated only in system 3. These observations suggest that, in system 4, the CRD-cholesterol interaction pattern is affected by the presence of cholesterol in the TMD binding site.

Table 5.4: H-bond between bound cholesterol and SMO residues.

System name	Replicate	Donor	Hydrogen	Acceptor
Cholesterol bound to SMO-TMD (System 2)				
System 2	Rep1	S387OG	S387HG1	CLR788O
		Y394OH	Y394HH	CLR788O
		R400NH1	R400HH11	CLR788O
		R400NH2	R400HH21	CLR788O
		CLR788O	CLR788H46	E518OE1
		CLR788O	CLR788H46	E518OE2
System 2	Rep2	S387OG	S387HG1	CLR788O
		Y394OH	Y394HH	CLR788O
		R400NH1	R400HH11	CLR788O
		R400NH2	R400HH21	CLR788O
		CLR788O	CLR788H46	D384OD1
		CLR788O	CLR788H46	D384OD2
		CLR788O	CLR788H46	E518OE1
		CLR788O	CLR788H46	E518OE2
System 2	Rep3	S387OG	S387HG1	CLR788O
		Y394OH	Y394HH	CLR788O
		R400NH1	R400HH11	CLR788O
		R400NH2	R400HH21	CLR788O
		CLR788O	CLR788H46	E518OE1
		CLR788O	CLR788H46	E518OE2
Cholesterol bound to SMO-CRD (System 3)				
System 3	Rep1	K105NZ	K105HZ1	CLR788O
		Y130OH	Y130HH	CLR788O
		R161N	R161HN	CLR788O

Continued on next page

Table 5.4 – continued from previous page

		R161NH2	R161HH21	CLR788O
		CLR788O	CLR788H46	D95OD2
		CLR788O	CLR788H46	L108O
		CLR788O	CLR788H46	W109NE1
		CLR788O	CLR788H46	G162O
System 3	Rep2	K105NZ	K105HZ1	CLR788O
		Y130OH	Y130HH	CLR788O
		CLR788O	CLR788H46	D95OD1
		CLR788O	CLR788H46	D95OD2
		CLR788O	CLR788H46	G162O
System 3	Rep3	K105NZ	K105HZ1	CLR788O
		CLR788O	CLR788H46	D95OD1
		CLR788O	CLR788H46	D95OD2
		CLR788O	CLR788H46	Y130OH
System4: TMD-bound cholesterol (index 788) SMO				
System 4	Rep1	Y394OH	Y394HH	CLR788O
		CLR788O	CLR788H46	S387OG
		CLR788O	CLR788H46	E518OE1
		CLR788O	CLR788H46	E518OE2
System 4	**Rep2**	Y394OH	Y394HH	CLR788O
		R400NH1	R400HH11	CLR788O
		R400NH2	R400HH12	CLR788O
		CLR788O	CLR788H46	H470ND1
		CLR788O	CLR788H46	H470ND2
		CLR788O	CLR788H46	E518OE2
		CLR788O	CLR788H46	N521ND2
System 4	Rep3	S387OG	S387HG1	CLR788O
		Y394OH	Y394HH	CLR788O
		R400NH1	R400HH11	CLR788O
		R400NH2	R400HH12	CLR788O

Continued on next page

Table 5.4 – continued from previous page

		CLR788O	CLR788H46	E518OE1
		CLR788O	CLR788H46	E518OE2
System4: CRD-bound cholesterol (index 789) SMO				
System 4	**Rep1**	K105NZ	K105HZ1	CLR789O
		Y130OH	Y130HH	CLR789O
		CLR789O	CLR789H46	D95OD1
		CLR789O	CLR789H46	D95OD2
System 4	**Rep2**	K105NZ	K105HZ1	CLR789O
		Y130OH	Y130HH	CLR789O
		CLR788O	CLR788H46	D95OD1
		CLR788O	CLR788H46	D95OD2
System 4	**Rep3**	K105NZ	K105HZ1	CLR789O
		Y130OH	Y130HH	CLR789O
		CLR788O	CLR788H46	D95OD1
		CLR788O	CLR788H46	D95OD2

5.3.4 Presence of cholesterol in SMO-TMD limits the flexibility of all domains

Domain-wise root mean square deviation (RMSD) of SMO was calculated for all simulated systems to see the influence of the bound cholesterol on the structure and stability of its domains. The density of domain-wise RMSD values (Figure 5.8) suggests that the non-TM domains (except LD) significantly contributed to the observed structural dynamics in SMO in all systems. Though the RMSD of LD was less compared to that for other domains, this decreased even more in the cholesterol-bound systems compared to that in the cholesterol-free system. The root mean square fluctuations (RMSF) plot also reflected the relatively decreased flexibility of LD in the cholesterol-bound systems than in the cholesterol-free system (Figure 5.9a). The overall RMSD for the TMD did not change significantly across the systems (the detailed TM-wise observations are reported in the next section). The RMSD values of CRD and ICD were highest in system 3 (holo SMO-CRD), followed by slightly lower values in system 4 (cholesterol bound to both the TMD and CRD binding sites). Interestingly, with the RMSD values of CRD and ICD being lowest in system 2 (holo SMO-TMD with no cholesterol in CRD),

the major changes in structural dynamics appears to take place in response to the binding of cholesterol to CRD. Further, the RMSF plot shows that the fluctuations of N-terminal residues of CRD were consistently higher in the absence of cholesterol (Figure 5.9a). The visual inspection of the trajectory of the system 1 (apo SMO) also showed the movement of CRD N-terminal residues toward the membrane. The RMSF of ICD was higher in the cholesterol-bound system 3 and 4 than that of the cholesterol-free system. The noticeable point, which also supports the RMSD analysis, is the decrease in the RMSF value of ICD in cholesterol-bound system 2, where cholesterol is present in TMD. The RMSD and RMSF patterns indicate that the presence of cholesterol in CRD makes CRD and ICD more flexible (in both systems 3 and 4), whereas the presence of cholesterol only in TMD limits the flexibility of CRD and ICD.

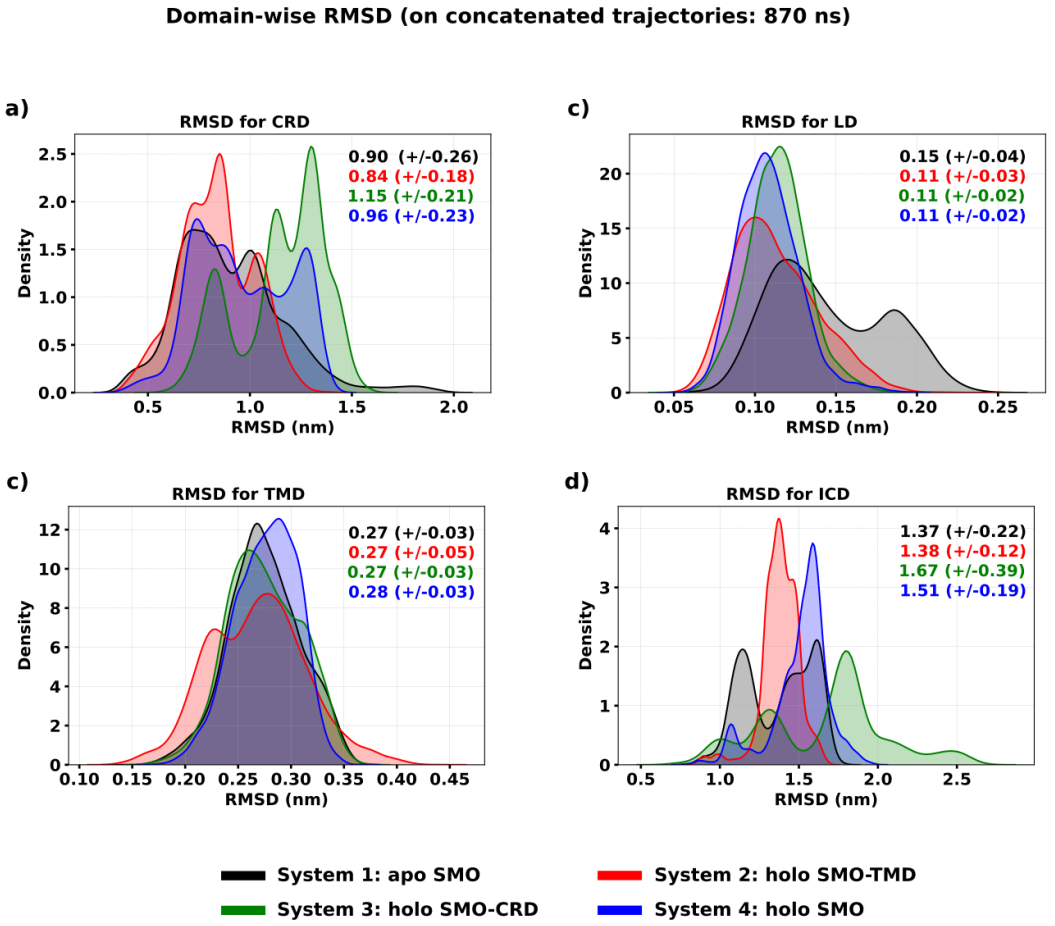


Figure 5.8: The histogram of domain-wise RMSD for a) CRD, b) LD, c) TMD, and d) ICD Cα atoms were calculated for the concatenated trajectories (870 ns) of system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). System 2 shows a decrease in the RMSD of CRD and ICD. The average RMSD value for each system is written in the corresponding color for that system.

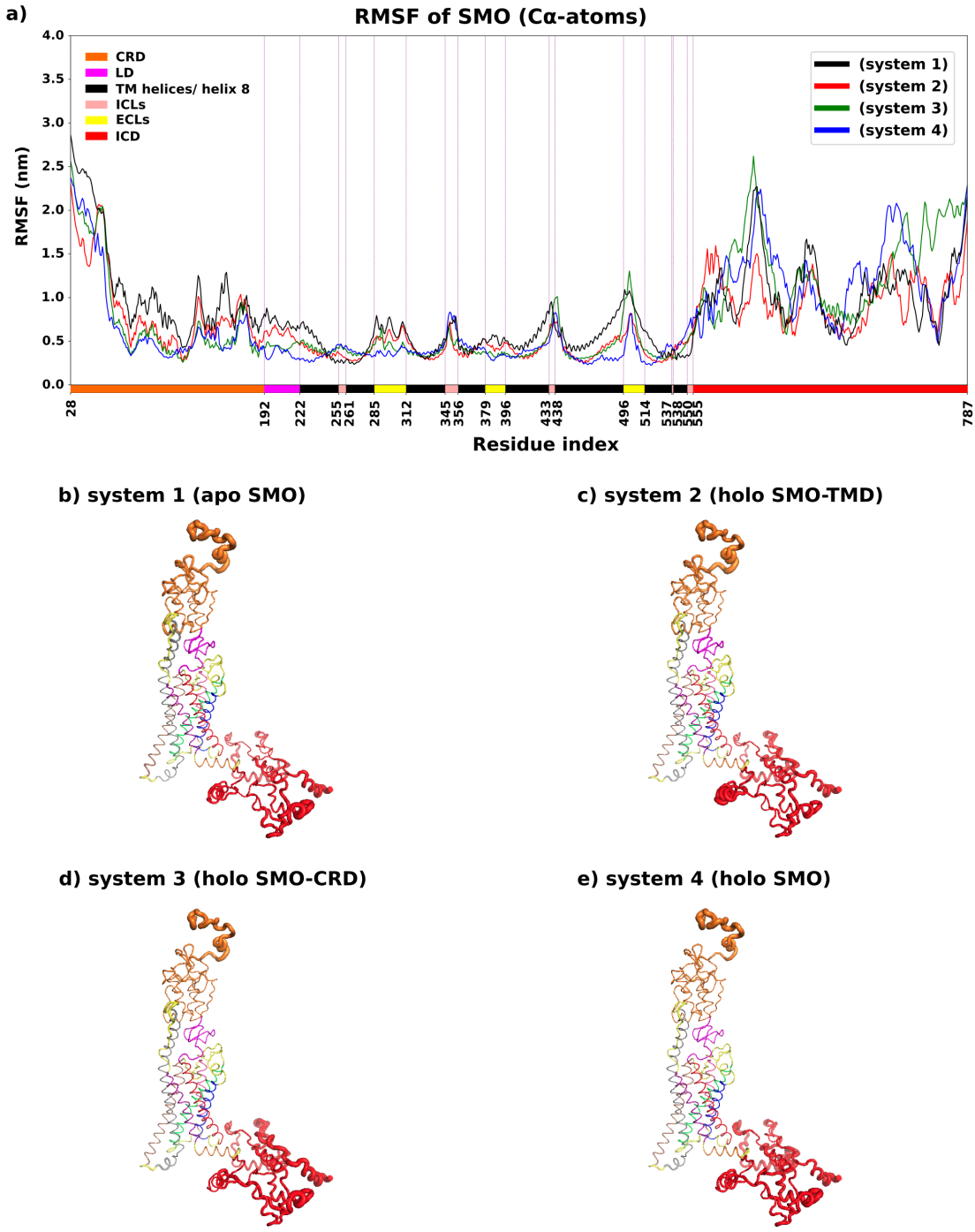


Figure 5.9: a) RMSF of SMO in system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). The residue-wise RMSF of SMO was mapped onto the modeled SMO structure and shown in panels b-e, with the thickness of the ribbon indicating low to high fluctuation. The TM helices are colored according to the scheme used in Figure 5.5.

5.3.5 Different dynamics of TM6, TM7, and helix 8

Individual TM RMSDs and helix tilt angles were studied in all four systems to capture subtle changes in TMD (Figures 5.10 and 5.11). No significant changes were observed in TM1, TM2, TM3, TM4, and TM5 across the systems (Figure 5.10). However, cholesterol binding induces significant changes in the dynamics patterns of TM6, TM7, and helix 8 (Figure 5.10). RMSD of TM6 increases upon cholesterol binding to CRD (system 3), or to both CRD and TMD (system 4), but there is no significant RMSD change upon cholesterol binding to TMD alone (system 2). In contrast, the RMSD of TM7 decreases upon cholesterol binding to either or both the binding sites (systems 2, 3, and 4).

Interestingly, the RMSD of helix 8 shows higher RMSD values in systems 2 and 4 as compared to systems 1 and 3. Tilt angle analysis showed a decrease in the average tilt angle of TM6 in cholesterol-bound systems (especially in single cholesterol-bound systems) compared to the cholesterol-free system, suggesting the straightening of TM6 relative to the membrane plane (Figure 5.11). In contrast, TM7 tilted more towards the membrane in cholesterol-bound systems, indicating a change in its conformation (Figure 5.11).

The dynamics of extracellular and intracellular loops (ECLs and ICLs) were examined by RMSF analysis (Figure 5.9). ECL1 and ECL3 dynamics were reduced in the presence of cholesterol in either or both binding sites. At the same time, ECL2 flexibility is limited when cholesterol is bound to the CRD binding site. There is no significant change in the flexibility of ICLs. Here, the change in the RMSD value in TM6, TM7, and helix 8, underline the structural changes observed in these helices with respect to the initial structure. This showed that cholesterol binding to TMD (system 2) led to an increase in the RMSD of helix 8. In contrast, cholesterol binding to CRD (system 3) resulted in an increase in the RMSD of TM6. Interestingly, the binding of cholesterol to both TMD and CRD (system 4) or either one of these (systems 2 and 3) caused a decrease in the RMSD of TM7.

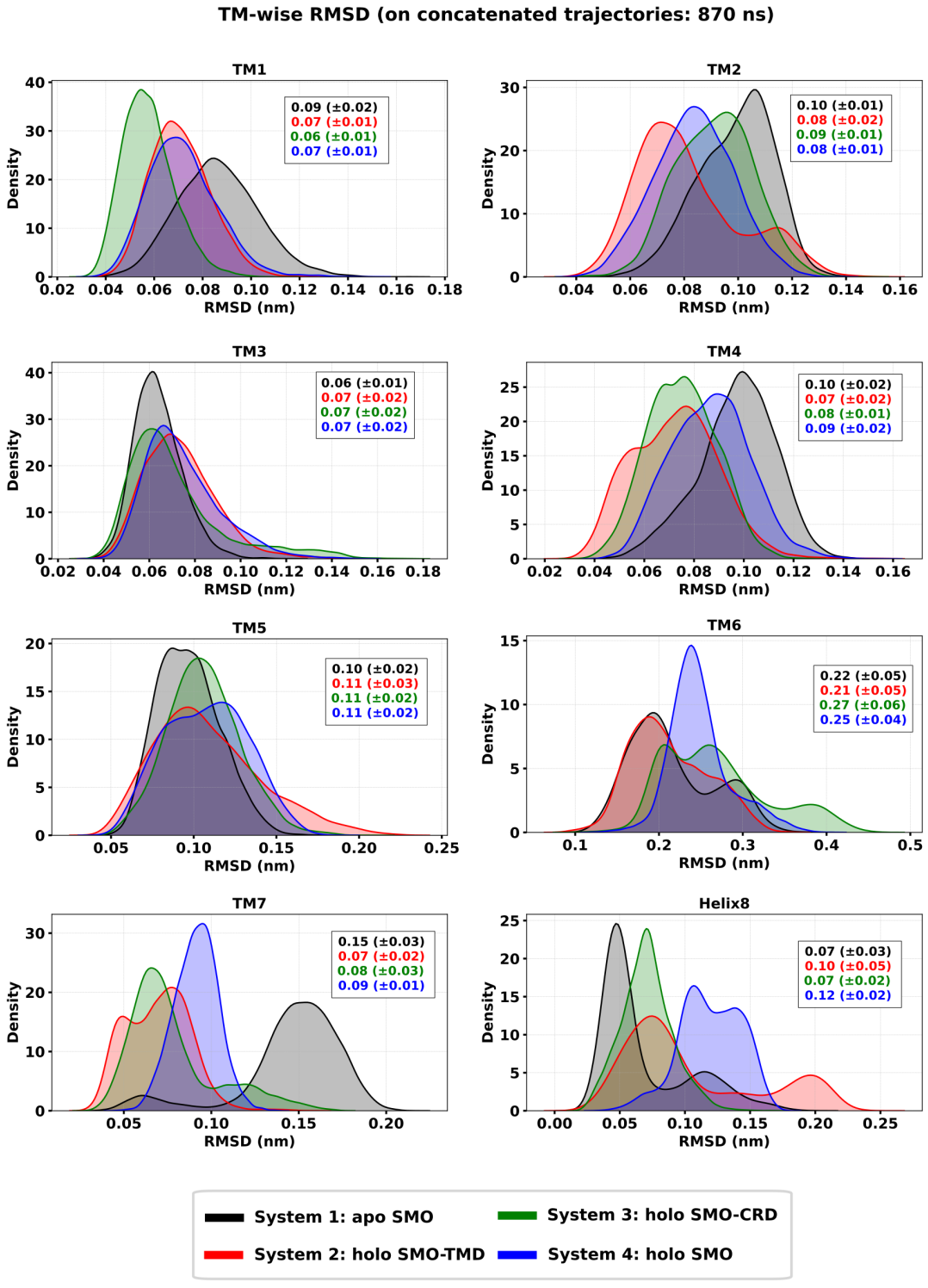


Figure 5.10: TM-wise RMSD histograms for a) TM1, b) TM2, c) TM3, d) TM4, e) TM5, f) TM6, g) TM7, and h) helix 8 calculated for the concatenated trajectories (870 ns) of system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). The average RMSD value for the respective system is written in the box for each TM helix, including helix 8.

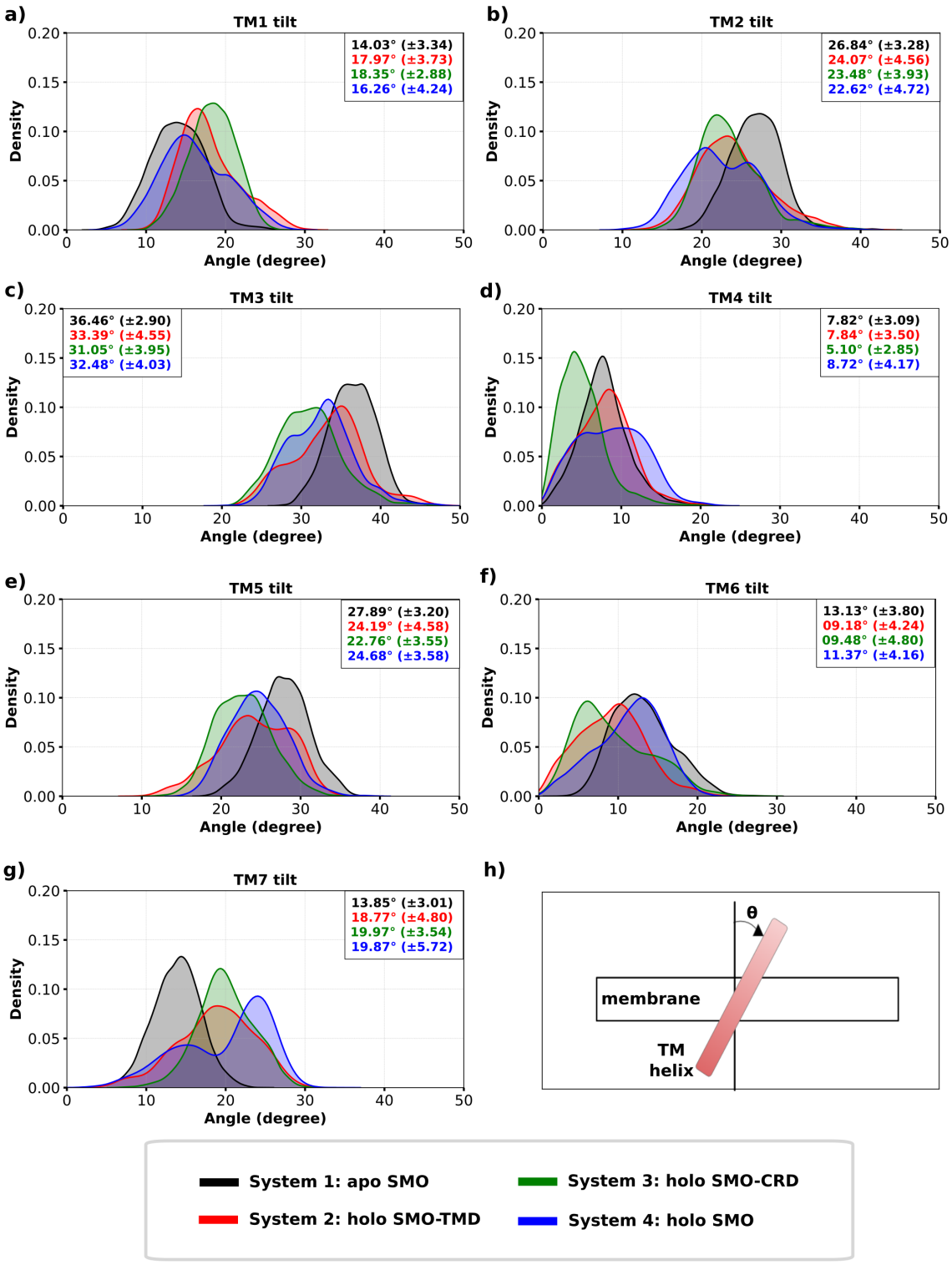


Figure 5.11: Axis tilts of TM helices. a-g) histogram of the tilts of the individual TM helices with respect to the z-axis for simulation system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue); and h) a schematic representation showing axis tilt calculation for TM helices. The average tilt value for the respective system is written in the box for each TM helix.

5.3.6 Loss of cation-π interaction observed in all systems

A cation-π interaction between R451 in TM6 and W535 in TM7, (Figure 3.1; Chapter 3), reported in some of the crystal and cryo-EM structures of SMO, has been hypothesized to characterize the cholesterol-free inactive state of SMO, and has been suggested to break during its cholesterol binding induced transition to the active state [13, 68, 158]. This 'ionic lock' was monitored by measuring the distance between the COMs of the positively charged guanidine group of R451 and the indole group of W535 (Figure 5.12). This distance was 0.44 nm in the modeled SMO structure, which was used for the preparation of all simulated systems. Interestingly, this ionic lock broke in systems 1 and 2, with the distance values increasing to 0.92 nm and 0.81 nm, respectively, during the equilibration step itself. Further, replica-wise variations notwithstanding, the observed average distance values of this interaction of 0.92 nm, 0.81 nm, 0.8 nm, and 0.87 nm in systems 1, 2, 3, and 4, respectively, (Figure 5.13 and Table 5.5), indicated that the ionic-lock was broken during the respective production runs in all the systems (Figure 5.12). Thus, it appears that the hypothesis, based on crystal and cryo-EM structures, correlating inactive and active states of SMO, respectively with the presence or absence of this ionic-lock is not supported by the current simulation studies (more on this in the discussion section).

Table 5.5: System and replica-wise average distance (in nm) between R451 and W535 of TM6 and TM7 for ionic-lock (cation-π interaction).

Ionic-lock				
System name	**Concatenated trajectory (870ns)**	**Rep1 (290ns)**	**Rep2 (290ns)**	**Rep3 (290ns)**
System 1	0.92 ($\pm$0.14)	0.94 ($\pm$0.11)	0.91 ($\pm$0.09)	0.9 ($\pm$0.18)
System 2	0.81 ($\pm$0.27)	0.48 ($\pm$0.13)	0.89 ($\pm$0.07)	1.07 ($\pm$0.08)
System 3	0.8 ($\pm$0.3)	0.45 ($\pm$0.07)	1.03 ($\pm$0.25)	0.91 ($\pm$0.15)
System 4	0.87 ($\pm$0.12)	0.79 ($\pm$0.14)	0.86 ($\pm$0.1)	0.93 ($\pm$0.11)

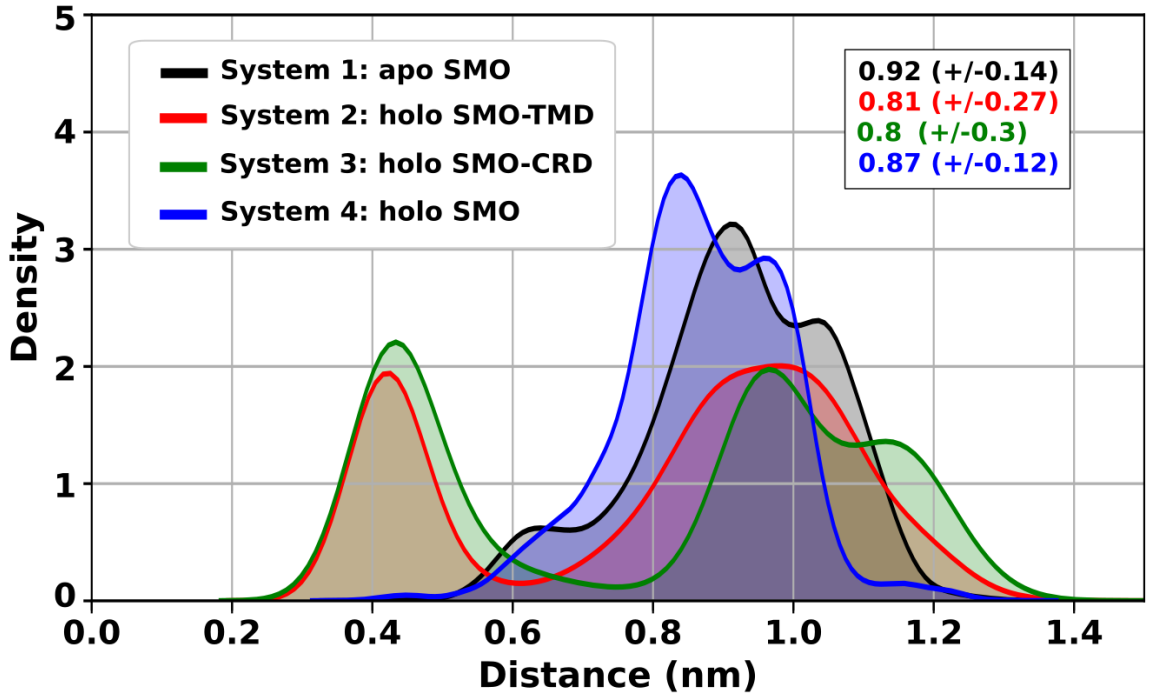


Figure 5.12: Cation-π (ionic lock) interaction between R451 (TM6) and W535 (TM7). Histogram of the distances between the COMs of the guanidine group of R451 (TM6) and the indole group of W535 (TM7) calculated for system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). The average distance value is near ∼0.9 nm for all systems, suggesting a loss of this interaction. A peak near ∼0.4 nm in systems 2 and 3 indicates a persistent interaction between R451 and W535 in replicate 1 of both systems (see Figures 5.13b and 5.13c).

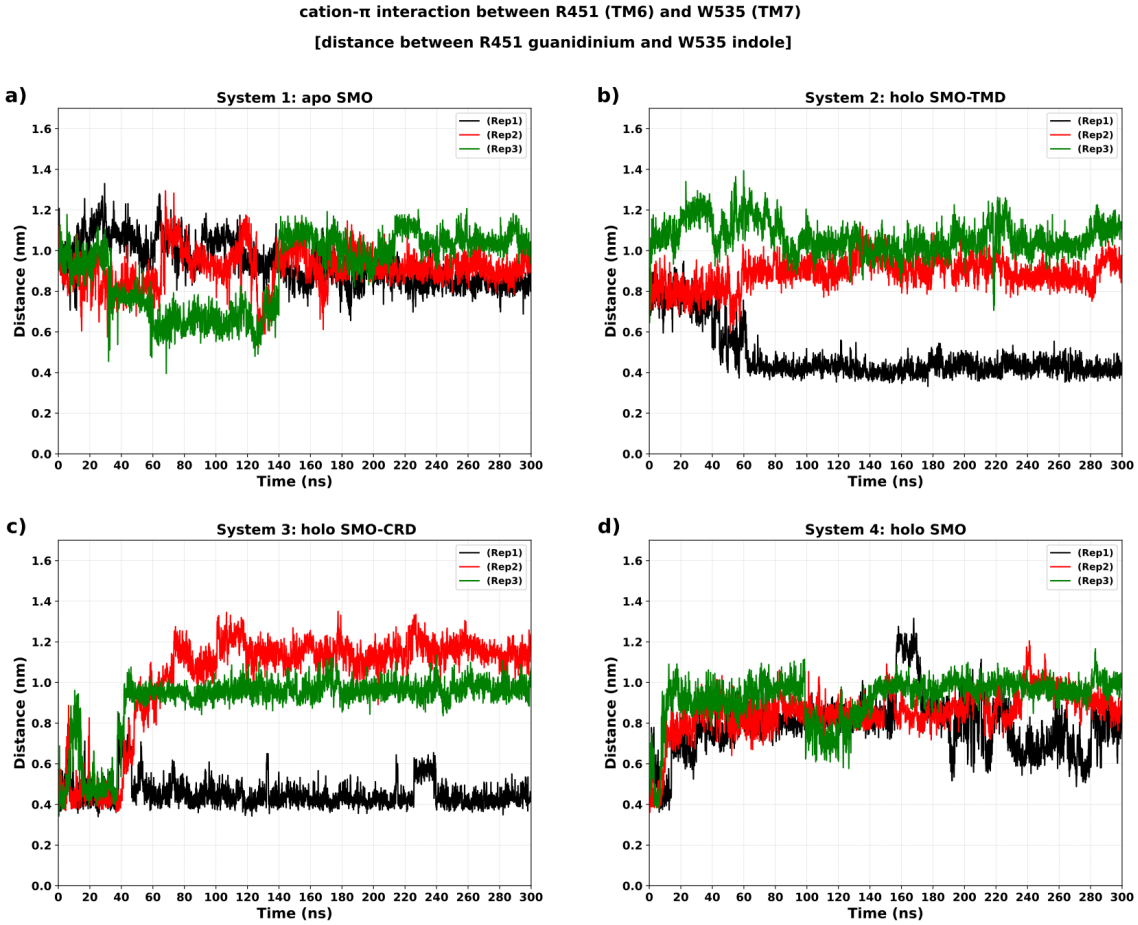


Figure 5.13: Replicate-wise distances between the COMs of the guanidine group of R451 (TM6) and the indole group of W535 (TM7), respectively, for a) system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). Replicate 1 in both systems 2 and 3 have less distance value between R451 and W535, contributing to the curve near ∼0.4 nm in the main Figure 5.12.

5.3.7 The primacy of interaction shifts from D-R to E-R within the DRE network in system 4 (holo SMO)

The salt-bridge interactions between D473 (TM6), R400 (TM5), and E518 (TM7), constitute a DRE network [13] situated at the extracellular end of TMD (Figure 3.1; Chapter 3). Variations in the distances between the COMs of the side chains of interacting residues, observed in the simulation trajectories of all the systems, were examined to understand the role of cholesterol binding on the DRE network (Figure 5.14). As shown in Figure 5.14a, the average interaction distance between D473 and R400, of the DRE salt bridge network, for the cholesterol-free system (system 1) and the single cholesterol-bound systems (systems 2 and 3) were respectively significantly shorter (∼0.2 - 0.25 nm) than that in the holo system (system 4, 0.44 nm). Interestingly, the histograms corresponding to the former set show peaks at around 0.2 and 0.4 nm respectively, which differ remarkably from the range spanning from around 0.3 nm to 0.6 nm in the holo system histogram. In contrast, as shown in Figure 5.14b, the average respective interaction distance between E518 and R400 in systems 1, 2, and 3 (∼ 0.6 nm) were significantly longer than that for the system 4 (0.39 nm). The enhanced strength of the E-R interaction in system 4 is highlighted by the presence of the two sharp peaks at ∼0.2 nm and ∼0.4 nm respectively, in the corresponding histogram.

The DRE interactions between D473, R400, and E518 in two representative snapshots from system 4 replicate 2 at 36.6 ns and 214.8 ns were shown by superimposing the TMD (Figure 5.14c). In the 36.6 ns structure, the close proximity of D473 and R400 was observed, with the main interactions formed between the oxygen atom of the negatively charged D473 residue and the nitrogen atom of the positively charged R400 residue. However, in the 214.8 ns snapshot, a shift of R400 towards E518 was observed. The existence of a salt-bridge interaction between R400 and D473 in system 4 was also indicated by a value around 0.4 nm, as shown in Figure 5.14a. It was previously reported that E518 also interacts with TMD cholesterol in system 4, which could possibly have affected the interaction between R400 and D473. Overall, this DRE network analysis suggests that the extracellular region of TM6-TM5 was close to each other in the case of cholesterol-free and single cholesterol-bound systems (systems 1, 2, and 3). In contrast, the extracellular region of TM5-TM7 was close to each other when cholesterol is bound to both the sites (system 4). It may thus be concluded that within the DRE network, the primacy of the interaction shifted from D-R to E-R in system 4 (holo SMO).

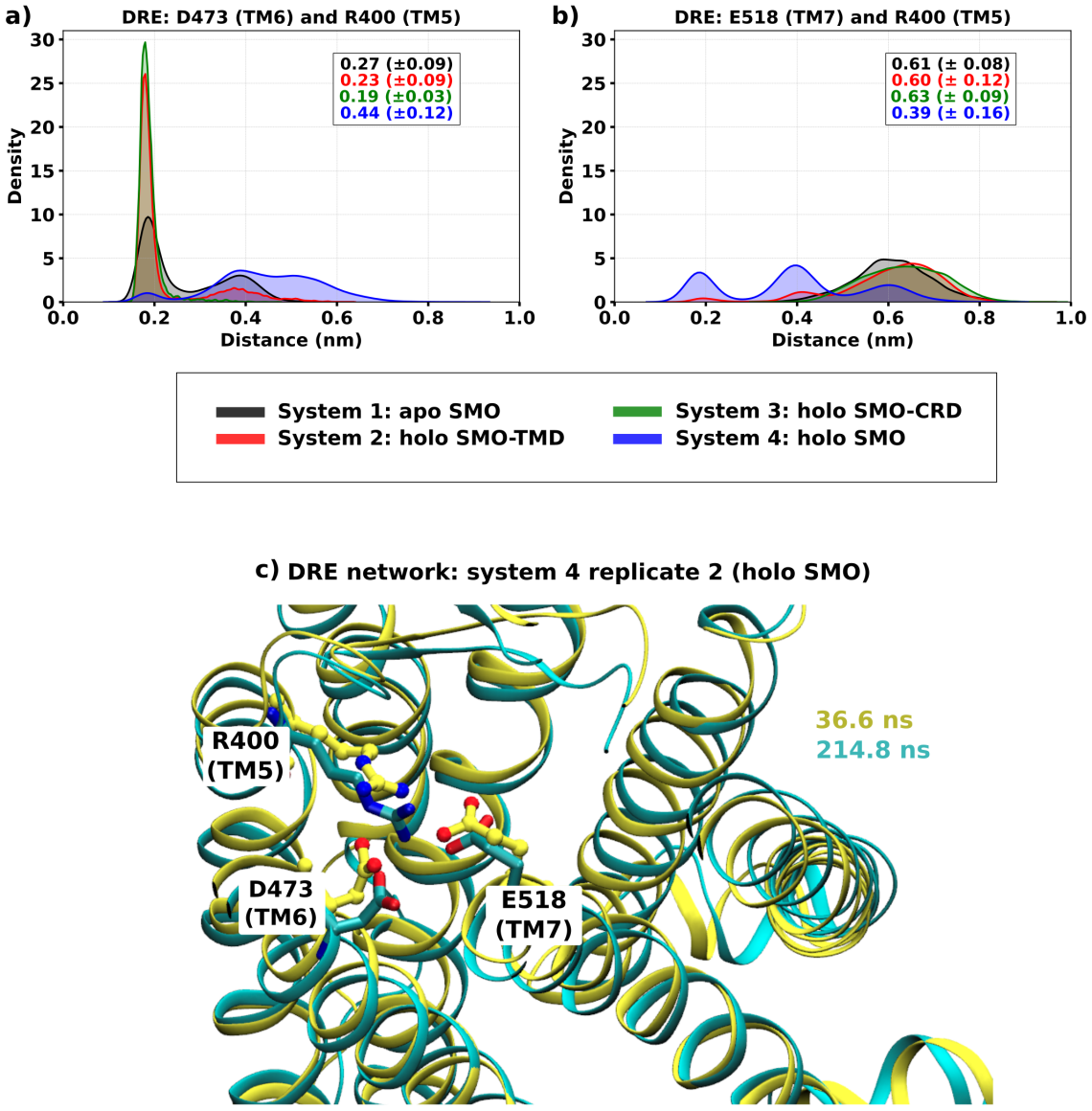


Figure 5.14: DRE network at extracellular TMD region. Histogram of the distances between the COMs of the side chains of interacting residues a) D473 (TM6) and R400 (TM5), b) E518 (TM7) and R400 (TM5) calculated for system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). c) The structural superimposition of two representative TMD snapshots from simulated system 4 at 36.6 ns (yellow) and 214.8 ns (cyan) shows salt bridge interaction between the D-R and R-E by the nearness of R400 and E518.

5.3.8 Membrane cholesterol interactions with cholesterol-binding motifs

The interaction occupancy of membrane cholesterol molecules with identified cholesterol-binding motifs (Figures 3.3 and 5.15a, Table 5.6, and Video 5.1) was calculated to study the changes in membrane-regulated dynamics of SMO from cholesterol-free system to cholesterol-bound systems. An interaction is categorized as high, moderate, and low, respectively, depending on whether the occupancy percentage is greater than 50, between 30 and 50, or less than 30. Table 5.6 suggests that the interactions of motifs vary across different systems. The motifs, *strict* CCM (lower TM4 and TM2), CRAC4 (middle TM5), CARC9 (lower TM6), CRAC-like4 (lower TM5), and `TMD_6O3C` (within TMD lower region) located at lower TMD regions have higher interaction with membrane cholesterol in system 1. Compared to system 1, these interactions in system 2 were reduced to moderate or low interactions. However, the interactions of these motifs increase to a high level in system 3, which further increases in occupancy in system 4. With the exception of CRAC4, interaction occupancy is lower than in system 3 but still shows high interaction in system 4, and the CRAC-like4 interaction becomes less significant. Another motif, the CRAC2 (ICL2) interaction with membrane cholesterol, falls in the moderate range in system 1 (50.89%), decreases in all cholesterol-bound systems, and is not significant in systems 3 and 4. Further, CARC7 (lower TM4) shows higher interaction in all systems. The interaction occupancy of CARC7 was lower in system 2 compared to system 1, but it gradually increased in systems 3 and 4. The interaction between the CBP (lower TMD) and CRAC-like1 (lower TM1 and ICL1) motifs is higher in systems 1 and 2, decreases in systems 3, and then rises to a high level in system 4. This implies that cholesterol binding to TMD has an impact on these motifs, increasing the interaction. In contrast, the interaction of `TMD_6O3C` binding residues (experimentally identified cholesterol binding site inside lower TMD) was found to be higher in system 1 but disappears in system 2, and increases in system 3. However, this interaction also becomes high in system 4. The other two cholesterol-binding sites within the TMD, sites 1 and 2, which are located in the middle and upper TMD regions, respectively, exhibit moderate interaction in system 1 and disappear in cholesterol-bound systems. Interestingly, CARC11, which is present in helix 8, exhibits moderate interaction in systems 1 and 3, diminishes to negligible interaction in system 2, but exhibits high interaction in system 4. Some motifs, such as CARC5 (ICL1 and TM2 lower), CARC12 (ICD), CARC13 (ICD), CARC4 (LD), and CARC10 (TM6 extended region), have consistently low interactions across all systems.

CRAC-like2 (ECL1 upper) interaction is interestingly high in system 1 and gradually declines in systems 2 and 3 (still exhibit high interaction), which unexpectedly shows moderate interaction in system 4. However, despite a slight decline in the level of interaction in system 3, the interaction of TM2/3e (the extracellular portion of TM1, TM2 helices, and ECL1) with membrane cholesterol is still strong. Finally, cholesterol-binding motifs present in ECL2, CRAC-like3 (ECL2), CRAC3 (ECL2 and upper TM5), and CARC8 (ECL2 and upper TM5)

show a moderate level of interaction with membrane cholesterol in system 1, which disappeared in system 2. But their interaction increases to a very high level in systems 3 and 4 (where cholesterol is bound to CRD). These results imply that the interactions between membrane cholesterol and cholesterol-binding motifs are highly system-dependent, which, in turn, is indicative of their correlation with cholesterol binding induced conformational changes in ECD and TMD, respectively.

In summary, it is apparent that most of the motifs interact with more than one cholesterol molecule, both simultaneously as well as asynchronously, in all systems. The observed trend of increase in membrane cholesterol interactions with these motifs located in the middle and lower TMD regions (CRAC4, CARC7, and CARC9) and the upper TMD region (CARC8, CRAC3, and CRAC-like3, TM2/3e) suggests that cholesterol binding to TMD hampers its interaction with membrane cholesterol. In contrast, the presence of cholesterol in CRD enhances these interactions.

Table 5.6: Averaged membrane cholesterol occupancy (in percentage) near cholesterol-binding motifs/sites within the cutoff of 0.5 nm.

S. no.	**Motif name**	**System 1**	**System 2**	**System 3**	**System 4**
1.	CRAC1	0	0	0	0
2.	CRAC2	50.89	40.2	5.28	9.77
3.	CRAC3	40.48	0.56	98.97	99.97
4.	CRAC4	100	27.82	86.63	76.88
5.	CARC1	0	0	0	0
6.	CARC2	0	0	0	0
7.	CARC3	0	0	0	0
8.	CARC4	0.29	0.37	0.02	0.01
9.	CARC5	11.91	1.22	0	11.06
10.	CARC6	48.75	10.49	0.28	0.51
11.	CARC7	83.79	70.28	75.07	99.44
12.	CARC8	48.86	8.49	99.99	100
13.	CARC9	80.61	19.54	71.16	100
14.	CARC10	0.29	0	0	0
15.	CARC11	36.79	25.88	36.86	56.84
16.	CARC12	11.69	0.75	1.21	16.32
17.	CARC13	0	0	0	0.43
18.	CRAC-like1	98.99	91.25	31.46	89

Continued on next page

S. no.	Motif name	System 1	System 2	System 3	System 4
19.	CRAC-like2	99.55	72.87	68.15	30.82
20.	CRAC-like3	39.13	0.52	71.07	99.87
21.	CRAC-like4	76.76	1.1	82.51	26.5
22.	CCM	100	43.01	63.15	100
23.	CRD_5L7D	0	0	0	0
24.	CRD_6O3C	0	0	0	0
25.	TMD_6O3C	97.64	1.99	78.22	99.99
26.	CRD_6D35	0	0	0	0
27.	TM2/3e	100	100	77.81	92.72
28.	TMD_site 1	42.7	0.26	7.22	7.44
29.	TMD_site 2	42.71	0.26	7.22	7.48
30.	TM_LD_site 3	0	0	0	0
31.	CRD_site 4	0	0	0	0
32.	CBP	100	80.55	43.87	100

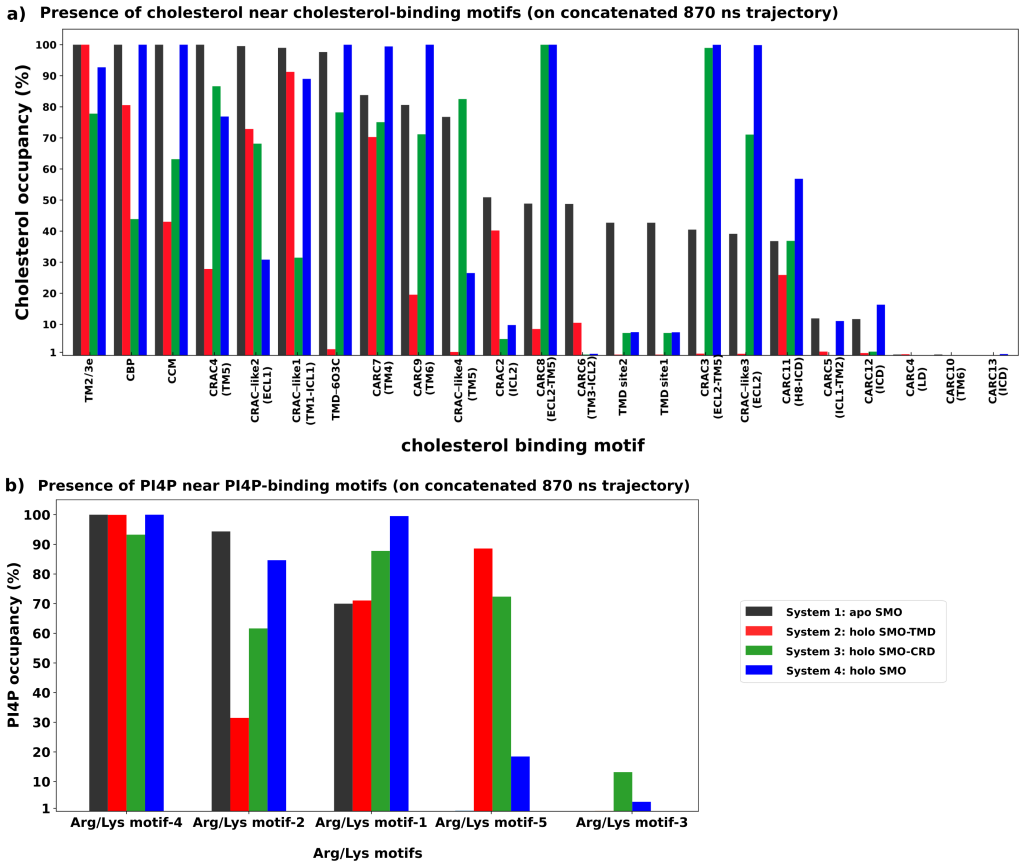


Figure 5.15: The occupancy of ciliary lipids was analyzed in the simulations of system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue). a) shows the occupancy of membrane cholesterols at cholesterol-binding motifs/sites, b) shows the occupancy of PI4P lipids at ICD Arg/Lys clusters. The occupancy values were calculated using the concatenated trajectories for each simulation system. Most cholesterol-binding motifs interact with membrane cholesterol in system 4, which is reduced in the single cholesterol-bound systems. All five Arg/Lys clusters interact with PI4P lipids in systems 3 and 4, where cholesterol is bound to CRD binding domains.

Figure 5.16 displays the atomistic details of the interactions of residues of the cholesterol-binding motifs (CARC7, CRAC4, and *strict* CCM) with membrane cholesterol molecules as observed in selected snapshots at 290 ns from replicate 3 of system 1. CARC7 (TM4) and CRAC4 (TM5) interact with two cholesterols (indexed 882 and 853), while cholesterol 882 is observed to be sandwiched between CARC7 and CRAC4. The OH of cholesterol 882 interacts with F360 of CARC7, and the ring of the same cholesterol interacts with L363. The OH of cholesterol 853 is seen to interact with R; and the ring of this cholesterol interacts with Y417.

Similarly, the OH of bound cholesterol 861 is interacted with by Y269 from TM2 of *strict* CCM, and the ring of this cholesterol interacts with W365 from TM4 of *strict* CCM.

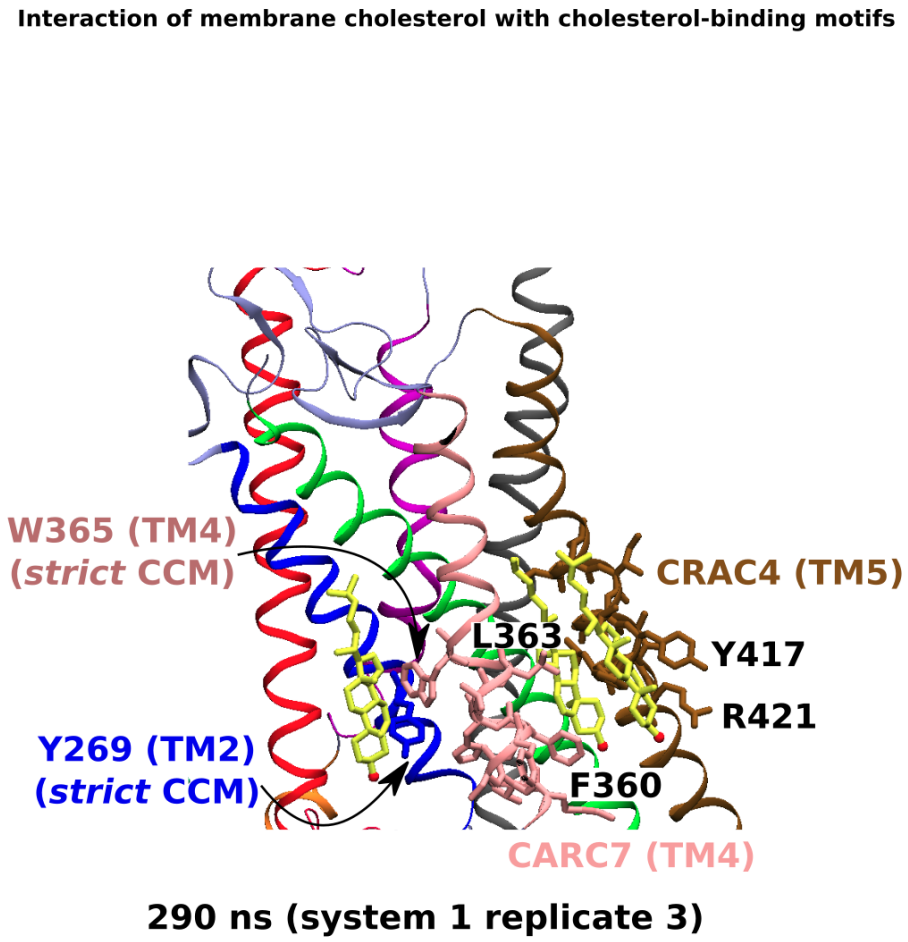


Figure 5.16: Interaction of membrane cholesterols with cholesterol-binding motifs. A representative snapshot at 290 ns from system 1 replicate 3 shows the interacting residues of CARC7 (TM4; pink), CRAC4 (TM5; brown), and strict CCM (W365 from TM4 in pink and Y269 from TM2 in blue) motifs are shown in stick representation. Membrane cholesterols are shown in stick (yellow) representation with OH as a red sphere.

5.3.9 PI4P interacts with all Arg/Lys clusters in systems 3 and 4

SMO has five Arg/Lys clusters in the ICD, and two clusters, Arg/Lys clusters 1 and 4, show higher occupancy for PI4P lipids (Figure 5.15b and Table 5.7). Arg/Lys cluster 2 has the highest PI4P occupancy for systems 1, 3, and 4. However, there is no interaction of Arg/Lys cluster 3 with PI4P in system 1 (apo SMO) and system 2 (cholesterol in the TMD binding site). Arg/Lys cluster 3 seems to have very less occupancy in system 4 (holo SMO) when both binding sites are occupied with cholesterol. But when cholesterol is in the CRD, the occupancy for PI4P is a little higher. Similarly, Arg/Lys cluster 5 shows PI4P occupancy in the presence of cholesterol in either or both cholesterol-binding sites. Interestingly, Arg/Lys cluster 1 also has higher occupancy for PI4P in all systems, especially in system 4 (holo SMO). Other than Arg/Lys clusters, PI4P also interacts with the lower TMD residues of SMO.

Residue-wise PI4P occupancy was calculated for TMD and ICD (Figure 5.17). The analysis shows PI4P occupancy near the lower regions of TM5 and TM3 in system 1. Whereas, in addition to TM5, the lower regions of TM6, TM7 were populated with PI4P in system 2. In contrast, while TM5 does not show higher occupancy for PI4P in system 3, the TM2, TM4, TM6, and TM7 lower TM regions show an affinity for PI4P. In system 4, most interactions appear to be governed by the lower TM region of TM1, TM4, TM5, and helix 8. It is important to note that in systems 3 and 4, the interaction of TM4 with PI4P was common. It can be linked to SMO interaction with PI4P lipid in the modeled ciliary membrane with all Arg/Lys clusters present in the ICD (discussed later).

Table 5.7: Averaged PI4P occupancy (in percentage) near Arg/Lys clusters within the cutoff of 0.5 nm.

S. no.	Motif name	System 1	System 2	System 3	System 4
1.	Arg/Lys motif-1	70.00	71.07	87.79	99.54
2.	Arg/Lys motif-2	94.36	31.48	61.65	84.66
3.	Arg/Lys motif-3	0.00	0.10	13.19	3.18
4.	Arg/Lys motif-4	100.00	99.95	93.27	100.00
5.	Arg/Lys motif-5	0.18	88.60	72.37	18.46

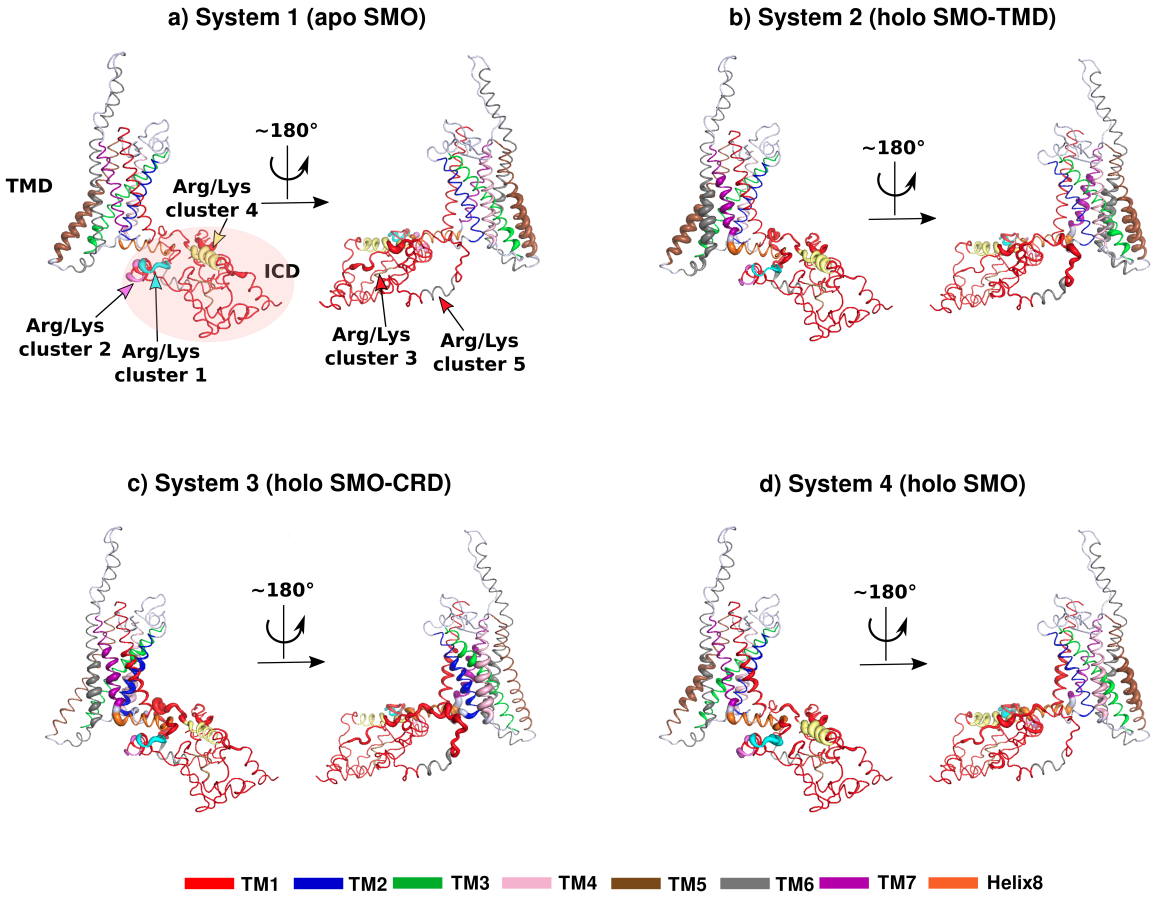


Figure 5.17: PI4P occupancy near SMO TMD and ICD, including Arg/Lys clusters mapped onto the modeled SMO structure for a) system 1, b) system 2, c) system 3, and d) system 4 (domain and TM helices colored according to the color scheme in Figure 5.9), the thickness of the ribbon shows low to high occupancy for PI4P lipid.

5.3.10 PCA analysis shows that the binding of cholesterol to TMD affects helix 8, and that to CRD affects ECL1, ECL3, and ICL3

PCA was performed for the backbone atoms of TMD on the concatenated trajectory for all simulated systems (Figures 5.18 and 5.19). The top three principal components (PCs) that accounted for more than 50% of the total motion were analyzed to estimate the effect of cholesterol binding to one or both binding sites on SMO dynamics. The percentage motion of each PC for each system is listed in Table 5.8. The major structural movements in the TMD are displayed in Figure 5.18, with the direction of movement indicated by arrows.

Table 5.8: The percentage of the top three PC modes for each system.

System name	**PC1** (%)	**PC2** (%)	**PC3** (%)	**Total** (%)
System 1	29.27	13.38	10.65	**53.3**
System 2	21.13	17.72	14.89	**53.74**
System 3	42.23	15.08	7.58	**64.89**
System 4	37.69	14.61	8.55	**60.85**

The PCA and associated RMSF analyses reveal the dynamics of ECLs, ICLs, and helix 8. In general, ECL3 is flexible (as shown in Figure 5.19 system 1), but its flexibility was enhanced when cholesterol was bound to CRD (system 3) and suppressed when bound to TMD (system 2), as indicated by the respective values in Figure 5.19. Similarly, ICL3 is generally flexible, but the flexibility was enhanced upon cholesterol binding to CRD (system 3). Further, ECL1 showed flexibility only in system 3, indicating that its dynamics are also affected by cholesterol binding to CRD. On the other hand, the RMSF values of helix 8 were, respectively, high in the top two PCs of system 2, slightly increased in system 4, and negligible in systems 1 and 3. As depicted in Figure 5.9, the motion of helix 8 in system 2 was mostly directed towards the membrane, possibly in response to cholesterol binding to TMD. Thus, it can be concluded that cholesterol binding to CRD directly influences the loop regions, specifically ECL3, ICL3, and ECL1, while cholesterol binding to TMD affects the dynamics of helix 8.

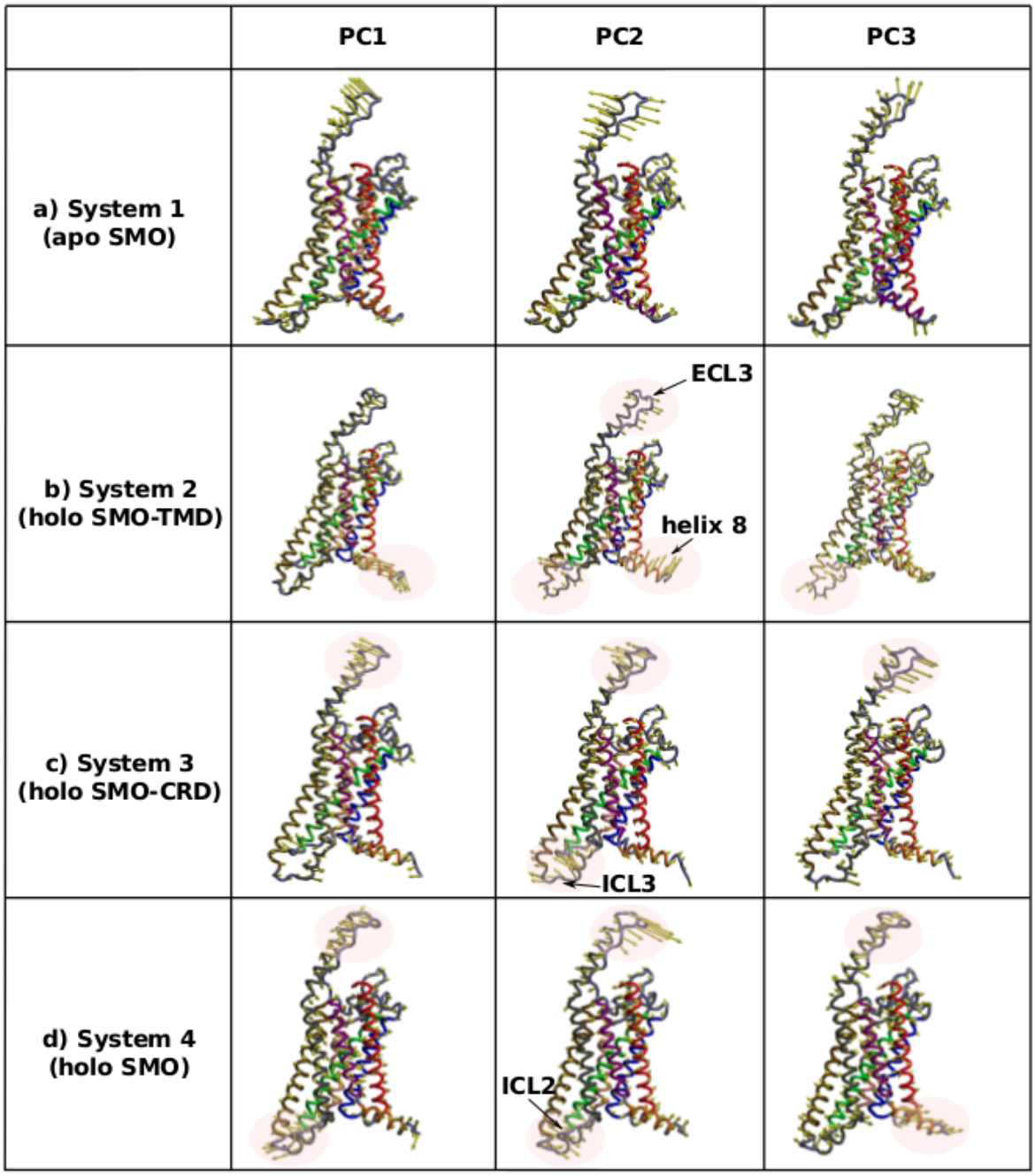


Figure 5.18: Essential dynamics of SMO-TMD. The top three PCs (PC1-3) of SMO-TMD with arrows showing the direction and magnitude of the motion of residues. System 2 shows motion in helix 8, whereas system 3 shows motion in ECL3 and ICL3.

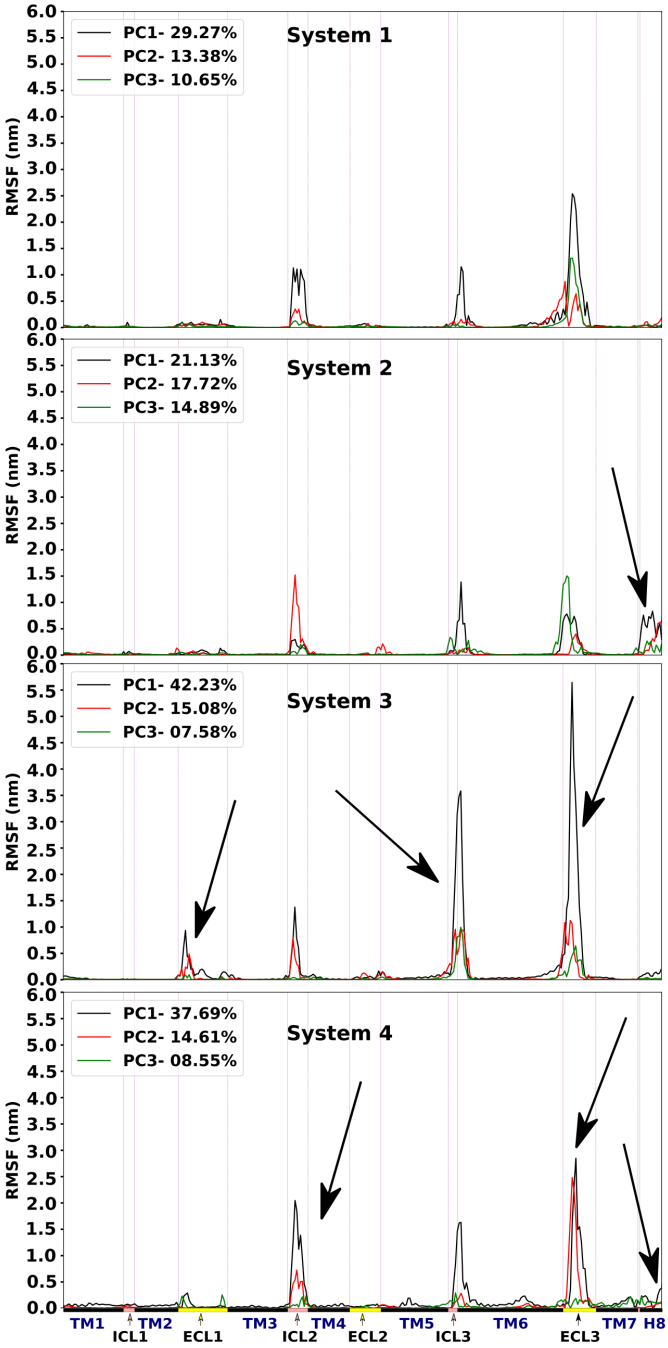


Figure 5.19: Essential dynamics of SMO-TMD. RMSF plot corresponding to the PC1, PC2, and PC3 of the TMD region for simulations a) system 1 (in black), system 2 (in red), system 3 (in green), and system 4 (in blue) with domain demarcation along the abscissa. The arrows point to the regions of each system that have greater flexibility.

5.3.11 Comparative analysis of the DCCM plots of the four SMO systems indicates varying extents of reinforcement of intradomain interactions resulting from cholesterol binding

DCCM plots reveal how the fluctuations of different residues correlate with each other, and the appearance of regions of positive or negative correlation suggests strong interactions between residues within these regions, respectively. Thus, domain-wise analysis of these regions in the DCCM plots can provide insights into the interactions between residues within and across domain boundaries. A quick overview of the DCCM plots corresponding to the four systems (Figure 5.20) suggested that in system 1 (apo SMO), interactions between residues were discontinuous within several domains, respectively, while including interactions with some of the residues in corresponding adjacent domains. However, in system 4 (holo SMO), and to varying extents in systems 2 (holo SMO-TMD) and 3 (holo SMO-CRD), the intra-domain interactions were mostly reinforced.

Therefore the relevant domain-wise DCCM subplots (Figures 5.21-5.27) were closely examined. Among the domains, ECD (CRD + LD), TMD (TM1 to TM7 + H8 + ICL1 to ICL4 + ECL1 to ECL3), and ICD, the relatively longish domain stretches were divided into sub-regions (Figure 5.21), for convenience of some of the analysis detailed below.

1. CRD was divided into two regions: N-terminal CRD (N-CRD, R28 – H63) and the rest of CRD (R-CRD, C64 – G191).

    (a) R-CRD was further subdivided into seven sub-regions: 1 (C64-C78), 2 (L79), 3 (G80-V107), 4 (L108-A128), 5 (V129-Q148), 6 (A149-C169) and 7 (T170-G191) labelled as R-CRD-x where x takes values 1-7 respectively.

2. Similarly, TM6 was divided into three subregions TM6-1 (438-452), TM6-2 (453-468), and TM6-3 (469-498) based on the two noticeable helix bend points at 452 and 468, respectively.

3. The ICD region has five Arg/Lys (R/K) clusters: R/K1 (561-567), R/K2 (571-576), R/K3 (624-629), R/K4 (667-680), and R/K5 (703-711) which are known to participate in different interactions during simulation.

    (a) Accordingly, the ICD domain were divided into six sub-regions: ICD-1 (555-567), ICD-2 (568-576), ICD-3 (577-629), ICD-4 (630-680), ICD-5 (681-711) and ICD-6 (712-787).

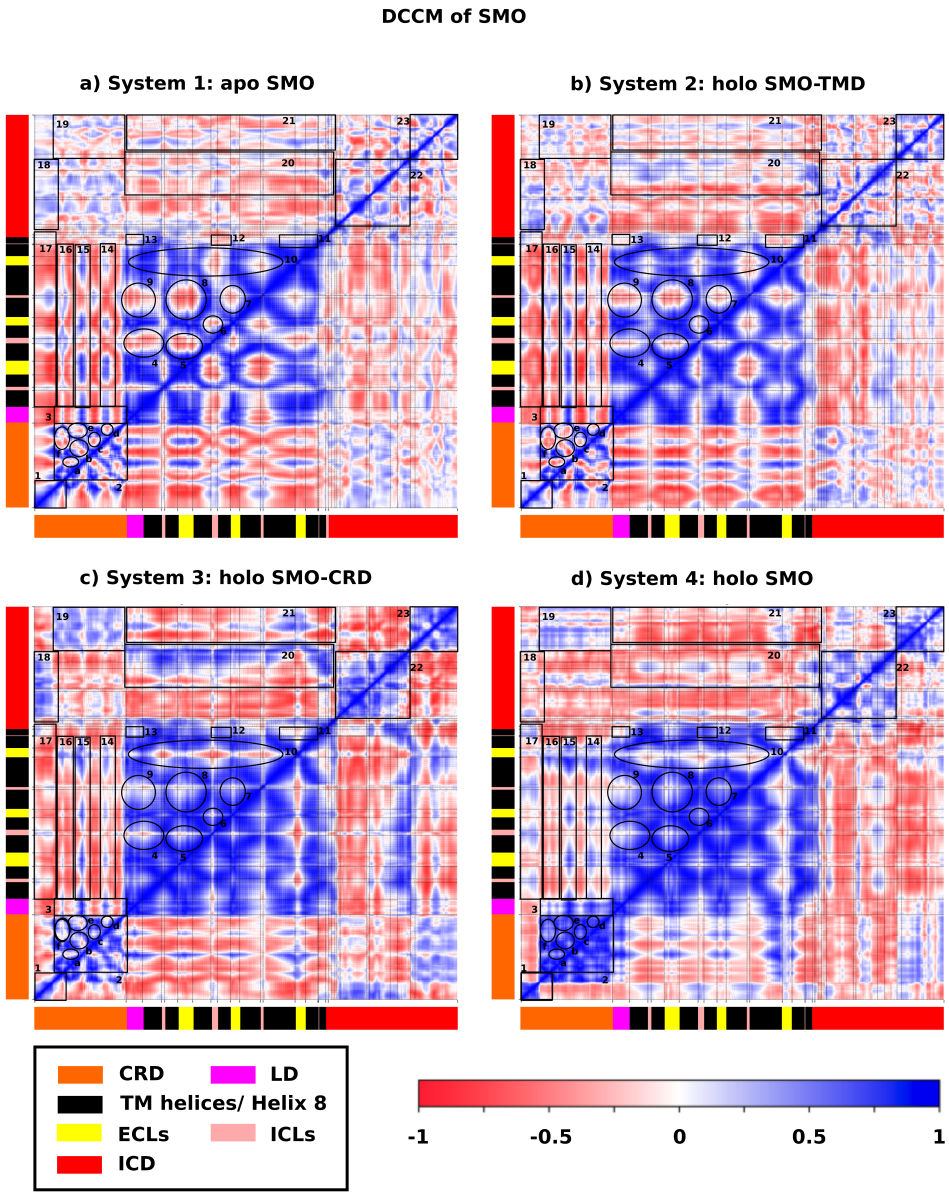


Figure 5.20: Time-averaged DCCM for the SMO Cα atoms during the simulation. DCCM of the concatenated trajectory of SMO in system 1, system 2, system 3, and system 4, showing the correlation between the domain movements with domain demarcations and TM boundaries indicated by dotted lines. Box 1 represents the self-correlations of ECD (CRD and LD), boxes 2 and 3 indicate the correlations of CRD with TMD, box 4 shows the correlations of CRD with ICD, boxes 5 and 7 indicate the intra-correlations of TMD and ICD, respectively, and box 6 shows the correlations of Arg/Lys clusters 4 and 5 (ICD) with TMD.

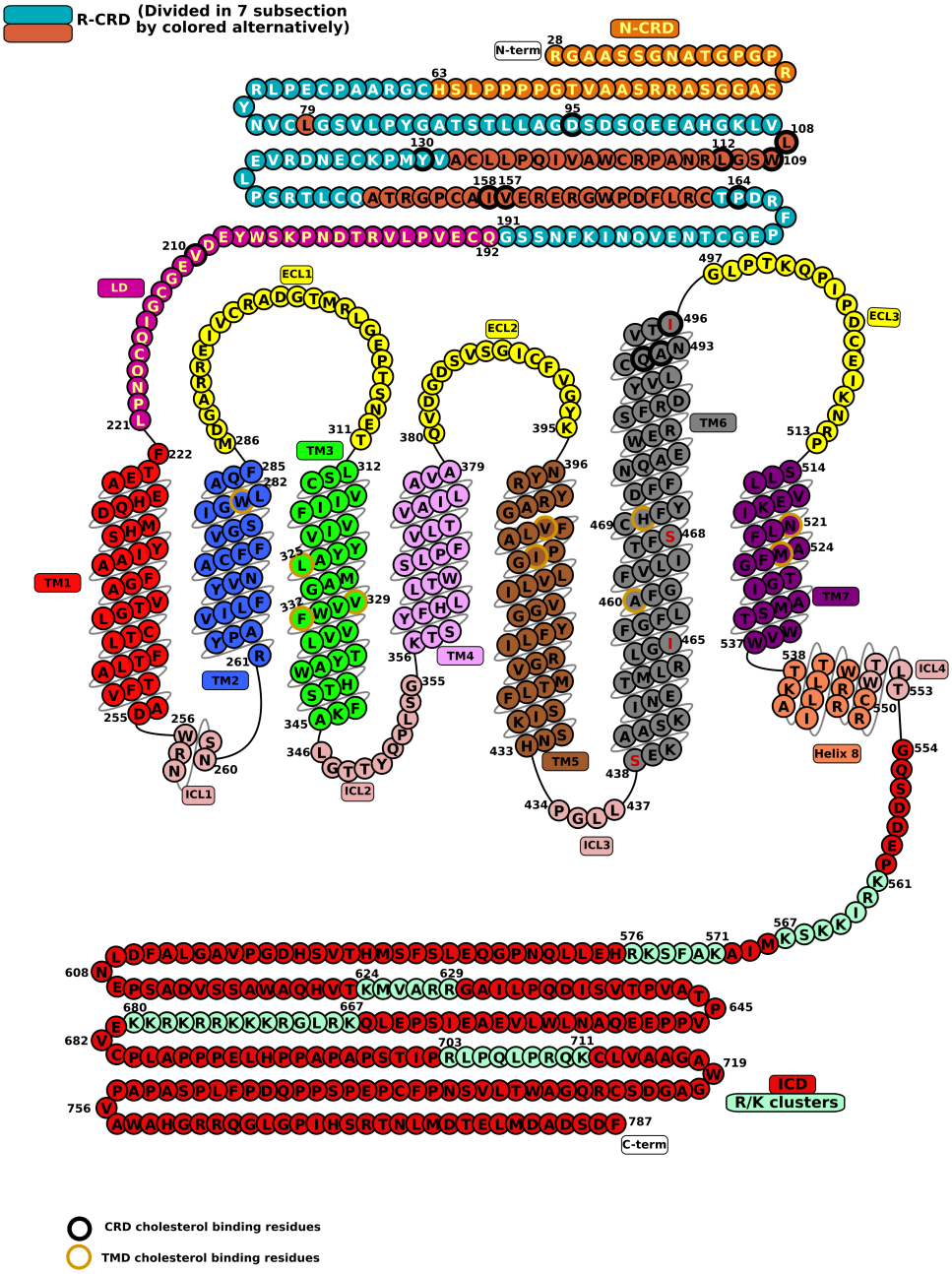


Figure 5.21: A snake plot of SMO (modified from www.gpcrdb.org) showing the main structural features of SMO, its domains, and sub-regions of domains. CRD (N-CRD and R-CRD regions and 7 sub-regions of R-CRD; R-CRD-1 to R-CRD-7), LD, TMD (7 TM helices, three ICLs, and three ECLs with helix 8 and short ICL4), and ICD (6 sub-regions of ICD; ICD1-6) are shown. Residues involved in cholesterol binding at CRD and TMD binding sites are circled in black and yellow.

5.3.11.1 Self-correlation within ECD

Figure 5.22 shows the DCCM for the self-correlation between the ECD (LD + CRD) residues. All the residues within LD are strongly positively (+ve) correlated with each other, and strongly negatively (-ve) correlated with all the N-CRD residues, in all the four systems, with no significant cholesterol binding induced variations. All the residues within N-CRD are also positively correlated with each other, albeit with some noticeable cholesterol binding induced variations. In systems 2 (holo SMO-TMD) and 3 (holo SMO-CRD), there appears to be two positively correlated sub-regions, while in system 4 there is a remarkable increase in positive correlation involving all the N-CRD residues.

The correlation patterns involving the seven sub-regions of R-CRD (R-CRD-1 to R-CRD-7) with all the LD residues generally show the following alternating positive and negative correlations: 1 (-ve), 2 (+ve), 3 (-ve), 4 (+ve), 5 (-ve), 6 (+ve) and 7 (-ve), in both the systems 1 and 2. The negative correlations noticeably decrease in system 3 and almost completely disappear in system 4. Given the negative correlation between N-CRD and LD residues, the corresponding correlation values with N-CRD has the following complementary pattern: 1 (+ve), 2 (-ve), 3 (+ve), 4 (-ve), 5 (+ve), 6 (-ve) and 7 (+ve), in systems 1 and 2. In this case, the spread of negative correlation regions increases in system 3, and all R-CRD residues become completely negatively correlated with all N-CRD residues in system 4.

As can be expected from the sub-region wise variations in the dynamics of R-CRD described above, in the DCCM of R-CRD with itself, there are large islands of negative correlation embedded within the positive correlation region in systems 1 and 2. These negative correlation islands reduce drastically in system 3, and in system 4 all the residues of R-CRD appear to be strongly positively correlated with each other.

To summarize, while cholesterol binding to TMD does not noticeably affect the DCC pattern within ECD residues, cholesterol binding to CRD does have a remarkable effect. Though there is no noticeable change in the negative correlation between N-CRD and LD residues, all residues of N-CRD and R-CRD, respectively, strongly correlate positively with each other within their respective sub-regions and negatively between the two sub-regions. These observations provide the space for hypothesizing that cholesterol binding to CRD strongly reinforces intra-region interactions within R-CRD.

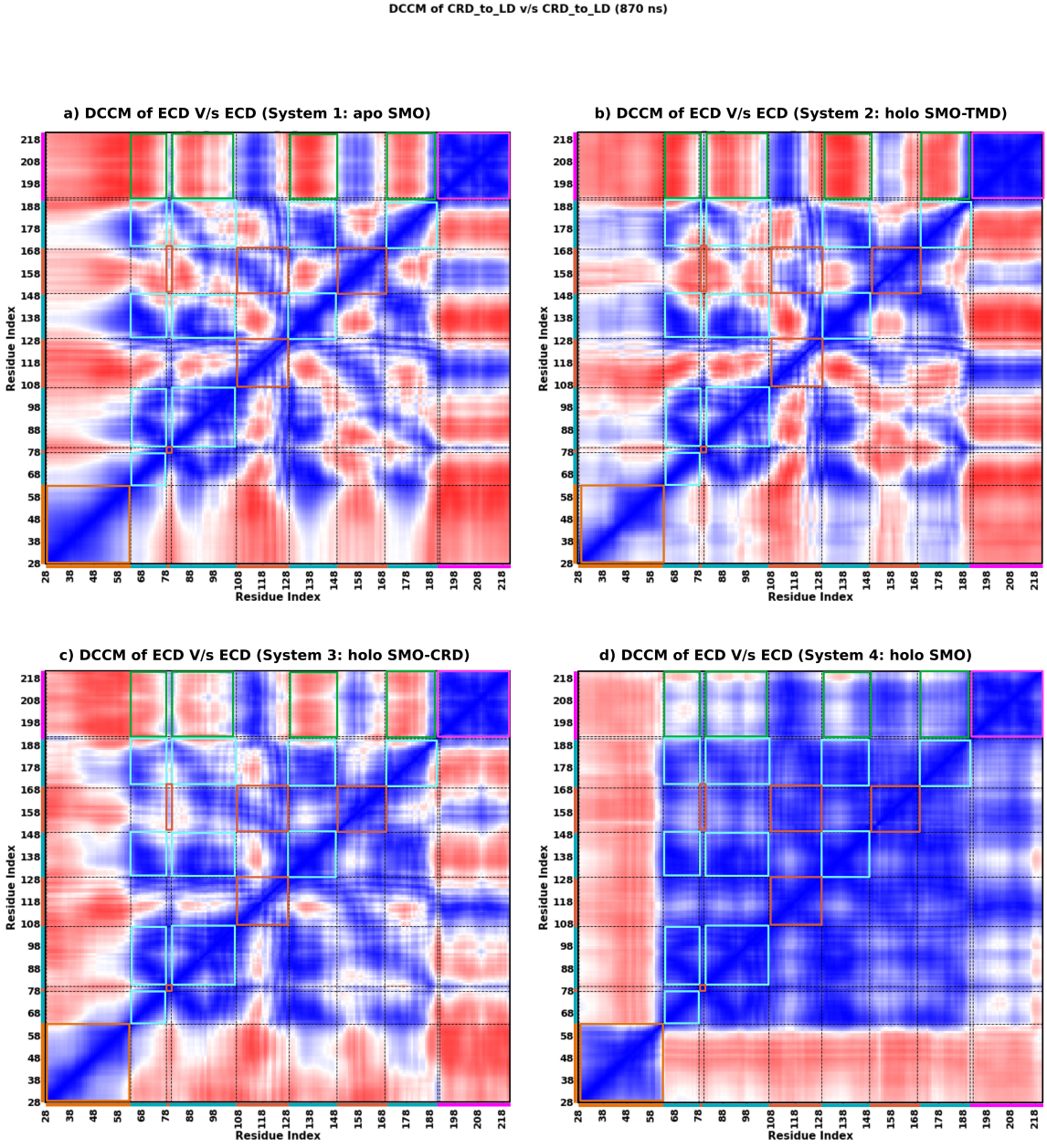


Figure 5.22: Intra-correlation of SMO-ECD (CRD and LD). CRD regions, sub-regions (N-CRD and R-CRD-1 to R-CRD-7), and LD are shown along the axes according to the color scheme described in Figure 5.21.

5.3.11.2 Self-correlation within TMD

A comparative analysis of the DCCM plots showing the self-correlation between TMD residues (Figure 5.23) reveals the effects of cholesterol binding. The plot for system 1 shows several clear regions of negative and/or zero correlation as white and red islands within an otherwise dense blue matrix. All of these involve residues from the loop regions (ICLs and ECLs) and often include helix residues (TM1 to TM7, H8), thereby disrupting the positive correlations within individual helix residues. These are:

1. ICL1 with: TM1, lower TM3 – ICL2 – lower TM4, lower TM5 – ICL3 – TM6-1, ECL1, ECL2, and ECL3

2. Lower TM3 – ICL2 – lower TM4 with: TM1, upper TM2 – ECL1 – upper TM3, ECL2, TM6-2 – TM6-3 – ECL-3 – TM7, helix 8 and ICL4

3. Lower TM5 – ICL3 – TM6-1 with: TM1, ECL2 – upper TM4, TM6-3 – ECL-3 and ICL4

The sizes of these negative correlation regions are mostly progressively reduced in cholesterol bound systems. In systems 2 and 3, respectively, several of these negative correlations get restricted to the loop regions only, without their adjoining helix stretches. Finally, barring notable exceptions such as negative correlations involving ECL3 with the four ICLs (note that the ECL3 – ICL4 negative correlation is not observed in system 1), most of the residues are positively correlated in system 4. Here too, cholesterol binding appears to reinforce mutual interactions among all the residues within TMD.

5.3.11.3 Self-correlation within ICD

Cholesterol binding builds up self-correlated blocks ICD1+ICD2, ICD3+ICD4+ICD5, and ICD6 especially in systems 3 and 4 (Figure 5.24).

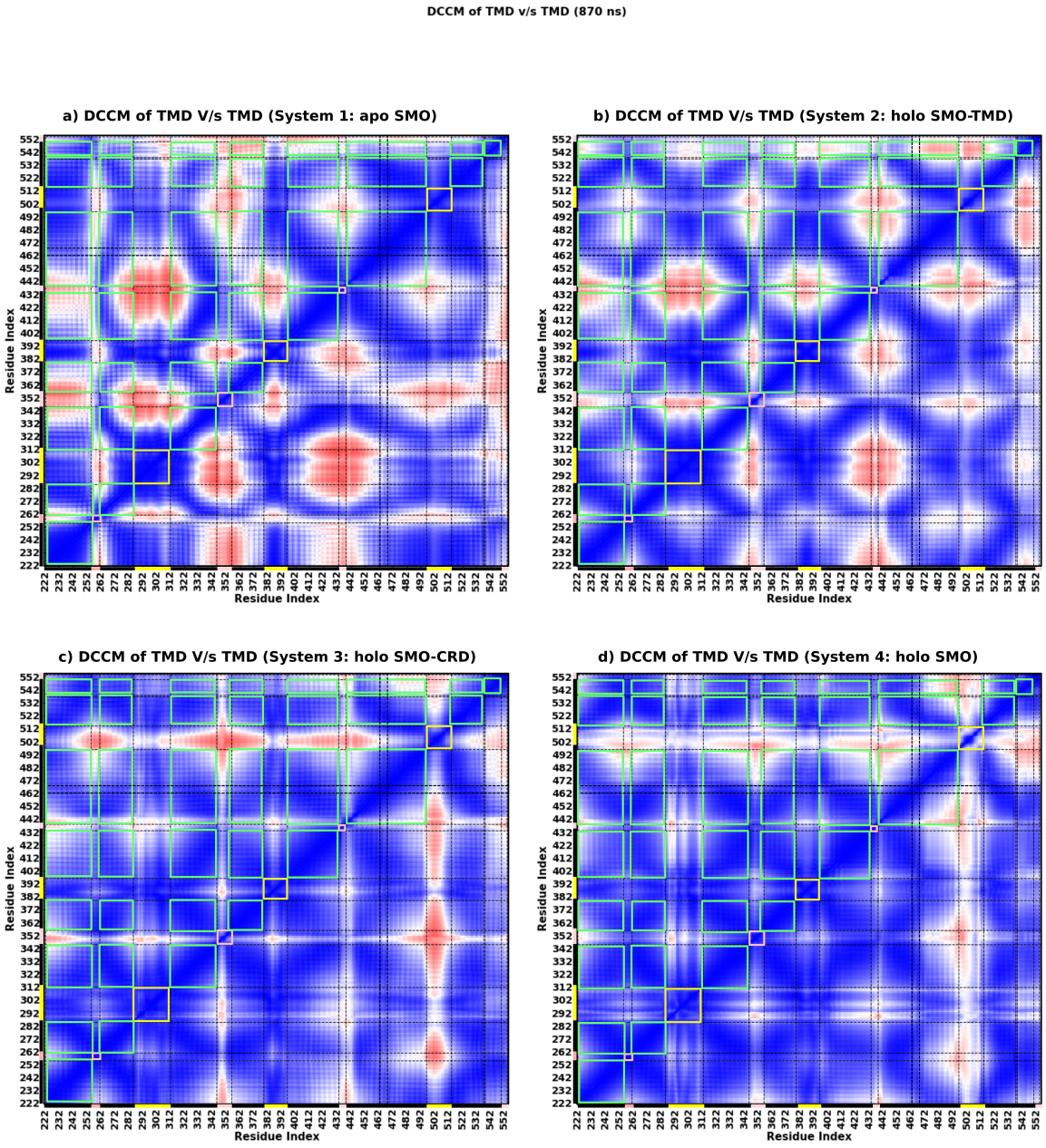


Figure 5.23: Intra-correlation of SMO-TMD. TM helices, ICLs, and ECLs are shown along the axes according to the color scheme described in Figure 5.21.

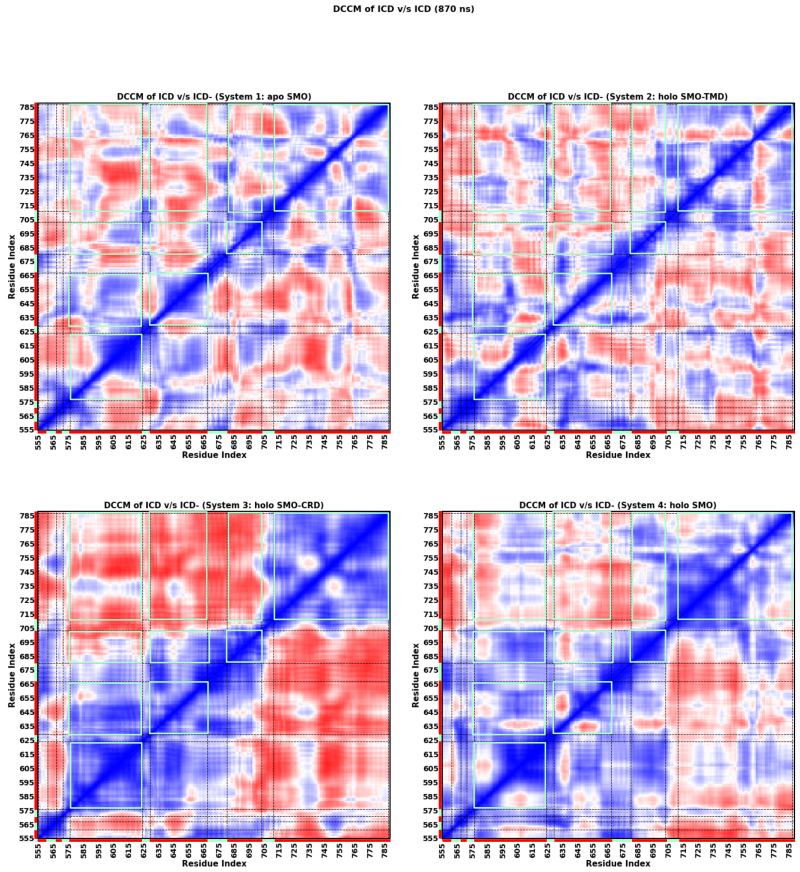


Figure 5.24: Intra-correlation of SMO-ICD. ICD regions are shown along the axes according to color scheme described in Figure 5.21. Arg/Lys clusters are colored in cyan.

5.3.11.4 Cross domain regions of the DCCM plots may provide clues to the mechanism of inter-domain communication

ECD vs TMD:

The ECD versus TMD plots combine features above and provide insights into how changes related to cholesterol binding might affect the communication between CRD and TMD via LD (Figure 5.25). In systems 1 to 3, N-CRD correlation with ICL1, ICL2 and ICL3 are positive, but correlations of R-CRD are mixed and are +ve, -ve and indifferent, respectively. In system-4, because of the cholesterol binding induced stabilization of interactions within ECD and TMD, positive correlations involving N-CRD increase, while those involving R-CRD become negative. Further, in systems 1 to 3 correlation of N-CRD with ICL4 is negative but becomes positive in system 4. R-CRD mainly correlates with TMD domain-wise with alternating signs, in systems 1 to 3. Interestingly, it shows strong positive correlation with upper TM6 – ECL3 – upper TM7. Helix 8 shows a positive correlation with N-CRD in all cholesterol-bound systems. However, it only shows some positive correlation with R-CRD-4, R-CRD-6, and R-CRD-7 in system 2. In system 1, LD correlates negatively with ICL1 to ICL4 and positively with the rest of R-CRD. The negative regions reduce progressively with cholesterol binding.

TMD vs ICD:

All cholesterol-bound systems show a positive correlation with Arg/Lys cluster 4, which seems to emphasize the significance of this cluster. Furthermore, ICD4 (some parts) +ICD5 (including Arg/Lys cluster 5) show a strong positive correlation with TMD (except TM6-3-ECL3) in system 3, but is absent in system 4 (Figure 23). It is important to note that ICL3 exhibits a positive correlation with ICD6 only in system 3, which is absent in systems 2 and 4 and weak in system 1.

ECD vs ICD:

In systems 1 and 2, the correlation between ECD and ICD shows scattered regions of +ve and -ve correlation that don't appear to follow any recognizable pattern. Interestingly, patterns begin to emerge as cholesterol binds to CRD. In system 3, N-CRD is negatively correlated with ICD6, and R-CRD exhibits a strong negative correlation with ICD4+ICD5. Further consolidation of the dispersed positive and negative correlations regions in the plots of systems 1 and 2, is observed in system 4, where there is a positive correlation between R-CRD+LD and Arg/Lys cluster 5+ICD6. Additionally, R-CRD and some parts of ICD4 (∼residue 635 to 660) also show positive correlation (Figure 5.27).

The specific correlation of helix 8 with some R-CRD subregions, only in system 2, and the correlation of ICL3 with R-CRD, only in system 3, are suggestive of variations in the flow of communication in single cholesterol-bound systems. Also, as discussed later, these changes can be linked to the interaction of membrane cholesterol and PI4P lipids with TMD and ICD (discussed further).

The specific correlation of helix 8 with some R-CRD subregions, only in system 2, and the correlation of ICL3 with R-CRD, only in system 3, are suggestive of variations in the flow of communication in single cholesterol-bound systems. Also, as discussed later, these changes can be linked to the interaction of membrane cholesterol and PI4P lipids with TMD and ICD (discussed further).

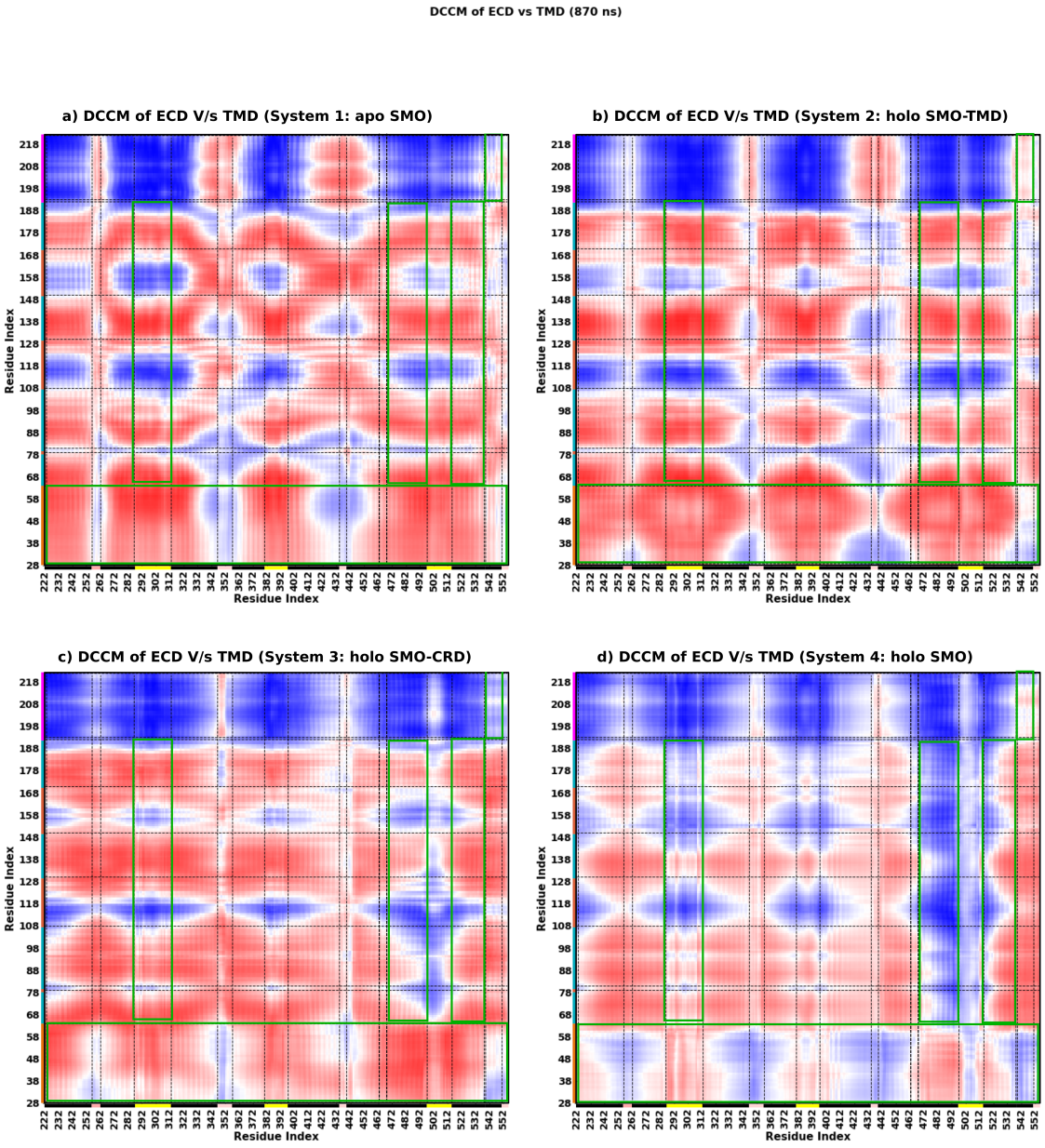


Figure 5.25: Correlation between SMO-ECD and SMO-TMD. ECD and TMD regions are shown along the axes according to the color scheme described in Figure 5.21.

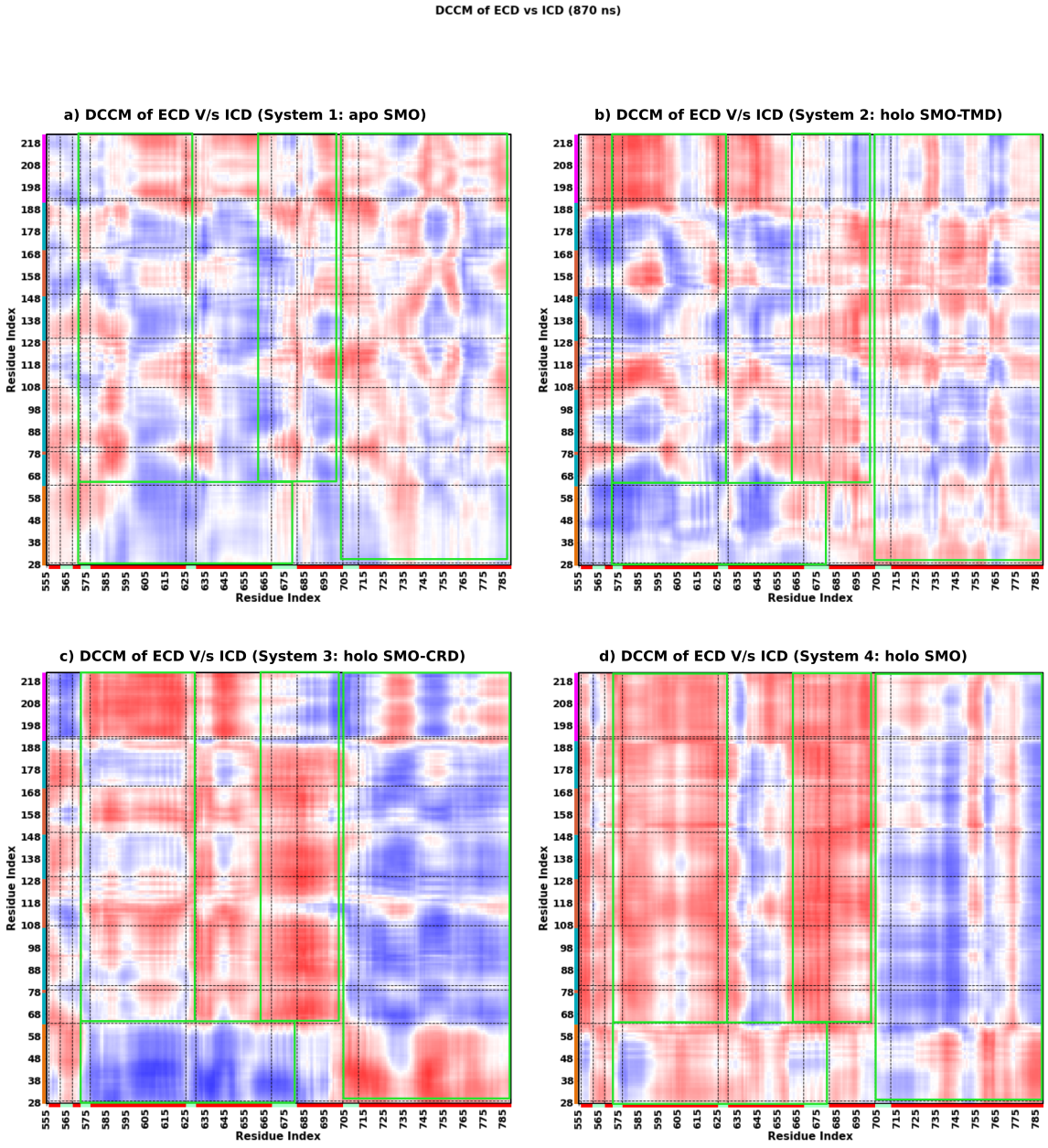


Figure 5.26: Correlation between SMO-TMD and SMO-ICD. TMD and ICD regions are shown along the axes according to the color scheme described in Figure 5.21.

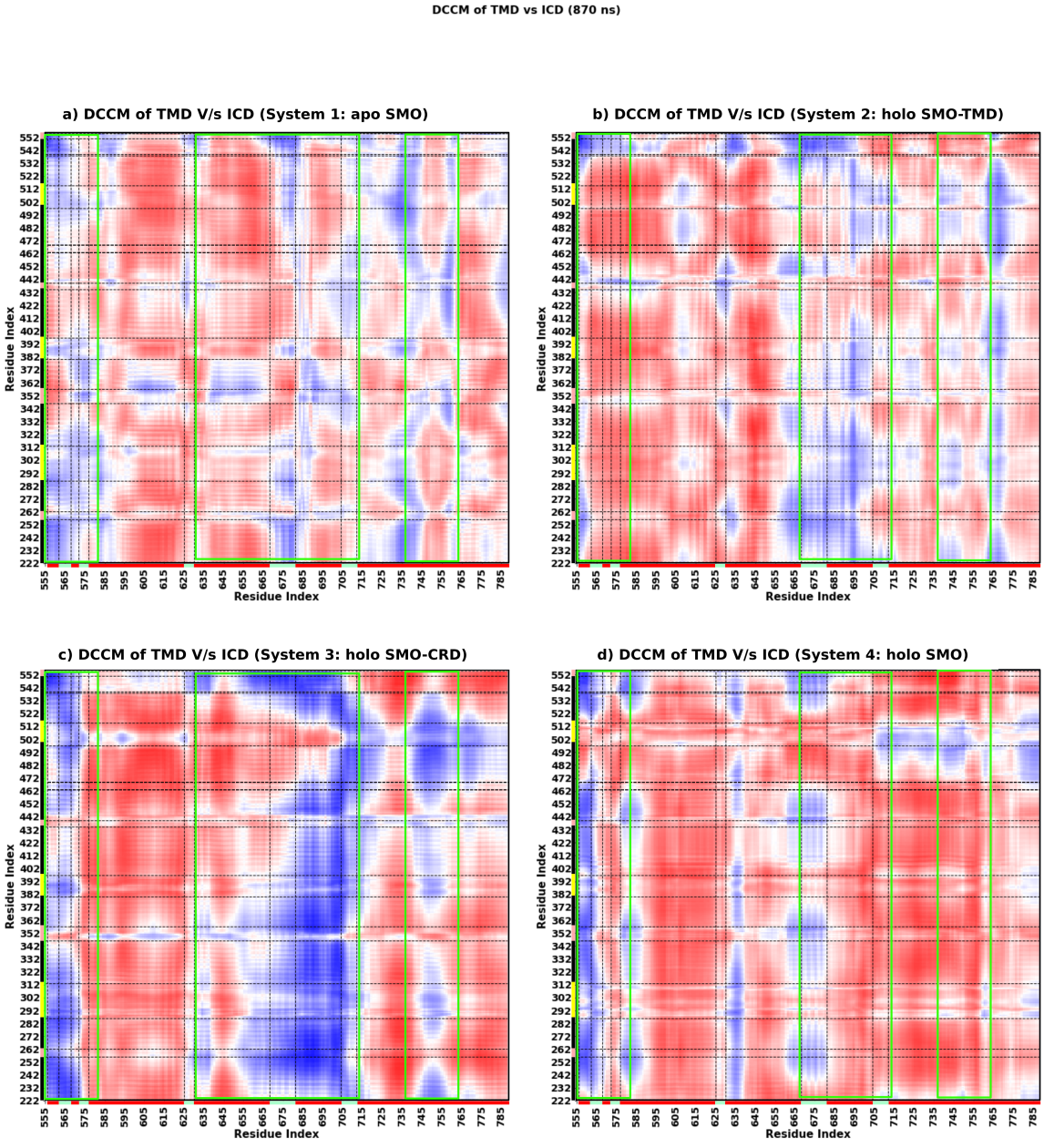


Figure 5.27: Correlation between SMO-ECD and SMO-ICD. ECD and ICD regions are shown along the axes according to the color scheme described in Figure 5.21.

5.3.11.5 Pairwise comparison of the real average structures of the cholesterol-bound systems with that of the apo system reveals associated conformational changes in domains that are distant from the respective binding sites

Does cholesterol binding only result in conformational changes in the neighborhood of the respective binding sites, or does it also lead to significant conformational changes in regions distant from the binding sites? If the latter is the case, identification of the cholesterol-binding induced conformational changes should in principle help in mapping the inter-region interaction network responsible for the transduction of the cholesterol binding signal. This was investigated by a pairwise comparison of the real average structure (defined in section 5.2.3.5) of each of the three cholesterol bound systems (Systems 2-4) with that of the apo system (System 1) (Video 5.2). Since TM4 was found to be relatively unaffected by cholesterol binding in either of the binding sites, the pairwise comparison was carried out by superimposing the TM4 helices of the two structures, followed by an analysis of the conformational variation of the other regions.

While TM1 and TM2 show conformational shifts in the presence of cholesterol in any of the binding sites, cholesterol binding to CRD revealed significant conformational changes in TM5, TM7, helix 8, and ICD, and TM6 developed a bend (residue 479; near the outer leaflet region) in the presence of cholesterol in both the binding sites. The outward movement of TM5, was maximum in system 3 though slightly reduced in system 2, and was very less in system 4 as compared to single cholesterol-bound systems. Also, the conformational shift in the extracellular region of TM7, was maximum in systems 2 and 3 though slightly reduced in system 4. These changes result in conformational changes in the ICD. Notably, the change in ICD conformation appears to correlate strongly with the conformational change at helix 8 in system 2.

5.3.12 Clustering analysis of the frames in the trajectories of the four systems provides further understanding of the dynamics of cholesterol-binding induced conformational changes in different domains

The comparative analysis involving real average structures, discussed above, provides a static view of the conformational changes across domains, in response to cholesterol binding. In order to compare the changes in the conformational dynamics, within and between systems, the frames belonging to the concatenated trajectories (870 ns) of the systems, respectively, were clustered by a 'k-means' clustering algorithm. To track changes in the TMD, clustering was done on the backbone atoms of TMD using a 0.2 nm RMSD cut off as input. This resulted in 20, 19, 19, and 22 clusters for systems 1, 2, 3, and 4, respectively. Among these, some clusters

were sparsely populated; while some structures did not belong to any of the clusters and were categorized separately. The clusters obtained for each of the four systems were respectively sorted based on their population (Table 5.9). The top four dominant clusters of the cholesterol-free state system 1 (apo SMO) added up to 58.90%, while the same for systems 2, 3, and 4 added up to 73.94%, 60.28%, and 62.00% of the total structure frames, respectively (Videos 5.3-5.6).

Visual analysis of the representative structures corresponding to the top 4 clusters within each system, respectively, (Videos 5.3-5.6) displayed significant changes in the CRD and ICD conformation, in tune with the correlated motion observed in the DCCM analysis. The up and down movement of TMD-bound cholesterol within the TMD, discussed earlier, could also be seen clearly in Videos 5.4 and 5.6. The top 4 cluster representatives of each system were then aligned by superimposing the TM4. The alignment shows the changes in CRD and ICD conformation in greater detail. For example, in system 2 (Videos 5.4), in cluster 1, the CRD residues were closer to the TM6 than in cluster 2. The TM6 extended region (residues connected to ECL3) shows a little bend in cluster 1, though not in any other cluster. TM6 also showed a sharper bend in its structure near the outer leaflet area in cluster 3 compared to in any other cluster. At the same time, the conformations of helix 8 in clusters 1 and 2 were different from those in clusters 3 and 4. Following that, the ICD reveals a conformational disposition.

For system 3 (Video 5.5), conformational changes are observed in CRD, ICD, as well as in ECL3 and ICL3. ICL3 is close to a small region of ICD in cluster 1, whereas this is absent in any other cluster. For system 4 (Video 5.6), in cluster 1, ICD is close to TMD, but moves away in cluster 2, alongside with conformational changes in the CRD (especially in the N-CRD) and the outer leaflet region of TM6 . While in cluster 3, ICD was again displaced near the membrane near TM4. In cluster 4, CRD and ICD conformations were different. All four clusters showed evidence of cholesterol movement within TMD. When comparing these with system 1 representative clusters, the four clusters show that CRD is highly dynamic, as compared to other system clusters.

Table 5.9: Information about the clusters in all four systems under study, along with the cluster populations and percentage populations (in parenthesis).

Cluster index	**System 1**	**System 2**	**System 3**	**System 4**
1	154 (17.68%)	234 (26.87%)	168 (19.29%)	228 (26.18%)
2	145 (16.65%)	195 (22.39%)	122 (14.01%)	117 (13.43%)
3	109 (12.51%)	136 (15.61%)	118 (13.55%)	98 (11.25%)

4	105 (12.06%)	79 (9.07%)	117 (13.43%)	97 (11.14%)
5	103 (11.83%)	70 (8.04%)	85 (9.76%)	83 (9.53%)
6	68 (7.81%)	54 (6.2%)	61 (7%)	60 (6.89%)
7	57 (6.54%)	33 (3.79%)	56 (6.43%)	49 (5.63%)
8	46 (5.28%)	24 (2.76%)	43 (4.94%)	31 (3.56%)
9	27 (3.1%)	23 (2.64%)	40 (4.59%)	31 (3.56%)
10	18 (2.07%)	10 (1.15%)	27 (3.1%)	25 (2.87%)
11	13 (1.49%)	4 (0.46%)	24 (2.76%)	23 (2.64%)
12	10 (1.15%)	2 (0.23%)	2 (0.23%)	9 (1.03%)
13	6 (0.69%)	1 (0.11%)	2 (0.23%)	5 (0.57%)
14	2 (0.23%)	1 (0.11%)	1 (0.11%)	4 (0.46%)
15	2 (0.23%)	1 (0.11%)	1 (0.11%)	2 (0.23%)
16	2 (0.23%)	1 (0.11%)	1 (0.11%)	2 (0.23%)
17	1 (0.11%)	1 (0.11%)	1 (0.11%)	2 (0.23%)
18	1 (0.11%)	1 (0.11%)	1 (0.11%)	1 (0.11%)
19	1 (0.11%)	1 (0.11%)	1 (0.11%)	1 (0.11%)
20	1 (0.11%)	—	—	1 (0.11%)
21	—	—	—	1 (0.11%)
22	—	—	—	1 (0.11%)

Along with the real average structure analysis discussed above, and in conjunction with other studies including PCA and DCCM, these observations provide interesting clues regarding the mechanism of transduction of the cholesterol binding signal.

5.4 Discussion

5.4.1 The HH signaling pathway and the two roles of PTCH1

In the HH signaling pathway, the binding of HH ligand to its receptor PTCH1 leads to the activation of SMO by relieving it from its PTCH1-mediated inhibition. Activated SMO, in turn, activates the GLI transcription factors by relieving them from their inhibition by SUFU and PKA, resulting in the expression of a wide range of target genes [2, 3, 12, 128]. Though the actual mechanism is far from being clear, as discussed in the introduction, current understanding on the process of SMO activation in response to HH signaling appears to converge on two roles of PTCH1. The first involves the translocation of SMO to the primary cilia, at the tip of which it can bind to the SUFU complex. The presence of SMO in the primary cilia is thus a prerequisite

to its activity. The current consensus in this regard is that, in the absence of HH signaling, PTCH1 inhabits the ciliary membrane. On HH binding, it gets displaced from the primary cilia and allows the translocation of SMO into the cilia. However, though a prerequisite, this translocation of SMO into the primary cilium, by itself does not lead to SMO activity. Here the second role of PTCH1, that as a cholesterol transporter which, depending on HH signaling, controls the availability of accessible cholesterol in the cilia, becomes important. It is now well accepted that SMO has two sterol (cholesterol) binding sites, one inside the TMD and the other in the CRD. The activity of SMO is linked to the binding of cholesterol as an agonist ligand to either or both the sites. The question being asked is "how?"

5.4.2 Kinnebrew's hypothesis provides a context

The crystal structures of SMO showing bound ligands in two ligand-binding sites, respectively, in TMD and CRD, are silent regarding where the ligand binds first. A recent study by Kinnebrew et al. suggested that by default, cholesterol binds to the TMD when SMO localizes to the site of activity, namely the primary cilia, and transitions SMO to a basal activity state (as in system 2). In the absence of sufficient levels of accessible cholesterol in the outer leaflet of the membrane, and hence, in the absence of CRD-bound cholesterol, SMO continues to show only basal activity. With the restoration of cholesterol levels, resulting from HH signaling, cholesterol binds to the CRD of SMO, which then transitions to the fully active state (system 4). It has been proposed that the TMD binding site in this context acts as an allosteric site that controls HH signaling and SMO activity, whereas the CRD site is the orthosteric site that serves as the primary target of PTCH1 regulation of SMO activity [84].

5.4.3 Investigating the dynamics of cholesterol free and cholesterol bound SMO in modeled ciliary membrane

Atomistic MD simulations of four systems (Table 5.1) were performed to understand the stabilization of SMO in primary cilia and its activation upon cholesterol binding to either or both CRD and TMD binding sites. The interactions of membrane lipids (cholesterol and PI4P) with SMO were seen in the atomistic simulation of system 1 (apo SMO), which provides insights into how they may be influencing the dynamics of each other. It may be noted that the observations, by and large, agree with those made in the earlier CG simulation studies of SMO in the modeled ciliary membrane [91]. The effect of cholesterol binding as a ligand to either or both of the cholesterol-binding sites was explored in systems 2, 3, and 4. The analysis revealed how this binding affects the interplay of SMO with its interacting membrane lipids, and particularly in the context of the functional states outlined in Kinnebrew's hypothesis, the structural and functional dynamics of SMO.

5.4.4 Cholesterol binding as a ligand to SMO impacts the interplay of its interactions with the PI4P enriched ciliary membrane and its functional dynamics

It is important to note that there are two types of interactions of lipids with SMO: (a) cholesterol, as an agonist ligand, binding to the CRD and/or TMD, which leads to conformational changes and is primarily responsible for the transduction of the signal regulating SMO activity; and (b) interaction of cholesterol (and other ciliary lipids, such as PI4P) from the ciliary membrane with the TMD and ICD. Unlike the CG simulation, no interaction between membrane cholesterol and CRD was observed in the atomistic simulation [91]. In order to achieve this, CRD must bend considerably toward the membrane in order to interact with membrane cholesterol. One reason might be that CG simulations enable the observation of larger movements that take place over longer timescales. However, similar membrane cholesterol dynamics were seen where the cholesterol molecules moved around within and between the upper and lower membrane leaflets. Additionally, PI4P also showed lateral movement within the inner leaflet to interact with SMO. It can be inferred that the cholesterol binding as a ligand to SMO impacts the interplay of its interactions with membrane cholesterol and PI4P in the ciliary membrane and, thus, its functional dynamics.

5.4.5 Changes in the SMO binding sites and in distant regions upon cholesterol binding to TMD

On cholesterol binding to TMD, there is a noticeable increase in TMD volume, accompanied by an increase in the interhelical distances of TM1-TM6 and TM5-TM6. Interestingly, the binding reduces the RMSD across all SMO domains, stabilizing the CRD, LD, and ICD. The noticeable RMSD change in TMD, such as a decrease in the RMSD of TM6 and TM7 and an increase in the RMSD of helix 8 with reference to the initial structure, are indicative of some structural changes in the region. These changes can be related to the straightening of TM6 (decrease in TM6 tilt angle) and the bending of TM7 with respect to the ciliary membrane, respectively. The structural superimposition of the real average TMD from systems 2 and 1, respectively, reveals a conformational change in helix 8, which is supported by an increase in RMSF from the PCA results. Although DCCM analysis does not show much of an impact on the self-correlation between residues within CRD, TMD, or ICD, respectively, the positive self-correlation within TMD begins to improve slowly. One significant observation in the dynamics of the CRD is the change of the secondary structure of the N-terminal CRD region from a loop to an α-helix (Figure 5.28b). This structural element change was not observed in system 3 (Figure 5.28c), but it was observed in system 4 (Figure 28d). These results support the hypothesis that the changes observed in TMD upon cholesterol binding to TMD are an allosteric effect, as

it also alters the CRD conformation. Apart from these changes, the interaction between membrane cholesterols and the cholesterol-binding motifs found in lower TMD regions (CARC7 in TM4, CRAC4 in TM5, CARC9 in TM6, strict CCM in TM4 and TM2) either declines to a moderate level or becomes very negligible. Furthermore, it is important to consider that cholesterol bound to the TMD shows high levels of dynamics and moves up and down in the TMD region. Additionally, the bound cholesterol spends more time in proximity to the upper TMD region, which is evidenced by the increased interaction between bound cholesterol and upper TMD residues, including ECL2. This may explain the observed lower level of interaction between membrane cholesterol and cholesterol-binding motifs that are located in the ECL2 region.

Subsequently, the loss of membrane cholesterol interaction in cholesterol-binding motifs present in TM5 and TM6 could be because of the high interaction of PI4P lipids with TM5, TM6, and TM7. Similarly, the cholesterol-binding motif present in helix 8 loses its interaction with membrane cholesterol, and this can be linked to high helix 8 flexibility. Definitely, cholesterol binding to TMD changes helix 8 dynamics, but again, it is unclear whether changes in the interaction of membrane lipids with helix 8 caused the helix 8 dynamics or the other way around. Finally, the cholesterol binding to TMD enhances the interaction between ICD and the membrane by causing a conformational change in ICD as well as a change in the conformation of helix 8.

5.4.6 Changes in the SMO binding sites and in distant regions upon cholesterol binding to CRD

On cholesterol binding to CRD in apo SMO, TMD shows a little increase in its volume. In contrast to cholesterol-binding to TMD, TM1-TM6, and TM5-TM6 interhelical distances decrease, which is consistent with the dynamics in TM6 (as the TM6 RMSD value increases). The remarkable effect of cholesterol binding to CRD results in strong self-correlation in the CRD (N-CRD and R-CRD), TMD, and ICD subregions (ICD 1-6). The observations reveal a strong positive correlation between all residues in the N-CRD and R-CRD subregions, and a negative correlation was observed between the two sub-regions. These findings provide the space for hypothesizing that cholesterol binding to CRD strongly reinforces intra-region interactions within R-CRD. Additionally, the TMD loops (aside from ECL3) begin to develop a positive correlation with the TM helices, and the ICD correlations also get better with CRD (region from Arg/Lys 2 to ICD-5 with N-CRD and ICD-6 to R-CRD). Furthermore, these dynamics in SMO domains are corroborated with RMSD and PCA results, as they show an increase in the flexibility of CRD, ECL2, ECL3, ICL2, ICL3, and ICD. The membrane lipids may have played an important role in the stabilization of TM helices, as the membrane cholesterol interaction with most of the upper and lower cholesterol-binding motifs has increased tremendously

compared to system 2. This is also the reason for the stability of helix 8, as it moderately interacts with membrane cholesterols. The interaction with the cholesterol-binding motif present in ICD completely vanishes. Instead, the PI4P interaction with all Arg/Lys clusters in ICD has improved. The Arg/Lys clusters 1, 4, and 5 even show a positive correlation with TMD, possibly an important reason for the ICD intra-positive correlation of this part. Though the cholesterol that is bound to the CRD is not much flexible at this binding site, it interacts with more residues from the CRD, LD, and TM6 extended regions during the simulation. The visual inspection of the trajectory and clustering analysis showed that the larger changes in CRD and ICD dynamics are linked to cholesterol binding to CRD. This ultimately results in the conformational disposition of ICD toward the ciliary membrane. The current data support the notion that CRD is an orthosteric binding site. This conclusion is supported by the observation that cholesterol binding to CRD causes a significant conformational change in SMO, which is a crucial attribute of an orthosteric binding site, and that this state has medium signaling activity [84].

5.4.7 Cooperativity between the allosteric TMD and orthosteric CRD sites in the fully functional state of SMO (system 4 - holo SMO) and characterizing the changes caused by cholesterol binding to TMD or CRD in holo SMO

The conversion of state 2 to state 4 has been hypothesized by Kinnebrew et al., which changes the medium activity state to the high activity state [84]. The findings indicate that SMO dynamics, its interaction with ciliary lipids, and its interactions with TMD- and CRD-bound cholesterol with their respective binding sites in states 2 and 3 cannot be used to properly understand the dynamics in state 4 by routine application of additivity principles. First of all, compared to system 2, the TMD volume is further increased when the second cholesterol is introduced to the CRD binding site. Similar to system 2, the upper and lower TM1-TM6 interhelical distances are increasing in this system as well (Figure 5.4 and Table 5.2). Surprisingly, the TM5-TM6 lower distance has decreased, while its upper distance has increased. This might be connected to a modification in the DRE interaction network seen only in system 4, where the interaction between D473 (TM6) and R500 (TM5) has been switched to E518 (TM7) and R500 (TM5), indicating that this is a cooperative effect of the presence of cholesterol in TMD and CRD binding sites. This also suggests that the dynamics of TM6 are affected by cholesterol binding to either the CRD or TMD or both sites. The ECL3 loop, connected to TM6, becomes more dynamic (high RMSF from PCA), and this is further supported by DCCM analysis as it exhibits a negative correlation with the rest of the TMD. DCCM of CRD again reveals the interesting fact that within N-CRD and R-CRD, respectively, self-correlation becomes stronger in the presence of the second cholesterol in the CRD site, and they correlate

negatively with each other. This is clearly the effect of the presence of cholesterol in CRD. Interestingly, LD also correlated with R-CRD in system 4, suggesting the cooperative effect of these binding sites. However, this is also correlated with the interaction pattern of CRD-LD residues with CRD-bound cholesterol (discussed earlier). In addition to this, a change in the structural elements of the N-terminal CRD region from loop to α-helix (Figure 5.28d) was observed, similar to what was observed in system 2 (although the extent of the change was less in system 2). These results agree with Kinnebrew et al. that when cholesterol is bound to both TMD and CRD sites of SMO, the RMSF values of CRD, especially the C-terminal CRD residues, were reduced (Figure 5.7a) as compared to the single cholesterol-bound SMO states [84]. TMD also shows different dynamics here; apart from the change in the DRE network, the positive self-correlation between TMD residues is stronger in system 4 than in any other system. Notably, this TMD-self correlation gradually gets better from system 1 to system 4, indicating that this might be the key to connecting the SMO states of apo and holo SMO (two intermediate states, holo SMO-TMD and holo SMO-CRD) with their corresponding activities (ranging from no activity to high activity, with intermediate states having basal and medium activity). Moreover, the function of membrane lipids is included as membrane cholesterol, and PI4P interaction is high in most of the cholesterol-binding motifs in upper and lower TMD. Helix 8 is directly impacted by cholesterol binding to TMD in systems 2 and 4 because it does not show significant changes from this analysis in system 3 (suggested by PCA). Helix 8 shows the interaction with membrane cholesterol and PI4P, which is observed independently in systems 3 and 2. Helix 8 has moderate interaction with membrane cholesterol in system 2 and not in system 3. While the PI4P interaction with helix 8 is observed in system 2 and not in system 3. PCA also reveals higher flexibility in ECL3, ECL1, and ICL3 in system 3, which are also observed in system 4, suggesting these dynamics are because of the CRD cholesterol binding. However, ECL3 flexibility is suppressed a little bit by TMD cholesterol binding as compared to its flexibility in system 3. In this instance, just like in system 3, PI4P lipids interact with all five Arg/Lys clusters present in the ICD, suggesting the role of CRD cholesterol binding again, leading to an increase in the membrane and ICD interaction. Last, but not least, the ICD receives these changes from the CRD to the TMD via loops (ECLs and ICLs), which can also be seen in clustering results.

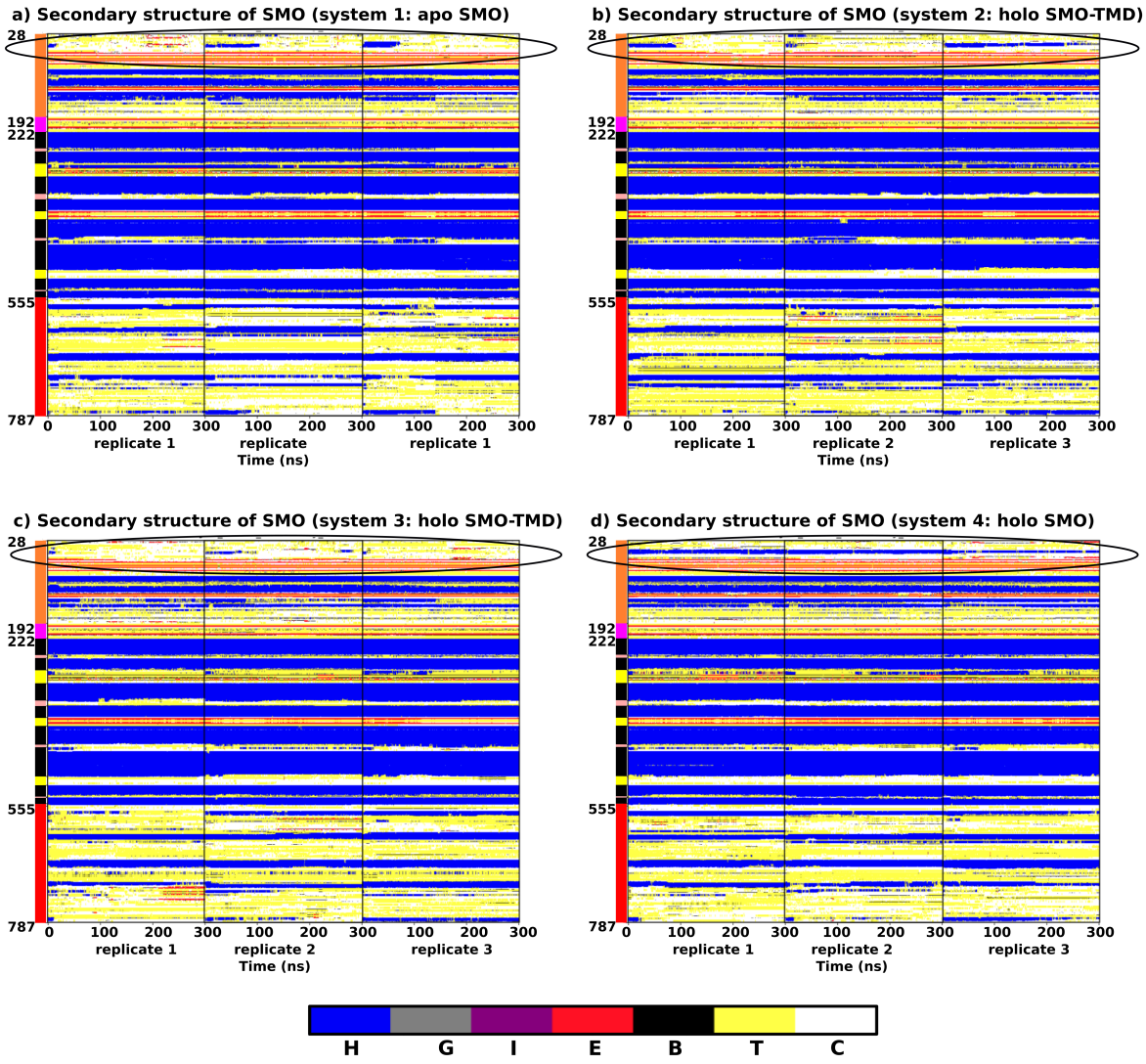


Figure 5.28: Change in the secondary structure elements of SMO during MD simulation. The CRD region is encircled for each system to highlight the formation of an α-helix in system 4 (also in system 2). Secondary structure content was determined using the STRIDE method available in VMD.

5.4.8 Ionic-lock between R451 and W535 present in TM6 and TM7 does not correlate with cholesterol-mediated SMO activation

One important mechanism for stabilizing the active conformation of GPCRs is the "ionic lock" [158]. Huang et al. reported that the inactive SMO is stabilized by a cation-π interaction between R451 and W535 (a lock), and this lock is broken in a cancer mutant [68]. They used mutagenesis and functional assays to determine the role of the cation-π lock in modulating SMO. Later, Bansal et al. also showed that this cation-π interaction is broken in active SMO

by using an active SMO structure for preparing the system [13]. However, no such correlation has been observed in the present data. This interaction was broken in the cholesterol-free system and in the system with cholesterol bound to both CRD and TMD. On the other hand, the two single cholesterol-bound systems showed the presence of the lock. Further, the distance of this interaction has been calculated in all hSMO structures available in the PDB. The distance values of this ionic lock interaction in 4O9R (inactive SMO) and 6OT0 (active SMO) are 0.74 nm and 0.48 nm, respectively (Table 5.10). This suggests that the distance values of this ionic lock interaction is not correlated with the active and inactive states of SMO.

Table 5.10: Distance value of ionic-lock interaction in hSMO available in PDB. The distance between the COMs of the guanidine group of R and the indole group of W was used to calculate the interaction.

S. no.	**PDB ID (Chain)**	**Species (State)**	**Method**	**Arginine position**	**Tryptophan position**	**distance value (Å)**	**distance value (nm)**
1.	4JKV (A)	Human (Inactive)	X-ray	451	535	5.22	0.52
2.	4O9R (A)	Human (Inactive)	X-ray	451	535	7.37	0.74
3.	4N4W (A)	Human (Inactive)	X-ray	451	535	6.01	0.6
4.	4QIM (A)	Human (Inactive)	X-ray	451	535	5.51	0.55
5.	4QIN (A)	Human (Inactive)	X-ray	451	535	4.94	0.49
6.	5L7D (A)	Human (Inactive)	X-ray	451	535	4.78	0.48
7.	5L7I (A)	Human (Inactive)	X-ray	451	535	4.83	0.48
8.	5V56 (A)	Human (Inactive)	X-ray	451	535	5.27	0.53
9.	5V57 (A)	Human (Inactive)	X-ray	451	535	5.24	0.52
10.	6OT0 (R)	Human (Active)	cryo-EM	451	535	4.79	0.48
11.	6XBJ (R)	Human (Active)	cryo-EM	451	535	6.68	0.67
12.	6XBK (R)	Human (Active)	cryo-EM	451	535	5.39	0.54

13.	6XBL (R)	Human (Active)	cryo-EM	451	535	4.81	0.48
14.	6XBM (R)	Human (Active)	cryo-EM	451	535	4.81	0.48
15.	7ZI0 (A)	Human (Inactive)	X-ray	555	639	5.48	0.55
16.	GPCR modeled structure (A)	Human (Inactive)	computa-tional modeling	451	535	4.48	0.45

5.4.9 TMD-bound cholesterol dynamics is restricted by the CRD-bound cholesterol in system 4 (holo SMO)

SMO converts from state 1 (apo state; system 1) to state 2 (holo SMO-TMD; system 2) upon cholesterol binding to TMD. The TMD-bound cholesterol shows a different dynamics pattern when it is only in the TMD compared to when a second cholesterol binds to the CRD binding domain. Multiple structures available in the PDB suggest multiple cholesterol-binding sites within TMD. Huang et. al proposed that the SMO TMD contains a tunnel that extends from the inner leaflet of the membrane. This tunnel serves as a conduit for cholesterol involved in SMO activation, allowing it to travel from the membrane to the CRD [68, 125]. Here, in system 2, cholesterol primarily moves to the upper TM region, with some movements also occurring towards the middle TM region. It oscillates up and down within the TMD binding site and rotates around its long axis. The residue that interacts with the cholesterol may be one reason why it stays here for a more extended period of time. The residue from the upper TM regions shows higher occupancy for this interaction (Table 5.3). The involvement of the ECL2 residue D384 is also seen in system 2 only. The bound cholesterol is highly dynamic, as suggested by RMSF analysis. When the second cholesterol binds to the CRD binding site (system 4), TMD cholesterol dynamics get restricted compared to those in system 2. TMD cholesterol is seen to be present more often in the TM region from the upper middle to the lower middle. Extra cholesterol in the CRD binding site may be causing the TMD-bound cholesterol to move towards the middle or lower middle TMD region from the upper TMD region. Again, this is supported by the high cholesterol occupancy of most middle and lower interacting residues, the loss of interaction with D384, and the lower RMSF value of TMD-bound cholesterol in system 4. However, the movement of TMD cholesterol to the CRD binding site was not observed, as suggested by Huang et al. and Qi et al. [68, 125], which is state 3, SMO with medium

activity. There is no experimental or computational evidence to support the hypothesis that cholesterol can move from the TMD to the CRD, i.e., from SMO state 2 to state 3. This question still remains open as to how cholesterol binds to the CRD, given the fact that the binding of cholesterol to the CRD alone is also sufficient for SMO activity.

5.4.10 Interaction pattern of membrane lipids with SMO in cholesterol-free and cholesterol-bound states

These results supported the previous findings from CG simulations [91] when considering the importance of membrane lipids, particularly membrane cholesterol. The membrane cholesterol occupancy around identified sequence and structure motifs present in the SMO TMD region shows interactions in atomistic simulations as well. In previous CG simulation of SMO in the modeled ciliary membrane [91], strict-CCM, TM2/3e, CARC7, CRAC4, CARC8, CARC9, CRAC-like2, and CRAC-like1 had higher occupancy with the membrane. The CG simulations (chapter 4) indicated the significance of the CARC7 and CRAC4 motifs, and a cholesterol was sandwiched between these two cholesterol-binding motifs at the lower TMD. These results are consistent with these observations in apo SMO system 1. However, in single cholesterol-bound systems (systems 2 and 3), the degree of interaction has decreased for both CARC7 and CRAC4 motifs as compared to system 1. Surprisingly, these interactions were regained in the presence of cholesterol in both the TMD and CRD domains. In contrast, the interaction of CRAC4 decreased to 27.82% in system 2 (compared to 96.82% in system 1), increased to 86.63% in system 3, and slightly decreased to 76.88% in system 4. Later, the importance of SMO-ICD is highlighted by the PI4P interactions with Arg/Lys clusters in ICD that are present in atomistic simulations. The pattern of PI4P interaction with Arg/Lys motifs suggests that PI4P binding is important for the functioning of SMO, as the apo SMO interacts with PI4P only via three Arg/Lys clusters out of five, while the holo SMO (system 4) interacts with PI4P with all five Arg/Lys clusters. In the earlier CG simulation, the presence of an interaction between the Arg/Lys cluster 5 and PI4P was reported, which is not observed in system 1. However, this interaction is present in all cholesterol-bound systems, with a lower occupancy in system 4 as compared to systems 2 and 3. It has to be noted that the CG simulation showed the movement of ICD toward the ciliary membrane and near TM4 through PCA analysis [91]. This supports the observation that all Arg/Lys clusters (ICD) interact with PI4P. TM4, which does not interact with PI4P in system 3, significantly interacts with membrane cholesterol and PI4P (especially in system 4, where holo SMO has high signaling activity). Helix 8 dynamics in system 2, ICL3 and ICD dynamics in systems 3 and 4, changes in DCCM pattern, and RMSF PCA analysis are all associated with changes in membrane lipid interaction with lower TMD and ICD. The conformational changes in CRD and ICD with respect to TMD that were seen

in CG simulations and in this atomistic simulation demonstrate that it is important for SMO to stabilize in the ciliary membrane.

5.5 Conclusion

The current study explored how information is transmitted from the extracellular region of SMO to its intracellular region at the molecular level. In state 2 of SMO, which has basal activity and has cholesterol bound to TMD, the signal is transmitted from the CRD to ICD through helix 8. Here, the N-terminal structure of the CRD changes from a loop to an α-helix with an increase in TMD volume caused by the dynamics of TM1, TM5, TM6, and TM7. This alters the ICD conformation by having an effect on the dynamics of helix 8. This increase in TMD volume may facilitate G-protein binding to the TMD lower region. While in state 3 of SMO, which has medium activity and cholesterol bound to CRD, the signal is primarily transmitted from CRD to ICD through ECL2, ECL3, and ICL3, with a slight increase in TMD volume. Here, there is an increase in the interaction between ciliary lipid and SMO TMD, and all five ICD Arg/Lys interact with PI4P lipids to aid in the conformational disposition of ICD, potentially creating a setting that is conducive to G-protein interaction. Finally, in state 4 of SMO, which has high activity, the signal is transduced from the CRD to ICD via ECL2, ECL3, ICL3, and TM helices. The N-terminal CRD residues structured as α-helix, and furthering the self positive correlation between CRD subregions helped, TMD-bound cholesterol to adjust in lower TMD region, reorganize the DRE network, and increase in TMD volume, making the TMD-intra correlation positive (except ECL3), increase in its interaction with membrane lipids. In this instance as well, all five Arg/Lys clusters of ICD interact with the PI4P lipid of the ciliary membrane, altering the conformation of ICD relative to the TMD and possibly making lower TMD available for G-protein.

Chapter *6*

Conclusion and Future Perspective

In this thesis, I discussed the results of molecular dynamics (MD) simulations of SMO under different membrane conditions and in various ligand dependent activity states to get molecular-level insights into the activation mechanism of SMO. I have analyzed coarse-grained (CG) and atomistic MD simulation results in conjunction with experimental data on SMO interactions with membrane lipids (cholesterol and PI4P), and have obtained reliable insights into the role of SMO interactions with ligands and membrane lipids in driving the signal transduction mechanism crucial for SMO activation. The significant results are summarized below.

6.1 Reading the sequences and structures of SMO

In the first section, I looked at the primary sequences of SMO from different species, carefully analyzing them to identify conserved residues and motifs. Here are my findings in the field of GPCR biology on SMO involved in the HH signaling pathway, along with previous research providing detailed insights into the sequence and structure of SMO: Identification of sequence cholesterol-binding motifs: I searched for the cholesterol-binding sequence motifs (CARC, CRAC, and CARC-like motifs) on full-length SMO sequences that are crucial in GPCRs. A total of 21 cholesterol-binding sequence motifs are present in SMO from the N-terminal cysteine-rich domain (CRD) to the C-terminal intracellular cytoplasmic domain (ICD), consisting of 4 CRAC, 13 CARC, and 4 CRAC-like motifs. Refining the definition of the *strict* cholesterol consensus motif (*strict*-CCM): I refined the definition of the *strict* cholesterol consensus motif (CCM) in SMO by performing multiple sequence alignments (MSA) of the SMO sequence with various class A receptors and then with class F receptors. Residues K356 (4.41)-W365 (4.50)-L362 (4.47), located in the TM4 of SMO, correspond to the cholesterol consensus motif (CCM). These residues, along with Y262 (2.39) from TM2, constitute the *strict* CCM observed in other GPCRs. Identification of sphingolipids binding motif (SBM): I also identified the SBM IQPLLCAVY (residue 122-130) present in the CRD domain, which overlaps with a CRAC motif. Identification of conserved motifs in TM helices: Conserved mo-

tifs in GPCRs have been reported to have structural and functional importance. They serve as signature sequences, aiding in their role in receptor activation mechanisms, and can be targeted for drug development. However, SMO does not have classical TM conserved motifs, which are present in class A GPCRs with functional significance. There are conserved motifs that are present in TM helices of all class F GPCRs. I have identified the class F conserved motifs in TM6 and TM7.

In addition, I collected data from various crystal and cryo-EM structures for all residues and their positions that interact with cholesterol. I have also highlighted the five stretches of positively charged residues (Arg/Lys clusters) in the ICD, reported in previous studies, that help the SMO to stabilize in primary cilia. These analyses helped me to understand the interaction of SMO with various lipid components. I have highlighted previously identified conserved SMO motif KATXXXW present in helix 8. I also focused on previously identified conserved switches networks, such as the DRE network present at the upper TM region, the WGM network present at the lower TM region, and the ionic lock between residues of TM6 and TM7. These networks and the ionic lock have been reported to be important for the dynamics of SMO.

There is a lot more information present in the sequence and structure of SMO that demands thorough investigation and careful analysis in order to obtain insights into their functional implications. However, it is crucial to highlight, that without sufficient experimental validation, connecting computational findings and conclusions with the actual function of the protein or receptor is difficult.

6.2 Simulating the functional dynamics of SMO

While molecular dynamics simulations can provide insights into the functional mechanism of proteins, an appropriate membrane model is essential for a proper investigation of transmembrane proteins such as SMO. Further, any investigation into the functional aspects of dynamics places the twin demands on the simulation setup. On the one hand, it needs the simulation duration to be long enough to observe the larger movements associated with the functional dynamics, and on the other hand, it also needs a detailed view of the intra- and intermolecular interactions which are involved in the structural and functional dynamics. Building upon previous research, I have designed a membrane consisting of five different lipid components, with variations in the composition of the outer and inner leaflets. Observations from this study revealed intriguing dynamics of SMO within the ciliary membrane.

(a) The two major domains, cysteine-rich domain (CRD) and ICD, show notable changes in the presence of cholesterol, PI4P, and other lipids. They display a tendency to drift toward the membrane and interact with membrane lipids. Thus, during the course of simulation,

CRD often interacts with membrane lipids involving both, occasional downward bending of CRD as well as upward movement of membrane cholesterol molecules.

(b) The configuration of TMD in the ciliary membrane is different from that in the POPC membrane. I have observed that the membrane cholesterols are attracted more toward the cholesterol-binding motifs in the lower TM regions compared to that in the upper region, especially to CARC7 (TM4) and CRAC4 (TM5). Notably, these motifs were found to engage in interactions not only with a single cholesterol molecule but also simultaneously with two or more cholesterol molecules.

(c) All four key residues of the *strict* CCM, which I redefined, also showed higher occupancy for membrane cholesterol.

(d) The sequence-dependent preference of cholesterol accumulation, as observed during my simulation, opens up the possibility for future mutational analysis for studying mutation dependent variations in SMO-cholesterol association patterns.

(e) In addition to TMD, ICD also plays a crucial role in shaping the functional dynamics of GPCRs. The movement of the ICD towards the membrane facilitates the accessibility of the TMD for G-protein coupling.

(f) I have also observed the interaction of cholesterol with various regions in the ICD. In addition to this, I have also highlighted the interaction of PI4P with the five Arg/Lys clusters present in ICD, especially with Arg/Lys clusters 4 and 5. The PI4P-mediated interactions associated with the resultant anchoring of SMO to the primary cilia may possibly have a stabilizing influence on SMO conformation. I have mentioned a few mutations in ICD (Chapter 1) that switched SMO into a continuously active form. In this scenario, the results and observations from my study can be helpful in further investigation of these experimental observations.

Finally, my study highlights the importance of studying the functional dynamics of SMO in conjunction with the dynamics of the membrane lipids. It is the intricate interplay of the dynamics of the protein and of the membrane lipids that together shape the functionality. CRD-membrane lipid interactions, PI4P-mediated SMO anchoring, sequestering cholesterol near TMD, and other sequence-dependent associative accumulation of cholesterol, etc., are only some examples of the significance of this integrative view of the functional dynamics of SMO. This CGMD study also provides the basis for further mechanistic studies using all-atom molecular dynamics.

6.3 Further mechanistic insights into SMO activity

The process of SMO activation, which occurs upon the release of PTCH1 inhibition following a signal from HH, has remained a long-standing mystery. It has been established from previous studies that cholesterol can activate SMO and that it has two binding sites of SMO in CRD and TMD. In addition to this, a very recent model of SMO activation suggests that related to the binding of cholesterol to either or both of these binding sites, SMO can exist in four different states, which may be correlated with their respective activity levels. SMO exhibits distinct activity states based on cholesterol binding: the inactive state (SMO in primary cilia without cholesterol binding), the basal activity state (cholesterol bound to TMD), the medium activity state (cholesterol bound to CRD), and the high activity state (cholesterol bound to both TMD and CRD).

I performed multiple atomistic simulations of different SMO states in the ciliary membrane, with a total simulation time of 3.6 μs across all systems. I modeled four systems, systems 1 - 4, having SMO with no bound cholesterol, cholesterol bound only to TMD, cholesterol bound only to CRD, and cholesterol bound simultaneously to both TMD and CRD, respectively, to mimic the four experimentally identified states. In addition to corroborating observations from my CG simulation, analysis of the trajectories from atomistic simulations of the four systems provided a lot of valuable insights into cholesterol binding induced function and activity of SMO. Observations from this study provide insights into the relationship between specific structural changes in different states of SMO and their corresponding functions.

(a) Compared to the inactive state of SMO, in the basal activity state, TMD volume has increased, which also correlates with the increase in lower TM-TM distance. The cholesterol bound to TMD was also seen to move up and down and rotate along its axis in the TMD pocket. However, while the occupancy of membrane cholesterol near most of the cholesterol-binding motifs in this state is reduced as compared to the inactive state, the membrane cholesterol interaction with CARC7, CRAC-like 1, CBP, and TM2/3e motifs remains high. I have reported the conformational change in helix 8 during the simulation and dislocation of the ICD domain towards the membrane and its interaction with PI4P lipids, especially the region where Arg/Lys clusters 1, 4, and 5 are present. One important observation is the change in the formation of the -helix of the N-terminal CRD region. It shows the allosteric behavior of the TMD binding site.

(b) In the medium activity state, where cholesterol is bound to CRD, some intriguing dynamics is observed. Though the cholesterol binding to CRD does not change the TMD volume much, it has increased the self-correlation within TMD residues as compared to the inactive and basal activity states of SMO. It also improves the positive correlation within CRD residues. The loop flexibility (ECL2, ECL3, ICL2, and ICL3) has increased,

and hence the level of fluctuation dynamics during the simulation. The interaction of membrane lipids with SMO has also increased. Most of the cholesterol-binding motifs have higher occupancy for membrane cholesterol, and all five Arg/Lys clusters interact with PI4P. This shows that the binding of cholesterol to CRD (at the extracellular site) has a significant effect observable across different regions of SMO.

(c) In the fully active SMO state, where cholesterol is bound to both TMD and CRD, the molecular level changes involve the formation of an -helix involving the N-terminal CRD residues. In addition, I also observed a self-positive correlation between residues within CRD subregions, similar to our observations in the basal activity state of SMO. This also shows the dynamics of TMD-bound cholesterol to be gravitating toward the lower TMD region, thus, possibly further increasing TMD volume. A reorganization of the conserved DRE network was also observed. In this state, the intra- and inter-correlation between SMO domains have significantly turned out to be positive. Further, I noticed a substantial increase in the interaction between SMO and the lipids present in the surrounding membrane. In the high-activity SMO state, most of the cholesterol-binding motifs (including two from ICD) interact with membrane cholesterol. Further, all five Arg/Lys clusters of ICD interact with the PI4P lipid of the ciliary membrane, which alters the conformation of ICD relative to the TMD and possibly making lower TMD available for G-protein.

The observations made from all three with cholesterol (cholesterol-bound) states, along with their comparison to the without cholesterol (cholesterol-free) state, provide insights into how the dynamics of SMO progressively change from state 1 to state 4, no activity state to high activity state, and hence correlate the structural-conformational change with the activity of SMO. This study offers valuable insights into the intricate relationship between structure and function by correlating the particular structural changes to the corresponding functional states.

6.4 Future directions

Research can never end. This topic has several outstanding open questions not only related to SMO but also related to the other components as well. Further computational studies have the potential to answer many of the unanswered concerns about the HH signaling system.

1. **SHH ligand transport:** The signal molecule SHH itself is a mystery of how it traveled from one cell to another though it has dual lipid modification (palmitoylation at the N-terminal and cholesteroylation at the C-terminal). Understanding the transport mechanism of SHH is an open question that requires further investigation. Computational modeling and simulations can be employed to study the diffusion and transport mechanisms of the SHH signal molecule. This could involve setting up a membrane bilayer

and placing the SHH in the extracellular part to investigate how SHH moves across the membrane.

2. **Molecular mechanism of PTCH1 regulation on SMO:** PTCH1 is a membrane protein that is hypothesized to transport cholesterol and regulate the activity of the SMO. It is also dependent on the concentration of cholesterol in the membrane. However, the molecular and atomic details of this mechanism are not understood. Exploring the mechanisms by which PTCH1 regulates SMO is an area of interest. Molecular dynamics simulations can be used to explore this objective. By simulating PTCH1 and SMO together in the primary cilia by varying the cholesterol concentration (low to high), it is possible to gain insights into their structures, dynamics, and conformational changes. Further, the results can be compared and correlated to understand the cholesterol-mediated PTCH1 regulation of SMO.

3. **Studying the coupling of G-protein with SMO:** The coupling of G-protein with SMO is another aspect that can be studied to understand one aspect of the downstream signaling process in HH signaling. Designing a computational system that contains SMO and G-protein in the ciliary membrane, where the signaling occurs, could provide valuable information about the binding interfaces, their interaction, and signal transduction from the cellular component to the intracellular component.

4. **Investigating the dynamics of SMO variants in particular cancers:** Several cancers have different SMO variants that influence SMO function. Incorporating these critical mutations into SMO and examining the resulting interaction and change in dynamics can illuminate their role in cancer development and progression. Computational modeling and simulations may be helpful in the study of the structural and functional impact of SMO variations associated with cancer. It is affordable to investigate changes in protein stability, binding affinity, and conformational dynamics induced by individual mutations by introducing them into the SMO structure and performing molecular dynamics simulations.

5. **Role of sphingomyelin (SM) in the HH pathway:** Another important lipid of the ciliary membrane is sphingomyelin, which is restricted to the upper leaflet. Recent research indicates that SM has a very high affinity for cholesterol, and hence it plays an important role in regulating the cholesterol accessibility for SMO. Additionally, SM depletion is a common method for increasing cholesterol accessibility at the plasma membrane. The identification of SBM opens the question of the selectivity of cholesterol or sphingomyelin in a particular state. By simulating SMO in cholesterol and SM containing ciliary membrane, two important questions can be addressed; i) SMO-SM interactions, its selectivity of cholesterol or sphingomyelin, and ii) Cholesterol-SM interaction and their cooperative effects in SMO dynamics.

6. **Understanding SMO dimerization:** GPCR dimerization enhances signaling. The role of lipids such as cholesterol in this process has been successfully studied for -2 adrenergic and 5-HT receptors. While drosophila and mammalian SMO are known to form dimers, the dimerization of SMO in humans remains relatively unexplored. This can be explained using a molecular dynamics simulation of two copies of SMO in the ciliary membrane. Furthermore, exploring the role of lipids in SMO dimerization can provide more valuable insights into this process.

7. **Simulating the entire HH pathway component:** Bringing all the components and interactions of the HH pathway together in a comprehensive simulation can help visualize and understand the pathway as a whole. Designing such simulations can provide a more comprehensive perspective on the dynamics and behavior of the HH pathway. This objective is computationally demanding, as this will integrate various components of the HH pathway, including HH ligands, PTCH1 and SMO, G-protein, and downstream proteins with the ciliary membrane. By constructing such a comprehensive computational model, simulations can be performed to study the dynamics, signaling outputs, and regulatory mechanisms of the pathway as a whole.

In summary, I believe that the work I have carried out for my thesis provides significant insights into SMO dynamics. Some events, such as details of interactions between two components, conformational change in specific structural motifs, and the transitions between active and inactive states, cannot be captured experimentally, but can be done by computational approach. Together the computational and experimental work can bridge the gaps in our understanding of the HH signaling pathway smoothly.

Appendix *A*

Supplementary Information

Supporting text for Supplementary Videos

A.1 Chapter 4

Structural dynamics of Smoothened (SMO) in the ciliary membrane and its interaction with membrane lipids

Video 4.1: Interaction of membrane cholesterol with CRD. The trajectory (t= 9.1 to 9.2 μs) depicts the movement of CRD towards the ciliary membrane and its interaction with two membrane cholesterols (index 5 and 52), extracted from replicate 1 of SMO in ciliary membrane CGMD simulation (Figure 4.5a, main text). The protein is rendered as a domain-wise color scheme described in Figure 4.3 (main text), CRD is highlighted in orange, and phosphate groups of all ciliary lipids are shown as silver transparent spheres. The interested cholesterol molecules are shown as a yellow surface, and its ROH bead is shown in black. Water, ions, and other ciliary lipids/cholesterol molecules are omitted for clarity.

Video 4.2: Interaction of membrane cholesterol with LD. The trajectory (t= 2.25 to 2.3 μs) depicts the movement of LD towards ciliary membrane and its interaction with membrane cholesterol (index 48), extracted from replicate 1 of SMO in ciliary membrane CGMD simulation (Figure 4.5b, main text). The protein is rendered as a domain-wise color scheme described in Figure 4.3 (main text), LD is highlighted in magenta, and phosphate groups of all ciliary lipids are shown as silver transparent spheres. The interested cholesterol molecules are shown as a yellow surface, and its ROH bead are shown in black. Water, ions, and other ciliary lipids/cholesterol molecules are omitted for clarity.

Video 4.3: Interaction of membrane cholesterol with CRAC4 and CARC7 motifs. The trajectory (t= 9.2 to 9.5 μs) depicts the interaction of three membrane cholesterols (index 27,

78, and 92) with CARC7 and CRAC4 motifs, extracted from replicate 1 of SMO in ciliary membrane CGMD simulation (Figure 4.7, main text). The protein is rendered as a gray surface, CARC7 are CRAC4 are highlighted in pink and green, respectively. The interested cholesterol molecules are shown as a yellow surface, and its ROH bead are shown in black. Phosphate groups of all ciliary lipids are shown as silver transparent spheres. Water, ions, and other ciliary lipids/cholesterol molecules are omitted for clarity.

Video 4.4: Interaction of membrane cholesterol with strict-CCM. The trajectory (t= 9.2 to 9.5 μs) depicts the interaction of four membrane cholesterols (index 27, 61, 78 and 96) with strict-CCM residues, extracted from replicate 1 of SMO in ciliary membrane CGMD simulation (Figure 4.17, main text). The protein is rendered as silver, TM4 and TM5 are highlighted in blue and pink, respectively. The strict-CCM residues, W365, K356, L362, and Y262, are shown in red, ice-blue, green, and orange beads. Residue Y269 is also shown in blue beads. The interested cholesterol molecules are shown as a yellow surface, and its ROH bead are shown in red. Phosphate groups of all ciliary lipids are shown as silver transparent spheres. Water, ions, and other ciliary lipids/cholesterol molecules are omitted for clarity.

Video 4.5: A cholesterol inside TMD of SMO in ciliary membrane. The 15 μs trajectory shows that the cholesterols positioned inside TMD near site 1 is stable and remains inside throughout the simulation. SMO in ciliary membrane CGMD simulation (Figure 4.36, main text). The protein is rendered as a domain-wise color scheme described in Figure 4.3 (main text), CRD, LD, TMD and ICD are colored in orange, magenta, blue and red. The interested cholesterol molecule is shown as a yellow surface, and its ROH bead are shown in red. Water, ions, and other ciliary lipids/cholesterol molecules are omitted for clarity.

A.2 Chapter 5

Putative role of cholesterol in shaping the structural and functional dynamics of Smoothened (SMO)

Video 5.1: Distance between individual membrane cholesterol and specific cholesterol binding motifs/sites for all three replicates of SMO in the modeled ciliary membrane for all four systems.

Video 5.2: Pairwise comparison of the real average structure of each of the three cholesterol-bound systems (systems 2-4) with that of the apo system (system 1).

Video 5.3: Shows three views of the representative structure from each cluster (20 clusters) of system 1 (apo SMO) and changes in the CRD and ICD conformation. SMO is rendered as a domain-wise color scheme described in Figure 4.3 (main text), with CRD in orange, LD in magenta, and ICD in red. The helices were displayed in transparent ribbon representation with red, blue, green, pink, brown, gray, purple, and orange colors, respectively.

Video 5.4: Shows three views of the representative structure from each cluster (19 clusters) of system 2 (holo SMO-TMD) and changes in the CRD and ICD conformation. SMO is rendered as a domain-wise color scheme described in Figure 4.3 (main text), with CRD in orange, LD in magenta, and ICD in red. The helices were displayed in transparent ribbon representation with red, blue, green, pink, brown, gray, purple, and orange colors, respectively. The TMD-bound cholesterol was shown as a VDW representation in yellow, with OH as a red sphere.

Video 5.5: Shows three views of the representative structure from each cluster (20 clusters) of system 3 (holo SMO-CRD) and changes in the CRD and ICD conformation. SMO is rendered as a domain-wise color scheme described in Figure 4.3 (main text), with CRD in orange, LD in magenta, and ICD in red. The helices were displayed in transparent ribbon representation with red, blue, green, pink, brown, gray, purple, and orange colors, respectively. The CRD-bound cholesterol was shown as a VDW representation in yellow, with OH as a red sphere.

Video 5.6: Shows three views of the representative structure from each cluster (20 clusters) of system 4 (holo SMO) and changes in the CRD and ICD conformation. SMO is rendered as a domain-wise color scheme described in Figure 4.3 (main text), with CRD in orange, LD in magenta, and ICD in red. The helices were displayed in transparent ribbon representation with red, blue, green, pink, brown, gray, purple, and orange colors, respectively. Both TMD and CRD bound cholesterols were shown as a VDW representation in yellow, with OH as a red sphere.

link: `https://figshare.com/s/e60903dfa3063bfe9dd0`

Related Publications

Journal Publications

1. Kumari, S.; Mitra, A.; Bulusu, G. Structural dynamics of Smoothened (SMO) in the ciliary membrane and its interaction with membrane lipids. Biochim Biophys Acta Biomembr. 2022;1864(8):183946.

2. Kumari, S.; Mitra, A.; Bulusu, G. Putative role of cholesterol in shaping the structural and functional dynamics of Smoothened (SMO). J Phys Chem B. 2023;127(44):9476-9495.

Poster Presentations

1. Kumari, S.; Mitra, A.; Bulusu, G. Investigating cholesterol binding to cholesterol-binding motifs of SMO-TMD. 44th Indian Biophysical Society Meeting (IBS 2022) 2022, organized by National Centre for Biological Sciences (NCBS), Bangalore (March 27-29, 2023).

2. Kumari, S.; Mitra, A.; Bulusu, G. Structural variations of different cholesterol-bound states of Smoothened (SMO). MBU50 Symposium 2022, organized by Indian Institute of Science (IISc), Bangalore January 23-25, 2023.

3. Kumari, S.; Mitra, A.; Bulusu, G. Investigating cholesterol binding to cholesterol-binding motifs of SMO-TMD. 44th Indian Biophysical Society Meeting (IBS 2022) 2022, organized by ACTREC, Tata Memorial Centre, Navi Mumbai (March 30-31 and April 1, 2022).

4. Kumari, S.; Mitra, A.; Bulusu, G. Structural modulation in Smoothened (SMO) in ciliary membrane and its interaction with membrane lipids. Molecular Modeling (MM) 2021. Organized by Association of Molecular Modelers of Australasia (December 6-8, 2021).

5. Kumari, S.; Mitra, A.; Bulusu, G. Structural modulation in Smoothened (SMO) in ciliary membrane and its interaction with membrane lipids". EMBO Workshop, Advances and Challenges in Biomolecular Simulations 2021. Organized by Alexandre Bonvin, Utrecht University, NL, October 18 – 21.

6. Kumari, S.; Mitra, A.; Bulusu, G. Structural modulation in Smoothened (SMO) in ciliary membrane and its interaction with membrane lipids. HySci 2020. Organized by CCMB Hyderabad, December 17-18.

Award/Achievements

1. Fee waiver for virtual EMBO Workshop-Advances and Challenges in Biomolecular Simulation (2021).
2. Best poster award in HySci 2020 organized by CSIR-CCMB.
3. Travel grant for IBS 2020 conference.

Bibliography

[1] S. A. Adcock and J. A. McCammon. Molecular dynamics: survey of methods for simulating the activity of proteins. *Chemical reviews*, 106(5):1589–1615, 2006.

[2] J. Alcedo and M. Noll. Hedgehog and its patched-smoothened receptor complex: a novel signalling mechanism at the cell surface. *Biological Chemistry Hoppe-Seyler*, 378:583–590, 1997.

[3] C. Alexandre, A. Jacinto, and P. W. Ingham. Transcriptional activation of hedgehog target genes in drosophila is mediated directly by the cubitus interruptus protein, a member of the GLI family of zinc finger DNA-binding proteins. *Genes & development*, 10(16):2003–2013, 1996.

[4] S.-B. T. Amos, A. C. Kalli, J. Shi, and M. S. Sansom. Membrane recognition and binding by the phosphatidylinositol phosphate kinase PIP5k1a: a multiscale simulation study. *Structure*, 27(8):1336–1346, 2019.

[5] A. M. Arensdorf, S. Marada, and S. K. Ogden. Smoothened regulation: a tale of two signals. *Trends in pharmacological sciences*, 37(1):62–72, 2016.

[6] G. E. Arnold and R. L. Ornstein. Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome p450bm-3. *Biophysical journal*, 73(3):1147–1159, 1997.

[7] W. B. Asher, P. Geggier, M. D. Holsey, G. T. Gilmore, A. K. Pati, J. Meszaros, D. S. Terry, S. Mathiasen, M. J. Kaliszewski, and M. D. McCauley. Single-molecule FRET imaging of GPCR dimers in living cells. *Nature methods*, 18(4):397–405, 2021.

[8] S. X. Atwood, K. Y. Sarin, R. J. Whitson, J. R. Li, G. Kim, M. Rezaee, M. S. Ally, J. Kim, C. Yao, and A. L. S. Chang. Smoothened variants explain the majority of drug resistance in basal cell carcinoma. *Cancer cell*, 27(3):342–353, 2015.

[9] K. L. Ayers and P. P. Thérond. Evaluating smoothened as a g-protein-coupled receptor for hedgehog signalling. *Trends in cell biology*, 20(5):287–298, 2010.

[10] C. B. Bai, W. Auerbach, J. S. Lee, D. Stephen, and A. L. Joyner. Gli2, but not gli1, is required for initial shh signaling and ectopic activation of the shh pathway. *Development*, 129(20):4753–4761, 2002.

[11] J. A. Ballesteros and H. Weinstein. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in g protein-coupled receptors. In *Methods in neurosciences*, volume 25, pages 366–428. 1995.

[12] F. Bangs and K. V. Anderson. Primary cilia and mammalian hedgehog signaling. *Cold Spring Harbor perspectives in biology*, 9(5):a028175, 2017.

[13] P. Bansal, S. Dutta, and D. Shukla. Activation mechanism of the human smoothened receptor. *Biophysical Journal*, 122(7):1400–1413, 2023.

[14] M. T. Barakat, E. W. Humke, and M. P. Scott. Learning from jekyll to control hyde: Hedgehog signaling in development and cancer. *Trends in molecular medicine*, 16(8):337–348, 2010.

[15] P. A. Beachy, S. S. Karhadkar, and D. M. Berman. Mending and malignancy. *Nature*, 431(7007):402–402, 2004.

[16] M. Bidet, O. Joubert, B. Lacombe, M. Ciantar, R. Nehmé, P. Mollat, L. Brétillon, H. Faure, R. Bittman, and M. Ruat. The hedgehog receptor patched is involved in cholesterol transport. *PloS one*, 6(9):e23834, 2011.

[17] P. Björkholm, A. M. Ernst, M. Hacke, F. Wieland, B. Brügger, and G. von Heijne. Identification of novel sphingolipid-binding motifs in mammalian membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1838(8):2066–2070, 2014.

[18] R. Blassberg, J. I. Macrae, J. Briscoe, and J. Jacob. Reduced cholesterol levels impair smoothened activation in smith–lemli–opitz syndrome. *Human molecular genetics*, 25(4):693–705, 2016.

[19] S. Blotta, J. Jakubikova, T. Calimeri, A. M. Roccaro, N. Amodio, A. K. Azab, U. Foresta, C. S. Mitsiades, M. Rossi, and K. Todoerti. Canonical and noncanonical hedgehog pathway in the pathogenesis of multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 120(25):5002–5013, 2012.

[20] R. Bradley and R. Radhakrishnan. Coarse-grained models for protein-cell membrane interactions. *Polymers*, 5(3):890–936, 2013.

[21] J. Briscoe and P. P. Thérond. The mechanisms of hedgehog signalling and its roles in development and disease. *Nature reviews Molecular cell biology*, 14(7):416–429, 2013.

[22] S. Buchoux. FATSLiM: a fast and robust software to analyze MD simulations of membranes. *Bioinformatics*, 33(1):133–134, 2017.

[23] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[24] E. F. Byrne, R. Sircar, P. S. Miller, G. Hedger, G. Luchetti, S. Nachtergaele, M. D. Tully, L. Mydock-McGrane, D. F. Covey, and R. P. Rambo. Structural basis of smoothened regulation by its extracellular domains. *Nature*, 535(7613):517–522, 2016.

[25] A. Callejo, J. Culi, and I. Guerrero. Patched, the receptor of hedgehog, is a lipoprotein receptor. *Proceedings of the National Academy of Sciences*, 105(3):912–917, 2008.

[26] A. Chattopadhyay, Y. D. Paila, S. Shrivastava, S. Tiwari, P. Singh, and J. Fantini. Sphingolipid-binding domain in the serotonin 1a receptor. In *Biochemical Roles of Eukaryotic Cell Surface Macromolecules: 2011 ISCSM Proceedings*, pages 279–293. Springer, 2012.

[27] H. Chen, Y. Liu, and X. Li. Structure of human dispatched-1 provides insights into hedgehog ligand biogenesis. *Life science alliance*, 3(8), 2020.

[28] J. K. Chen, J. Taipale, K. E. Young, T. Maiti, and P. A. Beachy. Small molecule modulation of smoothened activity. *Proceedings of the National Academy of Sciences*, 99(22):14071–14076, 2002.

[29] Y. Chen, S. Li, C. Tong, Y. Zhao, B. Wang, Y. Liu, J. Jia, and J. Jiang. G protein-coupled receptor kinase 2 promotes high-level hedgehog signaling by regulating the active state of smo through kinase-dependent and kinase-independent mechanisms in drosophila. *Genes & development*, 24(18):2054–2067, 2010.

[30] Y. Chen, N. Sasai, G. Ma, T. Yue, J. Jia, J. Briscoe, and J. Jiang. Sonic hedgehog dependent phosphorylation by CK1α and GRK2 is required for ciliary accumulation and activation of smoothened. *PLoS biology*, 9(6):e1001083, 2011.

[31] S. Y. Cheng and J. M. Bishop. Suppressor of fused represses gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. *Proceedings of the National Academy of Sciences*, 99(8):5442–5447, 2002.

[32] Z. Choudhry, A. A. Rikani, A. M. Choudhry, S. Tariq, F. Zakaria, M. W. Asghar, M. K. Sarfraz, K. Haider, A. A. Shafiq, and N. J. Mobassarah. Sonic hedgehog signalling pathway: a complex network. *Annals of neurosciences*, 21(1):28, 2014.

[33] M. Chávez, S. Ena, J. Van Sande, A. de Kerchove d'Exaerde, S. Schurmans, and S. N. Schiffmann. Modulation of ciliary phosphoinositide content regulates trafficking and sonic hedgehog signaling output. *Developmental cell*, 34(3):338–350, 2015.

[34] A. F. Cooper, K. P. Yu, M. Brueckner, L. L. Brailey, L. Johnson, J. M. McGrath, and A. E. Bale. Cardiac and CNS defects in a mouse with targeted disruption of suppressor of fused. *Development*, 132(19):4407–4417, 2005.

[35] K. C. Corbit, P. Aanstad, V. Singla, A. R. Norman, D. Y. Stainier, and J. F. Reiter. Vertebrate smoothened functions at the primary cilium. *Nature*, 437(7061):1018–1021, 2005.

[36] A. C. E. Dahl, M. Chavent, and M. S. Sansom. Bendix: intuitive helix geometry analysis and abstraction. *Bioinformatics*, 28(16):2193–2194, 2012.

[37] P. Dai, H. Akimaru, Y. Tanaka, T. Maekawa, M. Nakafuku, and S. Ishii. Sonic hedgehog-induced activation of the gli1promoter is mediated by GLI3. *Journal of Biological Chemistry*, 274(12):8143–8152, 1999.

[38] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of chemical physics*, 98(12):10089–10092, 1993.

[39] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, and J. S. Richardson. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research*, 35:W375–W383, 2007.

[40] D. H. de Jong, G. Singh, W. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schafer, X. Periole, D. P. Tieleman, and S. J. Marrink. Improved parameters for the martini coarse-grained protein force field. *Journal of chemical theory and computation*, 9(1):687–697, 2013.

[41] D. L. DeCamp, T. M. Thompson, F. J. de Sauvage, and M. R. Lerner. Smoothened activates gαi-mediated signaling in frog melanophores. *Journal of Biological Chemistry*, 275(34):26322–26327,

2000.

[42] N. Denef, D. Neubüser, L. Perez, and S. M. Cohen. Hedgehog induces opposite changes in turnover and subcellular localization of patched and smoothened. *Cell*, 102(4):521–531, 2000.

[43] I. Deshpande, J. Liang, D. Hedeen, K. J. Roberts, Y. Zhang, B. Ha, N. R. Latorraca, B. Faust, R. O. Dror, and P. A. Beachy. Smoothened stimulation by membrane sterols drives hedgehog pathway activity. *Nature*, 571(7764):284–288, 2019.

[44] G. J. Dijkgraaf, B. Alicke, L. Weinmann, T. Januario, K. West, Z. Modrusan, D. Burdick, R. Goldsmith, K. Robarge, and D. Sutherlin. Small molecule inhibition of GDC-0449 refractory smoothened mutants and downstream mechanisms of drug ResistanceMechanisms of resistance to a hedgehog pathway inhibitor. *Cancer research*, 71(2):435–444, 2011.

[45] X. Ding, Y. Yang, B. Han, C. Du, N. Xu, H. Huang, T. Cai, A. Zhang, Z.-G. Han, and W. Zhou. Transcriptomic characterization of hepatocellular carcinoma with CTNNB1 mutation. *PLoS One*, 9(5):e95307, 2014.

[46] Y.-L. Ding, Q.-S. Wang, W.-M. Zhao, and L. Xiang. Expression of smoothened protein in colon cancer and its prognostic value for postoperative liver metastasis. *Asian Pacific Journal of Cancer Prevention*, 13(8):4001–4005, 2012.

[47] J. M. Duarte, N. Biyani, K. Baskaran, and G. Capitani. An analysis of oligomerization interfaces in transmembrane proteins. *BMC structural biology*, 13(1):1–11, 2013.

[48] M. Ferraiolo, H. Atik, R. Ponthot, I. B. do Nascimento, P. Beckers, C. Stove, and E. Hermans. Receptor density influences ligand-induced dopamine d2l receptor homodimerization. *European Journal of Pharmacology*, 911:174557, 2021.

[49] N. Fuse, T. Maiti, B. Wang, J. A. Porter, T. M. T. Hall, D. J. Leahy, and P. A. Beachy. Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proceedings of the National Academy of Sciences*, 96(20):10992–10999, 1999.

[50] N. Garcia, M. Ulin, A. Al-Hendy, and Q. Yang. The role of hedgehog pathway in female cancers. *Journal of cancer science and clinical therapeutics*, 4(4):487, 2020.

[51] F. Ghirga, M. Mori, and P. Infante. Current trends in hedgehog signaling pathway inhibition by small molecules. *Bioorganic & medicinal chemistry letters*, 28(19):3131–3140, 2018.

[52] E. D. Gigante and T. Caspary. Signaling in the primary cilium through the lens of the hedgehog pathway. *Wiley Interdisciplinary Reviews: Developmental Biology*, 9(6):e377, 2020.

[53] X. Gong, H. Qian, P. Cao, X. Zhao, Q. Zhou, J. Lei, and N. Yan. Structural basis for the recognition of sonic hedgehog by human patched1. *Science*, 361(6402):eaas8935, 2018.

[54] E. Gracia, E. Moreno, A. Cortés, C. Lluís, J. Mallol, P. J. McCormick, E. I. Canela, and V. Casadó. Homodimerization of adenosine a1 receptors in brain cortex explains the biphasic effects of caffeine. *Neuropharmacology*, 71:56–69, 2013.

[55] B. J. Grant, A. P. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. Caves. Bio3d: an r package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696, 2006.

[56] B. Guleng, K. Tateishi, M. Ohta, Y. Asaoka, A. Jazag, L. Lian-Jie, Y. Tanaka, M. Tada, M. Seto, and F. Kanai. Smoothened gene mutations found in digestive cancer have no aberrant hedgehog signaling activity. *Journal of Gastroenterology*, 41(12):1238, 2006.

[57] S. Gupta, N. Takebe, and P. LoRusso. Targeting the hedgehog pathway in cancer. *Therapeutic advances in medical oncology*, 2(4):237–250, 2010.

[58] M. A. Hanson, V. Cherezov, M. T. Griffith, C. B. Roth, V.-P. Jaakola, E. Y. Chien, J. Velasquez, P. Kuhn, and R. C. Stevens. A specific cholesterol binding site is established by the 2.8 \AA structure of the human β2-adrenergic receptor. *Structure*, 16(6):897–905, 2008.

[59] A. Hasbi, B. F. O'Dowd, and S. R. George. Heteromerization of dopamine d2 receptors with dopamine d1 or d5 receptors generates intracellular calcium signaling by different mechanisms. *Current opinion in pharmacology*, 10(1):93–99, 2010.

[60] G. Hedger, H. Koldsø, M. Chavent, C. Siebold, R. Rohatgi, and M. S. Sansom. Cholesterol interaction sites on the transmembrane domain of the hedgehog signal transducer and class FG protein-coupled receptor smoothened. *Structure*, 27(3):549–559, 2019.

[61] J. A. Hern, A. H. Baig, G. I. Mashanov, B. Birdsall, J. E. Corrie, S. Lazareno, J. E. Molloy, and N. J. Birdsall. Formation and dissociation of m1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules. *Proceedings of the National Academy of Sciences*, 107(6):2693–2698, 2010.

[62] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, 18(12):1463–1472, 1997.

[63] S. A. Hollingsworth and R. O. Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.

[64] A. Hu and B.-L. Song. The interplay of patched, smoothened and cholesterol in hedgehog signaling. *Current opinion in cell biology*, 61:31–38, 2019.

[65] A. Hu, J.-Z. Zhang, J. Wang, C.-C. Li, M. Yuan, G. Deng, Z.-C. Lin, Z.-P. Qiu, H.-Y. Liu, and X.-W. Wang. Cholesterylation of smoothened is a calcium-accelerated autoreaction involving an intramolecular ester intermediate. *Cell Research*, 32(3):288–301, 2022.

[66] J. Huang and A. D. MacKerell Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry*, 34(25):2135–2145, 2013.

[67] P. Huang, D. Nedelcu, M. Watanabe, C. Jao, Y. Kim, J. Liu, and A. Salic. Cellular cholesterol directly activates smoothened in hedgehog signaling. *Cell*, 166(5):1176–1187, 2016.

[68] P. Huang, S. Zheng, B. M. Wierbowski, Y. Kim, D. Nedelcu, L. Aravena, J. Liu, A. C. Kruse, and A. Salic. Structural basis of smoothened activation in hedgehog signaling. *Cell*, 174(2):312–324, 2018.

[69] D. Huangfu and K. V. Anderson. Signaling from smo to Ci/Gli: conservation and divergence of hedgehog pathways from drosophila to vertebrates. *Development*, 133(1):3–14, 2006.

[70] C.-c. Hui and S. Angers. Gli proteins in development and disease. *Annual review of cell and developmental biology*, 27(1):513–537, 2011.

[71] W. Humphrey, A. Dalke, and K. Schulten. VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1):33–38, 1996.

[72] A. R. i Altaba. Gli proteins and hedgehog signaling: development and cancer. *Trends in genetics*, 15(10):418–425, 1999.

[73] H. I. Ingólfsson, M. N. Melo, F. J. Van Eerden, C. Arnarez, C. A. Lopez, T. A. Wassenaar, X. Periole, A. H. De Vries, D. P. Tieleman, and S. J. Marrink. Lipid organization of the plasma membrane. *Journal of the american chemical society*, 136(41):14554–14559, 2014.

[74] M. Jafurulla and A. Chattopadhyay. Sphingolipids in the function of g protein-coupled receptors. *European journal of pharmacology*, 763:241–246, 2015.

[75] K.-S. Jeng, I. Sheen, C.-M. Leu, P.-H. Tseng, and C.-F. Chang. The role of smoothened in cancer. *International Journal of Molecular Sciences*, 21(18):6863, 2020.

[76] J. Jiang and C.-c. Hui. Hedgehog signaling in development and cancer. *Developmental cell*, 15(6):801–812, 2008.

[77] K. Jiang, Y. Liu, J. Fan, J. Zhang, X.-A. Li, B. M. Evers, H. Zhu, and J. Jia. PI (4) p promotes phosphorylation and conformational change of smoothened through interaction with its c-terminal tail. *PLoS biology*, 14(2):e1002375, 2016.

[78] S. Jo, T. Kim, V. G. Iyer, and W. Im. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of computational chemistry*, 29(11):1859–1865, 2008.

[79] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*, 79(2):926–935, 1983.

[80] S. Kaesler, B. Lüscher, and U. Rüther. Transcriptional activity of GLI1 is negatively regulated by protein kinase a. *Biological Chemistry*, 381(7):545–551, 2000.

[81] J. Kim, E. Y. Hsia, J. Kim, N. Sever, P. A. Beachy, and X. Zheng. Simultaneous measurement of smoothened entry into and exit from the primary cilium. *PLoS One*, 9(8):e104070, 2014.

[82] M. Kinnebrew, E. J. Iverson, B. B. Patel, G. V. Pusapati, J. H. Kong, K. A. Johnson, G. Luchetti, K. M. Eckert, J. G. McDonald, and D. F. Covey. Cholesterol accessibility at the ciliary membrane controls hedgehog signaling. *Elife*, 8:e50051, 2019.

[83] M. Kinnebrew, G. Luchetti, R. Sircar, S. Frigui, L. V. Viti, T. Naito, F. Beckert, Y. Saheki, C. Siebold, and A. Radhakrishnan. Patched 1 reduces the accessibility of cholesterol in the outer leaflet of membranes. *Elife*, 10, 2021.

[84] M. Kinnebrew, R. E. Woolley, T. B. Ansell, E. F. Byrne, S. Frigui, G. Luchetti, R. Sircar, S. Nachtergaele, L. Mydock-McGrane, and K. Krishnan. Patched 1 regulates smoothened by controlling sterol binding to its extracellular cysteine-rich domain. *Science Advances*, 8(22):eabm5563, 2022.

[85] Y. Kise, A. Morinaka, S. Teglund, and H. Miki. Sufu recruits GSK3β for efficient processing of gli3. *Biochemical and biophysical research communications*, 387(3):569–574, 2009.

[86] J. H. Kong, C. Siebold, and R. Rohatgi. Biochemical mechanisms of vertebrate hedgehog signaling. *Development*, 146(10):dev166892, 2019.

[87] A. J. Kooistra, S. Mordalski, G. Pándy-Szekeres, M. Esguerra, A. Mamyrbekov, C. Munk, K. G. M., and D. E. Gloriam. GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Research*, 49:D335–D343, 2021.

[88] T. Kotani. Protein kinase a activity and hedgehog signaling pathway. *Vitamins & Hormones*, 88:273–291, 2012.

[89] A. Kotulak-Chrząszcz, Z. Kmieć, and P. M. Wierzbicki. Sonic hedgehog signaling pathway in gynecological and genitourinary cancer. *International Journal of Molecular Medicine*, 47(6):1–20, 2021.

[90] V. Kumar, A. K. Chaudhary, Y. Dong, H. A. Zhong, G. Mondal, F. Lin, V. Kumar, and R. I. Mahato. Design, synthesis and biological evaluation of novel hedgehog inhibitors for treating pancreatic cancer. *Scientific reports*, 7(1):1–15, 2017.

[91] S. Kumari, A. Mitra, and G. Bulusu. Structural dynamics of smoothened (SMO) in the ciliary membrane and its interaction with membrane lipids. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, page 183946, 2022.

[92] C.-W. Lam, J. Xie, K.-F. To, H.-K. Ng, K.-C. Lee, N. W.-F. Yuen, P.-L. Lim, L. Y.-S. Chan, S.-F. Tong, and F. McCormick. A frequent activated smoothened mutation in sporadic basal cell carcinomas. *Oncogene*, 18(3):833–836, 1999.

[93] N. R. Latorraca, A. J. Venkatakrishnan, and R. O. Dror. GPCR dynamics: structures in motion. *Chemical reviews*, 117(1):139–155, 2017.

[94] S. P. Lee, B. F. O'Dowd, R. D. Rajaram, T. Nguyen, and S. R. George. D2 dopamine receptor homodimerization is mediated by multiple sites of interaction, including an intermolecular interaction involving transmembrane domain 4. *Biochemistry*, 42(37):11023–11031, 2003.

[95] S. P. Lee, Z. Xie, G. Varghese, T. Nguyen, B. F. O'Dowd, and S. R. George. Oligomerization of dopamine and serotonin receptors. *Neuropsychopharmacology*, 23(1):S32–S40, 2000.

[96] T. Li, X. Liao, P. Lochhead, T. Morikawa, M. Yamauchi, R. Nishihara, K. Inamura, S. A. Kim, K. Mima, and Y. Sukawa. SMO expression in colorectal cancer: associations with clinical, pathological, and molecular features. *Annals of surgical oncology*, 21(13):4164–4173, 2014.

[97] G. Luchetti, R. Sircar, J. H. Kong, S. Nachtergaele, A. Sagner, E. F. Byrne, D. F. Covey, C. Siebold, and R. Rohatgi. Cholesterol activates the g-protein coupled receptor smoothened to promote hedgehog signaling. *Elife*, 5:e20304, 2016.

[98] D. Maier, S. Cheng, D. Faubert, and D. R. Hipfner. A broadly conserved g-protein-coupled receptor kinase phosphorylation mechanism controls drosophila smoothened activity. *PLoS genetics*, 10(7):e1004399, 2014.

[99] F. Mancia, Z. Assur, A. G. Herman, R. Siegel, and W. A. Hendrickson. Ligand sensitivity in dimeric associations of the serotonin 5ht2c receptor. *EMBO reports*, 9(4):363–369, 2008.

[100] A. Manglik, A. C. Kruse, T. S. Kobilka, F. S. Thian, J. M. Mathiesen, R. K. Sunahara, L. Pardo, W. I. Weis, B. K. Kobilka, and S. Granier. Crystal structure of the -opioid receptor bound to a morphinan antagonist. *Nature*, 485(7398):321–326, 2012.

[101] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries. The MARTINI force field: coarse grained model for biomolecular simulations. *The journal of physical chemistry B*, 111(27):7812–7824, 2007.

[102] V. Martín, G. Carrillo, C. Torroja, and I. Guerrero. The sterol-sensing domain of patched protein seems to control smoothened activity through patched vesicular trafficking. *Current Biology*, 11(8):601–607, 2001.

[103] J. M. McCabe and D. J. Leahy. Smoothened goes molecular: new pieces in the hedgehog signaling puzzle. *Journal of Biological Chemistry*, 290(6):3500–3507, 2015.

[104] E. C. Meng, E. F. Pettersen, G. S. Couch, C. C. Huang, and T. E. Ferrin. Tools for integrated sequence-structure analysis with UCSF chimera. *BMC bioinformatics*, 7(1):1–10, 2006.

[105] G. Milligan. A day in the life of a g protein-coupled receptor: the contribution to function of g protein-coupled receptor dimerization. *British journal of pharmacology*, 153:S216–S229, 2008.

[106] T. Mirzadegan, G. Benkö, S. Filipek, and K. Palczewski. Sequence analyses of g-protein-coupled receptors: similarities to rhodopsin. *Biochemistry*, 42(10):2759–2767, 2003.

[107] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink. The MARTINI coarse-grained force field: extension to proteins. *Journal of chemical theory and computation*, 4(5):819–834, 2008.

[108] R. C. Moraes, X. Zhang, N. Harrington, J. Y. Fung, M.-F. Wu, S. G. Hilsenbeck, D. C. Allred, and M. T. Lewis. Constitutive activation of smoothened (SMO) in mammary glands of transgenic mice leads to increased proliferation, altered differentiation and ductal dysplasia. *Development*, 134(6):1231–42, 2007.

[109] S. Nachtergaele, L. K. Mydock, K. Krishnan, J. Rammohan, P. H. Schlesinger, D. F. Covey, and R. Rohatgi. Oxysterols are allosteric activators of the oncoprotein smoothened. *Nature chemical biology*, 8(2):211–220, 2012.

[110] F. Nakatsu. A phosphoinositide code for primary cilia. *Developmental cell*, 34(4):379–380, 2015.

[111] T. D. Nguyen, M. E. Truong, and J. F. Reiter. The intimate connection between lipids and hedgehog signaling. *Frontiers in Cell and Developmental Biology*, 10, 2022.

[112] A. Nicheperovich and A. Townsend-Nicholson. Towards precision oncology: The role of smoothened and its variants in cancer. *Journal of Personalized Medicine*, 12(10):1648, 2022.

[113] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

[114] J. Oates and A. Watts. Uncovering the intimate relationship between lipids, cholesterol and GPCR activation. *Current opinion in structural biology*, 21(6):802–807, 2011.

[115] Y. Pan, C. B. Bai, A. L. Joyner, and B. Wang. Sonic hedgehog signaling regulates gli2 transcriptional activity by suppressing its processing and degradation. *Molecular and cellular biology*, 26(9):3365–3377, 2006.

[116] H. L. Park, C. Bai, K. A. Platt, M. P. Matise, A. Beeghly, C. C. Hui, M. Nakashima, and A. L. Joyner. Mouse gli1 mutants are viable but have defects in SHH signaling in combination with a gli2 mutation. *Development*, 127(8):1593–1605, 2000.

[117] J. H. Park, P. Scheerer, K. P. Hofmann, H.-W. Choe, and O. P. Ernst. Crystal structure of the ligand-free g-protein-coupled receptor opsin. *Nature*, 454(7201):183–187, 2008.

[118] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

[119] S. Pathi, S. Pagan-Westphal, D. P. Baker, E. A. Garber, P. Rayhorn, D. Bumcrot, C. J. Tabin, R. B. Pepinsky, and K. P. Williams. Comparative biological responses to human sonic, indian, and desert hedgehog. *Mechanisms of development*, 106(1):107–117, 2001.

[120] R. B. Pepinsky, P. Rayhorn, E. S. Day, A. Dergay, K. P. Williams, A. Galdes, F. R. Taylor, P. A. Boriack-Sjodin, and E. A. Garber. Mapping sonic hedgehog-receptor interactions by steric interference. *Journal of Biological Chemistry*, 275(15):10995–11001, 2000.

[121] M. Persson, D. Stamataki, P. te Welscher, E. Andersson, J. Böse, U. Rüther, J. Ericson, and J. Briscoe. Dorsal-ventral patterning of the spinal cord requires gli3 transcriptional repressor activity. *Genes & development*, 16(22):2865–2878, 2002.

[122] L. T. Potter, D. D. Flynn, H. E. Hanchett, D. L. Kalinoski, J. LuberNarod, and D. C. Mash. Independent m1 and m2 receptors-ligands, autoradiography and functions. *Trends in Pharmacological Sciences*, pages 22–31, 1984.

[123] X. Prasanna, A. Chattopadhyay, and D. Sengupta. Cholesterol modulates the dimer interface of the β2-adrenergic receptor via cholesterol occupancy sites. *Biophysical journal*, 106(6):1290–1300, 2014.

[124] X. Prasanna, M. Jafurulla, D. Sengupta, and A. Chattopadhyay. The ganglioside GM1 interacts with the serotonin1a receptor via the sphingolipid binding domain. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1858(11):2818–2826, 2016.

[125] X. Qi, L. Friedberg, R. De Bose-Boyd, T. Long, and X. Li. Sterols in an intramolecular channel of smoothened mediate hedgehog signaling. *Nature chemical biology*, 16(12):1368–1375, 2020.

[126] X. Qi, H. Liu, B. Thompson, J. McDonald, C. Zhang, and X. Li. Cryo-EM structure of oxysterol-bound human smoothened coupled to a heterotrimeric g i. *Nature*, 571(7764):279–283, 2019.

[127] Y. Qi, H. I. Ingólfsson, X. Cheng, J. Lee, S. J. Marrink, and W. Im. CHARMM-GUI martini maker for coarse-grained simulations with the martini force field. *Journal of chemical theory and computation*, 11(9):4486–4494, 2015.

[128] A. Radhakrishnan, R. Rohatgi, and C. Siebold. Cholesterol access in cellular membranes controls hedgehog signaling. *Nature chemical biology*, 16(12):1303–1313, 2020.

[129] D. R. Raleigh, N. Sever, P. K. Choksi, M. A. Sigg, K. M. Hines, B. M. Thompson, D. Elnatan, P. Jaishankar, P. Bisignano, and F. R. Garcia-Gonzalo. Cilia-associated oxysterols activate smoothened. *Molecular cell*, 72(2):316–327, 2018.

[130] T. K. Rimkus, R. L. Carpenter, S. Qasem, M. Chan, and H.-W. Lo. Targeting the sonic hedgehog signaling pathway: review of smoothened and GLI inhibitors. *Cancers*, 8(2):22, 2016.

[131] N. A. Riobo, B. Saucy, C. DiLizio, and D. R. Manning. Activation of heterotrimeric g proteins by smoothened. *Proceedings of the National Academy of Sciences*, 103(33):12607–12612, 2006.

[132] K. D. Robarge, S. A. Brunton, G. M. Castanedo, Y. Cui, M. S. Dina, R. Goldsmith, S. E. Gould, O. Guichert, J. L. Gunzner, and J. Halladay. GDC-0449—a potent inhibitor of the hedgehog pathway. *Bioorganic & medicinal chemistry letters*, 19(19):5576–5581, 2009.

[133] J. Robert. Textbook of cell signalling in cancer. *Cham: Springer International Publishing*, pages 27–41, 2015.

[134] D. Rognan and I. Mus-Veteau. Three-dimensional structure of the smoothened receptor: Implications for drug discovery. *The smoothened receptor in cancer and regenerative medicine*, pages 127–146, 2014.

[135] R. Rohatgi, L. Milenkovic, and M. P. Scott. Patched1 regulates hedgehog signaling at the primary cilium. *Science*, 317(5836):372–376, 2007.

[136] Rowlinson. The maxwell–boltzmann distribution. *Molecular Physics*, 103(21):2821–2828, 2005.

[137] M. Saqui-Salces and J. L. Merchant. Hedgehog signaling and gastrointestinal cancer. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1803(7):786–795, 2010.

[138] P. Sarkar and A. Chattopadhyay. Cholesterol interaction motifs in g protein-coupled receptors: Slippery hot spots? *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 12(4):e1481, 2020.

[139] D. Sengupta and A. Chattopadhyay. Identification of cholesterol binding sites in the serotonin1a receptor. *The journal of physical chemistry B*, 116(43):12991–12996, 2012.

[140] D. Sengupta, X. Prasanna, M. Mohole, and A. Chattopadhyay. Exploring GPCR–lipid interactions by molecular dynamics simulations: excitements, challenges, and the way forward. *The Journal of Physical Chemistry B*, 122(22):5727–5737, 2018.

[141] F. Shen, L. Cheng, A. E. Douglas, N. A. Riobo, and D. R. Manning. Smoothened is a fully competent activator of the heterotrimeric g protein gi. *Molecular pharmacology*, 83(3):691–697, 2013.

[142] D. Shi, X. Lv, Z. Zhang, X. Yang, Z. Zhou, L. Zhang, and Y. Zhao. Smoothened oligomerization/higher order clustering in lipid rafts is essential for high hedgehog activity transduction. *Journal of Biological Chemistry*, 288(18):12605–12614, 2013.

[143] Q. Shi, S. Li, J. Jia, and J. Jiang. The hedgehog-induced smoothened conformational switch assembles a signaling complex that activates fused by promoting its dimerization and phosphorylation. *Development*, 138(19):4219–4231, 2011.

[144] S. Shrivastava, M. Jafurulla, S. Tiwari, and A. Chattopadhyay. Identification of sphingolipid-binding motif in g protein-coupled receptors. *Biochemical and Biophysical Roles of Cell Surface Molecules*, pages 141–149, 2018.

[145] J. K. Sicklick, Y.-X. Li, A. Jayaraman, R. Kannangai, Y. Qi, P. Vivekanandan, J. W. Ludlow, K. Owzar, W. Chen, and M. S. Torbenson. Dysregulation of the hedgehog pathway in human hepatocarcinogenesis. *Carcinogenesis*, 27(4):748–757, 2006.

[146] R. D. Skeel. Variable step size destabilizes the störmer/leapfrog/verlet method. *BIT Numerical Mathematics*, 33(1):172–175, 1993.

[147] J. P. Slotte. Biological functions of sphingomyelins. *Progress in lipid research*, 52(4):424–437, 2013.

[148] P. J. Stansfeld, R. Hopkinson, F. M. Ashcroft, and M. S. Sansom. PIP2-binding site in kir channels: definition by multiscale biomolecular simulations. *Biochemistry*, 48(46):10926–10933, 2009.

[149] G. Straface, T. Aprahamian, A. Flex, E. Gaetani, F. Biscetti, R. C. Smith, G. Pecorini, E. Pola, F. Angelini, and E. Stigliano. Sonic hedgehog regulates angiogenesis and myogenesis during post-natal skeletal muscle regeneration. *Journal of cellular and molecular medicine*, 13(8):2424–2435, 2009.

[150] J. L. Sussman, D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Ritter, and E. E. Abola. Protein data bank (pdb): database of three-dimensional structural information of biological macromolecules. *Acta Crystallographica Section D: Biological Crystallography*, 54(6):1078–1084, 1998.

[151] J. Svärd, K. H. Henricson, M. Persson-Lek, B. Rozell, M. Lauth, . Bergström, J. Ericson, R. Toftgård, and S. Teglund. Genetic elimination of suppressor of fused reveals an essential repressor function in the mammalian hedgehog signaling pathway. *Developmental cell*, 10(2):187–197, 2006.

[152] G. J. Taghon, J. B. Rowe, N. J. Kapolka, and D. G. Isom. Predictable cholesterol binding sites in GPCRs lack consensus motifs. *Structure*, 29(5):499–506, 2021.

[153] J. Taipale and P. A. Beachy. The hedgehog and wnt signalling pathways in cancer. *nature*, 411(6835):349–354, 2001.

[154] J. Taipale, J. K. Chen, M. K. Cooper, B. Wang, R. K. Mann, L. Milenkovic, M. P. Scott, and P. A. Beachy. Effects of oncogenic mutations in smoothened and patched can be reversed by cyclopamine. *Nature*, 406(6799):1005–1009, 2000.

[155] J. Taipale, M. K. Cooper, T. Maiti, and P. A. Beachy. Patched acts catalytically to suppress the activity of smoothened. *Nature*, 418(6900):892–896, 2002.

[156] C. S. Tautermann, D. Seeliger, and J. M. Kriegl. What can we learn from molecular dynamics simulations for GPCR drug design? *Computational and Structural Biotechnology Journal*, 13:111–121, 2015.

[157] 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*, 102(13):5451–5459, 1995.

[158] B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski, and S. Filipek. Action of molecular switches in GPCRs-theoretical and experimental studies. *Current medicinal chemistry*, 19(8):1090–1109, 2012.

[159] A. S. Tsao, I. Wistuba, D. Xia, L. Byers, L. Diao, J. Wang, V. Papadimitrakopoulou, X. Tang, W. Lu, and H. Kadara. Germline and somatic smoothened mutations in non-small-cell lung cancer are potentially responsive to hedgehog inhibitor vismodegib. *JCO precision oncology*, 1:1–10, 2017.

[160] H. Tukachinsky, L. V. Lopez, and A. Salic. A mechanism for vertebrate hedgehog signaling: recruitment to cilia and dissociation of SuFu–gli protein complexes. *Journal of Cell Biology*, 191(2):415–428, 2010.

[161] P. J. Turner. Xmgrace. *Version*, 5:19, 2005.
[162] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. Berendsen. *GROMACS: fast, flexible, and free*, volume 26. 2005.
[163] M. Varjosalo, S.-P. Li, and J. Taipale. Divergence of hedgehog signal transduction mechanism between drosophila and mammals. *Developmental cell*, 10(2):177–186, 2006.
[164] L. Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.
[165] A. P. Visbal, H. L. LaMarca, H. Villanueva, M. J. Toneff, Y. Li, J. M. Rosen, and M. T. Lewis. Altered differentiation and paracrine stimulation of mammary epithelial cell proliferation by conditionally activated smoothened. *Developmental biology*, 352(1):116–127, 2011.
[166] T. Von Ohlen and J. E. Hooper. Hedgehog signaling regulates transcription through gli/ci binding sites in the wingless enhancer. *Mechanisms of development*, 68(1):149–156, 1997.
[167] B. Wang, J. F. Fallon, and P. A. Beachy. Hedgehog-regulated processing of gli3 produces an anterior/posterior repressor gradient in the developing vertebrate limb. *Cell*, 100(4):423–434, 2000.
[168] C. Wang, H. Wu, T. Evron, E. Vardy, G. W. Han, X.-P. Huang, S. J. Hufeisen, T. J. Mangano, D. J. Urban, and V. Katritch. Structural basis for smoothened receptor modulation and chemoresistance to anticancer drugs. *Nature communications*, 5(1):1–11, 2014.
[169] C. Wang, H. Wu, V. Katritch, G. W. Han, X.-P. Huang, W. Liu, F. Y. Siu, B. L. Roth, V. Cherezov, and R. C. Stevens. Structure of the human smoothened receptor bound to an antitumour agent. *Nature*, 497(7449):338–343, 2013.
[170] Q. T. Wang and R. A. Holmgren. Nuclear import of cubitus interruptus is regulated by hedgehog via a mechanism distinct from ci stabilization and ci activation. *Development*, 127(14):3131–3139, 2000.
[171] Y. Wang, Z. Zhou, C. T. Walsh, and A. P. McMahon. Selective translocation of intracellular smoothened to the primary cilium in response to hedgehog pathway modulation. *Proceedings of the National Academy of Sciences*, 106(8):2623–2628, 2009.
[172] T. A. Wassenaar, H. I. Ingólfsson, R. A. Bockmann, D. P. Tieleman, and S. J. Marrink. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *Journal of chemical theory and computation*, 11(5):2144–2155, 2015.
[173] T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of chemical theory and computation*, 10(2):676–690, 2014.
[174] U. Weierstall, D. James, C. Wang, T. A. White, D. Wang, W. Liu, J. C. Spence, R. B. Doak, G. Nelson, and P. Fromme. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature communications*, 5(1):1–6, 2014.
[175] L. E. Weiss, L. Milenkovic, J. Yoon, T. Stearns, and W. E. Moerner. Motional dynamics of single patched1 molecules in cilia are controlled by hedgehog and cholesterol. *Proceedings of the National*

Academy of Sciences, 116(12):5550–5557, 2019.

[176] X. Wen, C. K. Lai, M. Evangelista, J.-A. Hongo, F. J. de Sauvage, and S. J. Scales. Kinetics of hedgehog-dependent full-length gli3 accumulation in primary cilia and subsequent degradation. *Molecular and cellular biology*, 30(8):1910–1922, 2010.

[177] G. Wheway, L. Nazlamova, and J. T. Hancock. Signaling through the primary cilium. *Frontiers in cell and developmental biology*, 6:8, 2018.

[178] J. F. White, J. Grodnitzky, J. M. Louis, L. B. Trinh, J. Shiloach, J. Gutierrez, J. K. Northup, and R. Grisshammer. Dimerization of the class AG protein-coupled neurotensin receptor NTS1 alters g protein interaction. *Proceedings of the National Academy of Sciences*, 104(29):12199–12204, 2007.

[179] T. Williams and C. Kelley. Gnuplot 5.0: An interactive plotting programm. *URL http://gnuplot.info*, 2016.

[180] C. W. Wilson, M.-H. Chen, and P.-T. Chuang. Smoothened adopts multiple active and inactive conformations capable of trafficking to the primary cilium. *PloS one*, 4(4):e5182, 2009.

[181] E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, and R. M. Venable. CHARMM-GUI membrane builder toward realistic biological membrane simulations, 2014.

[182] X. Wu, J. Walker, J. Zhang, S. Ding, and P. G. Schultz. Purmorphamine induces osteogenesis by activation of the hedgehog signaling pathway. *Chemistry & biology*, 11(9):1229–1238, 2004.

[183] X. Xiao, J.-J. Tang, C. Peng, Y. Wang, L. Fu, Z.-P. Qiu, Y. Xiong, L.-F. Yang, H.-W. Cui, and X.-L. He. Cholesterol modification of smoothened is required for hedgehog signaling. *Molecular cell*, 66(1):154–162, 2017.

[184] Y. Xu and M. Havenith. Perspective: Watching low-frequency vibrations of water in biomolecular recognition by THz spectroscopy. *The Journal of chemical physics*, 143(17):170901, 2015.

[185] R. L. Yauch, G. J. Dijkgraaf, B. Alicke, T. Januario, C. P. Ahn, T. Holcomb, K. Pujara, J. Stinson, C. A. Callahan, and T. Tang. Smoothened mutation confers resistance to a hedgehog pathway inhibitor in medulloblastoma. *Science*, 326(5952):572–574, 2009.

[186] J. Zhang, Z. Liu, and J. Jia. Mechanisms of smoothened regulation in hedgehog signaling. *Cells*, 10(8):2138, 2021.

[187] J. Zhang, J. Yang, R. Jang, and Y. Zhang. GPCR-i-TASSER: a hybrid approach to g protein-coupled receptor structure modeling and the application to the human genome. *Structure*, 23(8):1538–1549, 2015.

[188] X. Zhang, F. Zhao, Y. Wu, J. Yang, G. W. Han, S. Zhao, A. Ishchenko, L. Ye, X. Lin, and K. Ding. Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. *Nature communications*, 8(1):1–10, 2017.

[189] Y. Zhang, D. P. Bulkley, Y. Xin, K. J. Roberts, D. E. Asarnow, A. Sharma, B. R. Myers, W. Cho, Y. Cheng, and P. A. Beachy. Structural basis for cholesterol transport-like activity of the hedgehog receptor patched. *Cell*, 175(5):1352–1364, 2018.

[190] Y. Zhao, C. Tong, and J. Jiang. Hedgehog regulates smoothened activity by inducing a conformational switch. *Nature*, 450(7167):252–258, 2007.

[191] Y. Zhu, R. M. James, A. Peter, C. Lomas, F. Cheung, D. J. Harrison, and S. A. Bader. Functional smoothened is required for expression of GLI3 in colorectal carcinoma cells. *Cancer letters*, 207(2):205–214, 2004.