

**STRUCTURE AND MORPHOLOGY OF MAGNESIUM  
CHANNEL BY COARSE-GRAINED MONTE CARLO AND  
MOLECULAR DYNAMICS SIMULATIONS**



A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy in Chemistry  
Department of Chemistry  
FACULTY OF SCIENCE  
Chulalongkorn University  
Academic Year 2020  
Copyright of Chulalongkorn University

โครงสร้างและสัณฐานวิทยาของแมงनीเซียมแซลนัลดโดยการจำลองมอนติคาโล และการจำลองเชิง  
พลวัตแบบคอสมอส



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต  
สาขาวิชาเคมี ภาควิชาเคมี  
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2563  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title                    STRUCTURE AND MORPHOLOGY OF  
MAGNESIUM CHANNEL BY COARSE-  
GRAINED MONTE CARLO AND  
MOLECULAR DYNAMICS SIMULATIONS  
By                                   Miss Warin Rangubpit  
Field of Study                 Chemistry  
Thesis Advisor                Professor PORNTHEP SOMPORNPI-  
SUT, Ph.D.  
Thesis Co Advisor            Professor Ras B. Pandey, Ph.D.

---

Accepted by the FACULTY OF SCIENCE, Chulalongkorn  
University in Partial Fulfillment of the Requirement for the Doctor of  
Philosophy

..... Dean of the FACULTY OF  
SCIENCE  
(Professor POLKIT SANGVANICH, Ph.D.)

#### DISSERTATION COMMITTEE

..... Chairman  
(Associate Professor VUDHICHAIR PARASUK,  
Ph.D.)

..... Thesis Advisor  
(Professor PORNTHEP SOMPORNPI-  
SUT, Ph.D.)

..... Thesis Co-Advisor  
(Professor Ras B. Pandey, Ph.D.)

..... Examiner  
(Professor SUPOT HANNONGBUA, Ph.D.)

..... Examiner  
(Professor Supason Wanichwecharungruang,  
Ph.D.)

..... Examiner  
(Assistant Professor Somsak Pianwanit, Ph.D.)

..... External Examiner  
(Assistant Professor Chalermopol Kanchanawarin,  
Ph.D.)



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

วารินทร์ ระวังพิศม์ : โครงสร้างและสัณฐานวิทยาของแมกนีเซียมแชนแนลโดยการจำลองมอนติ  
คาโล และการจำลองเชิงพลวัตแบบคอสมอส. ( **STRUCTURE AND  
MORPHOLOGY OF MAGNESIUM CHANNEL BY  
COARSE-GRAINED MONTE CARLO AND MOLECULAR  
DYNAMICS SIMULATIONS**) อ.ที่ปรึกษาหลัก : ศ. ดร.พรเทพ สมพรพิสุทธิ์,  
อ.ที่ปรึกษาร่วม : ศ. ดร.ราส บี ปานเคย์

แบบจำลองแบบหยาบ (CG) เป็นเครื่องมือที่มีประสิทธิภาพสำหรับการศึกษาระบบที่ซับซ้อนทางชีวโมเลกุลด้วยการใช้งานที่หลากหลาย เช่น การเคลื่อนที่ของโปรตีนขนาดใหญ่ การพับโปรตีน การประกอบตัวเอง เป็นต้น การจำลองพฤติกรรมดังกล่าวต้องใช้เวลาซึ่งไม่สามารถทำได้ในการจำลองแบบทุกอะตอม การศึกษานี้แนะนำวิธีการสำหรับการประยุกต์ใช้การจำลอง CG เพื่อตรวจสอบคุณสมบัติทางโครงสร้างและสัณฐานวิทยาของช่องแมกนีเซียมที่ทนต่อโคบอลต์ (CorA) จาก *Thermotoga maritima* แมกนีเซียม ( $Mg^{2+}$ ) มีหน้าที่สำคัญในกระบวนการทางสรีรวิทยาที่หลากหลาย การขาดแมกนีเซียมสามารถเพิ่มความเสี่ยงต่อปัญหาสุขภาพหลายอย่าง เช่น โรคเบาหวานโรคหัวใจ โรคหลอดเลือดและโรคกระดูกพรุน การขนส่ง  $Mg^{2+}$  ผ่านเมมเบรนนั้นอาศัยกลุ่มโปรตีนเมมเบรนเฉพาะที่เรียกว่าแมกนีเซียมแชนแนล ในขณะที่กลไกระดับโมเลกุลที่อยู่ภายใต้การควบคุมแคทไอออนไดวาเลนต์นี้ยังคงเป็นเรื่องที่ถกเถียงกันอยู่ แต่การทำความเข้าใจโครงสร้างและสัณฐานวิทยาของแมกนีเซียมแชนแนลในระดับโมเลกุลจะช่วยให้เข้าใจถึงพื้นฐานของสภาวะสมดุลของ  $Mg^{2+}$  ในเซลล์ งานนี้แบ่งออกเป็นสี่ส่วน ส่วนแรกสำหรับการตอบสนองเชิงอุณหภูมิต่อโครงสร้างของนาโนคิสก์ไลโปโปรตีนสองชนิดที่แตกต่างกันโดยใช้การจำลองพลวัตโมเลกุลแบบหยาบ ส่วนที่สองและสามแสดงให้เห็นถึงการประยุกต์ใช้การจำลองด้วยวิธีมอนติ คาร์โลแบบหยาบ เพื่อตรวจสอบการจัดระเบียบรวมกลุ่มด้วยตัวเองของ CorA และผลของเมทริกซ์ตัวถูกละลายที่มีต่อโครงสร้างของโปรตีนตามลำดับ สำหรับส่วนสุดท้ายใช้การจำลองพลวัตโมเลกุลแบบหยาบเพื่อตรวจสอบการตอบสนองเชิงโครงสร้างของ CorA เมื่อมีการปลดปล่อย  $Mg^{2+}$

สาขาวิชา เคมี

ลายมือชื่อนิติ

ปีการศึกษา 2563

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ลายมือชื่อ อ.ที่ปรึกษาร่วม

# # 6072827223 : MAJOR CHEMISTRY

KEYWORD: magnesium channel, coarse-grained, nanodiscs, coarse-grained molecular dynamics simulations, coarse-grained Monte Carlo simulation

Warin Rangubpit : STRUCTURE AND MORPHOLOGY OF MAGNESIUM CHANNEL BY COARSE-GRAINED MONTE CARLO AND MOLECULAR DYNAMICS SIMULATIONS.

Advisor: Prof. PORNTHEP SOMPORNPIST, Ph.D. Co-advisor: Prof. Ras B. Pandey, Ph.D.

Coarse-grained (CG) models have become a powerful tool for studying biomolecular complex systems with a wide range of potential applications such as large-scale protein motions, protein folding, self-assembly etc. Simulating such behaviors requires time scales which are not practically possible in all-atomistic simulations. This study presents a strategy for the application of CG simulations to investigate structural and morphological properties of cobalt-resistant magnesium channel (CorA) from *Thermotoga maritima*. Magnesium ( $Mg^{2+}$ ) plays a crucial function in a variety of physiological processes. The deficiency of magnesium can increase the risk of several health problems, for instance, diabetes, cardiovascular disease and osteoporosis. The transport of  $Mg^{2+}$  across the membrane is primarily facilitated by a specialized family of membrane proteins, called  $Mg^{2+}$  channels. While the molecular mechanism underlying the divalent cation regulation is still a subject of debate, understanding of structure and morphology of  $Mg^{2+}$  channels at the molecular level would provide insight into the fundamental basis of cellular  $Mg^{2+}$  homeostasis. This work has been divided into four parts. The first part explores thermal responses to structural organization of two different lipoprotein nanodiscs using CG molecular dynamics (MD) simulations. The second and third parts demonstrate the applications of CG Monte Carlo simulations to investigate self-organized globular bundles of CorA and the effect of solute matrix on the protein conformation, respectively. For the last part, CGMD simulations were conducted to investigate conformational responses of CorA upon the release of  $Mg^{2+}$ .

Field of Study: Chemistry

Student's Signature

Study:

.....

Academic Year: 2020

Advisor's Signature

Year:

.....

Co-advisor's Signature

.....

## ACKNOWLEDGEMENTS

This never-ending thesis ended during the almost as never-ending COVID-19 pandemic. Times were hard and many stories could be told, between these events though I managed to find the path and finish this research. This study would not be possible without the guidance and tremendous help of those who contributed and increased value to the completion of this study.

I wish to express my sincere thanks to my advisor Prof. Pornthep Sompornpisut not only for his super academic guidance but also for his words about learning, life, inspiration, motivation, and the chances he gave that helped to push me up. He never let any excellent opportunity pass me by if I wanted to try them. Aoyama Gakuin University in Japan is one of those great chances and I would like to thank Prof. Masashi Miyano and Dr. Hiromichi Saino sensei. They taught me to work smart and play hard in protein expression, X-ray crystallography field and fulfill my scientific passion. Moreover, my co-advisor, Prof. Ras B. Pandey who advise me to simulate biology in a physical style and open my vision further than ever. He was always supporting and took care of me when I was in Thailand and when in the United states of America. My gratitude also extends to: Prof. Peter Wolschann, Prof. Supot Hannongbua, and Prof. Vudhichai Parasuk who was my supporter, Assoc.Prof. Viwat Vchirawongkwin, Assoc.Prof. Somsak Pianwanit, Asst.Prof. Thanyada Rungrotmongkol, Dr. Nattapong Paiboonvarachart, Dr. Prompong pienpinijtham, Dr. Andrew William King, Dr. Brian G. Olson, Dr. Nopporn Keiyawet, Dr. Sirilak Kongkaew, Dr. Sunan Kitjaruwankul, Dr. Panisak Boonamnaj their scientific advice and sacrificed their time to help me to learn and gave me many kind

suggestions. All of my dissertation committees are also acknowledged for their helpful suggestion in my examination.

My dissertation is supported by Chulalongkorn university dusadi phipat scholarship. The funding from AGU CSE Short Internship Program. I would like to acknowledge HPC resources from The Center of Excellence in Computation Chemistry (CECC), Chulalongkorn University, Thailand, and The University of Southern Mississippi, USA.

Last but not least, I would like to give big thanks to Rita Porter, Thanawit Kuamit, Panyakorn Taweechat, Pisit Ierttanakij, Chayan Charoenpakdee, Pitcha Singhaphan, Mongkol Sapankaew, James A. Baker, Watcharit Kerdchuen, Thailand colleagues and all friends from Japan and USA for sharing academics, experiences, encouragements, and for being good friends to each other. I am glad to see you around!

Warin Rangubpit



# TABLE OF CONTENTS

	<b>Page</b>
ABSTRACT (THAI) .....	iii
ABSTRACT (ENGLISH).....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	viii
LIST OF FIGURES .....	ix
LIST OF EQUATIONS .....	xiv
CHAPTER I: INTRODUCTION.....	1
1.1 Significance of the studied protein .....	1
1.1.1 Ion Channels.....	2
1.1.2 Magnesium channel.....	4
1.2 Theoretical Background .....	7
1.2.1 Coarse-grained models for biomolecular systems.....	7
1.2.2 Martini Models .....	9
1.2.3 Molecular Mechanics Potential Energy Functions.....	9
1.2.4 Molecular dynamics and Monte Carlo simulations.....	11
1.3 Objectives .....	12
1.4 Relation research topics.....	12
1.5 Expected beneficial outcome.....	13
CHAPTER II.....	14
Thermally induced structural organization of nanodiscs by coarse-grained molecular dynamics simulations .....	14
2.1 Abstract.....	15
2.2 Introduction.....	15
2.3 Methodology.....	17
2.3.1 Molecular modeling of nanodisc .....	17

2.3.2 Coarse-grained molecular dynamics simulation .....	18
2.3.3 Trajectory analysis .....	18
2.4 Results and discussion .....	19
2.4.1 Nanodisc structural organization in response to temperature .....	20
2.4.2 Structural changes of MSP1 .....	24
2.4.3 Principal Component Analysis (PCA) .....	28
2.4.4 Solvation profile .....	29
2.4.5 Solvent accessible surface area .....	31
2.4.6 Radius of gyration of MSP1 .....	33
2.5 Conclusions.....	34
2.6 Supplementary figures .....	35
CHAPTER III .....	40
Globular bundles and entangled network of proteins (CorA) by a coarse-grained Monte Carlo simulation .....	40
3.1 Abstract.....	41
3.2 Introduction.....	41
3.3 Model.....	45
3.4 Results and discussion .....	48
3.4.1 Snapshots, contact map and profiles .....	49
3.4.2 RMS displacement and Radius of gyration.....	52
3.4.3 Structure factor .....	55
3.5 Conclusions.....	59
CHAPTER IV .....	61
Pinning the conformation of a protein (CorA) in a solute matrix with selective binding .....	61
4.1 Abstract.....	62
4.2 Introduction.....	62
4.3 Methodology.....	64
4.4 Results and discussion .....	66
4.5 Conclusions.....	72

4.6 Supporting information .....	73
CHAPTER V .....	76
Magnesium release-induced conformational changes of CorA magnesium channel .....	76
5.1 Abstract.....	77
5.2 Introduction.....	77
5.3 Methodology .....	79
5.3.1 All-atom and coarse-grained models.....	79
5.3.2 All-atom MD simulations.....	80
5.3.3 Coarse-grained MD simulations.....	81
<i>Distance restraints mimic the attraction between protomers</i> .....	81
<i>Simultaneous versus stepwise restraint removals</i> .....	83
5.3.4 MD trajectories analysis .....	83
5.4 Results and discussion .....	84
5.4.1 Validation: AAMD vs CGMD of the CorA closed conformation .....	84
5.4.2 Simultaneous vs sequentially stepwise removals.....	85
5.4.3 Large conformational changes in cytosolic domains upon the restraint removal.....	85
5.5 Conclusions.....	90
CHAPTER VI: CONCLUSION .....	92
6.1 Thermal responses of lipoprotein nanodiscs.....	92
6.2 Self-organized globular bundles and entangle networks of CorA.....	93
6.3 CorA conformations in a solute matrix .....	93
6.4 Magnesium release-induced conformational changes of CorA.....	93
REFERENCES .....	95
VITA.....	109

## LIST OF TABLES

	Page
<b>TABLE 1:</b> Solvent accessible surface areas per lipid of dppc and popc nanodiscs .....	32
<b>TABLE 2:</b> Distance restraints and the reference distance ( $r_{\text{ref}}$ ) between backbone beads of CG residues. Restraints between residue pairs are denoted .....	81



## LIST OF FIGURES

	Page
Figure I- 1: Magnesium and cell cycle progression.....	1
Figure I- 2: Three types of gated-ion channel.....	3
Figure I- 4: CorA protein components.....	7
Figure I- 5: Topologies of amino acid in Martini model .....	9
Figure II- 1: Nanodisc sideviews taken from MD snapshots to show a variation of the DPPC- and POPC-nanodisc shapes. Average order parameter, of the three terminal bonds of lipid acyl chains as a function of temperature.....	22
Figure II- 2: The flatness angle of nanodiscs as a function of temperature. Error bars represent the standard deviations. The digedral angle between two triangular planes is defined by four centerpoints located on the backbone arcs of the MSP1s. The SASA of the lipid tail at the rim of the bilayer disks for DPPC and POPC.. ..	23
Figure II- 3: The average RMSD of MSP1 in DPPC and POPC nanodiscs versus the temperature. The CG-based interaction energy of MSP1 with the peripheral lipids. For comparison, the values in the red bar graph are percentages relative to the values (100%) of the black bar graph at the same temperature.....	25
Figure II- 4: The per-residue RMSD of MSP1 versus simulation time. A color ranges from blue, white and red corresponds to the low-to-high RMSD value. ....	27
Figure II- 5: Root mean square fluctuation (RMSF) of each residue in protein chains P1 and P2 of MSP1 encapsulating DPPC and POPC membranes averaged over time .....	28
Figure II- 6: PCA plots of DPPC and POPC nanodisc projected between PC1 and PC2 that relate with structure of MSP1 snapshot.....	29
Figure II- 7: First solvation pair correlation function $g(r)$ and the number of underlying water molecules in DPPC and POPC nanodics at temperatures.....	31

Figure II- 8: Variation of the average solvent accessible surface area (SASA) of DPPC and POPC nanodisc with the temperature. ....	32
Figure II- 9: The average radius of gyration of MSP1 in DPPC and POPC nanodiscs. ....	33
Figure II- S1: Top and side views of the CG DPPC nanodisc and POPC nanodisc. MSP1 consisting of P1 and P2 encircles DPPC and POPC. ....	35
Figure II- S2: The RMSD of the backbone profile of MSP1. Snapshots of representative nanodiscs are shown at 250K and 300K .....	36
Figure II- S3: RMSD of MSP1 as a function of time of all-atomistic and coarse-grained simulations of DPPC nanodisc and POPC nanodisc. All-atom and coarse-grained snapshots. ....	37
Figure II- S4: Radial distribution function of water surrounding MSP1 and number of water represent in different temperature for DPPC and POPC nanodiscs. ....	38
Figure II- S5: The average number of water molecules surrounding Nanodisc and only MSP1 in the first shell .....	39
Figure II- S6: The radius of gyration profile of nanodiscs for DPPC and POPC at different temperature .....	39
Figure III- 1: Snapshots of iCorA and oCorA at the end of $10^7$ $\mu$ s. ....	42
Figure III- 2: Pentamer of CorA with inner core confined by a nanodisc. ....	44
Figure III- 3: Snapshots of CorA and iCorA with 100 protein chains and a high temperature at the end of $10^7$ time steps. Five proteins are shown in different colors in each simulation box. ....	50
Figure III- 4: Average number of residues around each of CorA at low and high temperatures.....	51
Figure III- 5: Average number of residues around each of iCorA residues in a crowded matrix at low and high temperatures.....	52
Figure III- 6: Variation of the root mean square (RMS) displacement ( $R_c$ ) with the time step of protein CorA and iCorA in dilute and crowded environment at representative low and high temperatures. ....	53
Figure III- 7: Variation of the average radius of gyration $R_g$ of CorA with the protein concentration $p$ at temperature $T = 0.020$ - $0.040$ . ....	54

- Figure III- 8: Variation of the average radius of gyration  $R_g$  of iCorA with the protein concentration  $p$  at temperature  $T = 0.015-0.040$ . ..... 55
- Figure III- 9: Structure factor  $S(q)$  versus the wave length ( $\lambda$ ) at temperature  $T = 0.030$  with a wide range of the number  $N_c$  (5-100) of CorA proteins in simulation box. .... 56
- Figure III- 10: Structure factor  $S(q)$  versus the wave length ( $\lambda$ ) at temperature  $T = 0.040$  with a wide range of the number  $N_c$  (5-100) of iCorA proteins in simulation box. The inset figure shows the variation of the radius of gyration of the protein with the number of protein chains. .... 58
- Figure III-11: Structure factor  $S(q)$  versus the wave length ( $\lambda$ ) in a crowded protein environment with  $N_c=100$  at a range of low to high temperatures  $T=0.015 - 0.040$ . Inset is the variation of the radius of gyration of iCorA with the temperature with number  $N_c = 100$  proteins in simulation box. .... 58
- Figure IV- 1: Snapshots of the protein chain CorA at the end of a million time steps. Bonds along the backbone of the protein contour; solute particles within the range of interaction of each residue. The interaction strength  $f = 2.0, 2.5, \text{ and } 3.5$  at  $T = 0.020$ . .... 67
- Figure IV- 2: Root mean square (RMS) displacement ( $R_c$ ) of the center of mass of the protein chain with the time step ( $t$ ) for a range of solute-residue interaction  $f = 1.00 - 4.00$  at a low temperature  $T = 0.020$ . The power-law exponent  $\nu$  ( $R_c \propto t^\nu$ ) in the short time regime ( $t \sim 10^2 - 10^4$ ) is also included for each interaction strength. .... 67
- Figure IV- 3: Binding profile of CorA for solute-residue interaction  $f=2.5, 3.0, 3.5, 4.0$  at a temperature  $T = 0.020$ . .... 69
- Figure IV- 4: Radius of gyration of CorA versus time step for solute-residue interaction  $f=1.0 - 7.0$  at  $T = 0.020$ . .... 70
- Figure IV- 5: Radius of gyration of the protein (CorA) versus solute-residue interaction strength  $f \equiv f_w$  at temperature  $T = 0.020, 0.032$ . .... 70
- Figure IV- 6: Structure factor ( $S$ ) of the protein versus wave length ( $\lambda$ ) for the interaction strength  $f_w = 2.5, 3.0, 3.5, \text{ and } 4.0$  at  $T = 0.020$ . The

inset is amplified version within a limited wavelength range spanning over the magnitude of the radius of gyration ( $R_g \sim 21$ ). The slope ( $1/\gamma = D$ ) is a measure of the effective dimension ( $D$ ) of the protein which is about 2 with lower values of $fw = 2.5, 3.0$ but varies with length scales at higher values of $fw = 3.5, 4.0$ . .....	71
Figure IV-S1: Variation of the x, y, z components ( $R_{gx}, R_{gy}, R_{gz}$ ) of the radius of gyration with the time step for representative interaction strength $f=1.0, 5.0$ at $T=0.032$ . .....	74
Figure S2: Variation of the average radius of gyration ( $R_g$ ) of the inner segment (iCorA) and outer segment (oCorA) with the solute-residue interaction strength ( $f$ ) at a temperature $T = 0.032$ .....	75
Figure V- 1: Backbone RMSD with respect to the reference CorA structure in the closed state versus simulation frames for AAMD and CGMD simulations.....	84
Figure V- 2 : RMSD with respect to the reference CorA structure in the closed state extracted from the 4 $\mu$ s CGMD simulations with simultaneous and stepwise removals. RMSD values were measured on the basis of individual protein chain, i.e., A, B, C, D and E and the overall pentamer. ....	85
Figure V- 3 : Contour maps representing backbone RMSD per residue as a function of simulation time obtained from CG simulations of simultaneous and stepwise restraint removals. The RMSD plot shows individual chain separately (i.e., A, B, C, D and E). .....	86
Figure V- 4: MD snapshots representing conformational changes of the cytosolic domain obtained from simultaneous and sequentially stepwise restraint removals at various simulation time. Cryo-EM structure of open state II (pdb 3jch). .....	87
Figure V- 5: Inter-subunit distance as a function of time obtained from MD trajectories of simultaneous and stepwise removal simulations. The distances between subunits were measured using the center of mass of a fragment corresponding to residues 278-280. ....	89



Figure V- 6: Inter-subunit distance as a function of time obtained from MD trajectories of Trial-III and Trial-IV systems. The pentamer arrangement of the residues 278-280 of Trial-III and Trial-IV models superimposed on the same fragment of the cryo-EM open state.....90



## LIST OF EQUATIONS

		Page
Equation 1:	The empirical potential energy function.....	10
Equation 2:	The bonds summation.....	11
Equation 3:	The angles summation .....	11
Equation 4:	The dihedrals summation .....	11
Equation 5:	The improper angles summation .....	12
Equation 6:	The phenomenological potential residue interacts .....	45
Equation 7:	The structure factor.....	55
Equation 8:	The relationship of structure factor and wave vector .....	55



## CHAPTER I: INTRODUCTION

---

Molecular simulations have increasingly become powerful tools in providing insight into the structure, dynamics and functional mechanisms of biological macromolecules. The great advance in high-performance computing technology offers opportunities for explorations of macromolecular complex systems with larger sizes and over longer simulation times. Although atomistic simulations can be achieved for systems containing several millions of atoms with milliseconds scale simulation times, it is still not possible to simulate several large-scale biomolecular complexes with their size requiring that the time go far beyond the limits of all-atom molecular simulations. During the past decades, Coarse-Grained (CG) models have continuously been developed with an increase in reliability and accuracy. Due to its lowered resolution and larger time steps, CG simulations have now emerged as an efficient and inexpensive approach to model large and complex systems. In this thesis, molecular dynamics and Monte Carlo simulations has been employed using CG models to investigate structure, morphology, self-assembly and interactions of a membrane protein ion channel called CorA magnesium channel. The works in this thesis shows the ability of CG simulations to provide detailed information on physical and thermodynamical properties compatible with all-atom simulations. This chapter contains necessary information on significance and literature reviews of the studied protein, theoretical background of simulations and research objectives.

### 1.1 Significance of the studied protein

Magnesium ion ( $\text{Mg}^{2+}$ ) is the most abundant intracellular divalent cation. It is involved in the regulation of cell cycle progression [1] which requires a huge number of molecules and interrelated pathways [2]. It is triggered by growth factors and hormones, as well as signaling from extracellular matrix components, all of which activate the transcription of genes that control cell cycle entry and progression, ultimately leading to cell division.  $\text{Mg}^{2+}$  plays an important role in maintaining tissue homeostasis and eventually driving tissue regeneration and repair. The cell cycle is

tightly regulated by activators, inhibitors, and sensors of DNA and chromosome damage. Moreover, Mg has a role in the formation of the mitotic spindle [3]. Evidence from prokaryotes and lower eukaryotes supports the idea that magnesium regulates the cell cycle. [4].

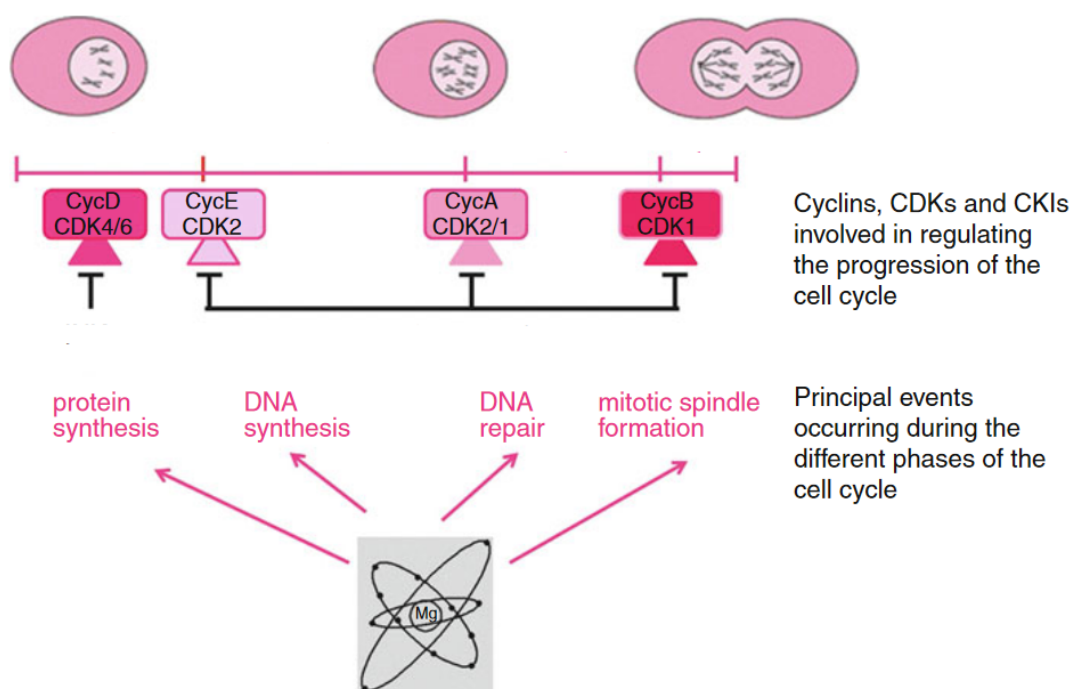


Figure I-1 Magnesium and cell cycle progression [1]

### 1.1.1 Ion Channels

Ion channels are protein that embed and span across a lipid bilayer (or phospholipid bilayer). They are classified into two basic types: leak channel and gated channel. The leak channel allows ions to pass through all the time. On the other hand, the gated channel is selective and controls the exchange of ions between extracellular and intracellular sides.

There are three categories of gated channel via external stimuli (Figure I-1): i) Voltage-gated, ii) Mechanically gated, and iii) Chemically gated. Firstly, the voltage-gated is a forced stimulus by electricity changes in the membrane potential or separation across the membrane such as  $\text{Na}^+$  channel,  $\text{K}^+$  channel by a condition zone that produces action potential. Secondly, the mechanically gated is associated with

mechanical stimulus such as stretch activated channels and sensory neurons of such as stretch reflex and tendon reflex. Lastly, the chemically gated or ligand-gated is activated by molecules or ions binding at a receptor site.

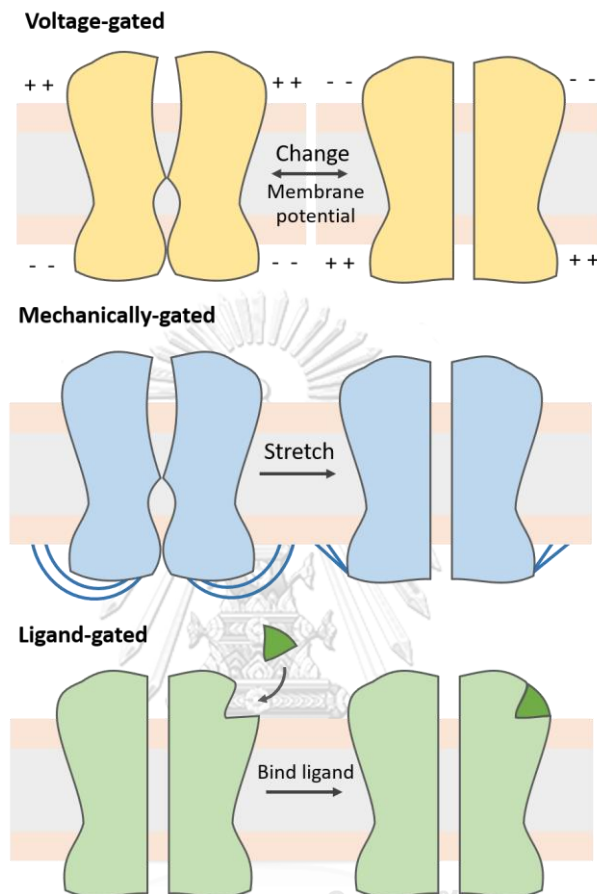


Figure I- 2: Three types of gated-ion channel

Ligand-gated channels react quickly to the binding of ligands and stimulus. There are two types of ligand-gated channels: extracellular and intracellular ligand gated channel. Only specific ligand can bind with specific channels, which can be modelled by "lock and key" or "induced fit", to control the channel in the close or open position. The binding site of a channel is not anywhere near the actual channel (allosteric binding) where the ligand binding location is called the allosteric site and is a place away from the ion channel. Once the ligand binds, it can control the opening and closing of the ion channel by altering the protein conformation of the entire

protein. An open channel can cause the ion permeability of the entire plasma membrane to quickly change.

The allosteric binding site of a ligand, is complementary to the ligand and can be on the intracellular side. Generally, membrane receptors respond to extracellular signals. Membrane receptors can have multiple allosteric binding sites for their ligands. These cases are different from voltage-gated ion channels which rely on the difference in membrane potential, but ligand-gated ion channels respond to the binding of a ligand. The stretch-activated ion channel depends on the deformation of the cell membrane, or the cell membrane being stretching and or pushed/stressed in some way.

### 1.1.2 Magnesium channel

$Mg^{2+}$  entry into cells via channels was first discovered in prokaryotes and protozoans. Several channels or channel-like structures responsible for  $Mg^{2+}$  entry in eukaryotic cells have been identified. Even when significant inward and outward gradients are imposed across the cell membrane, eukaryotic cells maintain  $Mg^{2+}$  content below the predicted electrochemical equilibrium potential of 50 mM [5]. This finding lends credence to the idea that eukaryotic cells have powerful and abundant  $Mg^{2+}$  transport mechanisms that keep total and free  $Mg^{2+}$  levels stable. The primary transport mechanisms involved in cellular  $Mg^{2+}$  homeostasis regulation is summarized in Figure I-3. Several  $Mg^{2+}$  entry mechanisms have been recently identified. They include: Claudins, MagT1, MMgTs, SLC41, ACDP2, NIPA, and Mrs2. These channels, which have a relatively high specificity for  $Mg^{2+}$ , are mostly found in the cell membrane, but they can also be found in the mitochondrial membrane or the Golgi cisternae.

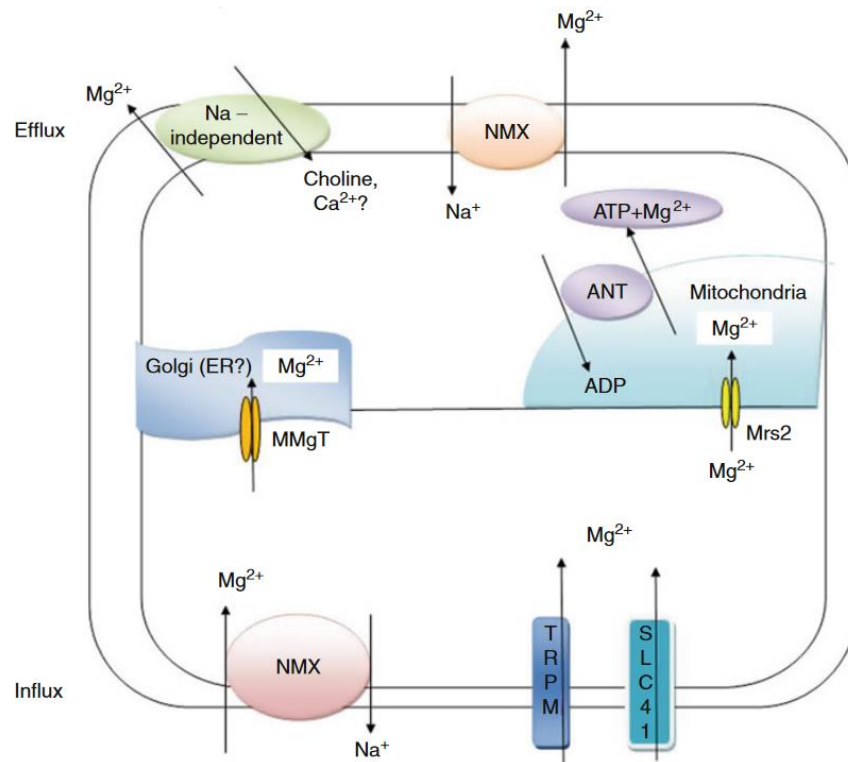


Figure I- 3  $Mg^{2+}$  transport mechanisms: The cartoon summarizes the main transport mechanisms responsible for  $Mg^{2+}$  extrusion and entry in eukaryotic cells. [1]

Mrs2 shares several structural homologies and a two-transmembrane domain topology with the bacterial transporter CorA. Mrs2 mutant yeasts have lower levels of total mitochondrial and matrix free  $Mg^{2+}$ , and they can be rescued by CorA fused to Mrs2's mitochondrial N-terminus leader sequence [6].

CorA is the major magnesium transporter in Salmonella; homologs have been identified in over 30 organisms from eubacteria, archaea, and eukaryotes. These play a crucial role in transporting ions and their disruption has been attributed to diseases such as abnormal neuronal tissue expression [7, 8], and irregular mitochondrial activity [9]. The 3D structures of CorA in both the closed and opened states were successfully solved using the x-ray [10, 11] and cryoEM techniques [12]. CorA is one of the  $Mg^{2+}$  gating channels and contains 351 residues. The CorA structure from *Thermatoga maritima* has a pentameric and funnel shaped structure (Figure I-4). The pentameric structure of CorA from the N-terminal to the C-terminal contains TM2, TM1, stalk ( $\alpha 7$ ),  $\alpha 6$ ,  $\alpha 5$ ,  $\alpha 4$ , and a hammer domain ( $\alpha 1$ -  $\alpha 3$ ). The rest of the N-

terminal part of CorA forms a separate domain outside the funnel. It consists of an  $\alpha\beta\alpha$  sandwich comprising two sets of three  $\alpha$ -helices ( $\alpha_1 \alpha_2 \alpha_3$  and  $\alpha_4 \alpha_5 \alpha_6$ ) on either side of a seven-stranded antiparallel  $\beta$ -sheets ( $\beta_2\uparrow \beta_1\downarrow \beta_3\uparrow \beta_7\uparrow \beta_6\downarrow \beta_5\uparrow \beta_4\downarrow$ ) as shown in in Figure I-4.

Crystal structures of  $Mg^{2+}$ -bound CorA reveals five  $Mg^{2+}$ -binding sites which are located at the interface between subunits in the cytosolic domain [12-14]. Each binding site which can host up to two  $Mg^{2+}$  (known as the M1 and M2 sites), is a divalent cation sensor, serving as a  $Mg^{2+}$ -sensing regulatory [14, 15]. The M1 and M2 metal binding sites are separated by a distance of  $\sim 7\text{\AA}$ .  $Mg^{2+}$  in the M1 site positions, these are in close proximity and form coordinate covalent bonds with the carboxylate group of D89 and D253' (" ' " indicates the adjacent subunit). The second M2 binding site lies close to the side chain of E88, D175 and D253. Electrophysiological studies showed that the channel activity is regulated by intracellular  $Mg^{2+}$  [15]. The protein conformation with fully Mg-occupied sites corresponds to the closed state while the channel transitions from closed to open conformations by the release of  $Mg^{2+}$  from the cytosolic binding sites [12].



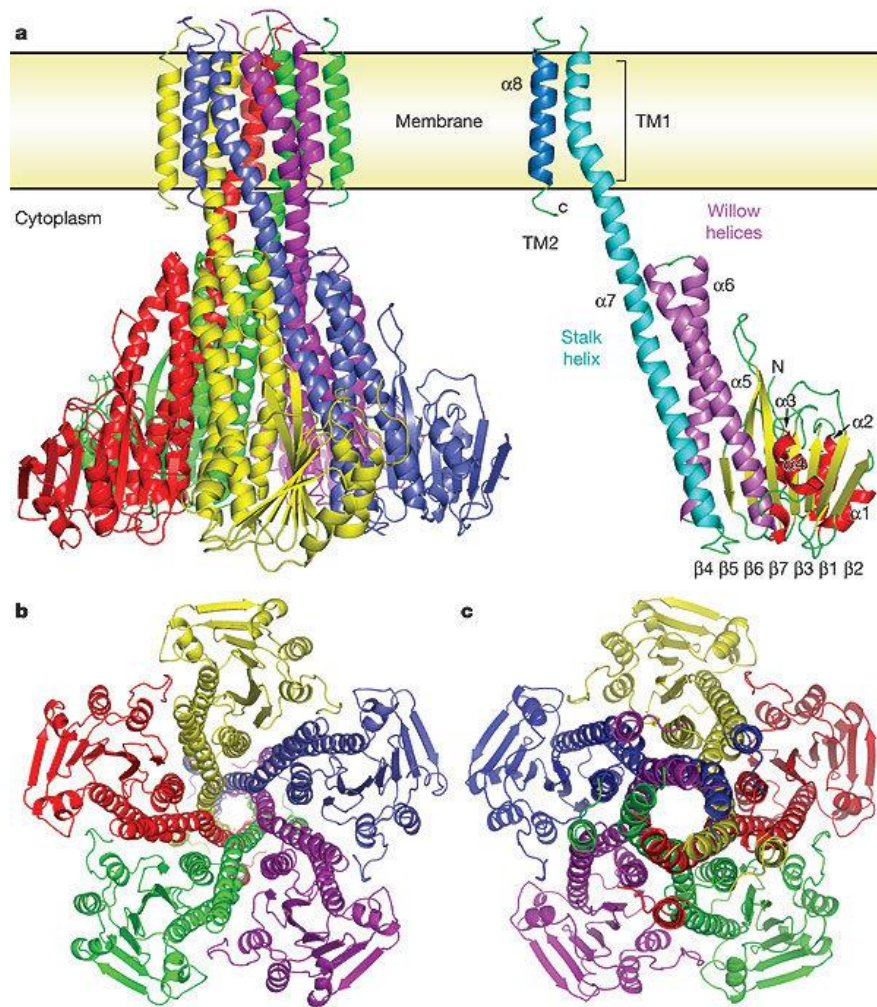


Figure I- 4: CorA protein components [11]

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

## 1.2 Theoretical Background

### 1.2.1 Coarse-grained models for biomolecular systems

Coarse-grained (CG) approaches have raised much interest in the study of large-scale biological processes on time scales inaccessible to all-atom simulations. Experiments can observe dynamics process to understand how life work at > millisecond scale. On the other hand, theoretical and computational works that operate on the atomistic scale for a protein of a few hundred amino acid residues typically operate in a range of nanosecond to microsecond time scale, due to the limits of computational

performance. CG serves as a competent method to address this and is still used at present, for example: A coarse-grained decomposition of surface geostrophic kinetic energy in the global ocean was conducted this year (2021), Modeling of protein-peptide interactions using the CABS-dock web server for a binding site search and flexible docking Methods in 2016, Simulations of protein aggregation: Insights from Atomistic and Coarse-Grained Models in 2014, Membrane-bending mechanism of amphiphysin N-BAR domains in 2009, Elucidating the mechanism behind irreversible deformation of viral capsids in 2009, Transport-related structures and processes of the nuclear pore complex studied through molecular dynamics in 2009, Disassembly of nanodiscs with cholera toxin in 2007, Coarse-grained models for proteins in 2005 [add references] and so on.

Coarse-grained models solve one of the main problems in computational biological science by addressing the time and length-scale gap between computational and experimental methods of biological systems. This model reduces the number of degrees of freedom to achieve a rapid simulation. CG simulations are typically conducted using either of these two main approaches:

- i) Residues-based Coarse-Graining [21, 22]: a protein-lipid CG model is cluster ~10 atoms including hydrogens and replaced by a CG bead. Every four waters also become one water bead. Each bonded bead is connected by harmonic springs. The harmonic angular potentials help to maintain the shape of molecular structure. The interactions are represented with the Lennard-Jones potential, and the long-range interaction with the Coulomb potential.
- ii) Shape-based Coarse-Graining [23, 24]: an approach to model large macromolecular assemblies that represent protein and other biomolecules as few point-like particles. A single protein comprised of compact domains, disordered, and elongated. This method allows compact domains and tails to have equal accuracy, with the advantage of efficient topology. The interactions also use the Lennard-Jones potential but more interaction details are possible depending on the program or the user assigned parameters to the model.

### 1.2.2 Martini Models

Martini coarse-grained models are one of the most widely used forcefield for CG-MD of biomolecular systems [25, 26]. In the Martini model, each amino acid consists of backbone beads and 1-3 sidechain beads as shown in Figure I-5. There are four types of sidechain beads: polar, intermediately polar, apolar, and charged which have subtypes based on hydrogen bonding capabilities or their polarity.

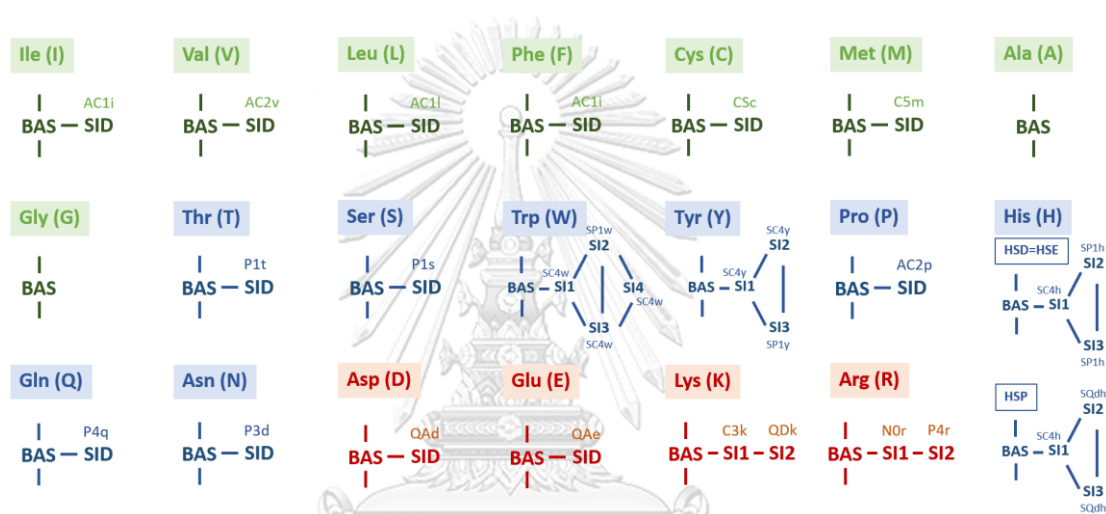


Figure I- 5: Topologies of amino acid in Martini model

Martini force fields have successfully been applied in the field of biological and materials research studies [26, 27]. The force field is still under active development to improve the model and bring it closer to the accuracy of all atom simulation for a wide variety of systems [28].

### 1.2.3 Molecular Mechanics Potential Energy Functions

The molecular mechanics force fields are used to calculate the potential energy of simulated systems. Atoms in molecular mechanics are treated as balls held together by springs. The empirical potential energy function can be written as follows:

*Equation 1: The empirical potential energy function*

$$V = V_{bonded} + V_{non-bonded}$$

Where  $V$  is a summation of bonded and non-bonded terms which can be expressed as:

$$V = (V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V_{\text{improper}})_{\text{bonded}} + (V_{\text{elec}} + V_{\text{vdW}})_{\text{non-bonded}}$$

The bonded energy term can be classified into four categories. The non-bonded energy term is derived from van der Waals and electrostatic interactions.  $V_{\text{bond}}$  represents the bond stretching energy whose the function is based on Hooke's law. The potential function of  $V_{\text{bond}}$  is given by

*Equation 2: The bonds summation*

$$V_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2$$

Where  $k_b$  is the force constant of the bond.  $r$  and  $r_0$  are bond length between two atoms at its current state and equilibrium state, respectively.

$V_{\text{angle}}$  represents the bond angle energy. The potential function of  $V_{\text{angle}}$  is expressed as:

*Equation 3: The angles summation*

$$V_{\text{angle}} = \sum_{\text{angle}} \frac{1}{2} k_\theta (\theta - \theta_0)^2$$

Where  $k_\theta$  is the force constant of the bond angle.  $\theta$  and  $\theta_0$  are bond angles at the current and equilibrium state, respectively.

$V_{\text{dihedral}}$  is a dihedral or torsion angle energy. It is the energy required to rotate the bonds. The torsional energy can be defined by Eq.4, where  $V_n$  is the rational barrier

height,  $n$  is periodicity of rotation,  $\tau$  is the torsion angle in the conformation of the molecule in that phase ( $\phi$ )

*Equation 4: The dihedrals summation*

$$V_{dihedral} = \sum_{dihedral} \frac{1}{2} V_n (1 + \cos[n\tau - \phi])$$

$V_{improper}$  is the out-of-plane potential energy associated with an improper torsion angle. It is defined as a small out-of-plane angle ( $\alpha$ ). This term measures angle from three atoms that form a plane and the angle of the bond from the top atom out of plane. The out-of-plane motional energy is often necessary of planar group, for example  $sp^2$  hybridized. The  $\psi$  is defined as an angle between the plane containing the first three atoms and the plane containing the last three atoms as represented by

*Equation 5: The improper angles summation*

$$V_{improper} = \sum_{improper} \frac{1}{2} V_n (1 + \cos[n\alpha - \psi])$$

The Lennard-Jones (LJ) function is used to describe van der waals attraction and repulsion between atoms. Electrostatic term is based on the Coulomb's law [29, 30].

#### 1.2.4 Molecular dynamics and Monte Carlo simulations

MD simulation is a deterministic method that can explain the "motion" of individual molecules in models and depends on the position, velocities, and orientation changes with time. The calculation is provided by solving differential equations, which rely on Newton's law.

In 2002, Martin Karplus and J. Andrew McCammon mentioned the roles of MD simulation in biomolecules over the last 25 years [31]. Concluding that it was widely used with many programs such as NAMD, CHARMM, AMBER, GROMACS and so on. The simulation methodology is continuously developed in parallel with computing performance resulting in an increase in the number of publications. Many applications from this method are possible allowing for a wide range of research topics including the monitoring of conformation changing of protein structural, study of the behavior

of gating channel, the response of temperature on biomolecules that cannot be observed directly by other means, interaction energy between molecules and/or protein. This technique can investigate implicit or explicit systems at femtosecond to microsecond scales depending on the issue of interest.

Monte Carlo (MC) simulation is a stochastic method, based on statistical physics, that performs a simulation where the number of molecules ( $N$ ), volume ( $V$ ), and temperature ( $T$ ) are fixed. The calculation is evaluated to an energy that is generated from the molecular configuration of the previous step and a random event called the "*Markov chain*" which is based on random walk theory.

### 1.3 Objectives

- I. To investigate structure and dynamics of phospholipids in nanodisc and bilayer forms through all-atom and coarse-grained molecular dynamics simulations
- II. To investigate the impact of lipid nanodiscs on the stability of the closed and open CorA conformations
- III. To investigate self-assembly morphology of CorA using coarse-grained simulations

### 1.4 Relation research topics

This dissertation is organized into six chapters. Chapter 1 provides an overview introduction related to the research work. Four research topics carried out during the study are presented in Chapter 2, 3, 4 and 5. Chapter 2 reports a CG-MD study to explore the structural response of lipoprotein nanodisc systems as a function of temperature. Nanodiscs are a promising tool and were used to encapsulate the membrane protein. They are discoidal protein-lipid particles that contain a phospholipid bilayer and wrap around a polypeptide. Nanodiscs have been used with several techniques such as electron microscopy, NMR, EPR, X-ray crystallography etc. [16-20]. A CG simulation of nanodisc structures was conducted to investigate the

effect of temperature since the 3D low-resolution structures of CorA in nanodiscs were determined using cryo-electron microscopy. This study aimed to understand the structure and dynamic responses in various temperatures. Chapter 3 and 4 presents CG-MC approach to study the structural behaviors, interactions and morphology of CorA magnesium channels. In this study, the CG models were conducted based on an amino acid sequence of CorA from *Thermotoga maritima* (UniProtKB - Q9WZ31). An in-house CG-MC program developed by Dr. Pandey, a co-advised professor from the University of Southern Mississippi. The findings in the study demonstrate the self-organization of the CorA protein in response to temperature and protein density, and the protein interactions in the solute matrix. There has been a debate whether the loss of  $Mg^{2+}$  in the binding sites causes significant conformational changes from the closed-bound to open-unbound states. Leaving some aspects of the protein mechanism of CorA still unclear and requiring further examination. Chapter 5 illustrates a CG-MD study to address the above issue. The results demonstrate the conformational response of the CorA closed state upon the release of  $Mg^{2+}$ . The conclusion of the thesis work is summarized in the final Chapter.

### **1.5 Expected beneficial outcome**

Obtain knowledge about the structure and mechanism of the protein function that could open opportunities for developing new tools or technology in bioscience.

CHULALONGKORN UNIVERSITY

## CHAPTER II

---

### **Thermally induced structural organization of nanodiscs by coarse-grained molecular dynamics simulations**

*Warin Rangubpit<sup>†</sup>, Pasawan Paritanon<sup>†</sup>, Ras B. Pandey<sup>‡</sup>, and Pornthep Sompornpisut<sup>\*,†</sup>*

*<sup>†</sup>Center of Excellence in Computational Chemistry, Department of Chemistry,  
Chulalongkorn University, Bangkok 10330, Thailand*

*<sup>‡</sup>School of Mathematics and Natural Sciences, University of Southern Mississippi,  
Hattiesburg, MS, 39406, USA*

Doi: 10.1016/j.bpc.2020.106464

Journal: Biophysical Chemistry, 267, 106464.

Received: 04 July 2020

Accepted: 24 August 2020

Published: 29 August 2020

---

**Keywords:** Self-organizing structure of proteins (CorA); protein folding; Coarse-grained model; Monte Carlo simulation; effective dimension of self-assembly



## 2.1 Abstract

Membrane scaffold proteins (MSP) nanodiscs have been extensively used in structural study of membrane proteins. In cryo-EM, an incorporation of target proteins into nanodiscs is conducted under a rapid change from cryogenic to ambient temperatures. We present a coarse-grained molecular dynamics (CGMD) study for investigating an effect of temperature on the structural organization of DPPC-nanodisc and POPC-nanodisc. A non-monotonic response of physical quantities (i.e. the lipid order parameter, nanodisc flatness, structural change, solvation property, radius of gyration) with increase in temperature ( $T=200-350$  K) is found to be associated with the gel-ripple-liquid crystalline phase change within nanodiscs. The reorganization of lipids upon temperature variation induced conformational changes of MSP to minimize hydrophobic mismatch. Structural response to temperature is different to a certain extent between the saturated DPPC and unsaturated POPC.

## 2.2 Introduction

Studying structure and function of membrane proteins presents more technical limitations than water soluble proteins because they are naturally embedded in the phospholipid bilayer with their hydrophobic transmembrane domains [2]. Therefore, it is difficult to obtain the correct structural data in its native environment through standard techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy. A typical solution is to solubilize membrane proteins in detergent micelles that can extract the membrane protein from its native lipid environment. Depending on the concentration, detergents can partition into the lipid bilayer, break interaction between membrane proteins and soluble factors, and solubilize membrane proteins by making a protein-detergent complex [34, 35]. However, some detergents are harsh and make the proteins unstable and/or distort its structure to a non-native state. Alternatively, liposome has been used to incorporate membrane proteins, which is beneficial to membrane protein assays. However, the liposome is too large, unstable, and difficult to prepare with exactly controlled size and stoichiometry [36].

In the past, liposome and bicelles were used in structural studies of membrane protein mainly for solid-state NMR [37] but they are destabilize the reconstituted membrane protein. To improve the resolution of membrane protein structure and increase the successfulness structure determination. The helper tool still searched for development. Lipoprotein nanodiscs have become a promising platform for the study of structure and function of membrane proteins in a native-state like condition. They have been successfully used in structure determination of over a dozen of membrane proteins using cryogenic electron microscopy (cryo-EM), solid-state and solution NMR [16, 18, 19, 38, 39]. The introduction of nanodiscs into single-particle cryo-EM has revolutionized in the field of structural biology for determination of membrane protein structure without crystallization. Nanodiscs are an engineered lipoprotein that is closely similar to high-density lipoprotein (also known as HDL) in structure. They consist of patches of phospholipids surrounded by two amphipathic membrane scaffold proteins (MSP), which can self-assemble the target membrane protein for structural investigation [40]. Although, the size and interfering limitations properties were try to fix it out and polymer based nanodisc emerged as a new choice [41]. By the way, the MSP based nanodisc was the originating successfully study membrane protein [19, 42]. This offers a greater advantage than detergents and liposomes in terms of size, solubility and stability at a single molecule level [43]. Nanodisc structure is a disc-like phospholipid arranged in a bilayer manner, encircled by two MSPs. The MSP in nanodiscs forms a discoidal double belt that wraps around the rim of the bilayer. There are several types of MSPs. Apolipoprotein A1 (Apo-A1) is a common mimetic peptides that engineer to several kinds of MSP to achieved the control size and flexible used as polymer-based nanodiscs. For the membrane components, dimyristoylphosphatidylcholine (DMPC), dipalmitoylphosphatidylcholine (DPPC) and 1-palmitoyl-2-oleoyl phosphatidylcholine (POPC) are the most common lipid used in nanodiscs [44].

Recently, the nanodisc formation was studied in DMPC and DLPC lipid using experimental data and coarse-grained molecular dynamics simulation [45]. To obtain the spatial arrangement of polymers in nanodiscs, the size, and thermal stability. In our study, the one of an important physical property of a lipid system is membrane

fluidity in response to temperature. The thermotropic phase transformation of phospholipid bilayers has been investigated by experimental and theoretical approaches [39, 46-49]. It has been shown that DPPC and POPC bilayers are rapidly converted from fluid phase to gel phase at  $\sim 295$  K [46] and 270 K [47], respectively. The gelation phase depended on a tilt angle with respect to bilayer normal [48]. Kariyazono showed that MSP1 of POPC nanodisc became an elliptical on increasing the temperature [50]. The lipid bilayer of nanodisc exhibits lateral thermal expansion on a rising the temperature [39]. For a typical cryo-EM experiment, nanodiscs containing the incorporated membrane protein are prepared under freeze-thawing temperature, ranging from cryogenic to room temperatures [51]. However, little is known about the effect of temperature variation on nanodisc structure. A detailed understanding the structure and dynamics of nanodisc particles in response to temperature will enhance imaging interpretation of the morphology of nanodiscs. In this study, we performed CG-MD simulations of two lipoprotein nanodiscs to investigate changes in conformation, flexibility, compactness and hydration number of the nanodiscs as a function of temperature. An aim of the study is to demonstrate an impact of temperature on the structure and dynamics of nanodiscs.

## 2.3 Methodology

### 2.3.1 Molecular modeling of nanodisc

A nanodisc consists of a bilayer lipid membrane encapsulated by lipo-proteins in a circular morphology (see Figure II-S1). The starting structures of lipoprotein nanodisc [52] with atomic scale resolution were designed from the Nanodisc Builder module in CHARMM-GUI [53, 54]. We consider two commercially available nanodiscs with membrane bilayer consisting of DPPC and POPC. The scaffold protein-1 (MSP1) encircles membranes in a diameter of  $98\text{\AA}$  involving 160 DPPC and 135 POPC lipids in 1:1 ratio of the upper- and lower-leaflets (see Figure II-S1). The number of lipids were based on the estimate of the area of the nanodisc, i.e.  $\pi \left(\frac{d}{2} - 5.5\right)^2$ , where  $d$  is the nanodisc diameter [54]. The encapsulating protein MSP1 consists of polypeptide chains, P1 and P2 in a double helical morphology.

The atomistic details of DPPC and POPC nanodiscs were converted to MARTINI coarse-grained (CG) models (Figure II-S1) using the residue-based CG script executed by the VMD program [55]. The MARTINI 2.2 force field topology was used for the CG models. The CG nanodiscs were solvated by MARTINI water beads. Na<sup>+</sup> and Cl<sup>-</sup> beads were subsequently added for charge neutralization of the systems. This simulation consists of 60 000 beads in a 20×20×19 nm<sup>3</sup> simulation box.

### 2.3.2 Coarse-grained molecular dynamics simulation

The CG-MD simulations were carried out at a temperature range 200 K - 350 K for 3 μs using the NAMD [56] involving MARTINI force field with a periodic boundary condition in three dimensions. The switching function is adapted for Lennard-Jones potential from 0.9 nm to 1.2 nm and Coulomb potentials were smoothly shifted to zero from 0 to 1.2 nm. The relative dielectric screening constant is set to 15.0. Initially, the MD systems were subjected to 6000 step energy minimizations prior to equilibration steps. The equilibration was performed with the following steps: (i) for the next 10<sup>5</sup> steps (0.2 ns), P1 and P2 proteins were constrained using 1.0, 0.5, 0.25 kcal/mol/Å<sup>2</sup> harmonic constraint energy function, respectively, in NVT-ensemble. (ii) Then, proteins were constrained using 0.125 and 0.0625 kcal/mol/Å<sup>2</sup> harmonic constraint energy function, respectively, in NVT-ensemble. (iii) Subsequently, these constraint were removed with NPT-ensemble using Langevin piston Nose-Hoover method [57, 58]. About 2-3 μs CG-MD simulations were performed with an integration time step of 20 fs.

### 2.3.3 Trajectory analysis

The trajectories of lipid membrane consisting of DPPC, POPC and the double helix proteins MSP1 were analyzed using TCL scripts in VMD. Unless otherwise specified, MD trajectories of each CG systems in the last 0.5-1.0 μs were used to extract structure and dynamics properties of the systems. The root-mean-square deviation (RMSD) and radius of gyration (R<sub>g</sub>) of MSP1 were computed as a function of time to illustrate the conformational fluctuation during the course of simulation. The root-

mean-square fluctuation (RMSF) of each residue of MSP1 was computed using the last 0.5  $\mu\text{s}$  of its MD trajectory. The orientation of lipid tails relative to the bilayer normal is described by an order parameter ( $S$ ), which was calculated by defining the orientation of the CG bond vectors with respect to the bilayer normal.

$$\langle S \rangle = \frac{1}{N} \sum_{i=1}^N \frac{1}{2} \langle 3\cos^2\theta - 1 \rangle$$

where  $\theta$  is the angle of a bond between two CG beads and the bilayer normal,  $N$  is the total number of time frames used for the calculations. The angular bracket indicates that  $S$  is averaged over all of the sampling time.  $S = 1$  corresponds to perfect alignment. In a random orientation,  $S$  takes zero value. The smaller value of an order parameter, the greater the degrees of motional freedom experienced by the lipid molecule.

The radial distribution function ( $g(r)$ ) of water around nanodisc was analyzed to investigate the hydration structure. The first solvation shell of the  $g(r)$  was used as a cut off for estimating the number of water molecules around MSP1 as well as the bilayer of the membrane. The average value of the hydration number and the solvent accessible surface area of the nanodisc was computed. The number of lipids per unit area was obtained by the surface area of the nanodiscs divided by number of lipids. Principal component analysis (PCA) was performed using about 4000 snapshots extracted from the last 3  $\mu\text{s}$  of MD trajectories at 200 K, 250 K and 300 K. To make a relevant analysis of the internal motions of MSP1, backbone beads of the snapshot structures were aligned against their starting structure. The first two principal components of motion, eigenvalue 1 and 2, which corresponded to the first and second eigenvectors of the covariance matrix were calculated using the WORDOM software [59] Visualization and all trajectory analyses were performed using TCL scripts executed by the molecular visualization program VMD [55].

## 2.4 Results and discussion

A number of physical quantities such as the order parameter of lipid tails, the flatness parameter of nanodiscs, RMSD of MSP1 and that of its residues, solvation profile,

solvent accessible surface area (SASA), radius of gyration of MSP1, etc. are studied as a function of temperature in both nanodiscs (DPPC, POPC). Analysis of the trends from the data on these physical quantities is presented in the following.

#### 2.4.1 Nanodisc structural organization in response to temperature

Each constitutive elements of the protein (MSP1), lipids, and water molecule execute their movements during the course of molecular dynamics simulation at each temperature dictated by the instantaneous force and the thermal agitation. The collective motion of these constituents may lead to global structural evolution of the nanodisc assembly in asymptotic time as system reaches its steady-state equilibrium. Snapshots of both nanodiscs (DPPC, POPC) assemblies towards the end of simulation (3  $\mu$ s) are presented in Figure II-1A for a range of temperature (T=200K-350K). Visual inspections reveal a remarkable change in conformation of the protein MSP1 encapsulating DPPC and POPC membrane on raising the temperature. The simulations at 200K, 300K and 350K show the regular MSP1 belt of the nanodisc oblate spheroid. However, other temperatures show much different structures in which the MSP1s have a periodic-like belt around the membrane bilayer, compatible with a specific type of non-raft lipids in the ripple phase theory (Figure II-1A). This finding suggests that the effect of conformational changes in the MSP1 might be associated with a change in the physical state of the bilayer membrane at least in part. At 200K, both nanodiscs are likely in a more gel phase structure where the lipid tails are closely packed in their extended chain. As temperature rises between 225K-275K, the nanodiscs show significant structural distortion in the discoidal bilayer, of which the shape adopts a corrugated-like structure consistent with the morphological characteristics of the metastable ripple state. However, at temperature T = 300K and 350K, DPPC-nanodisc maintains its discoidal structure similar to the simulations under the gel-to-liquid transition temperature. It should be noted that the crystalline order of the bilayer is lost at these temperatures. The acyl chains of the DPPC-nanodisc simulations became more randomly dispersed, indicating that the lipids are in the fluid phase. An analysis of the averaged order parameter,  $\langle S \rangle$ , as a function of temperature supports the change in the physical state of both nanodiscs (Figure II-

1B). The order parameter for the three terminal bonds of the CG lipid model was used to illustrate the degree of hydrophobic tail tilting away from the bilayer normal. As shown in Figure II-1B, the acyl chains in the gel phase ( $T=200\text{K}$ ) have  $\langle S \rangle$  relatively higher than in the fluid phase ( $T=300\text{K}$ ,  $350\text{K}$ ). This implies that the hydrophobic tails are on average essentially more extended than in the lipid-disordered state. For the gel-to-liquid crystalline transition temperature ( $T=225\text{K}-275\text{K}$ ), the wavy sinusoidal shape of the nanodisc is the results of a mixture of ordered and disordered lipids. This can be seen by a decreasing tendency of  $\langle S \rangle$  with increasing temperature, except for  $T$  between  $225\text{K}$  and  $250\text{K}$ .

The structural organization of POPC-nanodisc in response to temperature is found to be similar to DPPC-nanodisc. Nevertheless, the simulations of POPC-nanodisc at  $300\text{K}$  show distorted discoidal bilayers, which is similar to the ripple-like structure. This might suggest that the phase transition of POPC-nandisc to a fluid phase occurs at a temperature higher than DPPC-nanodisc. This observation seems to contradict the nature of saturated/unsaturated phosphocholines and the melting temperature ( $T_m$ ) between the two lipids reported from earlier experiments [60]. It is well-known that the saturated fatty acids (DPPC) have a gel phase  $T_m$  higher than the unsaturated fatty acids (POPC) because the saturated DPPC bilayer is more tightly packed than the unsaturated POPC. As can be seen in Figure II-1B, the order parameter of the three terminal bonds of POPC-nanodisc takes on the lower value in comparison with DPPC-nanodisc. This indicates that lipid molecules experience a greater freedom of motion since the unsaturated double bond causes a bend in the lipid tails. Thus, POPC containing nanodisc lost their orientation easier than DPPC-nanodisc. This explains an increase in the lipid tilt angle as the lipids become more disordered. In addition, it has experimentally demonstrated that the POPC-nanodisc has a molecular surface area larger than the DPPC-nanodisc [18, 61]. Therefore, POPC is more loosely packed than DPPC in nanodiscs. Because of less well-packed lipids, the discoidal distortion of the POPC-nanodisc at  $300\text{K}$  unlikely represents the ripple phase structure. Instead the MSP1 forms less stable nanodiscs with POPC.

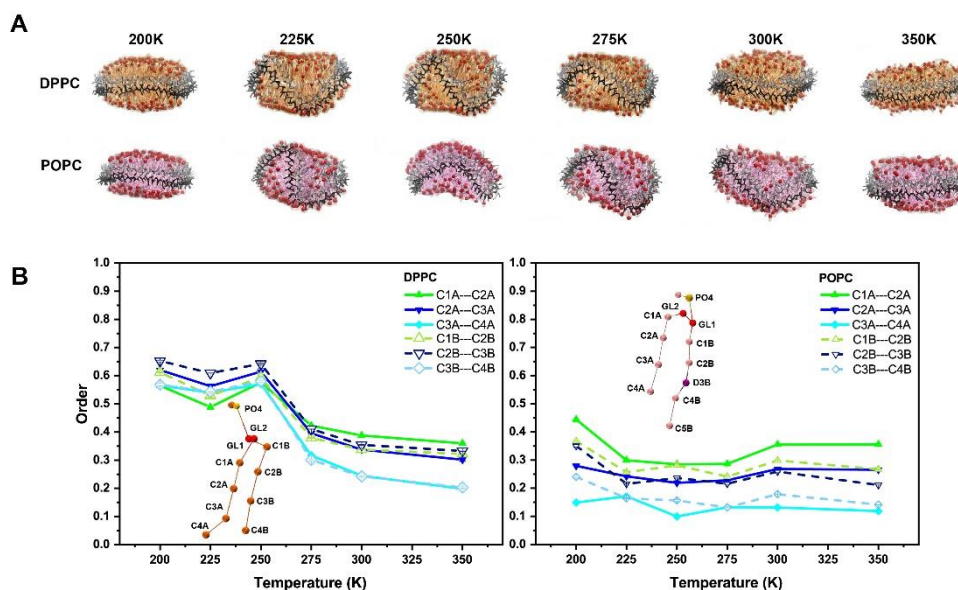


Figure II- 1: (A) Nanodisc sideviews taken from MD snapshots to show a variation of the DPPC- and POPC-nanodisc shapes at temperature  $T = 200\text{K} - 350\text{K}$ .

Differences in the MSP1 shape between the discoidal arrangement in the gel and liquid crystalline states, and the corrugation-like structure in the ripple state were observed for both nanodiscs. (B) Average order parameter,  $\langle S \rangle$ , of the three terminal bonds of lipid acyl chains as a function of temperature. (A) DPPC-nanodisc and (B) POPC-nanodisc. Two different chains of fatty acids were depicted as solid and dashed lines corresponding to A- and B-chain, respectively.

An inset illustrates the bead names and the molecular frames used in the S calculations

We have pursued a detailed analysis of the reorganization of nanodisc structure upon temperature variation. Figure II-2 showed a temperature-dependent flatness in terms of the dihedral angle ( $\phi$ ) of the MSP1 plane. We found that, at 200K, DPPC- and POPC- nanodiscs have the minimum deviation from the plane with  $\langle \phi \rangle$  of  $10^\circ$  and  $5^\circ$ , respectively, indicating that the MSP1 has a flat belt form. Distortions from the standard MSP1 belt are observed as temperature rises. The flatness angle showed a sharp increase at 225K, and then approached a steady state at 250K before declining as the temperature increases. Such structural changes of MSP1 were induced by the structural phase transition of the lipids as a result of protein-lipid interactions to relieve the hydrophobic mismatch. It should be noted that POPC-nanodisc has a



deviation of the flatness parameter larger than DPPC-nanodisc. This has been explained by the relatively loose packing of the unsaturated hydrocarbon chain of POPC. These results are nicely fit to the observations of lipid phase transition described previously.

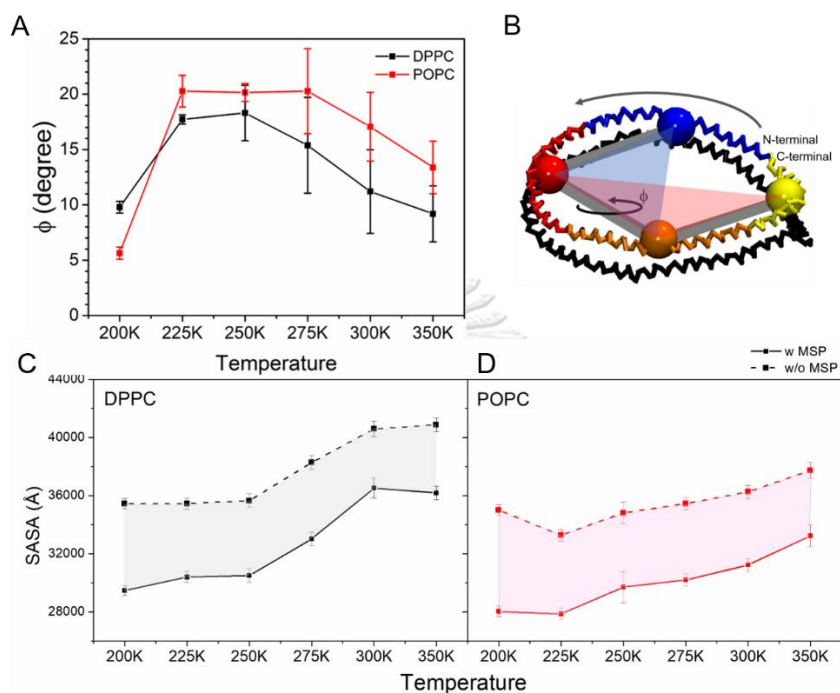


Figure II- 2: (A) The flatness angle of nanodiscs as a function of temperature. Error bars represent the standard deviations for 0.5 $\mu$ s MD trajectories. (B) Schematic definition of the dihedral angle ( $\phi$ ) for measuring the nanodisc flatness. The  $\phi$  between two triangular planes is defined by four centerpoints located on the backbone arcs of the MSP1s. Each MSP1 chain consisting of residue number 55-243 was divided into four parts, each of which is used to determine the center of mass based on the CG backbone beads. An angle of 0° corresponds to a completely planar MSP1. The level of distortion from the planarity depends on the amount of deviation from  $\phi = 0^\circ$ . The SASA of the lipid tail at the rim of the bilayer disks for (C) DPPC and (D) POPC. The solid and dashed lines correspond to the surface area computed in the presence and absence of MSP1, respectively.

### 2.4.2 Structural changes of MSP1

The structural response to temperature variation was assessed by analyzing the root-mean-square deviation (RMSD) of MSP1 with respect to its starting structure from the snapshot of the simulation data taken every 0.2 ns. The variation of relative backbone RMSD of MSP1 with the time ( $t= 0.0$  to  $3.0 \mu\text{s}$ ) at a range of temperature ( $T=200\text{K}-350\text{K}$ ) presented in Figure II-S2 show that it has reached a steady-state equilibrium at almost the entire temperature range. The average RMSDs were evaluated from the last  $0.5 \mu\text{s}$  of MD trajectories at each temperature. Variation of the average RMSDs of the encapsulating protein MSP1 with the temperature is presented in Figure II-3A. We see that the average RMSDs of MSP1 encapsulating both nanodiscs DPPC and POPC increases on raising the temperature from 200 K to 250 K followed by a decay on increasing the temperature further (250 K to 350 K). The thermal response of the average RMSD of MSP1 encapsulating both nanodiscs are non-monotonic. The RMSD of MSP1 encapsulating the unsaturated POPC membrane is higher than that of the saturated DPPC membrane due to the loose lipid packing. Figure II-3B illustrates the interaction energy of MSP1 with the peripheral lipids averaged over the trajectory of each simulation. As can be seen in figure, the interaction energy decreases as temperature increases. This suggests that protein-lipid interactions are significantly weaker when the lipids are in the gel phase. Furthermore, the interaction energy in the POPC-nanodisc system ranges between 80-90% compared to that in the DPPC-nanodisc across all temperature ranges. The analysis indicates that interactions between the protein belt and the lipid bilayer are stronger in the DPPC-nanodisc than the POPC-nanodisc. The CG-MD results supports the idea that DPPC nanodisc tends to resist its structural changes more than the POPC nanodisc at almost all temperature.

Note that, size and structure of the CG models for DPPC and POPC lipids are not very different; POPC lipid has an additional CG bead (Figure II-S1) with a double bond character in one of the tails than that of the DPPC. It is interesting to see that only a small change in the constitutive element of the membrane (DPPC versus POPC) leads to a global response of the corresponding nanodiscs. Thus by making a small change in lipid, one can modulate the structural response of the underlying

nanodisc. Results from the coarse-grained and atomic scale simulation of POPC nanodiscs were in good agreement with the area per lipid from annealing and fix temperature simulation observations [47, 62]. With rising temperature, the overall expansion of DPPC nanodisc size from the simulations still maintains the same trend as the experimental observations from the small-angle x-ray scattering measurements [39].

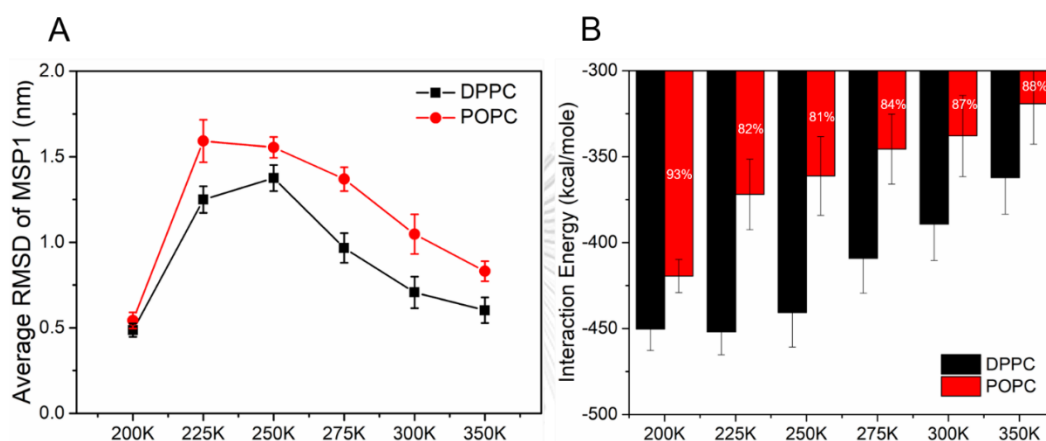


Figure II- 3: (A) The average RMSD of MSP1 in DPPC and POPC nanodiscs versus the temperature. (B) The CG-based interaction energy of MSP1 with the peripheral lipids. For comparison, the values in the red bar graph are percentages relative to the values (100%) of the black bar graph at the same temperature.

To verify such an unusual structural response of the RMSDs of the MSP1 with the temperature in CG-MD simulation, it would be desirably to carry out an all-atom MD simulation at a temperature such as 250 K where maximum change in conformations seem to appear. One of the major problems in including atomistic details in an MD simulation is the high cost of computing. It is difficult to carry out all-atom MD simulations for such a long-time scale as that with the coarse-grained MD. Therefore, an all-atomistic MD simulation is performed on the DPPC and POPC nanodiscs at 250 K for only 0.1  $\mu$ s to identify the trend as shown in Figure II-S3. It is encouraging to note that the evolutionary trends in data from all-atom MD simulation of RMSD of the protein MSP1 encapsulating DPPC and POPC membranes are consistent with that from coarse-grained MD simulations (see Figure II-S3).

It is worth re-iterating that the structural response of MSP1 is non-monotonic as a function of temperature with a maximum at a characteristic temperature ( $T = 250$  K) in both nanodiscs. The encapsulating protein MSP1 consists of two polypeptide chains P1 and P2 in a helical conformation. It may be interesting to examine how the local structures evolve in both chains P1 and P2 at different temperatures. We analyze the structural profile of these chains by examining the RMSDs of each residue (node) as a function of time at each temperature. Figure II-4 shows the RMSD contour profiles of chains P1 and P2 at representative gel, ripple and liquid crystalline temperatures ( $T=200$  K, 250 K, and 300 K) in both DPPC and POPC nanodiscs. The evolution of the RMSD profiles of both chains in both nanodiscs is relatively small and similar at the low temperature ( $T=200$  K, where the structure is almost frozen). Most distortion in RMSD profile of P1 and P2 appears during the phase transition (at 250 K) due to lipid-induced protein conformational changes. In DPPC nanodisc, changes in profile vary along the contour including the N- and C-termini and the mid-range of the P1 and P2 chains (Figure II-4); the evolutions of RMSD profiles of P1 and P2 are however similar. Regions with high RMSD span almost along entire contour of P1 and P2 in POPC nanodisc while retaining the similarity in its patterns. The profile of structural response of P1 and P2 in DPPC nanodisc differs considerably from that in POPC nanodisc. The RMSD profiles of P1 and P2 at 300 K are similar with relatively low structural distortions in DPPC nanodisc. The distortion profiles of P1 and P2 at 300 K in POPC nanodisc are also similar but exhibit higher distortions from that in DPPC nanodisc. Thus, the unsaturated lipid containing nanodisc has a greater temperature effect on the structure of MSP1.

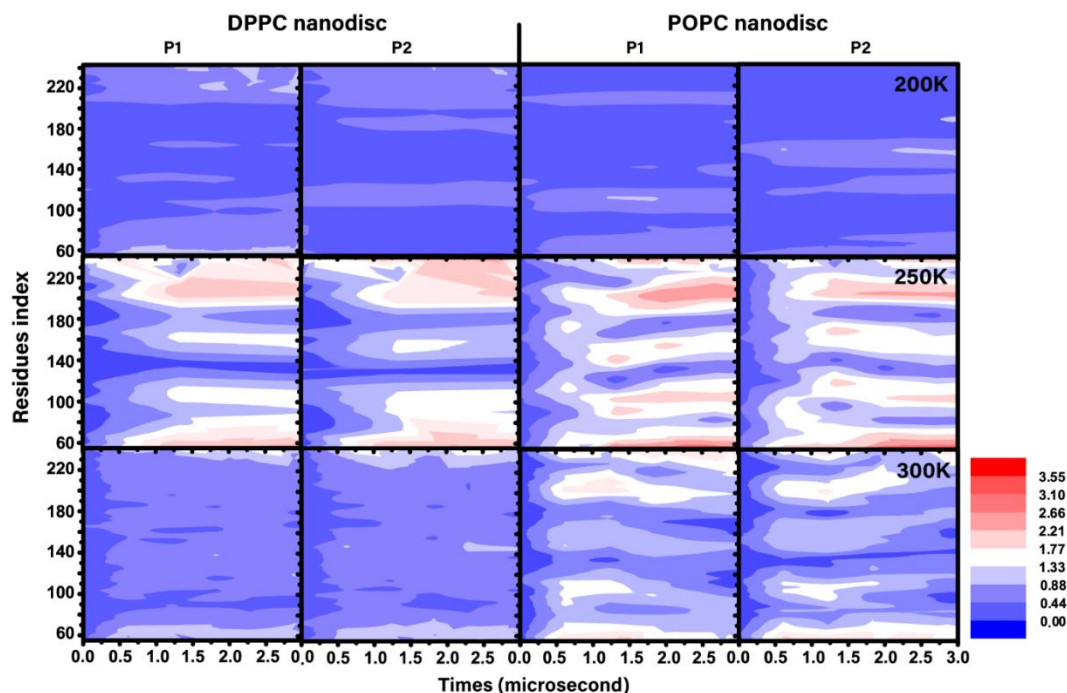


Figure II- 4: The per-residue RMSD of MSP1 versus simulation time. A color ranges from blue, white and red corresponds to the low-to-high RMSD value.

RMSF of the protein chains P1 and P2 of the encapsulating protein MSP1 with respect to mean structure (rather than the initial structure as discussed above) is evaluated to gain insight into their local structural fluctuations. RMSFs of protein chains P1 and P2 encapsulating DPPC and POPC membranes are presented in Figure II-5 at temperatures 200 K, 250 K and 300 K. We see that (i) the RMSFs of all residues are small and similar at the low temperature ( $T=200$  K) but vary along the contour at higher temperatures ( $T=250$  K, 300 K) with increasing magnitude. Onset of an oscillatory pattern seems to appear in RMSF profiles of P1 and P2 particularly in DPPC nanodisc. (ii) Variation patterns of RMSFs of each residue in both chains P1 and P2 are similar in each nanodiscs. (iii) RMSF of each residue increases with the temperature in both chains P1 and P2. The rate of increase in DPPC nanodisc appears to be linear while that in POPC nanodisc is non-linear (i.e. a relatively small increase at low temperatures is followed by large increase at higher temperatures).

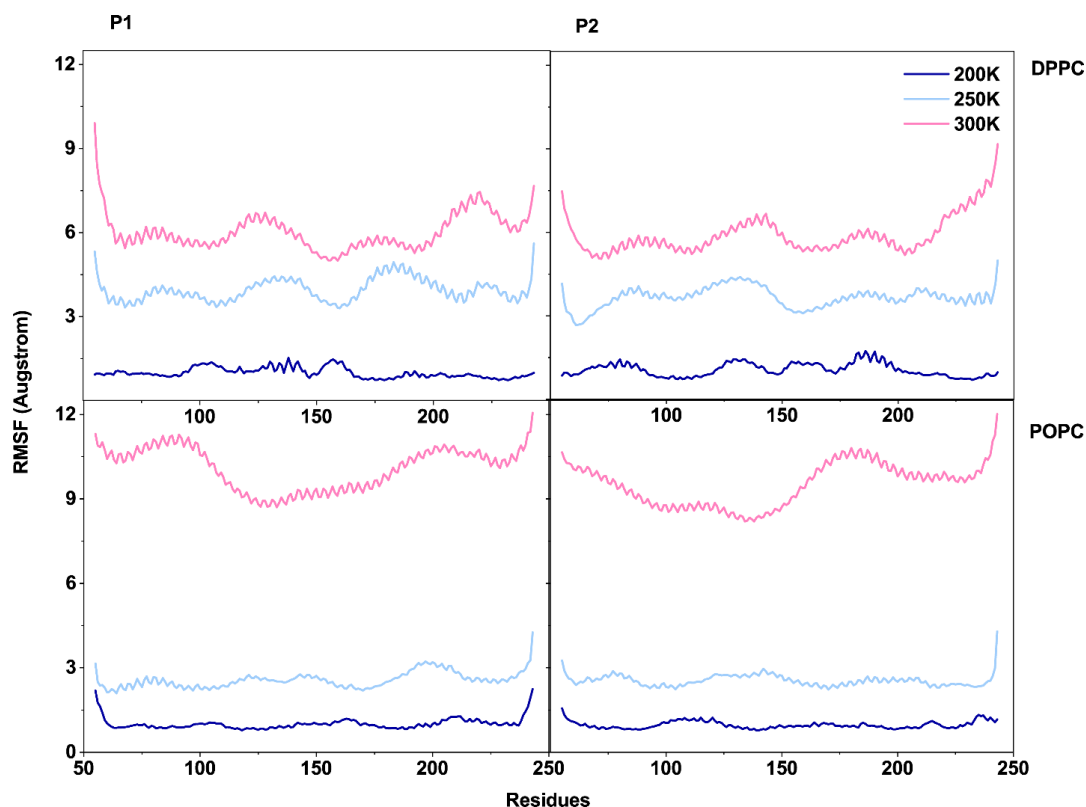


Figure II- 5: Root mean square fluctuation (RMSF) of each residue in protein chains P1 and P2 of MSP1 encapsulating DPPC and POPC membranes averaged over time 0.6-2.0  $\mu$ s at 200 K, 250 K, and 300 K

#### 2.4.3 Principal Component Analysis (PCA)

To identify large amplitude motions of the nanodiscs, especially MSP1, the collective protein motions about 1  $\mu$ s of simulation time were calculated using PCA. As shown by the two PCA modes at 200 K, there is no major motion of the both type nanodiscs. A small distortion from the ideal circular shape without apparent change in conformation was observed at the frozen temperature (Figure II-6). When temperature rose to 250 K, we found large-scale conformational changes. DPPC and POPC nanodiscs twisted and expanded laterally. At this temperature, the nanodiscs undergo a transformation from a circular shape to a ripple structure. The major motion around the N- and C-termini and the middle of the ring was observed. At 300 K, the MSP1 ring underwent a conformational twist, in which POPC nanodisc represents a more bended structure than DPPC nanodisc. These results are in good agreement with the out-of-plane deformation of the nanodiscs previously reported [63].

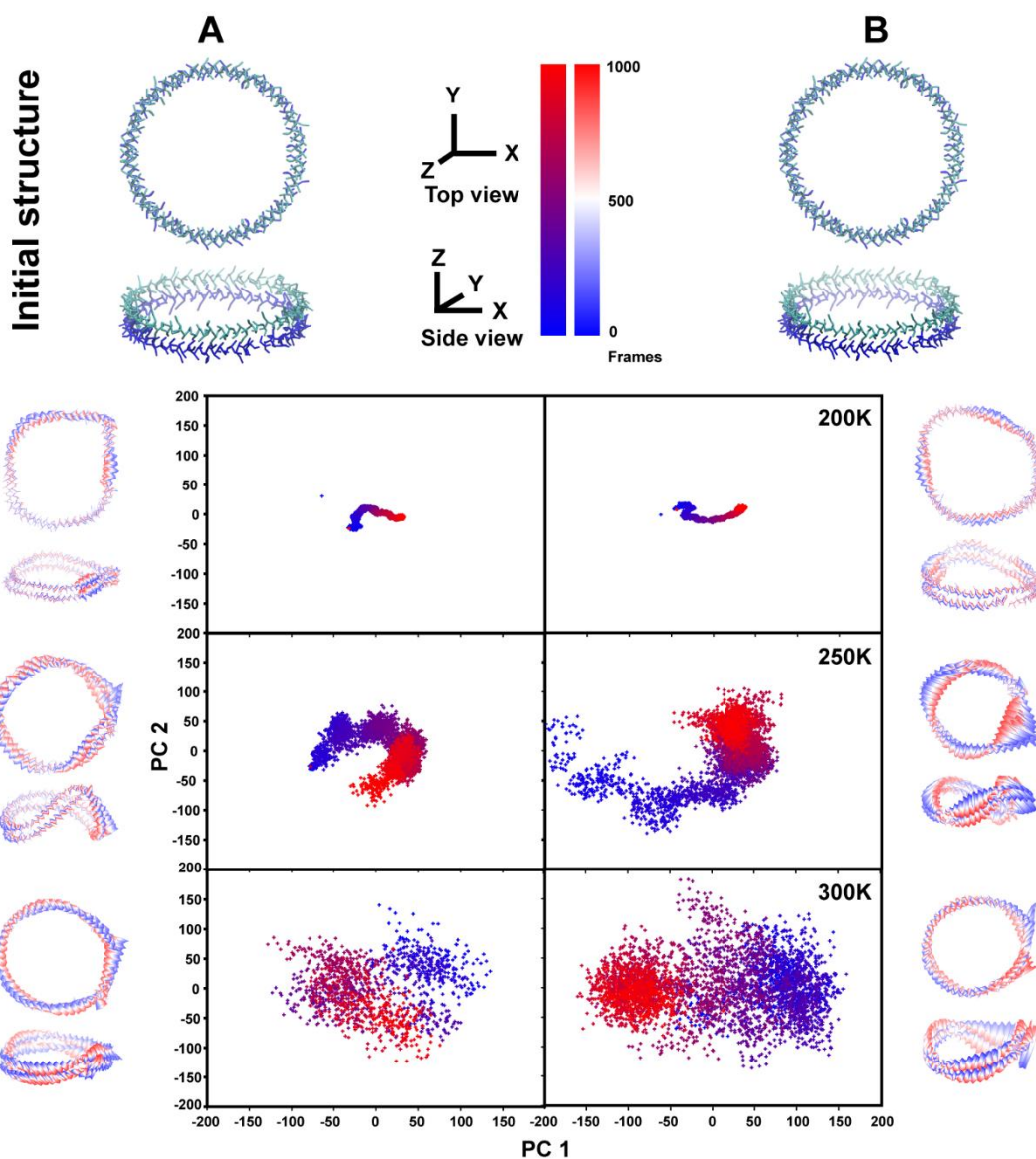


Figure II- 6: PCA plots of DPPC (A) and POPC (B) nanodisc projected between PC1 and PC2 that relate with structure of MSP1 from initial (blue) to final simulation (red) snapshot.

#### 2.4.4 Solvation profile

The solvation profiles of the nanodiscs were investigated by analyzing the pair correlation function  $g(r)$  between the nanodisc and surrounding water. The solvation pair correlation function (see Figure II-S4)  $g(r)$  in both nanodiscs (DPPC and POPC) showed a narrow and sharp first peak followed by a broad second peak (with splitting at low temperatures) and broader peaks at larger distances in the bulk. The maximum

positions of the first peak in each  $g(r)$  were located around 5.5 Å at all temperature ( $T=200$  K – 350 K). The non-zero value of the first minimum of pair correlation functions in both nanodiscs (DPPC, POPC) however, indicates the lability of water molecules in the first hydration shell of the nanodisc. The second peak is located around 10.0 Å. Based on the  $g(r)$  plots (Figure II-S4), a cut off distance of 7.0 Å was used to estimate the number of nearest neighbor water molecules in the first coordination shell of nanodiscs.

Note that the first peak of the solvation pair correlation function (Figure II-S4) dominates the solvation profile as it represents the most interacting water molecules. Therefore, it is worth examining the first peak of the solvation correlation function and the underlying population of water molecules closely. The data of the first solvation pair correlation function and the number of water molecules in the first shell are presented in Figure II-7 for a range of temperature ( $T=200$  K – 350 K) in both DPPC and POPC nanodiscs. We see that the maxima of the peak decreases on increasing the temperature. The nature of decay in DPPC nanodisc is similar to that in POPC except an anomaly in POPC at  $T = 250$  K (Figure II-7). The number of water molecules increases with the range with the maximum growth in the range of about 0.50 – 0.55 nm. Raising the temperature decreases the number of water molecules in the first hydration shell in both nanodiscs. The response of solvation profile appears to be slightly more sensitive in POPC nanodisc than that in DPPC.

CHULALONGKORN UNIVERSITY



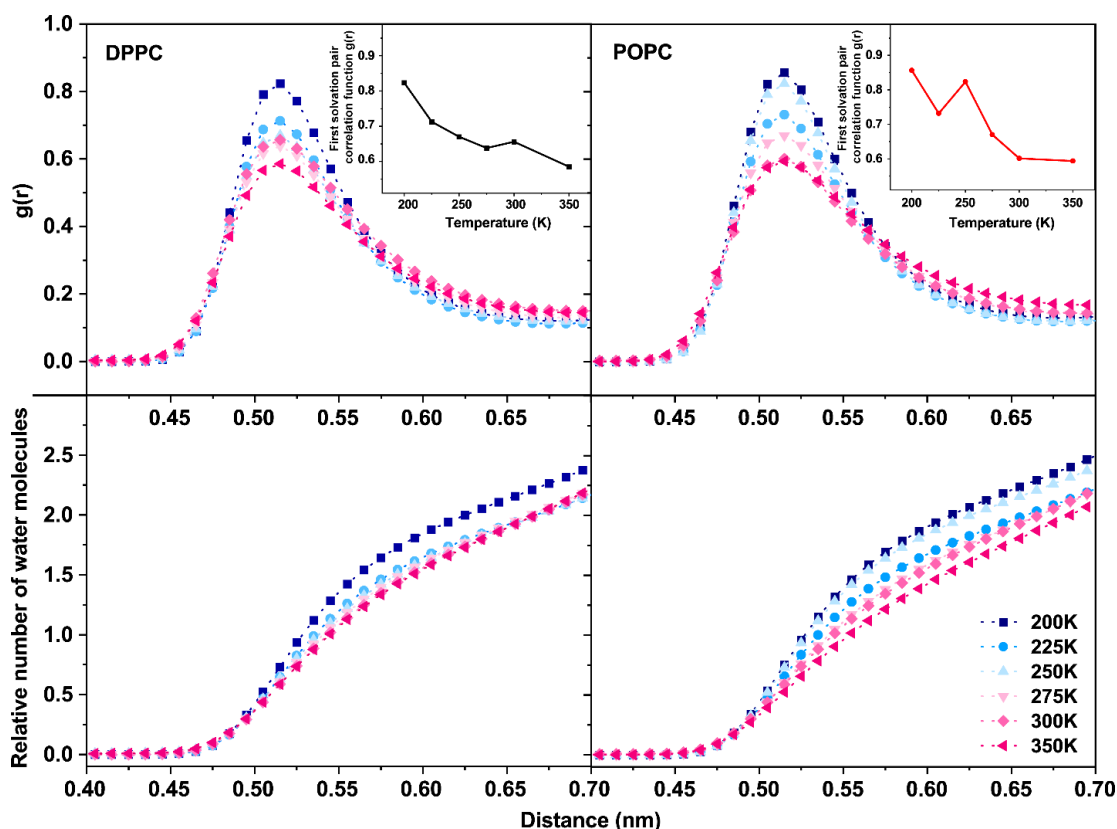


Figure II- 7: First solvation pair correlation function  $g(r)$  and the number of underlying water molecules in DPPC and POPC nanodiscs at temperatures  $T=200$  K – 350 K. The inset Figure II-S6 show the variation of the maxima of  $g(r)$  with the temperature.

The average number of water molecules surrounding the protein MSP1 and the entire nanodisc remains almost constant at all temperatures ( $T=200$  K – 350 K) except a drop around 250 K in POPC nanodisc (see Figure II-S5). Apart from anomaly at 250 K in POPC nanodisc, the hydration profiles of both nanodiscs are comparable within the error bars.

#### 2.4.5 Solvent accessible surface area

Solvent accessible surface area (SASA) of the lipids and the protein (MSP1) belt were calculated. Variation of the solvent accessible surface area with the temperature is presented in Figure II-8. We see that SASA increases with the temperature in both nanodisc and that the magnitude of SASA is larger in DPPC than that in POPC at all

temperatures. Constitutive elements, i.e. lipids expand on increasing the temperature. Since the number of DPPC lipids (160) is larger than number of POPC lipids (135), the larger value of SASA in DPPC nanodisc is not surprising. The surface accessible area per lipid in POPC nanodisc is therefore larger than that in DPPC (see table 1).

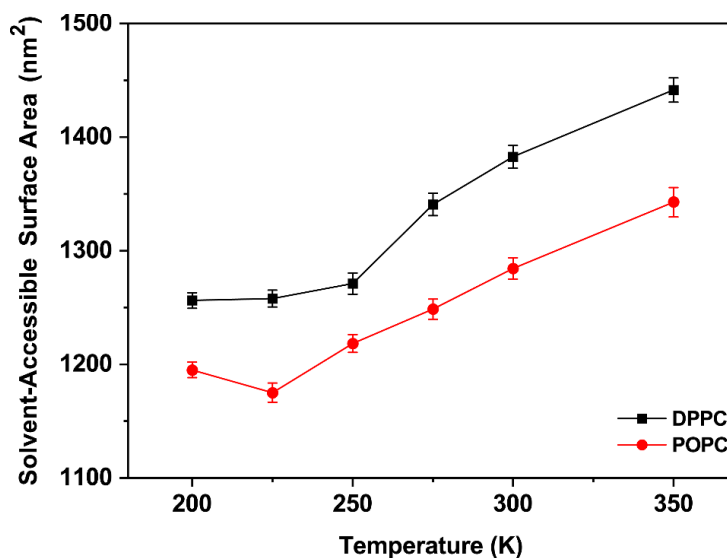


Figure II- 8: Variation of the average solvent accessible surface area (SASA) of DPPC (black) and POPC (red) nanodisc with the temperature; the average is evaluated from the last 500 ns simulation time.

**Table 1.** Solvent accessible surface areas per lipid of DPPC and POPC nanodiscs

Temperature (K)	SASA per lipid (nm <sup>2</sup> )	
	DPPC	POPC
200	7.85	8.85
225	7.86	8.70
250	7.94	9.02
275	8.37	9.25
300	8.64	9.51
350	9.01	9.95

#### 2.4.6 Radius of gyration of MSP1

The radius of gyration ( $R_g$ ) of the encapsulating protein MSP1 is a measure of the radius of the nanodisc. Therefore, the variation of  $R_g$  with the temperature may provide some insight into the thermal response of the nanodisc size. Figure II-9 shows the variation of  $R_g$  with the temperature in both DPPC and POPC nanodiscs. At low temperature ( $T=200$  K), the morphology of the nanodisc appears to be almost frozen. A small increase in temperature ( $T=225$  K), enhances the movement of lipid and proteins. As a result, the membrane and the protein (i.e. nanodisc) relaxes and becomes somewhat compact (with lower  $R_g$ ). The radius of gyration increases on further increasing the temperature ( $T=200$  K – 350 K), i.e. the nanodiscs expand. The radius of gyration of the DPPC nanodisc is higher than that of the POPC nanodisc. Expansion of nanodiscs with the temperature leads to increase in its surface accessible area as seen before (Figure II-8).

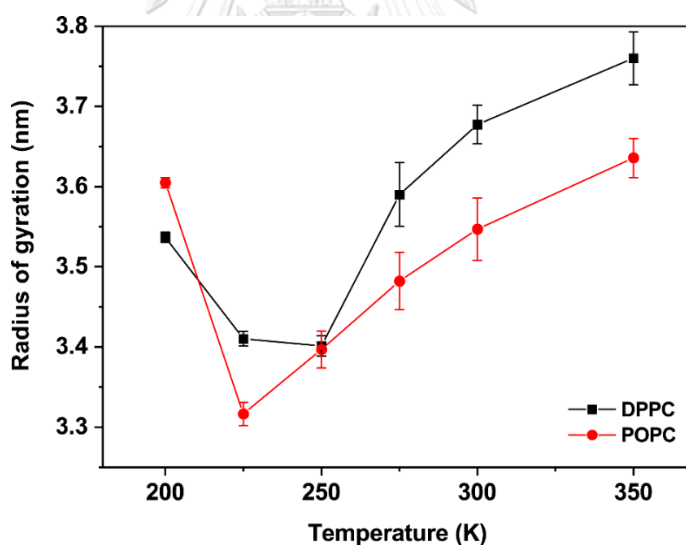


Figure II- 9: The average radius of gyration of MSP1 in DPPC and POPC nanodiscs at 200 - 350 K evaluated from the last 500 ns trajectory of simulation.

## 2.5 Conclusions

A nanodisc is designed from an assembly of lipid bilayer membrane encapsulated by a lipoprotein in a circular morphology. Two nanodiscs are considered involving a DPPC lipid membrane and a POPC lipid membrane in aqueous solution. Each membrane is encapsulated by a double helical lipoprotein MSP1 consisting of two polypeptide chains P1 and P2. Structural fluctuations of the nanodiscs are investigated as a function of temperature with a Molecular Dynamics simulation with a coarse-grained representation of its constitutive elements involving MARTINI force field.

Root mean square deviation (RMSD) of MSP1 and its residues, solvent accessible surface area (SASA) of the nanodiscs and their solvation profiles, radius of gyration of MSP1 include some of the physical quantities examined in detail for a range of temperature ( $T=200 - 350$  K). We find that the average RMSDs of MSP1 encapsulating both nanodiscs DPPC and POPC responds non-monotonically to temperature. The RMSDs in both nanodiscs increases on raising the temperature (200 K-250 K) followed by a decay on elevating the temperature further (250 K to 350 K) with a maximum around a characteristic value (250 K). We demonstrated that the non-monotonic temperature response is associated with the lipid bilayer phase transition. CG-MD data showed dramatic changes from the circular to twisted or non-circular shape at metastable ripple state (225K-275K). In addition, the magnitude of RMSD of MSP1 encapsulating POPC membrane is higher than that of the DPPC membrane. In the fluid phase, the protein belt of POPC-nanodisc appears to be more flexible than that of DPPC-nanodisc. Lipid molecules DPPC and POPC are similar in size and structure, though there is an important difference where POPC lipid has an additional CG bead with a double bond (C=C) character in one of the tails than that of the DPPC. The simulations showed that the interaction energy of the lipid with MSP1 is greater for DPPC-nanodisc than POPC-nanodisc. Thus, the DPPC nanodisc tends to resist its structural changes more than the POPC nanodisc at almost all temperature.

Solvent accessible surface area (SASA) of the nanodiscs increases with the temperature and that the magnitude of SASA is higher in DPPC than that in POPC. The radius of gyration ( $R_g$ ) of MSP1 also shows similar thermal response, i.e. it expands with the temperature and that the growth of  $R_g$  in DPPC nanodisc is higher

than that in POPC. These findings suggest that a small change in lipid structure leads to modulation the structural fluctuation of the nanodiscs with the temperature systematically. The variation of structure morphology and flexibility of nanodiscs with different lipid compositions should be taken into consideration in applying structural parameters during model fitting and data analysis. These precautions could eliminate ambiguities and improve the accuracy of the structural model. We hope that this study may help in designing appropriate nanodiscs and interpreting data in underlying experiments in future.

## 2.6 Supplementary figures

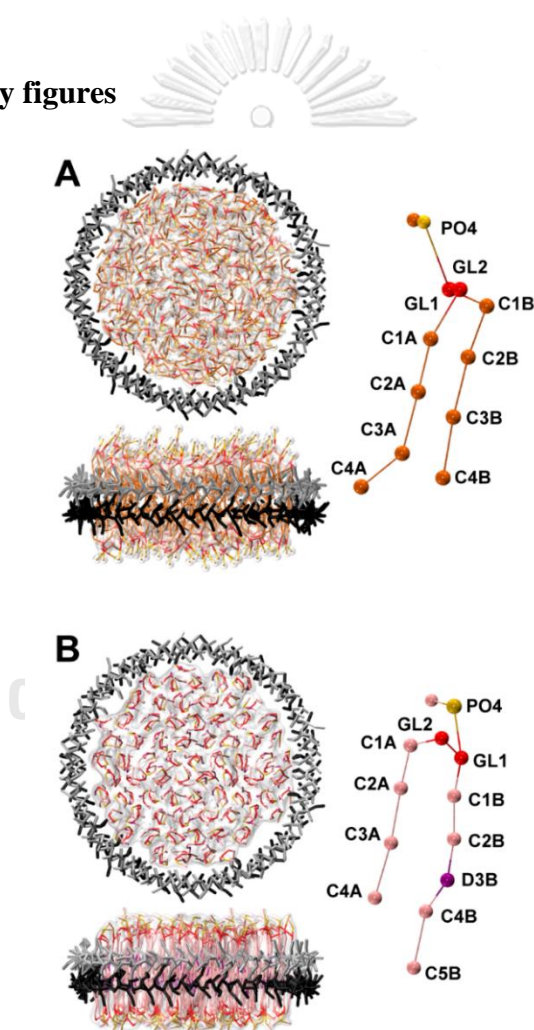


Figure II- S1: Top and side views of the CG DPPC nanodisc (A) and POPC nanodisc (B). MSP1 consisting of P1 and P2 (back and grey sticks) encircles DPPC and POPC. The corresponding MARTINI lipid topology of the CG model is shown on the right.

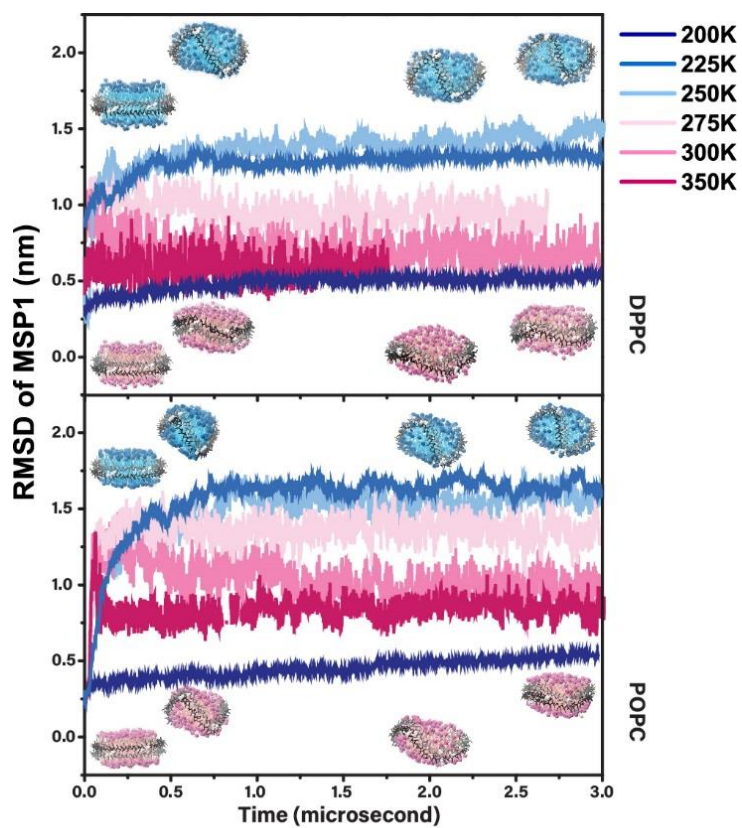


Figure II- S2: The RMSD of the backbone profile of MSP1 at temperatures 200K - 350K. Snapshots of representative nanodiscs are shown at 250K (blue) and 300K (pink)

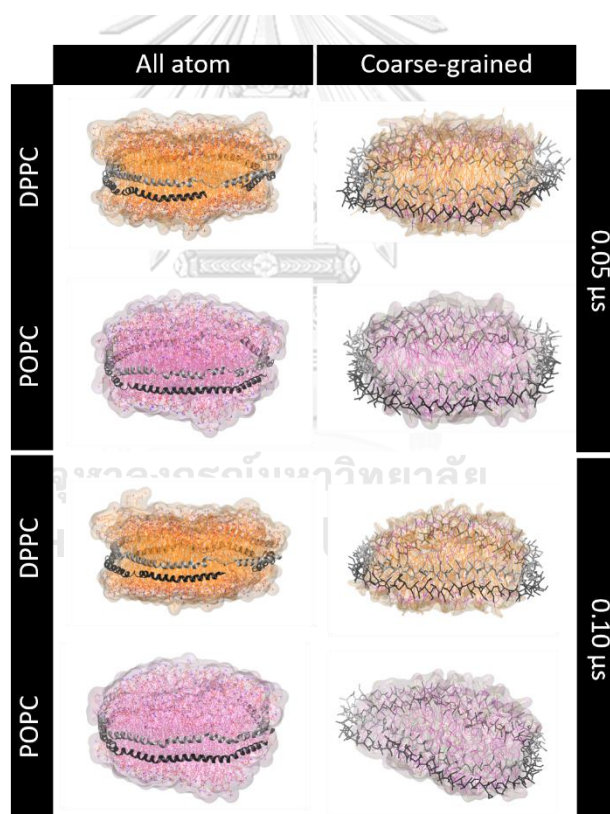
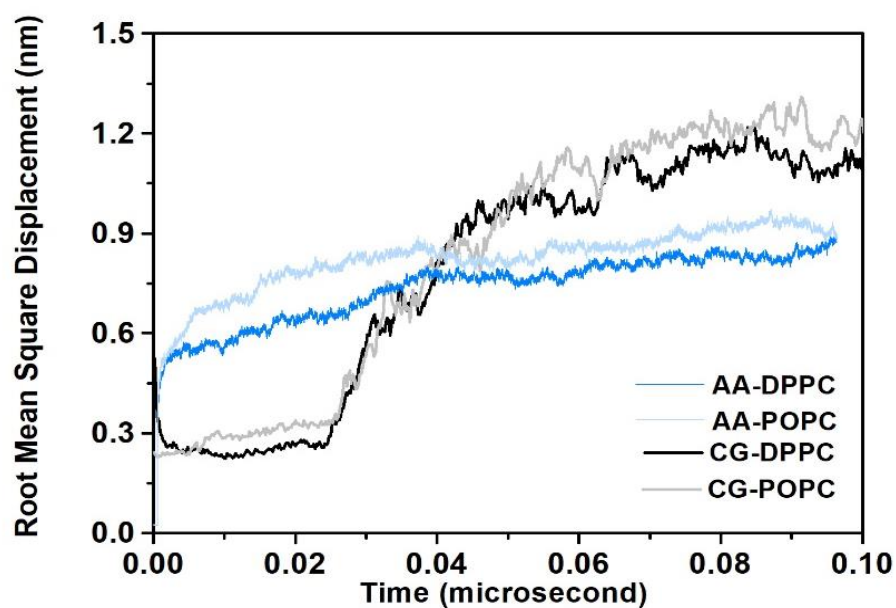


Figure II- S3: RMSD of MSP1 as a function of time of all-atomistic and coarse-grained simulations of DPPC nanodisc and POPC nanodisc at 250 K (Upper). All-atom and coarse-grained snapshots at time 0.05 and  $\sim 0.1 \mu\text{s}$  (Lower).

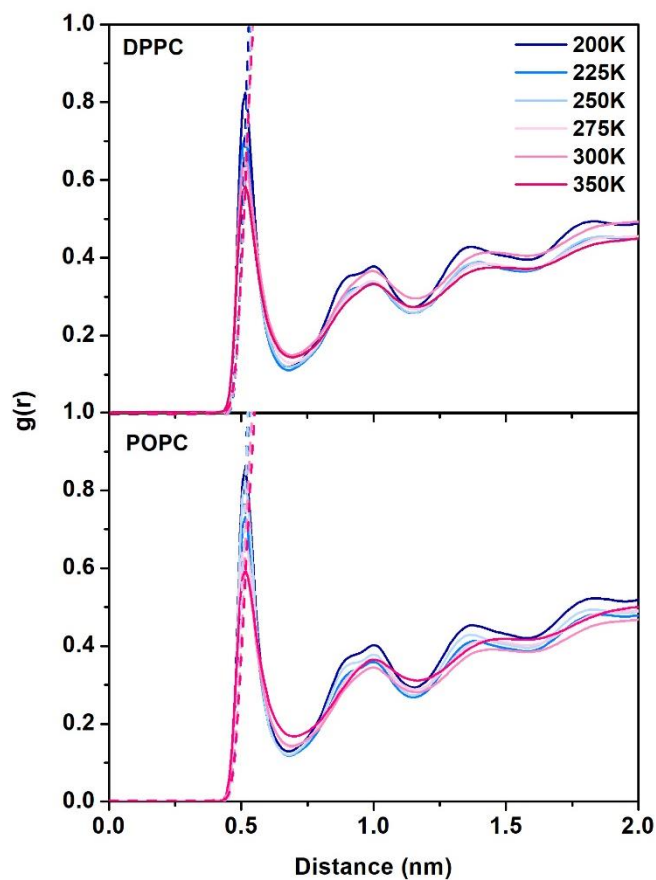


Figure II- S4: Radial distribution function of water surrounding MSP1 (line) and number of water (dash line) represent in different temperature for DPPC and POPC nanodiscs.

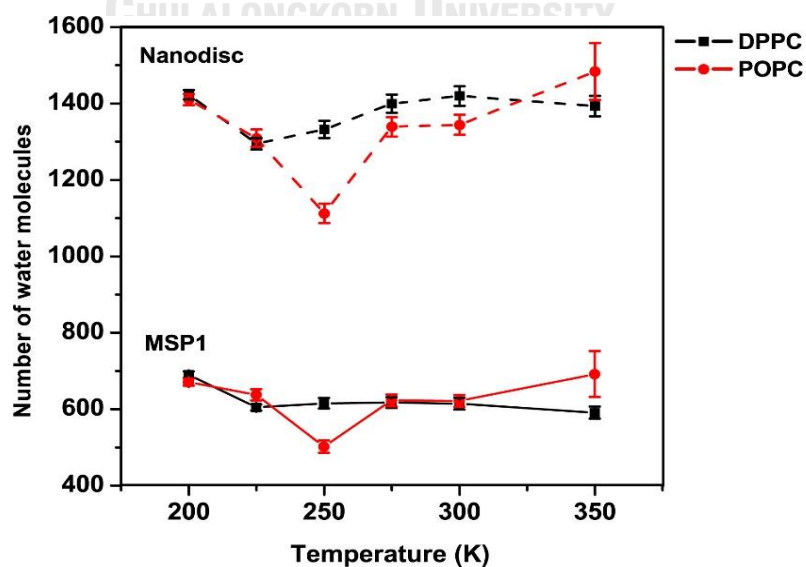




Figure II- S5: The average number of water molecules surrounding Nanodisc (dash line) and only MSP1 (continuous line) in the first shell of 0.7 nm at the last 500 ns of simulations.

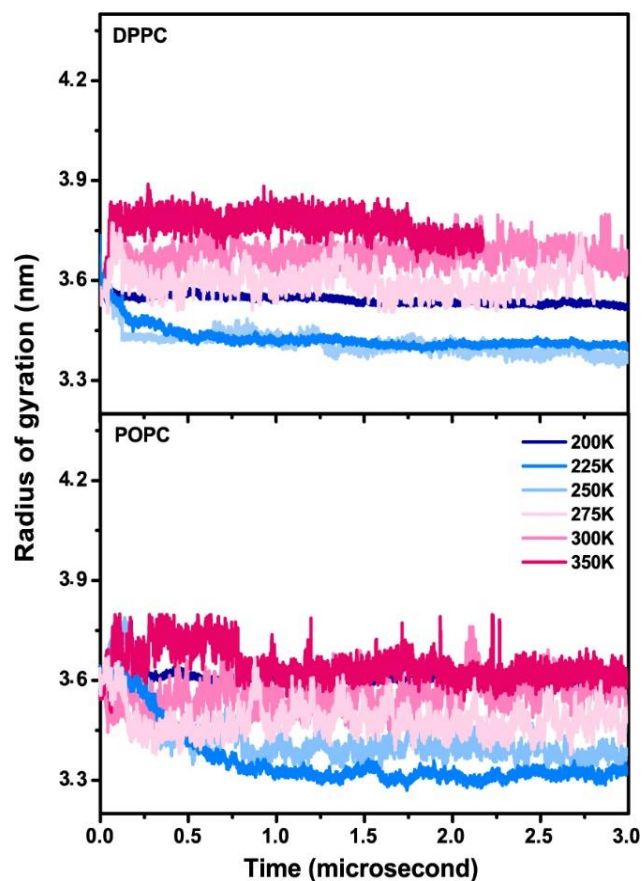


Figure II- S6: The radius of gyration profile of nanodiscs for DPPC and POPC at different temperature

## CHAPTER III

---

### **Globular bundles and entangled network of proteins (CorA) by a coarse-grained Monte Carlo simulation**

*Warin Jetsadawisut<sup>1</sup>, Sunan Kitjaruwankul<sup>2</sup>, Panisak Boonamnaj<sup>1</sup>, Pornthep Sompornpisut<sup>1</sup>, R.B. Pandey<sup>3</sup>*

*<sup>1</sup>Center of Excellence in Computational Chemistry, Department of Chemistry, Chulalongkorn University, Bangkok 10330, Thailand*

*<sup>2</sup>Faculty of Science at Sriracha, Kasetsart University Sriracha Campus, Chonburi 20230, Thailand*

*<sup>3</sup>Department of Physics and Astronomy, University of Southern Mississippi, Hattiesburg, MS 39406, USA*

Doi: 10.3934/biophy.2019.2.68

Journal: AIMS Biophysics, 2019, 6(2):68–82.

Received: 26 April 2019

Accepted: 24 June 2019

Published: 18 July 2019

---

**Keywords:** Self-organizing structure of proteins (CorA); protein folding; Coarse-grained model; Monte Carlo simulation; effective dimension of self-assembly

### 3.1 Abstract

Using a coarse-grained model, self-organized assembly of proteins (e.g., CorA and its inner segment iCorA) is studied by examining quantities such as contact profile, radius of gyration, and structure factor as a function of protein concentration at a range of low (native phase) to high (denature phase) temperatures. Visual inspections show distinct structures, i.e., isolated globular bundles to entangled network on multiple length scales in dilute to crowded protein concentrations. In native phase, the radius of gyration of the protein does not vary much with the protein concentration while that of its inner segment increases systematically. In contrast, the radius of gyration of the protein shows enormous growth with the concentration due to entanglement while that of the inner segment remains almost constant in denatured phase. The multi-scale morphology of the collective assembly is quantified by estimating the effective dimension  $D$  of protein from scaling of the structure factor: collective assembly from inner segments remains globular ( $D \sim 3$ ) at almost all length scales in its native phase while that from protein chains shows sparsely distributed morphology with  $D \sim 2$  in entire temperature range due to entanglement except in crowded environment at low temperature where  $D \sim 2.6$ . Higher morphological response of chains with only the inner-segments due to selective interactions in its native phase may be more conducive to self-organizing mechanism than that of the remaining segments of the protein chains.

### 3.2 Introduction

Self-assembly of protein [64-71] provides better mechanical strength and more stable dynamical response to the underlying environment such as membrane via its hierarchical morphology. It can provide a reliable and responsive pathway in ion channels [10-13, 15, 72-81] for selective transport of such specific elements as potassium ions. There are many examples of protein self-assembly with conflicting (adverse and cooperative) effects [66, 67]. For example, the self-assembly of proteins into a viral capsid is critical in preserving the genome from non-conducting external factors such as digesting enzymes of the host cells, undesirable pH, temperature, etc. [66]. On the other hand, the self-assembly of proteins triggered by the conformational

changes of the proteins may lead to undesirable aggregation such as amyloid beta-proteins into fibrils. Of a diverse range of proteins with unique and universal response properties, we consider a transmembrane protein CorA [10-13, 15, 72-80] which spans across the membrane with well-defined inner (iCorA) and outer (oCorA) segments [81]. Recently, we have observed an unusual thermal response of CorA particularly of its inner segment (iCorA) in its native phase [82]. For example, we find that the size of iCorA reduces on heating while that of the oCorA shows less organized response in its response of its overall size in its native phase as illustrated in Figure III-1.

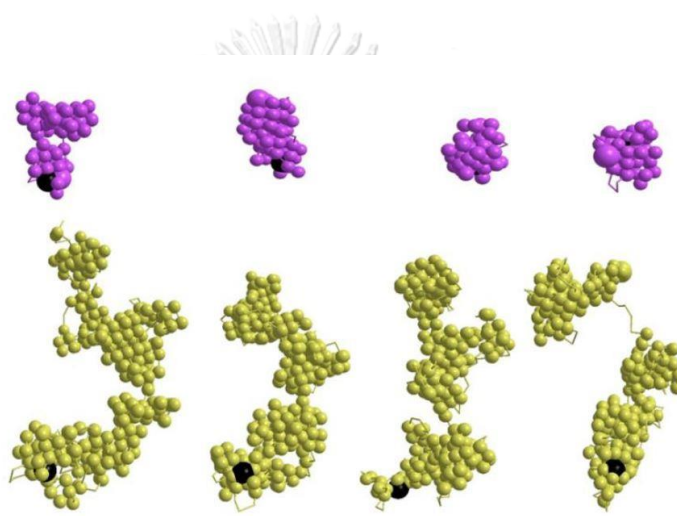


Figure III- 1: Snapshots of iCorA (top row) and oCorA (bottom row) at the end of  $10^7$   $\mu$ s time at  $T=0.010, 0.013, 0.018, 0.020$  [82].

In denatured phase, iCorA expands continuously and oCorA shows abrupt increase in its size in a narrow range of temperature

The functional structure of CorA is known to exist as a homo-pentamer [78] to provide coordinated open and close states for the selective transport of  $Mg^{2+}$  across the ion channels. Transport of ions depends on the permeation pathways which depend on the conformation of individual proteins and their interacting network. How CorA proteins assemble collectively into a well-defined morphology is not well understood.

Selective transport across the ion channels has been investigated extensively for many years by examining the binding of  $Mg^{2+}$  and identifying the role of special metal

binding motifs such as Gly-Met-Asn. Despite enormous efforts, particularly by laboratory measurements, understanding of the mechanism of ion transport and the underlying pathways remain speculative. Computer simulations become a viable tool to probe the speculative hypothesis and help understanding the structural response of proteins and its assembly. Extensive efforts have been recently made to understand the underlying structure of CorA embedded in a realistic membrane model with solvent by all-atom MD simulations [78]. Such investigations are insightful in probing the small-scale structural details; the dynamical changes generally appear too small to probe large-scale structural responses despite large-scale computer simulations. For example, the morphology of a set of five proteins with inner segment confined by a nanodisc seems to deform out somewhat if they start from a symmetric configuration (see Figure III-1). Even with a large-scale all-atom MD simulation, it is rather difficult to see a significant dissociation besides fusion of the symmetry. Therefore, a large-scale coarse-grained analysis is needed to augment and clarify the distinctions (see below) if feasible. Perhaps protein-protein interaction is not enough to direct five CorA proteins into a stable pentamer; other factors such as a nanodisc or membrane matrix may be needed for the stability. Before incorporating additional constitutive components to probe directed assembly, it would be important to understand the self-assembly of the proteins first. We see that it is not feasible to find appreciable structural variations in a large-scale all-atom MD simulation. Some degree of coarse-graining [81, 82] is thus almost impossible to avoid in such a large-scale investigation.

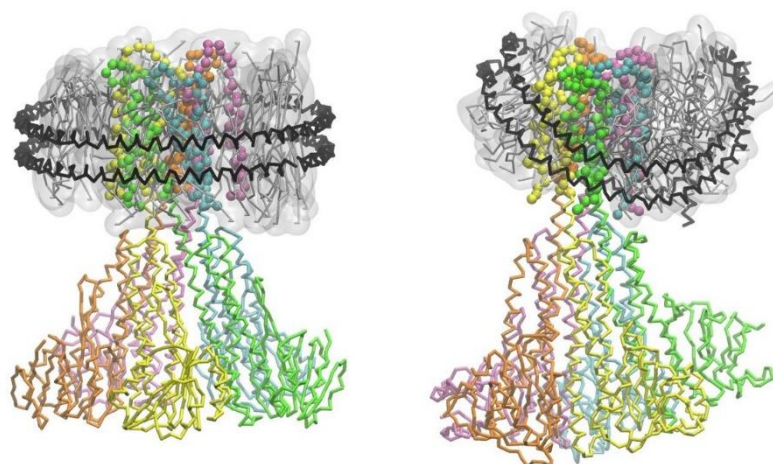


Figure III- 2: Pentamer of CorA (each protein shown in different colors) with inner core (iCorA) confined by a nanodisc shown in black. The initial configuration is on the left and final configuration on the right after 2  $\mu$ s.

We know that the structures of the inner (iCorA) and outer (oCorA) membrane segments of the protein (CorA) respond differently to temperature in its native and denatured phase. Their response also differs due to the underlying matrix (a crowded environment) with different internal (hydrophobic) and external (hydrophilic) solvent and solute environments. It would be interesting to probe how the proteins and the underlying matrix assemble on a large-time scale in presence of many proteins. The large-scale computer simulations become too intense with all the constitutive components [78] due to interplay between their cooperative and competing interactions. Therefore, it is practical to focus on a simplified model [81, 82] first before adding complexity of the underlying matrix systematically. Since the protein channel involves cooperative response of many proteins, it would be desirable to probe the collective structures in a crowded protein environment. Our goal is to understand the stability of the multi-scale morphologies of interacting proteins in its native and denatured phases. The model is described in the next section, followed by results and discussion with a concluding remark at the end.

### 3.3 Model

We consider a coarse-grained model [82] of the protein chain on a cubic lattice for simplicity, efficiency, and practical utility: it incorporates specificity of each residue, ample degrees of freedom (unlike minimalist lattice models) for residues to perform their stochastic movements and their covalent bonds to fluctuate. The internal structures of residue (i.e. fine-grained details at atomistic scale) are ignored, however, the specificity of residues is considered via their unique residue-residue interactions (see below) which is critical in capturing the unique characteristics of the protein. In our coarse-grained representation, CorA is a chain of 351 nodes (residues) tethered together on a cubic lattice by flexible peptide bonds. A node represents a residue and occupies a unit cell of size  $(2a)^3$ , with lattice constant  $a$ ); the bond length between consecutive residues varies between 2 and  $\sqrt{10}$  in unit of lattice constant ( $a$ ) [33]. Typically, a protein chain is placed in the simulation box in a random configuration, initially with minimum bond length  $(2a)$  between the consecutive nodes; this initial configuration is further randomized by allowing each node to perform its stochastic movement (to its 26 neighboring cells with varying distance) with the strictly implemented excluded volume constraint. The bond-fluctuating mechanism has been used extensively in addressing a range of complex problems in polymers [33]; we have extended its utility to model protein chains [81, 82]. Coarse-graining of the simulation box (cubic lattice) and the residues representation by a lattice cell make it one of the efficient computational method for modeling such systems as protein while retaining the potential for fine-graining [83] to further enhance the degrees of freedom.

Each residue interacts with the neighboring residues within a range ( $r_c$ ) of interaction with a generalized Lennard-Jones potential,

*Equation 6 the phenomenological potential residue interacts*

$$U_{ij} = \left[ |\varepsilon_{ij}| \left( \frac{\sigma}{r_{ij}} \right)^{12} + \varepsilon_{ij} \left( \frac{\sigma}{r_{ij}} \right)^6 \right], r_{ij} < r_c$$

where  $r_{ij}$  is the distance between the residues at site  $i$  and  $j$ ;  $r_c = \sqrt{8}$  and  $\sigma = 1$  in units of lattice constant. We use a knowledge-based residue-residue interaction as input in our

phenomenological potential Eq.6 for  $\varepsilon_{ij}$ . The knowledge-based residue-residue interactions [84, 85] are derived from the distribution of the amino acids in a growing ensemble of frozen protein structures in protein data bank (PDB); the underlying solvent environment is therefore taken into account implicitly. Various assumptions and approximations [84, 85] are further made in deriving these contact potentials which makes it somewhat difficult to calibrate the scales of the physical quantities in absolute units. The relative strength of residue-residue interactions is critical for protein to adopt its specific conformations. Thus, the potential strength,  $\varepsilon_{ij}$ , is unique for each interaction pair with appropriate positive (repulsive) and negative (attractive) values selected from the knowledge-based contact interactions [84, 85]. The knowledge-based residue-residue interactions have been used extensively in modeling protein structures for decades [86-89]. We use the residue-residue contact matrix by Betancourt-Thirumalai (BT) [84], an improved version of classic Miyazawa-Jernigan (MJ) interaction [85]. It is worth pointing out that there can be alternate methods to incorporate the specificity of residue interactions in a phenomenological potential Eq.6 including the computed interaction involving all-atom MD simulation.

Structural evolution of protein chains (CorA) are analyzed in detail as they interact, associate, and dissociate in their native (relatively low temperature) to denatured (high temperature) phases. The sample is prepared by inserting protein chains randomly first in the simulation box; protein chains are moved around with excluded volume constraints to randomize their conformation and distribution further. The protein chain CorA consists of 351 residue, inserting many chains, each of length ( $L_c = 351$ ) in a simulation box of size  $L^3$  ( $= 350^3$ ) become difficult beyond a certain number ( $N_c = 5 - 100$ ) of chains due to onset of jamming. The computer processing unit (CPU) also increases with the number of chains. Therefore, the number of protein chains is restricted to perform computer simulations with reasonable CPUs (order of days to week) while capturing the cooperative and competing effects of residue-residue, segmental, and overall protein-protein interactions on the local and global association and dissociation of protein chains. Further, relaxation time to reach steady-state (by examining the temporal evolution of various physical quantities) also increases which makes it



difficult to assess the thermal response of physical quantities in equilibrium. The protein concentration is defined by the fraction  $p$  of the lattice sites occupied by the residues of the protein chains, i.e.,  $p = 8 \times N_c \times L_c / L^3$ . The number of protein chains  $N_c = 5 - 100$  each with length  $L_c = 351$  (protein chain of CorA) and  $L_c = 61$  (inner segment of protein) are considered with simulation box of size  $L = 350$ , and  $150$  respectively. Note that the simulation box becomes crowded at a lower protein concentrations with longer chain lengths.

Each residue performs its stochastic movements with the Metropolis algorithm which involves selecting a residue randomly at a site say  $i$  of a randomly selected protein chain and attempting to move it to one of its neighboring site say  $j$ . If the excluded volume constraints and limits on the bond length for the proposed move are satisfied then we calculate the corresponding change in energy  $\Delta E_{ij} = E_j - E_i$  between its old ( $E_i$ ) and new ( $E_j$ ) configurations, and move the node with the Boltzmann probability  $\exp(-\Delta E_{ij} / T)$  where  $T$  is the temperature in reduced unit of the Boltzmann constant ( $k_B$ ). Attempts to move each node once defines the unit Monte Carlo step (MCS), i.e.  $N_c \times L_c$  attempts to move randomly selected nodes in a simulation box with  $N_c$  protein chains each with  $L_c$  residues defines the unit MCS. Simulations were performed for sufficiently long-time steps (typically 10 million-time steps) to make sure that system has reached its steady state equilibrium with a number of independent samples (10-100) to estimate the average values of the physical quantities. A number of local and global physical quantities were determined such as the energy of each residue and protein chain, contact map, their mobility, mean square displacement of the center of mass of the protein, radius of gyration, and its structure factor. These physical quantities are in arbitrary units, i.e., the length is in units of the lattice constant which is different from many all-atom simulations where realistic units for size and time scales are used via calibration of  $\sigma$  and  $\epsilon_{ij}$  from experimental data for simplified systems. It is difficult to quantify physical quantities in absolute units due to the phenomenological nature of the interactions Eq.6 used here. The trend in response of the physical quantities to such parameters as the temperature and network density should however be qualitatively comparable with appropriate observations.

As mentioned above, the transmembrane protein CorA consists of 351 residues ( $^1\text{M}^2\text{E} \dots ^{351}\text{L}$ ) with a well-defined outer (oCorA: residues  $^1\text{M}^{290}\text{V}$ ) and inner (iCorA:  $^{291}\text{M} - ^{351}\text{L}$ ) core. The inner (iCorA) and outer (oCorA) core of the protein exhibit contradictory thermal response in their native and denatured phases, negative (globularization on increasing the temperature) versus positive (random-coil) responses. The thermal response (variation of the radius of gyration with the temperature) of the protein CorA chain is dominated by the outer corA which is much larger than the inner core, it would be better to analyze the physical quantities of both CorA and its inner core, iCorA. Simulations are therefore carried out in a larger box ( $350^3$ ) with CorA proteins and smaller box ( $150^3$ ) with iCorA with the number of proteins 5-100. We find a remarkable difference in self-organized morphology of the protein CorA from that of its inner core (iCorA) as a function of temperature (follows).

### 3.4 Results and discussion

We study the effects of protein-protein interaction on a number of local and global physical quantities as proteins organize, associate and disassociate as a function of temperature and their concentration (i.e. the number of proteins). Our system has reached the steady-state equilibrium during the course of simulations (i.e. in  $10^7$  time steps) for almost entire temperature regime and concentrations except in extreme limits (i.e. high temperature and higher concentrations) where the relaxation time is too large due to entanglement and jamming of large proteins. Nevertheless, it is illustrative to understand the self-organizing morphologies as the protein chains self-assemble, entangle, and disassociate. As mentioned above CorA consists of inner (iCorA) and outer (oCorA) components which seem to exhibit different thermal response in their native and denatured phases. Since the outer core (oCorA), being much larger (with 291 residues  $^1\text{M}^2\text{E} \dots ^{290}\text{V}$ ) dominates the conformation of the entire protein chain, it is difficult to identify the role of specific structural response of the inner core (iCorA with only 61 residues  $^{291}\text{M}^{293}\text{V} \dots ^{351}\text{L}$ ) in analyzing the self-assembly of the whole protein CorA. Therefore, we investigate the structural response

of CorA and iCorA in presence of many proteins separately to assess its role in self-assembly.

### 3.4.1 Snapshots, contact map and profiles

Representative snapshots of CorA and that of its inner segments (iCorA) with 100 protein chains each in the simulation box is presented in Figure III-3 at a low ( $T = 0.020$ ) and a high ( $T = 0.040$ ) temperature. Both, the protein chains and its inner segments appear to disperse almost uniformly. A closer inspection shows that, in native (low temperature) phase, CorA proteins dissociate more on the scale comparable to the size of a protein chain in general while the clusters of inner segments appear to form isolated globular bundles. The residue-residue interaction is more conducive to agglomeration of chain with the inner segments than that of the full protein (CorA). Corresponding snapshots at a higher temperature ( $T = 0.040$ ) show different morphologies where some degree of phase-separation seem to occur with CorA as the proteins continue to dissociate, expand, and entangle. The spreading of extended inner segments at  $T = 0.040$  reduces the segregation of protein clusters seen at the low temperature. Attempts are made to quantify our qualitative observations of complex morphologies by analyzing the structure factor (see below).

Contact maps (Figures III-S1–S4) of the assembly of CorA chains and inner segments at a low temperature ( $T = 0.020$ ) seem to suggest a more compact morphology of iCorA in crowded protein environment than that of CorA. At high temperature ( $T = 0.040$ ), frequency of segmental contacts is much lower in their crowded protein matrices than that at the low temperature; segmental contacts in chains with inner segment is still higher than that with CorA. The average number ( $N_n$ ) of residues around each is a measure of the contact profile which is presented in Figures III-4 and -5 for a range of temperatures ( $T = 0.020$ – $0.040$ ). Obviously, the residue contact density decreases on increasing the temperature with more contacts at low temperatures (native phase) and less in denatured phase. Variation in pattern of the contact profiles with the temperature however, reveals how the protein segments organize during the self-assembly that leads to a global morphology. At the low temperature ( $T = 0.020$ ), most part of the protein chains seem to be surrounded by

proteins leading to a compact (globular) morphology. Onset of preferential contacts emerges on raising the temperature ( $T = 0.025$ ). For example, second half of residues in chains with only inner segment (iCorA) of the protein (residues  $^{316}\text{F} - ^{351}\text{L}$ ) have higher contact density than the first half (residues  $^{291}\text{M} - ^{315}\text{N}$ ) (Figure III-5). Preferential contact becomes more localized ( $^{316}\text{F} - ^{325}\text{W}$ ) on increasing the temperature further ( $T = 0.030$ ). Even though the contact density of many residues is relatively high in self-assembled morphology in CorA, its random distribution along the backbone makes less defined at higher temperatures (Figure III-4). Further, the fraction of segments with relatively low contact density remains appreciably high. Thus, the protein-protein interaction in inner segments is likely to provide a more stable support for a collective morphology of CorA protein

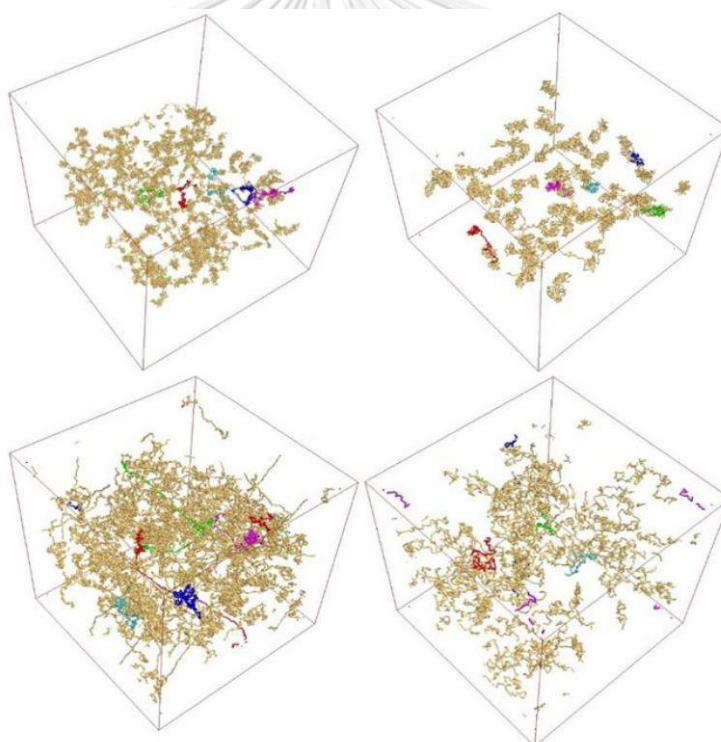


Figure III- 3: Snapshots of CorA (left column) and iCorA (right column) with 100 protein chains at a low ( $T=0.020$ , top) and a high ( $T=0.040$ , bottom) temperature at the end of  $10^7$  time steps. Five proteins are shown in different colors in each simulation box, remaining (95 chains) are golden khaki.

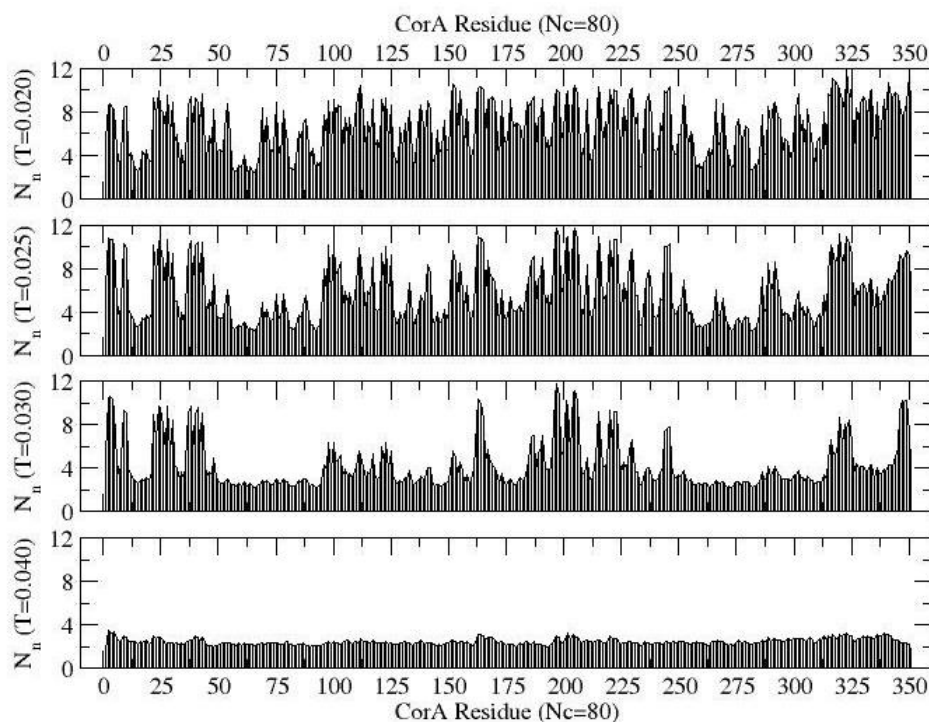


Figure III- 4: Average number of residues around each of CorA ( $^1M$ ,  $^2E$ , ...,  $^{351}L$ ) residues in a crowded matrix with 80 chains at low ( $T = 0.020, 0.025$ ) and high ( $T = 0.030, 0.040$ ) temperatures.

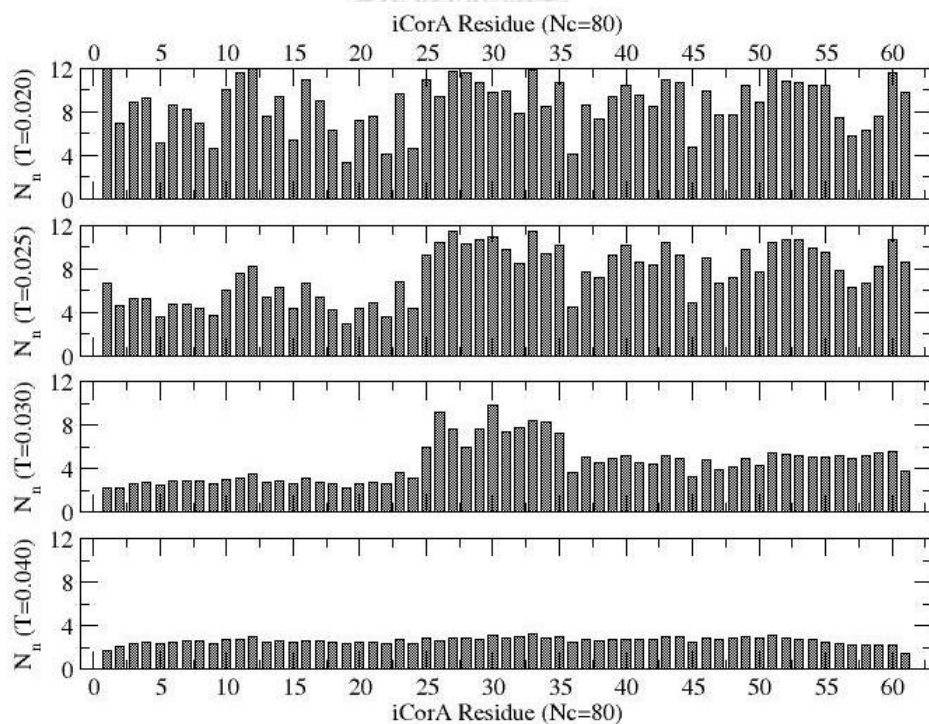


Figure III- 5: Average number of residues around each of iCorA ( $^{291}\text{M}$ ,  $^{293}\text{V}$ , ...,  $^{351}\text{L}$ ) residues in a crowded matrix with 80 chains at low ( $T = 0.020, 0.025$ ) and high ( $T = 0.030, 0.040$ ) temperatures.

### 3.4.2 RMS displacement and Radius of gyration

‘How protein chains move’ is analyzed by monitoring the variation of the root mean square (RMS) displacement ( $R_c$ ) of the center of mass of each protein with the time step ( $t$ ) as they associate and dissociate during their stochastic movements. Variations of  $R_c$  with  $t$  for CorA and iCorA with  $N_c = 5$  (dilute), 80 (crowded) are presented in Figure III-6 at representative low and high temperatures. Note that the dynamics of protein chains is almost diffusive,  $R_c \propto t^\nu$  with  $\nu = 1/2$  in asymptotic limit at lower temperatures ( $T=0.020-0.030$ ). At high temperature ( $T=0.040$ ), the protein CorA moves slightly faster than diffusion with  $\nu \sim 0.6$  in dilute ( $N_c = 5$ ) as well as crowded ( $N_c = 80$ ) matrix. The inner core protein (iCorA) chains move almost diffusively in dilute matrix ( $N_c = 5$ ) and super-diffusively ( $\nu \sim 0.7$ ) in a crowded environment ( $N_c = 80$ ) at high temperature ( $T=0.040$ ). In dilute protein matrix, inner core proteins appear to coagulate even at high temperatures while they move faster in a crowded environment, a concentration induced dynamics.

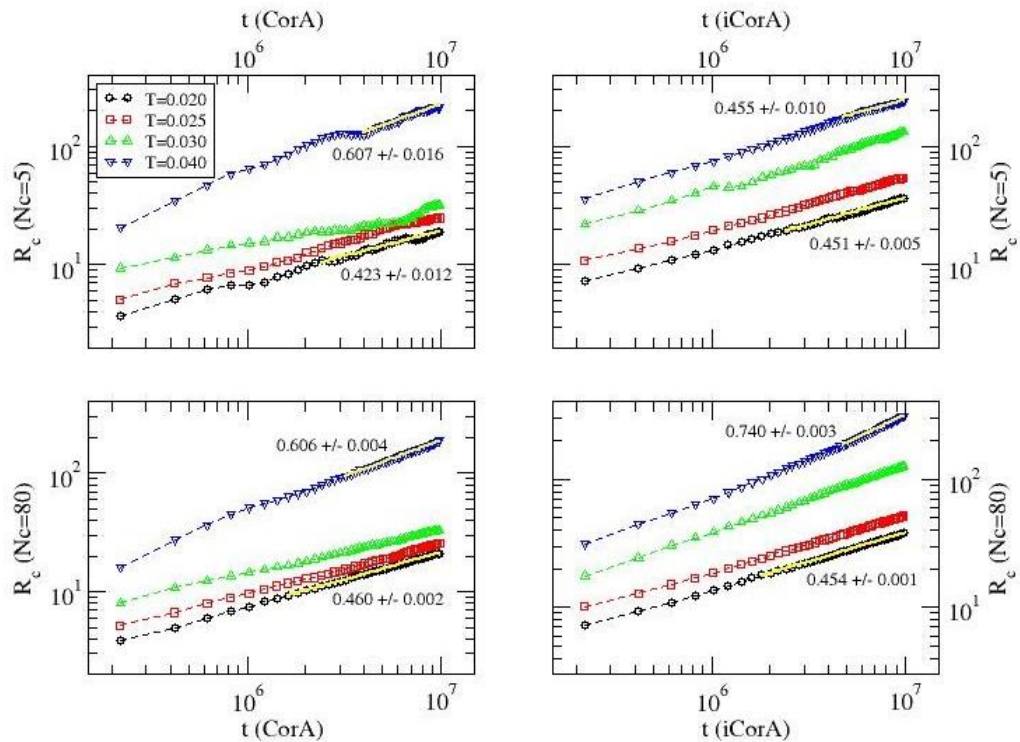


Figure III- 6: Variation of the root mean square (RMS) displacement ( $R_c$ ) with the time step of protein CorA and iCorA in dilute ( $N_c = 5$ ) and crowded ( $N_c = 80$ ) environment at representative low and high temperatures. Slopes of the asymptotic data points are included at a low and high temperatures for a guide.

How does the size of protein chains depend on the concentration of the crowded environment in native and denatured phases? How are residues distributed as the protein chains associate and dissociate? We address it by analyzing the radius of gyration of each protein and structure factor of the self-organized assembly. Variations of the average radius of gyration of the protein, CorA with the concentration is presented in Figure III-7 at representative low and high temperatures. The radius of gyration of the protein does not show much variation with the protein concentration at low temperatures (inset Figure III-7). At the high temperature ( $T=0.040$ ) the radius of gyration of CorA increases continuously with  $p$  and reaches a steady-state in the crowded regime. It should be pointed out that, onset of entanglement sets in as the protein chains begins to interact and entangle at higher protein concentrations; the radius of gyration has not reached the steady-state (see

below) at  $p \geq 0.0014$  due to competition between entanglement and their thermal expansion.

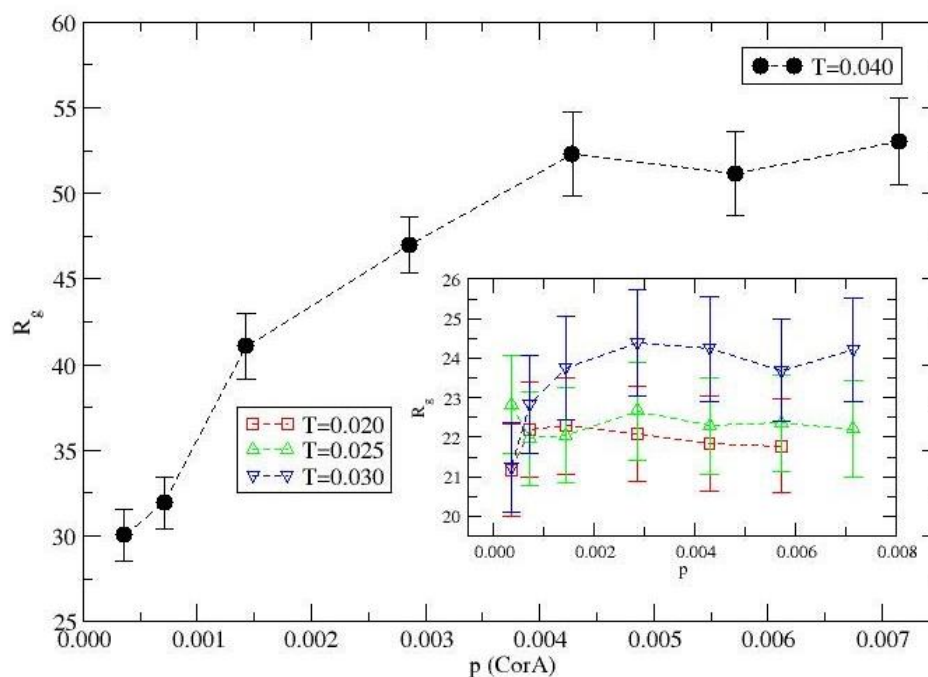


Figure III- 7: Variation of the average radius of gyration  $R_g$  of CorA with the protein concentration  $p$  at temperature  $T = 0.020-0.040$ .

Figure III-8 shows that the radius of gyration of iCorA increases with the protein concentration  $p$  systematically at low temperatures ( $T=0.015, 0.020$ ) in its native phase and exhibits little variation with  $p$  at higher temperatures ( $T \geq 0.025$ ) in its denatured phase. We therefore believe that the protein-protein interaction among the inner-segments of the protein in its native phase may be conducive to self-organized pathways for the ion-channel.



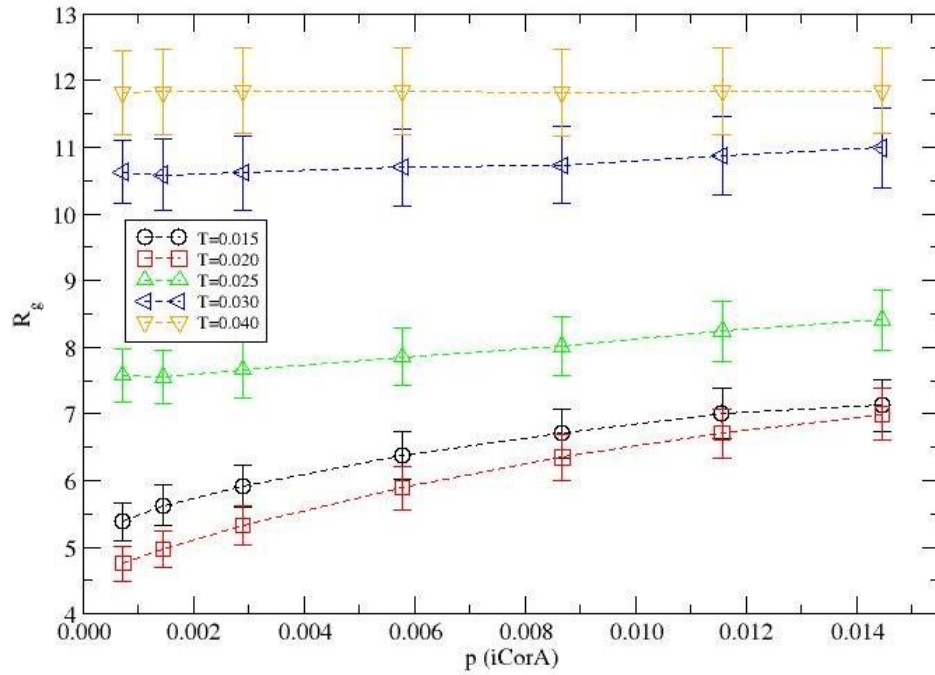


Figure III- 8: Variation of the average radius of gyration  $R_g$  of iCorA with the protein concentration  $p$  at temperature  $T = 0.015-0.040$ .

### 3.4.3 Structure factor

Overall distribution of residues over multiple length scales can be assessed by analyzing the structure factor  $S(q)$  which is defined as,

$$S(q) = \left\langle \frac{1}{N} \left| \sum_{j=1}^N e^{-i\vec{q} \cdot \vec{r}_j} \right|^2 \right\rangle_{|\vec{q}|} \quad \text{Equation 7 the structure factor}$$

where  $r_j$  is the position of each residue in all protein chains and  $|q| = 2\pi/\lambda$  is the wave vector of wavelength  $\lambda$ . Using a power-law scaling of the structure factor with the wave vector, i.e.,

*Equation 8 The relationship of structure factor and wave vector*

$$S(q) \propto q^{-1/\gamma}$$

one may study the spread of residues over the length scale  $\lambda$  by evaluating the exponent  $\gamma$  which describes the mass (residue) distribution. Since we know the

overall size of each protein chain via its radius of gyration (Figure III-7, -8), we can estimate the dependence of the number of residues on multiple length scales ( $\lambda$ ), spanning over the size of a residue to the simulation box. On the spatial length scale comparable to radius of gyration ( $\lambda \sim R_g$ ), we can estimate the scaling exponent  $\nu$  from the power-law,  $R_g \propto N^\nu$ , where  $N$  is the number of residues (a measure of the molecular weight of the protein); the effective (fractal) dimension ( $D$ ) of each protein  $D = 1/\nu$ ,  $\gamma = \nu$ . Similarly we can estimate the mass distribution of the protein self-assembly at  $\lambda > R_g$ .

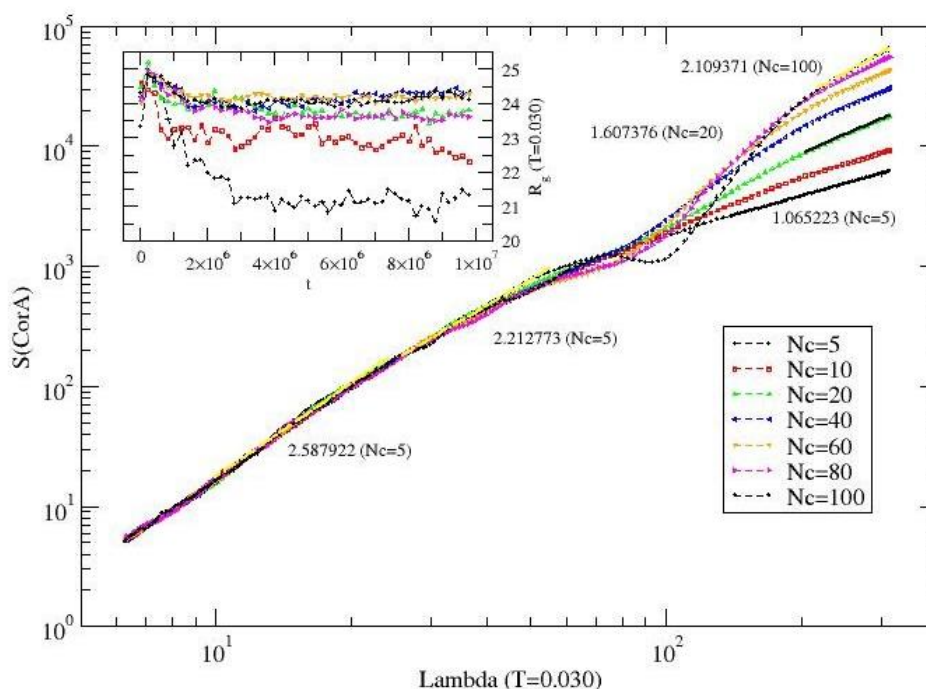


Figure III- 9: Structure factor  $S(q)$  versus the wave length (Lambda,  $\lambda$ ) at temperature  $T = 0.030$  with a wide range of the number  $N_c$  (5-100) of CorA proteins in simulation box.

Variation of the structure factor ( $S(q)$ ) with the wave length ( $\lambda$ ) of the self-assembly of the protein (CorA) is presented in Figure III-9 at temperature  $T = 0.030$  for a wide range of protein concentrations, dilute with  $N_c=5$  to crowded with  $N_c=100$ . Note that the radius of gyration has reached a steady-state value for all concentrations of the protein at  $T = 0.030$ . We see that  $D \sim 2.6$  for each protein ( $\lambda \sim R_g$ ),  $D \sim 2.2$  at  $R_g \leq \lambda \leq 50$  (in unit of lattice constant). For all residues distributed over the entire simulation

box,  $\lambda \sim L$ , the effective dimension decreases systematically from  $D \sim 2$  with  $N_c=100$  to  $D \sim 1$   $N_c=5$  where each protein chain appear to be isolated and very thinly distributed. At high temperature ( $T=0.040$ ), it is difficult to reach steady state during the course of simulations for the radius of gyration of CorA proteins at higher protein concentrations ( $p \geq 0.0014$ ,  $N_c \geq 20$ ) due to entanglements. Variation of the structure factor with the wave vector (see Figure III-S5) reveals a rather tenuous morphology ( $\lambda$  from size of a residue to about third of the simulation box,  $L/3$  much larger than  $R_g$  of protein chains) of the entangled fibrous protein chains with an effective dimension  $D \sim 1.7$ .

The structure factor of residues' distribution in a dilute-to-crowded protein environment ( $N_c=5-100$ ) of iCorA is presented in Figure III-10 at high temperature  $T = 0.040$ . Each protein chain appear to maintain globular conformation ( $D \geq 3$ ), at  $\lambda \leq R_g$ ; on larger scale  $\lambda > R_g$ , the density varies with length scale, i.e.  $D \sim 2.6, 3.4$  with increasing length scale up to about half the size of the simulation box ( $\lambda = L/2$ ). The overall morphology ( $\lambda \sim L$ ), seem to become more tenuous on reducing the number of protein chains;  $D \sim 1.6$  with  $N_c = 100$ . Note that the radius of gyration of the protein increases on increasing the number  $N_c$  of protein chains (inset Figure III-10).

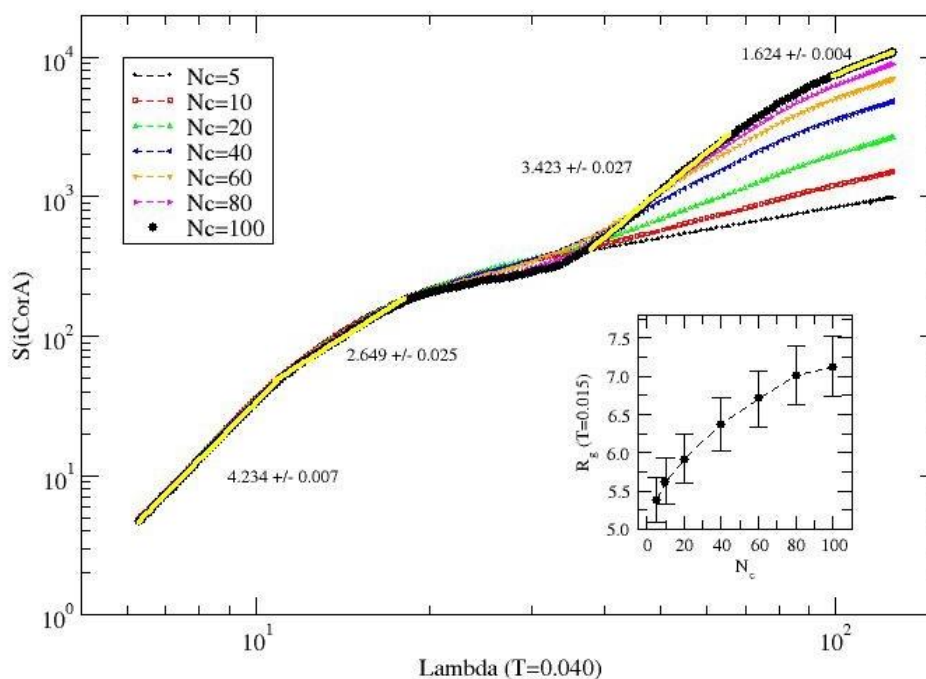


Figure III- 10: Structure factor  $S(q)$  versus the wave length ( $\Lambda$ ,  $\lambda$ ) at temperature  $T = 0.040$  with a wide range of the number  $N_c$  (5-100) of iCorA proteins in simulation box. The inset figure shows the variation of the radius of gyration of the protein with the number of protein chains.

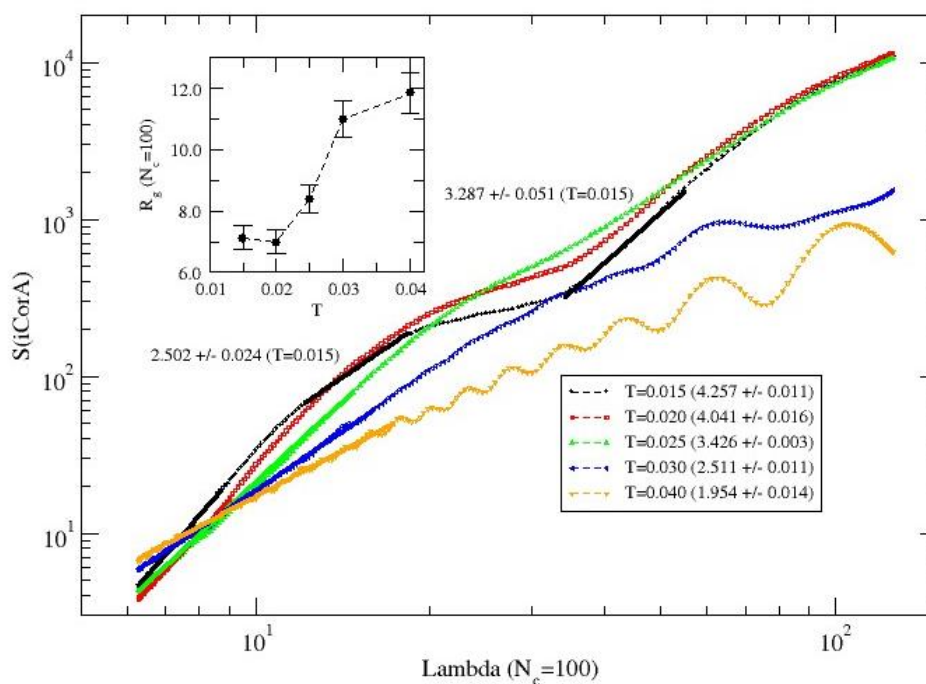


Figure III-11: Structure factor  $S(q)$  versus the wave length ( $\Lambda$ ,  $\lambda$ ) in a crowded protein environment with  $N_c=100$  at a range of low to high temperatures  $T=0.015 - 0.040$  with the estimates of slopes in brackets; slopes of larger scale data at  $T = 0.015$  are also included. Inset is the variation of the radius of gyration of iCorA with the temperature with number  $N_c = 100$  proteins in simulation box.

The variation of the structure factor with the wavelength in the most crowded environment ( $N_c=100$ ) is presented in Figure III-11 for a range of temperatures. The globular morphology ( $D \geq 3$ ) of individual protein chain seems to persist at lower temperatures ( $T=0.015, 0.025$ ) at length scale comparable to its size ( $\lambda \sim R_g$ ). On a larger-scale ( $R_g < \lambda < 3 R_g$ ), the assembled structure is not as compact ( $D \sim 2.5$ ), but the overall assembly remains compact (solid with  $D \sim 3$ ) for its spread over the simulation box at the low temperature ( $T=0.015$ ). The multi-scale morphology of self-organizing proteins is heterogenous regardless of its globular conformations in its

native phase. The morphology remains relatively dense  $D \sim 2.5$  even at a much higher temperature, i.e.  $T=0.030$ . On large length scale, ( $\lambda \sim L$ ), the self-assembled morphology appears tenuous. The morphology adopts to a random coil  $D \sim 1.8$  at high temperature  $T = 0.040$  where onset of a regular long range structure appear to sets in (with an oscillation in  $S(q)$ ).

### 3.5 Conclusions

Self-organized structures of interacting proteins (CorA) in crowded environments are investigated by a coarse-grained Monte Carlo simulation as a function of protein concentration at a range of low to high temperatures. Since the physical properties of protein (CorA) is overwhelmingly dominated by the outer membrane component (oCorA:  $^1M - ^{290}V$ ), it is difficult to assess the relative effect of the inner core (iCorA:  $^{291}M \dots ^{351}L$  which is about 5 fold smaller) in self-assembly. Therefore, we have examined the structural evolution with iCorA in addition to identify distinctions if any. Visual inspection show clear distinctions in morphology of the assembly in dilute solution and that in the crowded (dense) matrix both at low and high temperatures. CorA proteins seem to dissociate more on the scale comparable to the size of the proteins while the clusters of inner core iCorA protein phase-separate in its native phase (low temperatures). It appears that the protein-protein segmental interaction is more conducive to agglomeration of iCorA than that of CorA.

In order to identify the role of iCorA quantitatively, we have analyzed a number of local (contact maps, and average contact profile, mobility profile) and global (rms displacement of the center of mass of the protein, radius of gyration, structure factor, etc.) physical properties. Variation in pattern of the contact profiles with the temperature reveals how the protein segments organize during the self-assembly that leads to a global morphology. We find that a relatively lower fraction of residues of oCorA (residues  $^1M - ^{290}V$ ) participate in segmental globularization (with a random distribution) in comparison to that of the inner core segment ( $^{291}M \dots ^{351}L$ ) with well-defined globular region. Response of the global physical quantities such as the variation of the rms displacement of each protein chain with the time steps and that of its radius of gyration at a range of temperatures help in gaining further insight into the

response of protein chains as a function of temperature and protein concentration. For example, the radius of gyration of the protein CorA does not show much variation with the protein concentration at low temperatures. The radius of gyration of iCorA on the other hand increases with the protein concentration  $p$  systematically at low temperatures ( $T=0.015, 0.020$ ) in its native phase. We therefore believe that the protein-protein interaction among the inner-segments of the protein in its native phase may be conducive to self-organized pathways.

The spread of multi-scale morphology is also quantified by an effective dimension  $D$  from the scaling analysis of the structure factor with the wave vector. In general, the effective dimension of the assembled morphology of CorA is much lower than that of iCorA. For example, the most compact morphology of CorA with  $D \sim 2.6$  for each protein ( $\lambda \sim R_g$ ) even at the highest protein concentrations at low temperature while remain almost extended in a random configurations with  $D \leq 2$ . In contrast, the self-assembled morphology of iCorA remains globular  $D \sim 3$  at almost all length scales in its native phase. Because of the complexity of the crowded proteins (a diverse range of inter- and intra-chain residue-residue interactions and competition between the self-organizing mechanism and steric constraints and entanglement) it is difficult to identify main cause of a stable pentamer configuration. However, based on the analysis presented in this investigation, we believe that a relatively stronger protein-protein interactions among the inner segments (iCorA) of the protein is conducting to its stable self-assembly in addition to steric constraints of the membrane and interaction with other matrix constituents.

## CHAPTER IV

---

### **Pinning the conformation of a protein (CorA) in a solute matrix with selective binding**

Warin Jetsadawisut<sup>1</sup>, Sunan Kitjaruwankul<sup>2</sup>, Pornthep Sompornpisut<sup>1</sup>, R.B. Pandey<sup>3</sup>

<sup>1</sup>*Center of Excellence in Computational Chemistry, Department of Chemistry,  
Chulalongkorn University, Bangkok 10330, Thailand*

<sup>2</sup>*Faculty of Science at Sriracha, Kasetsart University Sriracha Campus, Chonburi  
20230, Thailand*

<sup>3</sup>*School of Mathematics and Natural Sciences, University of Southern Mississippi,  
Hattiesburg, MS, 39406, USA*

Doi: 10.1016/j.physa.2020.124823

Journal: Physica A: Statistical Mechanics and its Applications, 2020, 556: 124823.

CHULALONGKORN UNIVERSITY

Received: 2 October 2019

Received in revised: 13 May 2020

Published: 12 June 2020

---

**Keywords:** proteins CorA; protein folding; Coarse-grained model; Monte Carlo simulation; Interacting solute matrix

#### 4.1 Abstract

Conformation of a protein (CorA) is examined in a matrix with mobile solute constituents as a function of solute–residue interaction strength ( $f$ ) by a coarse-grained model with a Monte Carlo simulation. Solute particles are found to reach their targeted residue due to their unique interactions with the residues. Degree of slowing down of the protein depends on the interaction strength  $f$ . Unlike a predictable dependence of the radius of gyration of the same protein on interaction in an effective medium, it does not show a systematic dependence on interaction due to pinning caused by the solute binding. Spread of the protein chain is quantified by estimating its effective dimension ( $D$ ) from scaling of the structure factor. Even with a lower solute–residue interaction, the protein chain appears to conform to a random-coil conformation ( $D \sim 2$ ) in its native phase where it is globular in absence of such solute environment. The structural spread at small length scale differs from that at large scale in presence of stronger interactions:  $D \sim 2.3$  at smaller length scale and  $D \sim 1.4$  on larger scale with  $f = 3.5$  while  $D \sim 1.4$  at smaller length scale and  $D \sim 2.5$  at larger length scales with  $f = 4.0$ .

#### 4.2 Introduction

Understanding the structure of proteins has been a subject of intense interest for decades with an exponential growth in number of publications particularly using Molecular Dynamics simulations. Despite enormous efforts, predicting the nature of ‘protein folding’ remains a challenging problem [90]. Addressing how proteins reach its equilibrium configurations leads to more questions than answers. For example, what is an equilibrium configuration (minimization of global energy, approaching a steady-state radius of gyration, etc.)? Is equilibrium configuration necessary for a protein to perform its functions? The complexity in predicting the structural properties arises due to competition between the interaction energy and thermal agitation.

The characteristics of an interacting matrix in which the protein is embedded affects the dynamics and structure of the protein. For example, Duncan et al. [91] have recently reported sub-diffusion of proteins and lipids and effects of crowding on their



sub-diffusive dynamics in general. Zeindlhofer and Schroeder [92] have reviewed the analysis of biomolecules in aqueous ionic liquid mixtures where they have pointed out that the viscosity of the imbedding matrix affect the structure of the protein. They have cautioned that such effects may be specific to individual protein and therefore should be tested before generalizing the main observation. In this article we focus on a specific protein (CorA) embedded in an interaction-specific matrix with mobile constitutive elements called solutes.

The transmembrane protein CorA consists of 351 residues with a well-defined inner (iCorA) and outer (oCorA) segments in cells [10-12, 15, 77, 78, 81]. Numerous elements in a cell such as lipid molecules, ions, osmolytes, different types of proteins, etc. constitute the crowded environment which delicately control the structure of CorA for selective transport of magnesium ions across the membrane in open and closed states of the ion-channel. Obviously, it is not feasible to incorporate each constitutive elements at once in order to assess its specific effects on different segments of CorA protein. However, it is possible to represent the solutes by simplified particles with appropriate interactions, suitable to capture their specificity. The stochastic movements of each residue (leading to a collective movement of the protein chain) in an effective dynamic matrix environment may provide chances for each residue to interact with other residues of the protein chain. Attractive interactions between some residues (e.g. hydrophilic and electrostatic) and solute may lead to their transient binding while repulsive interaction between other residues may enhance the self-organizing segmental structures (globular or fibrous). Note that the strength of both attractive and repulsive interactions vary from one residue to another. The interplay between the solute-residue interaction and thermal agitation may lead to unique structural evolution. This article is focused on the effect of the strength of interaction on the selective binding of residues with the mobile solute particles and the structural evolution of CorA at a low (native phase) and a high (denatured phase) temperature [93].

It is worth pointing out that we have already examined the structure and dynamics of CorA in absence of environmental complexity and found interesting thermal response of inner and outer segments of the protein [81, 93]. For example, the thermal response

of the inner segment shows a continuous transition from globular to random-coil structure on raising the temperature while the outer segment exhibits an abrupt (nearly discontinuous) thermal response in a narrow range of temperature. Unlike in denatured phase, the conformation of the inner segment contracts on raising the temperature in its native phase where the outer segment appears less organized [93]. In an implicit effective medium, the size of the inner segment of the protein decreases in native phase and increases in denatured phase before reaching saturation with the residue-matrix interaction strength; the outer segment shows opposite response to effective medium [94]. In presence of explicit effective solute matrix the protein structure is pinned without a systematic trend with the interaction strength due to rapid binding of solutes with the selective residues as discussed in this article. The model is introduced in the next section followed by results and discussion and a concluding remark.

### 4.3 Methodology

A bond-fluctuation mechanism [95] is used to model the protein [81, 93, 94], a chain of 351 nodes, each representing unique specificity of corresponding residue. The simulation is performed on a cubic lattice with ample degrees of freedom for each residue to perform its stochastic moves and corresponding peptide bonds to fluctuate. Simulation box consists of a protein chain and a large number of solute particles (with a volume fraction  $c=0.002$ ) where the size of a solute is the same as that of a residue [96]. Initially the protein chain is placed in a random configuration and the explicit solute particles are distributed randomly. Each residue and solute molecule interacts with surrounding residues and solute molecules within a range ( $r_c$ ) with a generalized Lennard-Jones potential,

$$U_{ij} = \left[ \epsilon_{ij} \left( \frac{\sigma}{r_{ij}} \right)^{12} + \epsilon_{ij} \left( \frac{\sigma}{r_{ij}} \right)^6 \right], r_{ij} < r_c \quad (1)$$

where  $r_{ij}$  is the distance between the residues at site  $i$  and  $j$  or between the residue at site  $i$  and residue at site  $j$ ;  $r_c = \sqrt[6]{8}$  and  $\sigma = 1$  in units of lattice constant. The range of interaction includes ample number of lattice sites that can be occupied by solute

molecules or residues of the protein. The degrees of freedom can be further enhanced with finer-grain representations of each residue. A knowledge-based interaction matrix [84, 85, 96] is used for the residue-residue pair interaction ( $\varepsilon_{ij}$ ), which is derived from an ensemble of a large number of protein structures from the protein data bank (PDB). The strength  $\varepsilon_{ij}$  of the potential (210 independent elements of a  $20 \times 20$  matrix for 20 amino acids) is unique for each interaction pair with appropriate positive (repulsive) and negative (attractive) values [84, 85].

A solute-solute and solute-residue interactions require 21 unique parameters [96] to capture its specificity in an aquatic solute environment, as considered here for simplicity. The solute at a site ( $i$ ) interacts with a residue at another site ( $j$ ) with a similar interaction as (1) but with  $\varepsilon_{ij}=f \varepsilon_r A_{h/p/e}$  which is based on the hydrophathy index ( $\varepsilon_r$ ) that controls the attractive and repulsive nature of the residue towards the solute. The hydrophathy index is binned into a repulsive interaction ( $\varepsilon_r = 0.1$ ) for all hydrophobic ( $H$ ) residues, attractive interaction ( $\varepsilon_r = -0.2$ ) for all polar ( $P$ ) residues, and more attractive ( $\varepsilon_r = -0.3$ ) for all electrostatic ( $E$ ) residues. The magnitudes of the hydrophathy index  $\varepsilon_r$  are further weighted with a factor  $A_{h/p/e}$  to capture the specificity of each residue within each group ( $H, P, E$ ). The empirical parameter  $f$  is introduced to modulates the interaction strength. For example, the interaction of a solute with a hydrophobic residue such as valine (V),  $\varepsilon_{ij}=f(0.1)0.933$ , the interaction with a hydrophilic residue, say Tryptophan (W)  $\varepsilon_{ij}=f(-0.2)0.257$ , and the interaction with an electrostatic residue such as lysine (K)  $\varepsilon_{ij}=f(-0.3)0.867$ . The solute-solute interaction is ignored ( $\varepsilon_{ij}=0$ ) apart from their excluded volume effect. In the figures and text the interaction strength  $f$  and  $f_w$  are used interchangeably.

Each residue and solute particle executes their stochastic motion with the Metropolis algorithm. Attempt to move each residue and solute particle once defines unit Monte Carlo step time. Simulations are carried out for sufficiently long time to generate conformational ensembles in steady-state at a low and high temperature regimes each with 5-10 independent samples to analyze a number of local and global physical quantities. Interaction strength  $f$  is varied. Different lattice sizes are used to make sure that the qualitative trends are independent of the sample size. The results presented

here are based on data generated on a  $350^3$  sample which provides ample sampling at long-time scales without using excessive computer resources. Reduced units are used for temperature, time step, and spatial length scales in this simulation since our focus is on changes in physical quantities in response to changing the solute interaction strength ( $f$ ) that affect the preferential binding.

#### 4.4 Results and discussion

Solute particles interact with each residues with unique attractive and repulsive interactions controlled by the interaction strength ( $f$ ). Each residue and solute particle perform their stochastic motion. Solute particles are generally more mobile (at least initially) than the residues as their mobility is constrained by the peptide (covalent) bonds. Thus the probability for the solute particle to reach attractive residue and stay there within the range of interaction of the target site for a longer time is higher for higher  $f$ . Representative snapshots of protein and the solute particles that bind to specific residues are presented in Figure IV-1. A set of representative animations (see the supplement) shows the stochastic dynamics of protein segments, the progression of solute binding to specific sites, and pinning down its movements. We see that the number of solute particles that bind to residue increases with the solute-residue interaction strength ( $f$ ). For example, at the low temperature ( $T=0.020$ ), the number of binding sites is relatively smaller at lower solute-residue interactions ( $f = 2.5$ ) than that with the stronger interactions ( $f=3.5, 4.5$ ). Similar patterns appear at the high temperature (i.e.  $T=0.032, f = 4.5$ ). Further, there is no appreciable change in size of the protein except some variations in segmental organization. It is therefore difficult to quantify overall changes in size of the protein in response to solute-residue interactions.

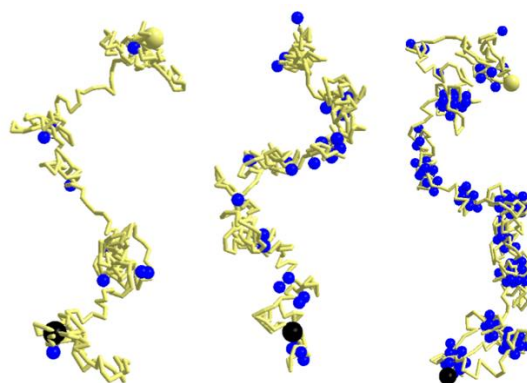


Figure IV- 1: Snapshots of the protein chain CorA at the end of a million time step. Bonds along the backbone of the protein contour are in yellow where the first residue is represented by the large black and the last residue by yellow sphere; solute particles within the range of interaction of each residue are shown in blue dots. The interaction strength  $f = 2.0, 2.5,$  and  $3.5$  from left to right at  $T = 0.020$ .

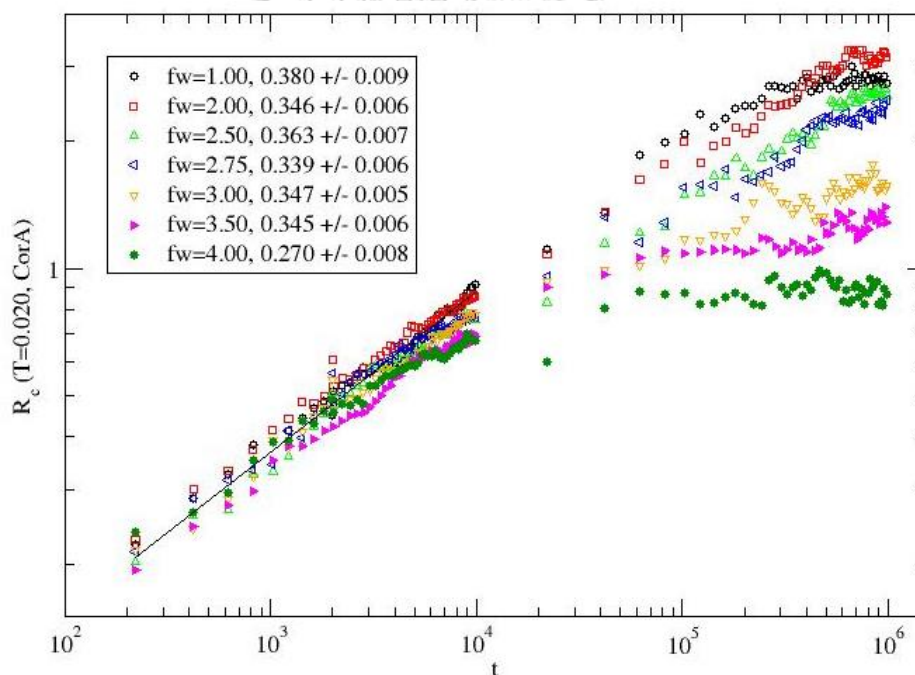


Figure IV- 2: Root mean square (RMS) displacement ( $R_c$ ) of the center of mass of the protein chain with the time step ( $t$ ) for a range of solute-residue interaction  $f = 1.00 - 4.00$  at a low temperature  $T = 0.020$ . The power-law exponent  $\nu$  ( $R_c \propto t^\nu$ ) in the short time regime ( $t \sim 10^2 - 10^4$ ) is also included for each interaction strength.

How does the protein chain move as the residues attempt to perform their stochastic motion? One may be able to assess the protein dynamics by examining the variation of the root mean square (RMS) displacement ( $R_c$ ) of the center of mass of the protein chain with the time step as shown in Figure IV-2. As the protein chain moves, its RMS displacement increases with the time step. Note that the movement is faster in short time (initially), followed by a slowdown asymptotically in long time regime. Asymptotic slowdown depends on the solute-residue interaction strength ( $f$ ); the higher the interaction, faster is the approach from very slow movement ( $f = 1.00 - 2.75$ ) to almost standstill ( $f = 3.00 - 4.00$ ). The dynamics of the protein chain in short time regime can be assessed by a power law dependence of its RMS displacement, i.e.,  $R_c \propto t^\nu$ , where the exponent  $\nu$  characterizes the nature of dynamics. We see that the dynamics of the protein is sub-diffusive  $\nu < 1/2$  for entire range of solute-residue interaction in the short time regime. In asymptotic regime the dynamics of the protein is not only sub-diffusive with low interactions but almost vanishes ( $\nu \sim 0$ ) with stronger interactions ( $f = 3.00 - 4.00$ ). Slowing down of the protein chain occurs as the solute particles bind to their target residues; the overall motion ceases as the conformation of the protein is pinned down by ample binding as seen in Figure IV-1.



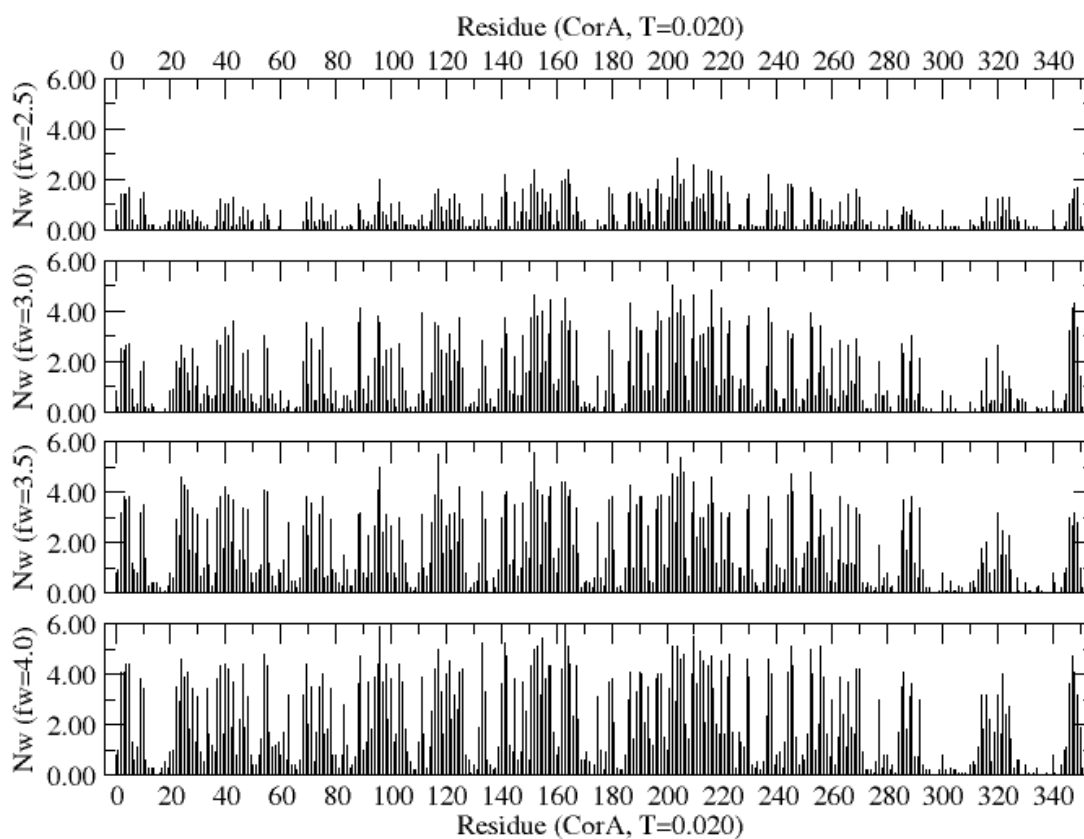


Figure IV- 3: Binding profile (average number  $N_w$  of solute around each residues) of CorA for solute-residue interaction  $f=2.5, 3.0, 3.5, 4.0$  at a temperature  $T = 0.020$ .

In order to assess the extent of binding and identify the targeted residues, let us examine the binding profile, i.e. the average number ( $N_w$ ) of solute particles around each residue within the range of interaction. Figure IV-3 shows the average number of solute particles around each residue for a range of solute-residue interaction strength  $f=2.5, 3.0, 3.5, 4.0$  at a temperature  $T = 0.020$ . Outer and inner transmembrane segments of the protein CorA consists of  $^1\text{M} \ ^2\text{E} \dots \ ^{291}\text{M}$  and  $^{292}\text{K} \ ^{293}\text{V} \dots \ ^{351}\text{L}$ . The extent of binding is enhanced with stronger solute-residue interaction. A large fraction of outer segment of CorA residues are pinned by high degree of solute binding specially at stronger interactions ( $f = 4.00$ ) while a considerable fraction (about half) of inner transmembrane segments remain unbounded (i.e.  $^{292}\text{K} \ ^{293}\text{V} \dots \ ^{309}\text{G}, \ ^{330}\text{V} \ ^{331}\text{L} \dots \ ^{343}\text{V}$ ). The inner segment of the protein CorA is more mobile than that of the outer transmembrane segment and perhaps more responsive in self-assembly as recently observed [97].

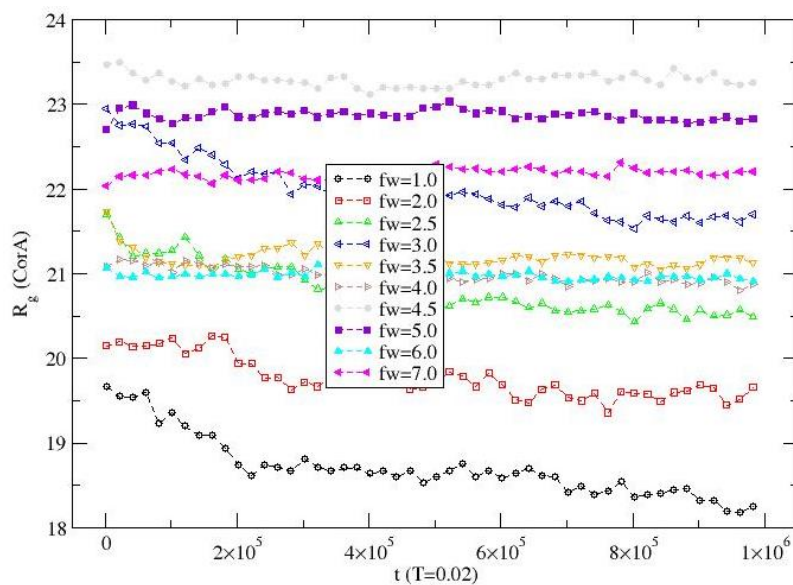


Figure IV- 4: Radius of gyration of CorA versus time step for solute-residue interaction  $f=1.0 - 7.0$  at  $T = 0.020$ .

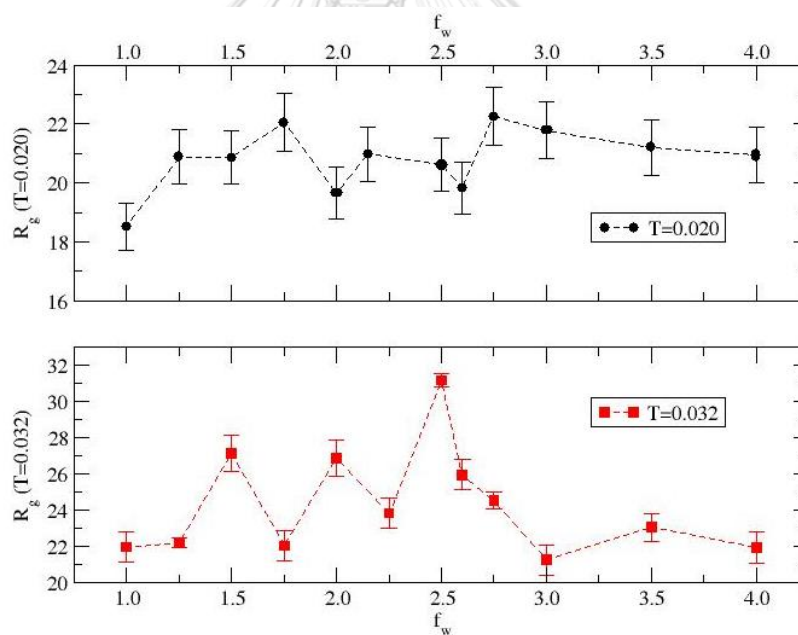


Figure IV- 5: Radius of gyration of the protein (CorA) versus solute-residue interaction strength  $f \equiv f_w$  at temperature  $T = 0.020, 0.032$ .

Because of the selective binding of solute particles, the protein conformation is arrested in a relatively short time step. Figure IV-4 shows the variation of the radius of gyration with the time step for a wide range of solute-residue interaction ( $f=1.0 - 7.0$ ) at a low temperature. We see that the radius of gyration becomes reaches its steady-state rather fast particularly at higher values of  $f$  and that the binding enhances



the stabilization. The variation of the average radius of gyration with the interaction strength (Figure IV-5) shows almost no systematic dependence on the interaction at temperatures  $T = 0.020, 0.032$  despite large fluctuations. Lack of a trend is due to pinning of the conformations of the protein as the solute particles move fast and bind the targeted residue. Note that lack of a systematic dependence on the interaction with the underlying matrix due to pinning is different from that of the protein in an effective medium [33] where the inner transmembrane segment exhibits a systematic dependence in both native and denatured phases. The x, y, and z-components of the radius of gyration appears to be of the same order of magnitude within the range of fluctuations (Figure IV-S1). It should be pointed out that the radius of gyration of the inner segment (iCorA) responds systematically to solute-residue interaction strength  $f$ , unlike the outer segment (oCorA). For example, the radius of gyration of iCorA decays continuously with the interaction strength  $f$  while that of the oCorA fluctuates wildly at temperature  $T = 0.032$  (see Figure IV-S2).

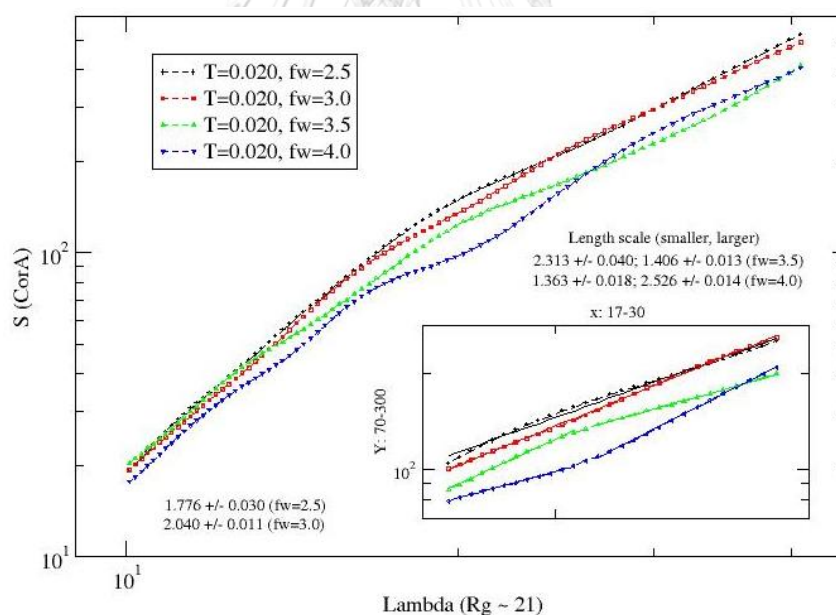


Figure IV- 6: Structure factor ( $S$ ) of the protein versus wave length ( $\Lambda$ ) for the interaction strength  $fw = 2.5, 3.0, 3.5,$  and  $4.0$  at  $T = 0.020$ . The inset is amplified version within a limited wavelength range spanning over the magnitude of the radius of gyration ( $R_g \sim 21$ ). The slope ( $1/\gamma = D$ ) is a measure of the effective dimension ( $D$ ) of the protein which is about 2 with lower values of  $fw = 2.5, 3.0$  but varies with length scales at higher values of  $fw = 3.5, 4.0$ .

To quantify the conformational spread of the protein, we analyze the structure factor  $S(q)$ ,

$$S(q) = \left\langle \frac{1}{N} \left| \sum_{j=1}^N e^{-i\vec{q}\cdot\vec{r}_j} \right|^2 \right\rangle_{|\vec{q}|} \quad (2)$$

where  $r_j$  is the position of each residue in all protein chains and  $|q| = 2\pi/\lambda$  is the wave vector of wavelength  $\lambda$ . Using a power-law scaling of the structure factor with the wave vector, i.e.,

$$S(q) \propto q^{-1/\gamma} \quad (3)$$

one may study the spread of residues over the length scale  $\lambda$  by evaluating the exponent  $\gamma$  which describes the mass (residue) distribution. The slope  $D = 1/\gamma$  of the data set spanning over the length scale comparable to radius of gyration is a measure of the effective dimension ( $D$ ) of the protein. At lower values of solute-residue interaction  $f = 2.5, 3.0$ , the conformation of the protein chain is like a random-coil ( $D \sim 2$ ) at  $T=0.020$  where it conform to a globular structure in absence of solute. Effective dimension of the protein depends on the length with stronger solute-residue interaction. For example,  $D \sim 2.3$  at smaller length scale and  $D \sim 1.4$  (very linear structure) on larger scale with  $f = 3.5$ . Reverse is the case with stronger interaction  $f = 4.0$ , where  $D \sim 1.4$  at smaller length scale and  $D \sim 2.5$  at larger length scales. Structural variability suggests that a wide range of fluctuating conformations can be pinned by appropriate solute.

#### 4.5 Conclusions

Conformational dynamics of a protein (CorA) in an interacting matrix with mobile solute particles is examined as a function of solute-residue interaction by a coarse-grained Monte Carlo simulation. Initially the protein chain is placed in the simulation box in a random configuration in presence of a random distribution of solute particles. Each residue and solute particle perform their stochastic movement. Because of the higher mobility of solute particles and specificity of the solute-residue interactions, they reach their specific target rather fast. As a result the conformation of the protein

is constrained by the solute particles that bind to specific protein sites and pinned with stronger solute-residue interactions.

The protein chain exhibits different dynamics in short and long time regime: the short time dynamics is sub-diffusive and not very sensitive to solute-residue interaction strength ( $f$ ), the long time (asymptotic) dynamics depends strongly on the interaction strength. The protein chain becomes almost immobile as solute particles bind to their target residues at higher solute-residue interactions. Lack of a systematic trend in conformational response to solute-residue interaction seems to occur due to pinning down a vast number of specific residues, particularly in the outer transmembrane segment of the protein. This observation is very different from that of the same protein chain in an effective medium [94] where there is a systematic dependence of the protein conformation on the residue-matrix interaction.

Spread of the protein chain can be quantified by estimating its effective dimension from scaling of the structure factor. In a relatively lower solute-residue interaction, the protein chain appears to conform to a random-coil conformation at a low temperature where it is globular in absence of such solute environment; this is obviously due to some degree of pinning of the protein conformation. Presence of stronger interacting solute leads to higher degree of pinning down the conformations that shows different spread at lower and larger length scales. For example,  $D \sim 2.3$  at smaller length scale and  $D \sim 1.4$  (very linear fibrous structure) on larger scale with the solute-residue interaction strength  $f = 3.5$ . Reverse is the case on increasing the interaction strength  $f = 4.0$ , where  $D \sim 1.4$  at smaller length scale and  $D \sim 2.5$  at larger length scales. Thus by selecting interacting matrix with specific solute-residue interaction, one may be able to achieve a desirable conformation of the protein.

#### 4.6 Supporting information

Representative short animations (presented below) show the stochastic dynamics of protein segments, the progression of solute binding to specific sites, and pinning down its movements. For example, at the low temperature ( $T=0.020$ ), the number of binding sites is relatively smaller at lower solute-residue interactions ( $f = 2.5$ ) than that with

the stronger interactions ( $f=3.5, 4.5$ ). Similar patterns appear at the high temperature (i.e.  $T=0.032, f=4.5$ ).

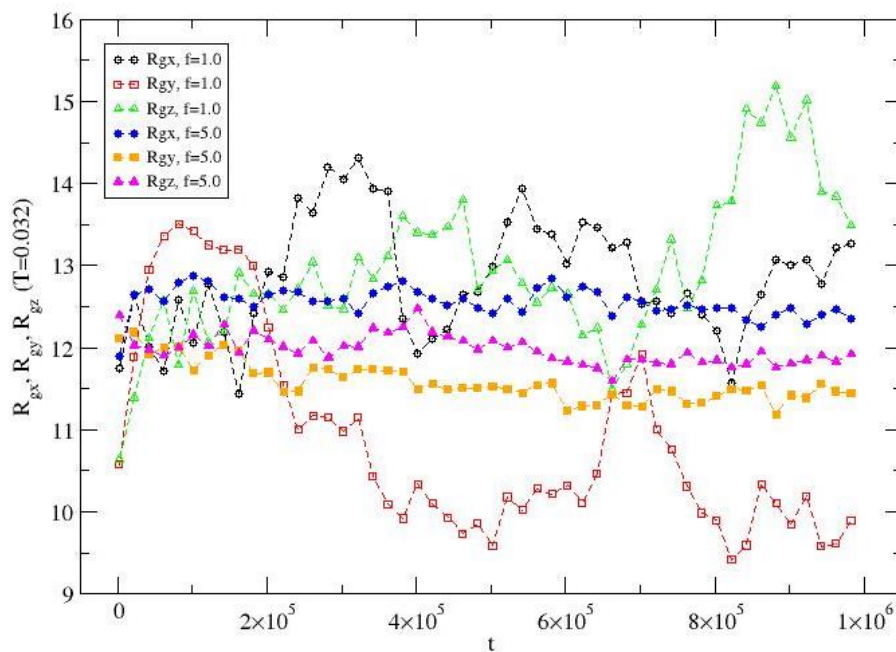


Figure IV-S1: Variation of the x, y, z components ( $R_{gx}$ ,  $R_{gy}$ ,  $R_{gz}$ ) of the radius of gyration with the time step for representative interaction strength  $f=1.0, 5.0$  at  $T=0.032$ .

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

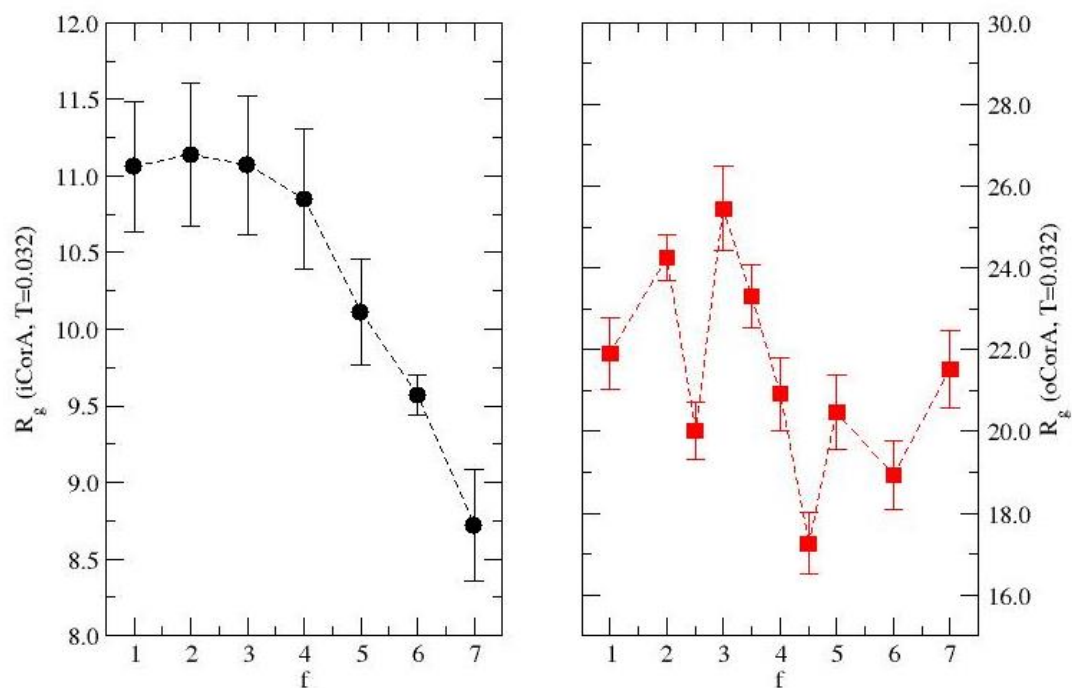
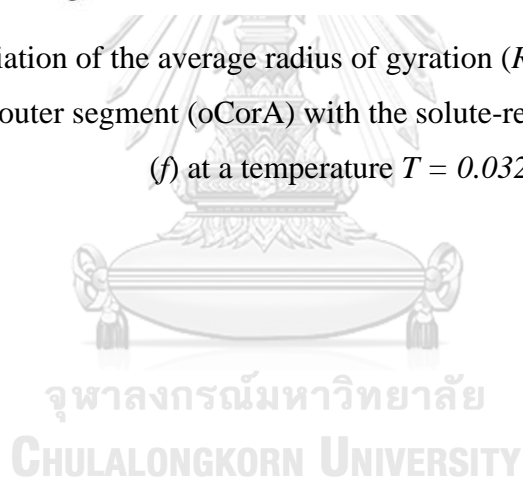


Figure S2: Variation of the average radius of gyration ( $R_g$ ) of the inner segment (iCorA) and outer segment (oCorA) with the solute-residue interaction strength ( $f$ ) at a temperature  $T = 0.032$ .



## CHAPTER V

---

### **Magnesium release-induced conformational changes of CorA magnesium channel**

*Warin Rangubpit<sup>†</sup>, Ras B. Pandey<sup>‡</sup>, and Pornthep Sompornpisut<sup>\*†</sup>*

*<sup>†</sup>Center of Excellence in Computational Chemistry, Department of Chemistry,  
Chulalongkorn University, Bangkok 10330, Thailand*

*<sup>‡</sup>School of Mathematics and Natural Sciences, University of Southern Mississippi,  
Hattiesburg, MS, 39406, USA*

(Publication has been submitted)

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

---

**Keywords:** Molecular Dynamics simulation; Coarse-grained model; CorA Magnesium channels; Divalent cation sensors

## 5.1 Abstract

CorA is prokaryotic  $Mg^{2+}$ -selective channels that mediate the ion flux across cell membranes. The CorA channel is a homopentamer possessing five magnesium binding sites, each of which is located at the intersubunit interface of the cytosolic domain. Earlier studies suggests that the protein conformation with fully Mg-occupied sites remains stable in the closed state while the channel is activated in the absence of intracellular  $Mg^{2+}$ . There has been a debate whether the loss of  $Mg^{2+}$  in the binding sites causes significant conformational changes from the closed-bound to open-unbound states. In this study, we have used coarse-grained MD simulations to examine the conformational response of the CorA closed state upon removal of  $Mg^{2+}$  restraints introduced as artificial bonds in coordination complex. The simulations with  $Mg^{2+}$  restraints that mimicked the attractive interactions between the protomer were carried out using two restraint removal approaches, simultaneous versus stepwise removals. The results of both approaches showed that the release of  $Mg^{2+}$  has induced a significant conformational rearrangement of the cytosolic domain. However, the movement was not large enough compared with the cryo-EM study. We found that the pentameric arrangement of the residues in the hinge-bending region obtained from the stepwise approach is found to be more consistent with experiments the, suggesting a possible mechanism for the  $Mg^{2+}$  release.

## 5.2 Introduction

CorA is  $Mg^{2+}$ -dependent gating channels that are essential for maintaining  $Mg^{2+}$  homeostasis within the cell [12]. The CorA gene was the first prokaryotic  $Mg^{2+}$  uptake system identified and recognized as the cobalt resistance phenotype in E.coli [98]. CorA is primarily found in bacteria [99]. Other known  $Mg^{2+}$  transport proteins in bacteria include MgtA, MgtB and MgtE [74, 99, 100]. CorA is a member of superfamily of  $Mg^{2+}$  transport proteins which has a functional homolog [101] similar to eukaryotic Mrs2 and Alr1 proteins [102, 103]. Although the CorA, Mrs2 and Alr1 superfamily exhibits low sequence similarity, they share significant conserved GMN motif [103]. The crystal [10, 11] and cryo-EM structures [12] of CorA determined under  $Mg^{2+}$ -bound conditions revealed a symmetric arrangement of pentamer with a

large funnel-shaped structure in the cytosol [11]. Each monomer is composed of a large N-terminal cytosolic domain, a stalk helix and two transmembrane segments. The N-terminal cytosolic domain consists of two sets of three  $\alpha$ -helices and seven parallel and antiparallel  $\beta$ -strands folded as an  $\alpha\beta\alpha$  sandwich-like structure.

Crystal structures of  $Mg^{2+}$ -bound CorA reveals five  $Mg^{2+}$ -binding sites which are located at the interface between subunits in the cytosolic domain [12-14]. Each binding site which can host up to two  $Mg^{2+}$  (known as M1 and M2 sites), is called divalent cation sensor, serving as  $Mg^{2+}$ -sensing regulatory [14, 15]. The M1 and M2 metal binding sites are separated by a distance of  $\sim 7\text{\AA}$ .  $Mg^{2+}$  in the M1 site positions in close proximity to form coordinate covalent bonds with the carboxylate group of D89 and D253' ("'" indicates the adjacent subunit). The second M2 binding site lies close to the side chain of E88, D175 and D253. Electrophysiological studies showed that the channel activity is regulated by intracellular  $Mg^{2+}$  [15]. The protein conformation with fully Mg-occupied sites corresponds to the closed state while the channel transitions from closed to open conformations by the release of  $Mg^{2+}$  from the cytosolic binding sites [12].

The  $Mg^{2+}$ -bound and -unbound CorA forms depends on the cytosolic concentration of  $Mg^{2+}$ . At high intracellular  $Mg^{2+}$  concentration, all binding sites are occupied by  $Mg^{2+}$ , and therefore the non-conductive ( $Mg^{2+}$  bound) CorA is more favorable. However, under an insufficiency of the divalent cation in the cell, there is an increase in the unoccupied sites due to releasing  $Mg^{2+}$  from the cytosolic binding sites. The release of  $Mg^{2+}$  induces conformational changes in the cytosolic domain, stalk helices and the pore which subsequently shift the population toward the open or conductive state [12]. The conformational differences between the closed and open states can be illustrated by the crystal and cryo-EM structures obtained in the  $Mg^{2+}$  bound (closed) and  $Mg^{2+}$  free (open) form, respectively [12]. A structural comparison revealed a break in the symmetry of the pentamer structure of the closed conformation.

Binding of  $Mg^{2+}$  to acidic residues located at the interface helps to hold protein protomers together. The release of  $Mg^{2+}$  leads to a loss of the interactions in this region, triggering conformational changes of the protein. However, there has been a



debate whether the loss of  $Mg^{2+}$  in the binding sites causes significant conformational changes from the closed-bound to open-unbound states. The crystal structures of two CorA mutants eliminating interactions of  $Mg^{2+}$  with D89 and D253 at M1 sites show no significant changes in the structure with respect to the structure in the closed state. In this study, we employed coarse-grained MD (CGMD) simulations to investigate the conformational response of the CorA closed state upon the release of  $Mg^{2+}$  from the binding sites. The closed-bound conformation was used the starting state. The simulations were carried out by incorporating the intersubunit distance restraints that mimicked the attractive interactions between  $Mg^{2+}$  and acidic residues of the two adjacent subunits. On the basis of  $Mg^{2+}$ -release, these restraints were subjected to be removed during the simulations. To observe the conformational changes in response to the  $Mg^{2+}$ -released restraints, the simulations were conducted under two approaches, simultaneous versus stepwise removal. The resulting structures were validated based on the available cryo-EM structure in the  $Mg^{2+}$  free condition.

### 5.3 Methodology

#### 5.3.1 All-atom and coarse-grained models

Initial structure was taken from the cryo-EM structure of *Thermotoga maritima* CorA (*TmCorA*) in the closed state with the resolution of 3.8Å (PDB code 3JCF) [12]. In this study, we performed all-atom (AA) and coarse-grained (CG) MD simulations of the CorA in phospholipid bilayer, starting from the structure in the closed state. The structure coordinates from the PDB file contain five protein chains subunits (chain A, B C, D and E). Each chain consists of residues from 3 to 351 with no missing loop. Before converting all-atom to coarse grained model, AAMD systems were first built according to the protocol described in our previous studies [78]. Briefly, Hydrogen atoms were added to the model using the psfgen plugin in VMD [55]. Ionization states of amino acid side chain were assigned based on their pKa values at a pH of 7 using PROPKA [104]. The phospholipid bilayer of POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) was built around the protein of which its initial position was based on the prediction of the spatial orientation of membrane proteins in

membrane [105]. TIP3P water molecules were added to solvate the entire system [106]. The total charge of the system was neutralized by adding counterions ( $\text{Na}^+$  and  $\text{Cl}^-$ ) at 0.1 M concentration using VMD's autoionize plugin. the simulation box consists of 477,934 atoms.

All-atom structures of the system were converted to MARTINI coarse-grained (CG) structure by modifying the residue-based CG scripts in VMD, as previously described [25, 26]. The MARTINI 2.2 force field topology was used for the CG models. The CG nanodiscs were solvated by MARTINI water beads.  $\text{Na}^+$  and  $\text{Cl}^-$  beads were subsequently added for charge neutralization of the systems. This simulation consists of 59,694 beads in a  $18 \times 18 \times 23 \text{ nm}^3$  simulation box.

### 5.3.2 All-atom MD simulations

AAMD simulation was performed as a control to check that the introduction of artificial constraints does not significantly alter the CorA structure. All simulations were carried out with the package NAMD 2.12 program [56]. For AAMD simulation, The CHARMM36 force fields were used for protein and lipid atoms [107]. TIP3P force field was used for water molecules. The simulations of the CorA closed conformation were carried out employing periodic boundary conditions. Langevin dynamics at a constant temperature of 310 K The pressure was kept constant at 1 atm using the Nose-Hoover Langevin piston method, with a piston period of 200 fs and a damping time of 50 fs [108]. The particle mesh Ewald summation was employed to treat long-range electrostatic interactions with a distance cut-off of  $12 \text{ \AA}$  and a pair-list distance of  $13.5 \text{ \AA}$  [109]. The water geometry was constrained as rigid using the SETTLE algorithm [110]. The SHAKE algorithm was employed to constrain the bonds to the other hydrogens [111]. To relax structural strains of the systems, energy minimization was performed to remove bad contacts between atoms, and then MD was conducted with the protein and lipid head group atoms kept fixed to their initial positions. A subsequent run then allowed the whole system (waters, lipids, and counterions) except for the protein to be relaxed. Finally, the equilibration and production runs were performed without any positional restraints. All-atom MD simulations were performed under NPT ensemble for  $4 \mu\text{s}$  with a time step of 2 fs.

### 5.3.3 Coarse-grained MD simulations

#### *Distance restraints mimic the attraction between protomers*

It has been suggested that all  $\text{Mg}^{2+}$  ions bound to residues in M1 and M2 sites remain partially hydrated [13, 80] as octahedral coordination is the preferred geometry for  $\text{Mg}^{2+}$  [101, 112]. This adds a particularly complicated issue for addressing specific interaction parameters of  $\text{Mg}^{2+}$  with CG models since Martini force field parameter is not available for accurate treatment of the divalent metal ion in binding sites. Other CG metal ions such as  $\text{Na}^+$  and  $\text{Cl}^-$  implicitly include a first hydration shell and are used qualitatively only. It can be thought that binding of  $\text{Mg}^{2+}$  to the M1 and M2 sites provides constraints to attract between protein protomers together. We, therefore, introduced a set of inter-subunit distance restraints as artificial forces that represent the  $\text{Mg}^{2+}$  interactions made between protein chains. Due to the unavailability of a CG model for  $\text{Mg}^{2+}$ , the restraints were alternatively introduced by distances between backbone beads of the following residues: E88, D89, D175, D179, and D253. The distance restraints were introduced as the “soft” harmonic potentials ( $U(r) = k(r - r_{ref})^2$ ) and incorporated into the simulations using “extraBonds” term implemented in NAMD.  $r_{ref}$  is the reference distance of the corresponding residues measured from the initial structure. There are five sets of restraining potentials between protein chains (labelled as A-B, B-C, C-D, D-E and E-A). For instance, six restraint distances of the A-B dataset were designated as R<sub>1</sub> (D175<sub>A</sub>/E88<sub>B</sub>), R<sub>2</sub> (D175<sub>A</sub>/D89<sub>B</sub>), R<sub>3</sub> (D179<sub>A</sub>/E88<sub>B</sub>), R<sub>4</sub> (D179<sub>A</sub>/D89<sub>B</sub>), R<sub>5</sub> (D253<sub>A</sub>/E88<sub>B</sub>), and R<sub>6</sub> (D253<sub>A</sub>/D89<sub>B</sub>). The restraints were summarized in Table2.

*Table 2: Distance restraints and the reference distance ( $r_{ref}$ ) between backbone beads of CG residues. Restraints between residue pairs (with residue numbers in parentheses) are denoted as R1 (D175/E88), R2 (D175/D89), R3 (D179/E88), R4 (D179/D89), R5 (D253/E88), and R6 (D253/D89)*

Restraints	$r_{ref}$
R1 <sub>A-B</sub>	15.7
R1 <sub>B-C</sub>	15.6
R1 <sub>C-D</sub>	15.4

R1 <sub>D-E</sub>	16.5
R1 <sub>E-A</sub>	15.8
<hr/>	
R2 <sub>A-B</sub>	16.4
R2 <sub>B-C</sub>	16.3
R2 <sub>C-D</sub>	16.0
R2 <sub>D-E</sub>	16.9
R2 <sub>E-A</sub>	16.4
<hr/>	
R3 <sub>A-B</sub>	15.9
R3 <sub>B-C</sub>	15.5
R3 <sub>C-D</sub>	15.2
R3 <sub>D-E</sub>	16.4
R3 <sub>E-A</sub>	15.7
<hr/>	
R4 <sub>A-B</sub>	16.0
R4 <sub>B-C</sub>	15.7
R4 <sub>C-D</sub>	15.3
R4 <sub>D-E</sub>	16.4
R4 <sub>E-A</sub>	15.8
<hr/>	
R5 <sub>A-B</sub>	11.0
R5 <sub>B-C</sub>	11.1
R5 <sub>C-D</sub>	10.9
R5 <sub>D-E</sub>	12.3
R5 <sub>E-A</sub>	11.0
<hr/>	
R6 <sub>A-B</sub>	10.7
R6 <sub>B-C</sub>	10.7
R6 <sub>C-D</sub>	10.6
R6 <sub>D-E</sub>	11.8
R6 <sub>E-A</sub>	10.6
<hr/>	

### *Simultaneous versus stepwise restraint removals*

For CG simulations, Martini force field were applied for protein, lipid and water [25, 26]. The CG-MD simulations were carried out at a constant temperature of 300 K with a periodic boundary condition. The martini force-field switching function was adapted for Lennard-Jones (LJ) potential with a switching distance is 9.0 Å, and the cutoff is 12.0 Å. A shifting function Coulombic potential with switching distance is 0.0 hardcoded in the NAMD source code and cutoff 12.0 Å. A time step of 20 fs was applied. Energy minimization with the restraints was performed, starting from the closed conformation. Prior to the simulation systems were equilibrated for ~40 ns with force constants of 0.5 kcal mol<sup>-1</sup> Å<sup>-2</sup>. To examine the effect of Mg<sup>2+</sup> depletion on the CorA conformation in the closed stated after equilibration, we have used two approaches of the subsequent simulations: a) simultaneous or all-at-once restraint removal; and b) stepwise restraint removal. The two examined approaches were based on the hypothesis that the CorA closed state releases Mg<sup>2+</sup> from all divalent cation sensor sites simultaneously or stepwise under physiological conditions. The detailed simulations after equilibration were described as follows:

- *Trial-I: Simultaneous removal*

All restraint datasets were removed all at once. The simultaneous removal simulation was performed for ~4 μs.

- *Trial-II: Sequentially stepwise removal*

The stepwise removal simulation was conducted by deleting sequentially each restraint dataset of the following order: R<sub>x</sub>(A-B), R<sub>x</sub>(B-C), R<sub>x</sub>(C-D), R<sub>x</sub>(D-E) and R<sub>x</sub>(E-A) (where x is 1, 2, 3 ... 6). The restraint deletion was made one at a time every ~0.5 μs of the simulation. The simulation was carried out for ~4 μs.

#### 5.3.4 MD trajectories analysis

Analysis of MD trajectories was conducted using TCL scripts in VMD 1.9.3. [55]. The root mean square displacement (RMSD) of backbone atoms (AAMD) or beads (CGMD) was computed with respect to a reference structure (i.e., starting

conformation of the MD simulation). The cryo-EM structure of the open state II (pdb 3jch) was used for comparing with the CG models obtained in the study. The model validation was achieved by structure superimposition and the distance analysis.

## 5.4 Results and discussion

### 5.4.1 Validation: AAMD vs CGMD of the CorA closed conformation

The backbone RMSDs with respect to the initial CorA pentamer structure of the closed conformations for AAMD and CGMD simulations were shown in Figure IV-1. Initially, the RMSD fluctuated below 0.1 nm. This was due to the position restraints imposed on the protein backbone for relaxing the surrounding environments in the systems. After decreasing the force constant of the position restraint potentials, the trajectory of both systems exhibited a rapid increase in RMSD to  $\sim 0.4$  nm after  $\sim 100$  frames corresponding to 0.4ns for AAMD and 4ns for CGMD. Then, the RMSD reaches a plateau with minimal deviations around  $\sim 0.5$  nm, indicating the protein was dynamically equilibrated. Both AAMD and CGMD systems showed a similar trend of the RMSD profile throughout the equilibration time of the simulations.

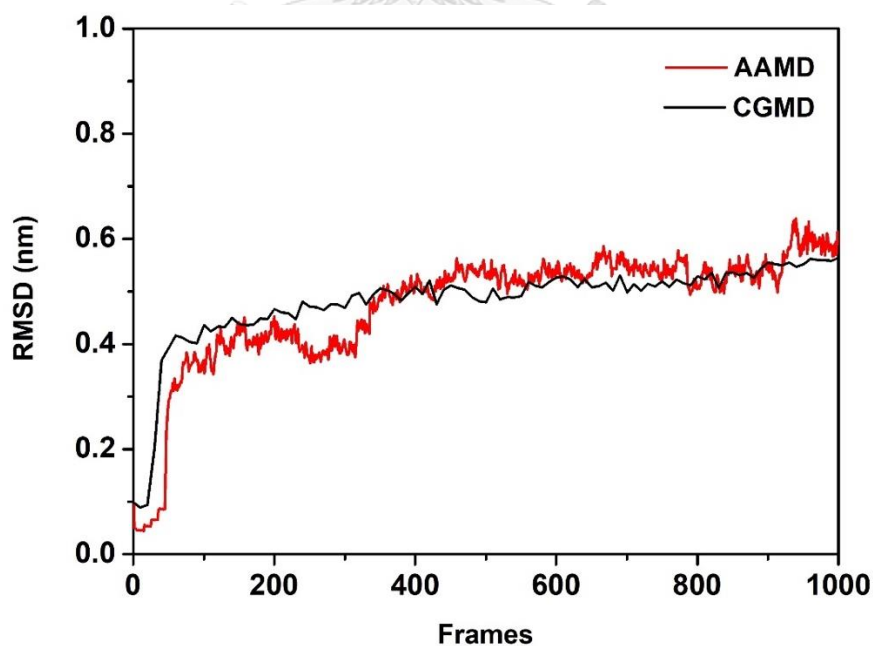


Figure V- 1: Backbone RMSD with respect to the reference CorA structure in the closed state versus simulation frames for AAMD and CGMD simulations.

#### 5.4.2 Simultaneous vs sequentially stepwise removals

Figure IV-2 illustrated the RMSD values for individual protein chain and the overall pentamer structure upon the release of  $Mg^{2+}$  restraints. The removal of the distance restraints leads to a continuous increase in overall RMSD around 0.7-0.8 nm. Notably, the simultaneous removal simulation showed RMSDs of the five protein chains (A, B, C, D and E) in range between 6 and 8 nm at the end of 4 $\mu$ s CGMD simulation (Figure IV-2A). However, the simulation with the sequentially stepwise removal showed a wider fluctuation range of RMSD per protein chain. The final RMSDs per protein chain reach between 0.5 to 0.9 nm at 4 $\mu$ s (Figure IV-2B).

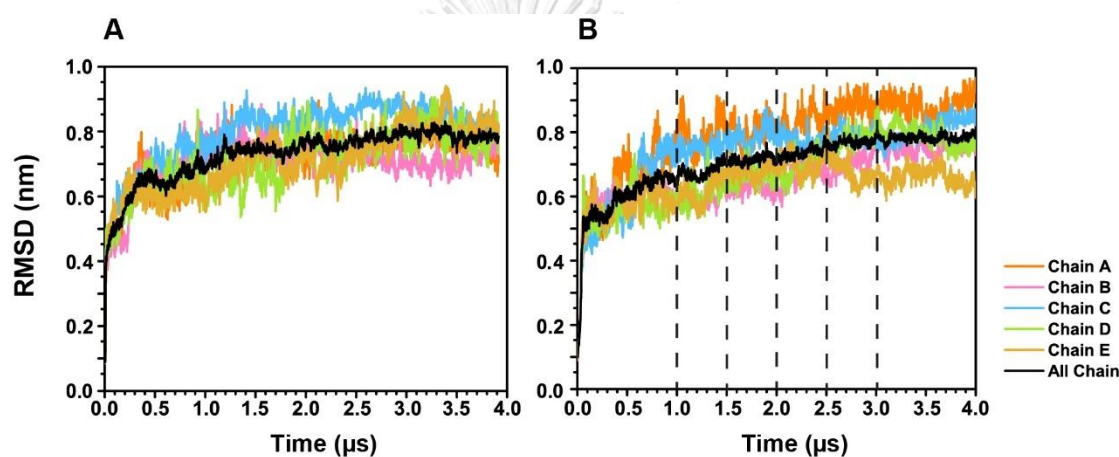


Figure V- 2 : RMSD with respect to the reference CorA structure in the closed state extracted from the 4  $\mu$ s CGMD simulations with simultaneous (A) and stepwise (B) removals. RMSD values were measured on the basis of individual protein chain, i.e., A, B, C, D and E and the overall pentamer. Dashed line indicates the time at which restraints were removed.

#### 5.4.3 Large conformational changes in cytosolic domains upon the restraint removal

To demonstrate structure changes upon the restraint removal, we plotted the backbone RMSD per residue as a function of simulation time as shown in Figure IV-3. We observed considerable differences with the RMSD with respect to the closed state conformation of two systems, in the simultaneous or sequentially stepwise restraint removals. As shown by the region in red color in Figure IV-3, the cytosolic domain containing residues 3-100 was the common region that showed high RMSD values

compared to the rest. As higher RMSD value indicates greater variation from the initial close conformation. Therefore, this suggested that the cytosolic domain appears as the most affected region upon the release of the restraints. However, structure changes in stalk and transmembrane domain were much smaller (shown by the blue region). This is consistent with the cryo-EM study of which conformational alterations were mainly detected in the cytosolic domain but the change in transmembrane regions was subtle. From the RMSD plot, we also found the stepwise removal system has high RMSD values spanning larger regions in the cytosolic domain than the simultaneous removal simulation. This implies that the stepwise removal has a greater impact on the overall protein structure. Furthermore, the stepwise removal simulation displayed large RMSD values in the cytosolic domain of chains A, B, C and D whereas that of chain E was least affected by the restraint removal. The results are in good agreement with a study by Perozo et al. [12]. It was reported that four of five chains of the CorA cytosolic domain underwent significant conformational changes in the absence of  $Mg^{2+}$ .

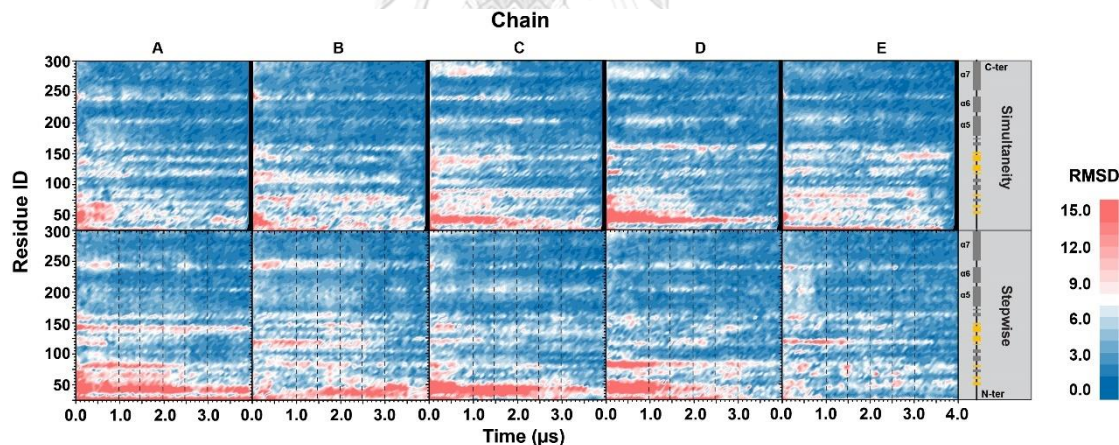


Figure V- 3 : Contour maps representing backbone RMSD per residue as a function of simulation time obtained from CG simulations of simultaneous and stepwise restraint removals. The RMSD plot shows individual chain separately (i.e., A, B, C, D and E). Dashed line indicates the time at which restraints were removed. Color gradient from red to blue corresponds high to low RMSD values.

As the RMSD analysis is not sufficient to clearly visualize difference observed between the two systems, we extracted the structure coordinates from MD trajectories at different time spans, i.e., 0.3  $\mu$ s, 0.5  $\mu$ s, 1.0  $\mu$ s, 2.0  $\mu$ s, 3.0  $\mu$ s, and 4.0  $\mu$ s, to better



understand the differences between the simultaneous (Trial-I) and stepwise (Trial-II) systems (Figure IV-4A). The backbone conformations of the CorA cytosolic domain at different time frames of both systems (shown in surface representation of Figure IV-4) displayed significant movement in the cytosolic region. However, the motion of the cytosolic domain from the simulations was not as large as that observed from the structure in the “open state II” based on the cryo-EM study (Figure IV-4B). This observation signifies that the loss of interactions between protein chains induced alteration in the domain conformation but not enough to drive the protein into the structure close to the cryo-EM experiments.

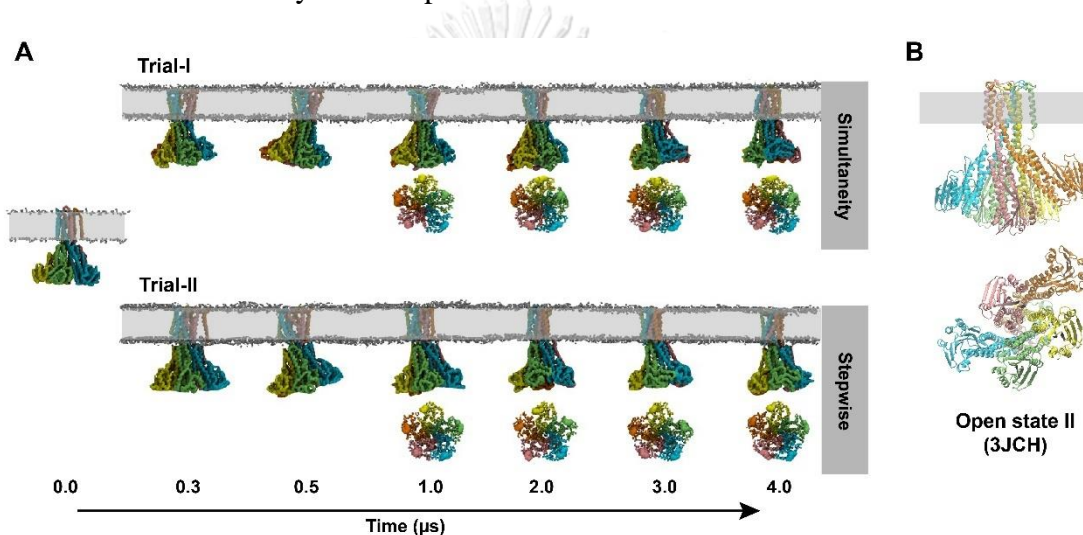


Figure V- 4: (A) MD snapshots representing conformational changes of the cytosolic domain obtained from simultaneous (Trial-I, upper) and sequentially stepwise (Trial-II, lower) restraint removals at various simulation time. (B) Cryo-EM structure of open state II (pdb 3jch). Structures show in side (upper) and bottom (lower) views. Area within grey line indicates membrane region.

A further validation of the simulation models was performed based on an analysis of inter-subunit distances and the inter-domain arrangement relative to the cryo-EM open-state structure. For a distance analysis, there are five sets of distances between protein chains (labelled as A-B, B-C, C-D, D-E and E-A) to be computed. It should, however, be noted that the cryo-EM density maps in the cytosolic domain were identified as a region of relatively low resolution (beyond  $6\text{\AA}$ ) and the conformation in this region was thus not well defined. We measured the distance between the two centers of masses for residues 278-280 located in the stalk helix instead. This was due

to in the cryo-EM study, the density maps of the residues in this fragment exhibited a relatively high resolution. L280 was also used as a reference position to demonstrate a difference between the closed and open state structures [12]. In addition, this fragment involves in an initiation of a “hinge-bending” motion in the absence of bound  $Mg^{2+}$ . As shown in Figure IV-5A, Trial-I models exhibited the time evolution of inter-subunit distances different from Trial-II model, implying a dissimilarity in the arrangement of the pentamer between the two models. Figure IV-5B showed the pentameric arrangement for the residues 278-280 of the simulation models superimposed on the cryo-EM structure in the open state. In the Trial-I model, the residues 278-280 of chain A moves away from the central axis of the channel whereas this fragment of other four chains (B, C, D and E) move towards to the channel axis. The structure superposition showed that chain D was only one out of five chains that was consistent with the cryo-EM open state model. Other the other hand, Trial-II model appears to agree better with the experimental data. As can be seen from Figure IV-5B, the arrangement of the fragment of chain B, C and E has a relatively good fit with the open state structure.

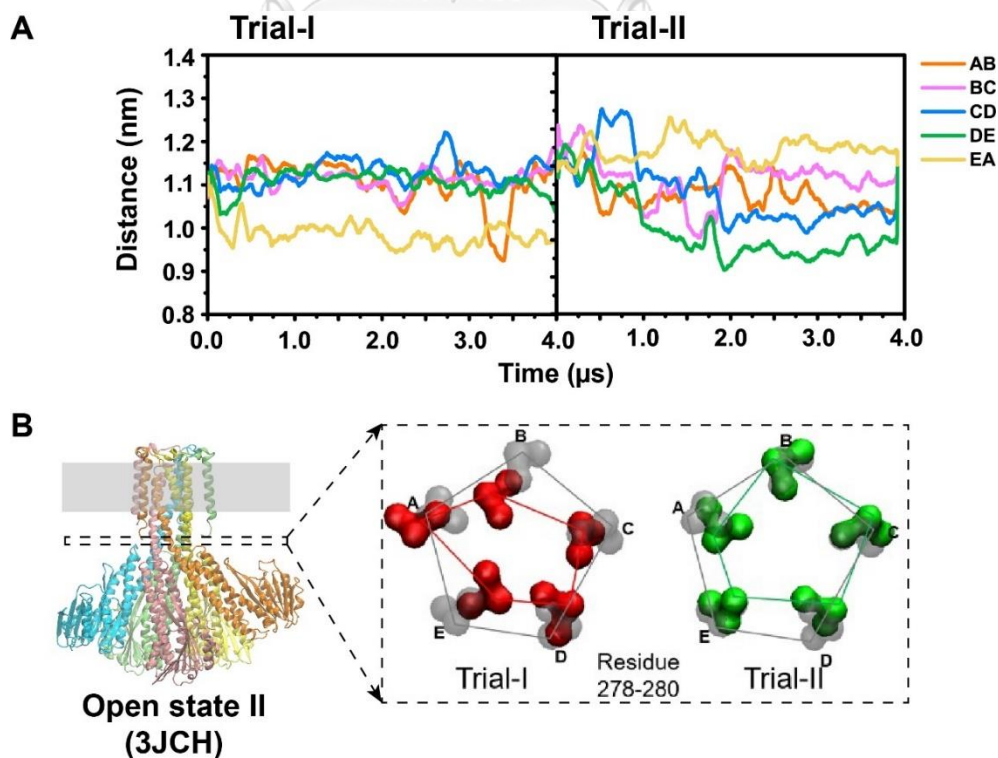


Figure V- 5: (A) Inter-subunit distance as a function of time obtained from MD trajectories of simultaneous (Trial-I) and stepwise (Trial-II) removal simulations. The distances between subunits (labelled as A-B, B-C, C-D, D-E and E-A) were measured using the center of mass of a fragment corresponding to residues 278-280 (indicated by dashed lines in the open state II (3jch) and box shown in B).

(B) Surface representation of the open state II (left) and the pentameric arrangement of the residues 278-280 of Trial-I and Trial-II models (red and green surface, respectively) superimposed on the same fragment of the cryo-EM open state (gray transparent surface).

Nevertheless, it is important to extend this study by conducting further simulations to address the problem about the time length of simulation as well as to vary the removal conditions of the simulations. The restraint deletion is no longer controlled by the simulation time. Instead, we introduced the upper-bound restraint violation as a criterion. The restraint violation was taken into consideration if a restrained distance exceeds the upper-bound distance of  $2\text{\AA}$ . The simulations were initiated by eliminating  $R_x(\text{E-A})$  restraint dataset. During the simulation, all restrained distances were monitored for a statistic analysis of restraint violations. The simulations were carried out with two restraint removal approaches designated as Trial-III and Trial-IV. Trial-III was the simulation with the deletion of single dataset having most violated restraints. For the Trial-IV simulation, the restraint deletion can be either single or double datasets if the violation criterion was met. The two additional simulations were extended up to  $\sim 9\ \mu\text{s}$ . The results were shown in Figure IV-6.

The plots in Figure IV-6A showed two different profiles of the center of mass distances between protein chains as a function of time. This suggested that the Trial-III model exhibited a different arrangement of subunits in the pentamer compared with the Trial-IV model. Figure IV-6B showed the pentameric arrangement for the residues 278-280 of both models superimposed on the same residues of the cryo-EM open state structure. As can be seen, the results of the Trial-IV simulation showed better agreement with the experiments compared to the Trial-III results.

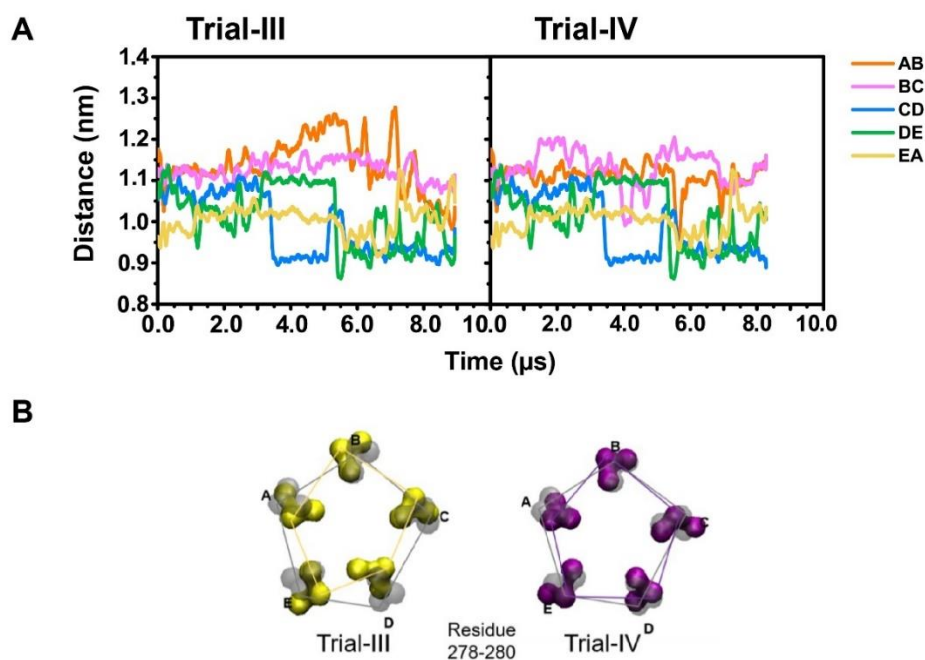


Figure V- 6: (A) Inter-subunit distance as a function of time obtained from MD trajectories of Trial-III and Trial-IV systems. (B) The pentamer arrangement of the residues 278-280 of Trial-III and Trial-IV models (yellow and purple surface, respectively) superimposed on the same fragment of the cryo-EM open state (gray transparent surface). All plots and images were defined same as in Figure

V-5.

## 5.5 Conclusions

This study employed coarse-grained molecular dynamics simulations to examine the conformational response of CorA magnesium channel upon the release of bound  $Mg^{2+}$ . The CG simulations start with the cryo-EM structure of CorA in the closed state. An effect of the  $Mg^{2+}$  deletion in the simulations was investigated using two restraint removal approaches, simultaneous and stepwise removals. We observed the conformational changes of CorA when the restraints representing binding of  $Mg^{2+}$  between protein chains were removed during the simulations. Although the restraint deletion showed a significant conformational rearrangement of the cytosolic domain, the movement in this region was not large enough compared with the cryo-EM structure of CorA in the open state. Nevertheless, the comparison of residues in the stalk helix arranged in pentamer offers an alternative validation of the simulation

studies. We found that the pentameric arrangement of the residues in the hinge-bending region obtained from the stepwise approach is found to be more consistent with experiments the, suggesting a possible mechanism for the  $Mg^{2+}$  release.



## CHAPTER VI: CONCLUSION

---

This thesis presents the use of coarse-grained models in MC and MD simulations to get detailed information on structural behaviors, self-assembly and interactions of the CorA magnesium channel. The results of this study show an ability of CG simulations to accurately reproduce physical and thermodynamical properties compatible with all-atom simulations that are limited by system sizes and time scales. The chapters provide necessary background information and include published and non-published works. The summary of the works is presented below.

### 6.1 Thermal responses of lipoprotein nanodiscs

MSP1 based lipoprotein nanodiscs are commonly used as encapsulating agents in cryo-EM structure determination of membrane proteins. Lipoprotein nanodiscs have been successfully used to encapsulate the CorA  $Mg^{2+}$  channel for structure determination using cryo-EM techniques. In this study, from the results of CGMD simulations, it was found that the structures of MSP1-DPPC and MSP1-POPC nanodiscs respond non-monotonically to temperature. We demonstrated that the non-monotonic temperature response is associated with the lipid bilayer phase transition. CG-MD data showed dramatic changes from the circular to twisted or non-circular shape at a metastable ripple state (225K-275K). In addition, the structure MSP1-POPC nanodiscs was found to be more sensitive to temperature than that of the MSP1-DPPC nanodisc. The interaction energy of the MSP1-DPPC nanodisc compared to MSP1-POPC nanodisc, show that the DPPC nanodisc tends to resist structural changes more than the POPC nanodisc upon an increase in temperature. It is hoped that this study will help in designing appropriate nanodiscs and interpreting data in underlying experiments in future.

## 6.2 Self-organized globular bundles and entangle networks of CorA

This study employed CGMC simulation to observe the self-organized structures of CorA in crowded environments as a function of protein concentration at a range of low to high temperatures. At both low and high temperatures, visual inspection displays obvious differences in the morphology of the assembly in dilute solution and that in the crowded matrix. The outer membrane region of CorA (oCorA) proteins appear to dissociate more on a scale proportional to their size, whereas the inner membrane region of CorA (iCorA) protein clusters phase-separate in their native phase (low temperatures). Based on the findings of this study, it was concluded that, in addition to steric limitations of the membrane and interactions with other matrix constituents, a considerably greater protein-protein connection among the protein's inner segments (iCorA) is conducive to its stable self-assembly.

## 6.3 CorA conformations in a solute matrix

This study used CGMC simulation to investigate the conformational dynamics of CorA in an interacting matrix with mobile solute particles. The protein chain is initially arranged in a random arrangement in the simulation box, with a random distribution of solute particles. Each residue and solute particle move in a random manner. In the short- and long-time regimes, the protein chain has different dynamics: the short-time dynamics is sub-diffusive and not highly sensitive to solute-residue interaction strength, whereas the long-time (asymptotic) dynamics is substantially dependent on the interaction strength. As solute particles attach to their target residues at increasing solute-residue interactions, the protein chain becomes practically immobile. The solute particles constrained protein conformation that bind to specific protein locations and pinned with stronger solute-residue interactions.

## 6.4 Magnesium release-induced conformational changes of CorA

A CGMD simulation was carried out to examine the conformational response of CorA magnesium channel upon the release of bound  $Mg^{2+}$ . Simultaneous and sequential deletions of restraints were used to assess the effect of the  $Mg^{2+}$  deletion in

the simulations. We observed conformational changes of CorA when the restraints binding the  $Mg^{2+}$  between the protein chains were removed during the simulations. Despite the constraint loss resulting in a major conformational rearrangement of the cytosolic domain, the movement in this region was insufficient when compared to cryo-EM structure in the open state. We found that the pentameric arrangement of the residues in the hinge-bending region determined using the stepwise approach is more consistent with observations, implying a plausible mechanism for  $Mg^{2+}$  release.





## REFERENCES

1. Kretsinger, R.H., *Magnesium in Biological Systems*, in *Encyclopedia of Metalloproteins*, R.H. Kretsinger, V.N. Uversky, and E.A. Permyakov, Editors. 2013, Springer New York: New York, NY. p. 1250-1255.
2. Alberts B, et al., *Molecular Biology of the Cell*, ed. 4. 2002, New York: Garland Science.
3. Wolf, F.I. and V. Trapani, *Cell (patho) physiology of magnesium*. Clinical science, 2008. **114**(1): p. 27-35.
4. Walker, G.M., *Magnesium and cell cycle control: an update*. Magnesium, 1986. **5**(1): p. 9-23.
5. Flatman, P.W., *Magnesium transport across cell membranes*. The Journal of membrane biology, 1984. **80**(1): p. 1-14.
6. Romani, A.M.J.A.o.b. and biophysics, *Cellular magnesium homeostasis*. 2011. **512**(1): p. 1-23.
7. Kuramoto, T., et al., *A Mutation in the Gene Encoding Mitochondrial Mg<sup>2+</sup> Channel MRS2 Results in Demyelination in the Rat*. PLOS Genetics, 2011. **7**(1): p. e1001262.
8. Leanza, L., et al., *Intracellular ion channels and cancer*. Frontiers in physiology, 2013. **4**: p. 227.
9. Weghuber, J., et al., *Mutational analysis of functional domains in Mrs2p, the mitochondrial Mg<sup>2+</sup> channel protein of Saccharomyces cerevisiae*. The FEBS Journal, 2006. **273**(6): p. 1198-1209.
10. Eshaghi, S., et al., *Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution*. Science, 2006. **313**(5785): p. 354-357.
11. Lunin, V.V., et al., *Crystal structure of the CorA Mg<sup>2+</sup> transporter*. Nature, 2006. **440**(7085): p. 833.
12. Matthies, D., et al., *Cryo-EM structures of the magnesium channel CorA reveal symmetry break upon gating*. Cell, 2016. **164**(4): p. 747-756.
13. Dalmas, O., et al., *Structural dynamics of the magnesium-bound conformation of CorA in a lipid bilayer*. Structure, 2010. **18**(7): p. 868-878.
14. Pfoh, R., et al., *Structural asymmetry in the magnesium channel CorA points to sequential allosteric regulation*. Proceedings of the National Academy of Sciences, 2012. **109**(46): p. 18809-18814.
15. Dalmas, O., et al., *Molecular mechanism of Mg<sup>2+</sup>-dependent gating in CorA*. 2014. **5**: p. 3590.
16. Barton, P.G. and F. Gunstone, *Hydrocarbon chain packing and molecular motion in phospholipid bilayers formed from unsaturated lecithins. Synthesis and properties of sixteen positional isomers of 1, 2-dioctadecenoyl-sn-glycero-3-phosphorylcholine*. Journal of Biological Chemistry, 1975. **250**(12): p. 4470-4476.
17. Borhani, D.W., et al., *Crystal structure of truncated human apolipoprotein AI suggests a lipid-bound conformation*. Proceedings of the National Academy of Sciences, 1997. **94**(23): p. 12291-12296.
18. Bayburt, T.H., Y.V. Grinkova, and S.G. Sligar, *Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins*. Nano Letters, 2002. **2**(8): p. 853-856.

19. Denisov, I.G., et al., *Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size*. Journal of the American Chemical Society, 2004. **126**(11): p. 3477-3487.
20. Denisov, I.G., et al., *Thermotropic phase transition in soluble nanoscale lipid bilayers*. The Journal of Physical Chemistry B, 2005. **109**(32): p. 15580-15588.
21. Shih, A.Y., et al., *Coarse grained protein– lipid model with application to lipoprotein particles*. The Journal of Physical Chemistry B, 2006. **110**(8): p. 3674-3684.
22. Shih, A.Y., et al., *Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations*. Journal of structural biology, 2007. **157**(3): p. 579-592.
23. Arkhipov, A., P.L. Freddolino, and K. Schulten, *Stability and dynamics of virus capsids described by coarse-grained modeling*. Structure, 2006. **14**(12): p. 1767-1777.
24. Arkhipov, A., Y. Yin, and K. Schulten, *Four-scale description of membrane sculpting by BAR domains*. Biophysical journal, 2008. **95**(6): p. 2806-2821.
25. de Jong, D.H., et al., *Improved parameters for the martini coarse-grained protein force field*. 2012. **9**(1): p. 687-697.
26. Marrink, S.J., et al., *The MARTINI force field: coarse grained model for biomolecular simulations*. The journal of physical chemistry B, 2007. **111**(27): p. 7812-7824.
27. Alessandri, R., F. Grünewald, and S.J. Marrink, *The Martini Model in Materials Science*. arXiv preprint arXiv:2012.07194, 2020.
28. Souza, P.C., et al., *Martini 3: a general purpose force field for coarse-grained molecular dynamics*. Nature methods, 2021. **18**(4): p. 382-388.
29. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*. Folding and Design, 1997. **2**(3): p. 173-181.
30. Boas, F.E. and P.B. Harbury, *Potential energy functions for protein design*. Current Opinion in Structural Biology, 2007. **17**(2): p. 199-204.
31. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. Nat Struct Biol, 2002. **9**(9): p. 646-52.
32. Biro, J., *Amino acid size, charge, hydrophathy indices and matrices for protein structure analysis*. Theoretical Biology and Medical Modelling, 2006. **3**(1): p. 15.
33. K., B., *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. 1995, Oxford University Press.
34. Kalipatnapu, S. and A. Chattopadhyay, *Membrane protein solubilization: recent advances and challenges in solubilization of serotonin1A receptors*. IUBMB Life, 2005. **57**(7): p. 505-12.
35. Linke, D., *Detergents: an overview*, in *Methods in enzymology*. 2009, Elsevier. p. 603-617.
36. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs*. FEBS letters, 2010. **584**(9): p. 1721-1727.
37. Liang, B. and L.K. Tamm, *NMR as a Tool to Investigate Membrane Protein Structure, Dynamics and Function*. Nature structural & molecular biology, 2016. **23**(6): p. 468.

38. Borhani, D.W., et al., *Crystal structure of truncated human apolipoprotein AI suggests a lipid-bound conformation*. Proceedings of the National Academy of Sciences, USA, 1997. **94**(23): p. 12291-12296.
39. Denisov, I.G., et al., *Thermotropic phase transition in soluble nanoscale lipid bilayers*. Journal of Physical Chemistry B, 2005. **109**(32): p. 15580-15588.
40. Koppaka, V., et al., *The structure of human lipoprotein AI Evidence for the "belt" model*. Journal of Biological Chemistry, 1999. **274**(21): p. 14541-14544.
41. Ravula, T., N.Z. Hardin, and A. Ramamoorthy, *Polymer nanodiscs: Advantages and limitations*. Chemistry and physics of lipids, 2019. **219**: p. 45-49.
42. Dörr, J.M., et al., *Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K<sup><sup>+</sup> channel: The power of native nanodiscs</sup>*. Proceedings of the National Academy of Sciences, 2014. **111**(52): p. 18607-18612.
43. Tall, A.R. and D.M. Small, *Solubilisation of phospholipid membranes by human plasma high density lipoproteins*. Nature, 1977. **265**(5590): p. 163.
44. Damiani, S., et al., *Albumin-bound nanodiscs as delivery vehicle candidates: Development and characterization*. Biophysical Chemistry, 2019. **251**: p. 106178.
45. Sahoo, B.R., et al., *Self-Assembly of Polymer-Encased Lipid Nanodiscs and Membrane Protein Reconstitution*. The Journal of Physical Chemistry B, 2019. **123**(21): p. 4562-4570.
46. Marrink, S.J., J. Risselada, and A.E. Mark, *Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model*. Chemistry and Physics of Lipids, 2005. **135**(2): p. 223-244.
47. Leekumjorn, S. and A.K. Sum, *Molecular characterization of gel and liquid-crystalline structures of fully hydrated POPC and POPE bilayers*. J Phys Chem B, 2007. **111**(21): p. 6026-33.
48. Kranenburg, M., M. Venturoli, and B. Smit, *Phase behavior and induced interdigitation in bilayers studied with dissipative particle dynamics*. Journal of Physical Chemistry B, 2003. **107**(41): p. 11491-11501.
49. Müller, S., et al., *Mixing behaviour of bilayer-forming phosphatidylcholines with single-chain alkyl-branched bolalipids: effect of lateral chain length*. Biophysical Chemistry, 2019. **244**: p. 1-10.
50. Kariyazono, H., et al., *Formation of stable nanodiscs by bihelical apolipoprotein A - I mimetic peptide*. Journal of Peptide Science, 2016. **22**(2): p. 116-122.
51. Shaw, A.W., M.A. McLean, and S.G. Sligar, *Phospholipid phase transitions in homogeneous nanometer scale bilayer discs*. Febs Letters, 2004. **556**(1-3): p. 260-264.
52. Jo, S., et al., *CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues*, in *Advances in protein chemistry and structural biology*. 2014, Elsevier. p. 235-265.
53. Lee, J., et al., *CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field*. Journal of Chemical Theory and Computation, 2015. **12**(1): p. 405-413.
54. Qi, Y., et al., *CHARMM - GUI Nanodisc Builder for modeling and simulation of various nanodisc systems*. Journal of computational chemistry, 2019. **40**(7): p.

- 893-899.
55. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.
  56. James C. Phillips, et al., *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 2005. **26**: p. 21.
  57. Martyna, G.J., D.J. Tobias, and M.L. Klein, *Constant pressure molecular dynamics algorithms*. Journal of Chemical Physics, 1994. **101**(5): p. 4177-4189.
  58. Feller, S.E., et al., *Constant pressure molecular dynamics simulation: the Langevin piston method*. Journal of Chemical Physics, 1995. **103**(11): p. 4613-4621.
  59. Seeber, M., et al., *Wordom: a program for efficient analysis of molecular dynamics simulations*. Bioinformatics, 2007. **23**(19): p. 2625-2627.
  60. Svetlovics, J.A., S.A. Wheaten, and P.F. Almeida, *Phase separation and fluctuations in mixtures of a saturated and an unsaturated phospholipid*. Biophys J, 2012. **102**(11): p. 2526-35.
  61. Nath, A., W.M. Atkins, and S.G. Sligar, *Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins*. Biochemistry, 2007. **46**(8): p. 2059-2069.
  62. Siuda, I. and D.P. Tieleman, *Molecular models of nanodiscs*. Journal of chemical theory and computation, 2015. **11**(10): p. 4923-4932.
  63. Shih, A.Y., et al., *Molecular dynamics simulations of discoidal bilayers assembled from truncated human lipoproteins*. Biophysical journal, 2005. **88**(1): p. 548-556.
  64. Furukawa, Y. and N. Nukina, *Functional diversity of protein fibrillar aggregates from physiology to RNA granules to neurodegenerative diseases*. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2013. **1832**(8): p. 1271-1278.
  65. Bai, Y., Q. Luo, and J. Liu, *Protein self-assembly via supramolecular strategies*. Chemical Society Reviews, 2016. **45**(10): p. 2756-2767.
  66. McManus, J.J., et al., *The physics of protein self-assembly*. Current opinion in colloid & interface science, 2016. **22**: p. 73-79.
  67. Sgarbossa, A., *Natural biomolecules and protein aggregation: emerging strategies against amyloidogenesis*. International journal of molecular sciences, 2012. **13**(12): p. 17121-17137.
  68. Sun, H., et al., *Nanostructures based on protein self-assembly: From hierarchical construction to bioinspired materials*. Nano Today, 2017. **14**: p. 16-41.
  69. Garcia-Seisdedos, H., et al., *Proteins evolve on the edge of supramolecular self-assembly*. Nature, 2017. **548**(7666): p. 244-247.
  70. Pandey, R.B., B.L. Farmer, and B.S. Gerstman, *Self-assembly dynamics for the transition of a globular aggregate to a fibril network of lysozyme proteins via a coarse-grained Monte Carlo simulation*. AIP Advances, 2015. **5**(9): p. 092502.
  71. Yang, L., et al., *Self - assembly of proteins: Towards supramolecular materials*. Chemistry - A European Journal, 2016. **22**(44): p. 15570-15582.
  72. Hmiel, S., et al., *Magnesium transport in Salmonella typhimurium: genetic characterization and cloning of three magnesium transport loci*. Journal of bacteriology, 1989. **171**(9): p. 4742-4751.

73. Maguire, M.E., *MgtA and MgtB: prokaryotic P-type ATPases that mediate Mg<sup>2+</sup> influx*. Journal of bioenergetics and biomembranes, 1992. **24**(3): p. 319-328.
74. KEHRES, D.G., C.H. LAWYER, and M.E. MAGUIRE, *The CorA magnesium transporter gene family*. Microbial & comparative genomics, 1998. **3**(3): p. 151-169.
75. Payandeh, J., et al., *Probing structure-function relationships and gating mechanisms in the CorA Mg<sup>2+</sup> transport system*. Journal of Biological Chemistry, 2008. **283**(17): p. 11721-11733.
76. Payandeh, J. and E.F. Pai, *A structural basis for Mg<sup>2+</sup> homeostasis and the CorA translocation cycle*. The EMBO journal, 2006. **25**(16): p. 3762-3773.
77. Neale, C., et al., *Hydrophobic gating of ion permeation in magnesium channel CorA*. PLoS computational biology, 2015. **11**(7).
78. Kitjaruwankul, S., et al., *Inner and Outer Coordination Shells of Mg<sup>2+</sup> in CorA Selectivity Filter from Molecular Dynamics Simulations*. The Journal of Physical Chemistry B, 2016. **120**(3): p. 406-417.
79. Chakrabarti, N., et al., *An iris-like mechanism of pore dilation in the CorA magnesium transport system*. Biophysical journal, 2010. **98**(5): p. 784-792.
80. Nordin, N., et al., *Exploring the structure and function of Thermotoga maritima CorA reveals the mechanism of gating and ion selectivity in Co<sup>2+</sup>/Mg<sup>2+</sup> transport*. Biochemical Journal, 2013. **451**(3): p. 365-374.
81. Kitjaruwankul, S., et al., *Asymmetry in structural response of inner and outer transmembrane segments of CorA protein by a coarse-grain model*. Journal of Chemical Physics, 2016. **145**(13).
82. Kitjaruwankul, S., et al., *Thermal-induced folding and unfolding of a transmembrane protein (CorA)*. Physica A: Statistical Mechanics and its Applications, 2018. **506**: p. 987-992.
83. Pandey, R.B. and B.L. Farmer, *Aggregation and network formation in self-assembly of protein (H3.1) by a coarse-grained Monte Carlo simulation*. J Chem Phys, 2014. **141**(17): p. 175103.
84. Betancourt, M.R. and D. Thirumalai, *Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes*. Protein Sci, 1999. **8**(2): p. 361-9.
85. Miyazawa, S. and R.L. Jernigan, *Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation*. Macromolecules, 1985. **18**(3): p. 534-552.
86. Miyazawa, S. and R.L. Jernigan, *Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading*. J Mol Biol, 1996. **256**(3): p. 623-44.
87. Tanaka, S. and H.A. Scheraga, *Medium- and Long-Range Interaction Parameters between Amino Acids for Predicting Three-Dimensional Structures of Proteins*. Macromolecules, 1976. **9**(6): p. 945-950.
88. Godzik, A., *Knowledge-based potentials for protein folding: what can we learn from known protein structures?* Structure, 1996. **4**(4): p. 363-366.
89. Huang, S.Y. and X. Zou, *Statistical mechanics - based method to extract atomic distance - dependent potentials from protein structures*. Proteins: Structure, Function, and Bioinformatics, 2011. **79**(9): p. 2648-2661.
90. Sorokina, I. and A. Mushegian, *Modeling protein folding in vivo*. Biology direct,

2018. **13**(1): p. 13.
91. Duncan, A.L., et al., *Protein crowding and lipid complexity influence the nanoscale dynamic organization of ion channels in cell membranes*. *Sci Rep*, 2017. **7**(1): p. 16647.
  92. Zeindlhofer, V. and C. Schroder, *Computational solvation analysis of biomolecules in aqueous ionic liquid mixtures : From large flexible proteins to small rigid drugs*. *Biophys Rev*, 2018. **10**(3): p. 825-840.
  93. Kitjaruwankul, S., et al., *Thermal-induced folding and unfolding of a transmembrane protein (CorA)*. *Physica A: Statistical Mechanics and its Applications*, 2018. **506**: p. 987-992.
  94. S. Kitjaruwankul, P.B., Sunita Subedi Paudel, Warin Jetsadawisut, Pornthep Sompornpisut, R.B. Pandey, *Conformational modulation of transmembrane segments of a protein (CorA) by effective media*. *Int. J. Vir. Inf. Dis.*, 2017. **2**: p. 8-13.
  95. Beveridge, D.L., *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, ed. K. Binder. 1995, New York: Oxford University Press.
  96. R. B. Pandey, D.J.J., and B. L. Farmer, *Preferential binding effects on protein structure and dynamics revealed by coarse-grained Monte Carlo simulation*. *J. Chem. Phys*, 2017. **146**: p. 195101.
  97. Warin Rangubpit, S.K., Panisak Boonamnaj, Pornthep Sompornpisut, Ras B. Pandey *Globular bundles and entangled network of proteins (CorA) by a coarse-grained Monte Carlo simulation*. *AIMS Biophysics*, 2019. **6**(2): p. 68-82.
  98. Kersey, C.M., P.A. Agyemang, and C.K. Dumenyó, *CorA, the magnesium/nickel/cobalt transporter, affects virulence and extracellular enzyme production in the soft rot pathogen Pectobacterium carotovorum*. *Molecular plant pathology*, 2012. **13**(1): p. 58-71.
  99. Maguire, M.E., *Magnesium transporters: properties, regulation and structure*. *Front Biosci*, 2006. **11**: p. 3149-3163.
  100. Smith, R.L. and M.E. Maguire, *Distribution of the CorA Mg<sup>2+</sup> transport system in gram-negative bacteria*. *Journal of bacteriology*, 1995. **177**(6): p. 1638-1640.
  101. Moomaw, A.S. and M.E. Maguire, *The unique nature of Mg<sup>2+</sup> channels*. *Physiology*, 2008. **23**(5): p. 275-285.
  102. Schweyen, R.J. and E.M. Froschauer, *CorA-Mrs2-Alr1 Superfamily of Mg<sup>2+</sup> Channel Proteins*, in *New Perspectives in Magnesium Research*. 2007, Springer. p. 46-54.
  103. Palombo, I., D.O. Daley, and M. Rapp, *Why is the GMN motif conserved in the CorA/Mrs2/Alr1 superfamily of magnesium transport proteins?* *Biochemistry*, 2013. **52**(28): p. 4842-4847.
  104. Rostkowski, M., et al., *Graphical analysis of pH-dependent properties of proteins predicted using PROPKA*. *BMC structural biology*, 2011. **11**(1): p. 1-6.
  105. Jo, S., T. Kim, and W. Im, *Automated builder and database of protein/membrane complexes for molecular dynamics simulations*. *PloS one*, 2007. **2**(9): p. e880.
  106. Lamoureux, G., et al., *A polarizable model of water for molecular dynamics simulations of biomolecules*. *Chemical Physics Letters*, 2006. **418**(1-3): p. 245-249.
  107. Huang, J. and A.D. MacKerell, *CHARMM36 all-atom additive protein force*

- field: Validation based on comparison to NMR data.* Journal of Computational Chemistry, 2013. **34**(25): p. 2135-2145.
108. Martyna, G.J., D.J. Tobias, and M.L. Klein, *Constant-Pressure Molecular-Dynamics Algorithms.* Journal of Chemical Physics, 1994. **101**(5): p. 4177-4189.
109. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems.* The Journal of chemical physics, 1993. **98**(12): p. 10089-10092.
110. Miyamoto, S. and P.A. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models.* Journal of computational chemistry, 1992. **13**(8): p. 952-962.
111. Kräutler, V., W.F. Van Gunsteren, and P.H. Hünenberger, *A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations.* Journal of computational chemistry, 2001. **22**(5): p. 501-508.
112. Wedekind, J.E., G.H. Reed, and I. Rayment, *Octahedral coordination at the high-affinity metal site in enolase: Crystallographic analysis of the MgII-enzyme complex from yeast at 1.9. Å resolution.* Biochemistry, 1995. **34**(13): p. 4325-4330.



1. Kretsinger, R.H., *Magnesium in Biological Systems*, in *Encyclopedia of Metalloproteins*, R.H. Kretsinger, V.N. Uversky, and E.A. Permyakov, Editors. 2013, Springer New York: New York, NY. p. 1250-1255.
2. Alberts B, et al., *Molecular Biology of the Cell*, ed. 4. 2002, New York: Garland Science.
3. Wolf, F.I. and V. Trapani, *Cell (patho) physiology of magnesium*. Clinical science, 2008. **114**(1): p. 27-35.
4. Walker, G.M., *Magnesium and cell cycle control: an update*. Magnesium, 1986. **5**(1): p. 9-23.
5. Flatman, P.W., *Magnesium transport across cell membranes*. The Journal of membrane biology, 1984. **80**(1): p. 1-14.
6. Romani, A.M.J.A.o.b. and biophysics, *Cellular magnesium homeostasis*. 2011. **512**(1): p. 1-23.
7. Kuramoto, T., et al., *A Mutation in the Gene Encoding Mitochondrial Mg<sup>2+</sup> Channel MRS2 Results in Demyelination in the Rat*. PLOS Genetics, 2011. **7**(1): p. e1001262.
8. Leanza, L., et al., *Intracellular ion channels and cancer*. Frontiers in physiology, 2013. **4**: p. 227.
9. Weghuber, J., et al., *Mutational analysis of functional domains in Mrs2p, the mitochondrial Mg<sup>2+</sup> channel protein of Saccharomyces cerevisiae*. The FEBS Journal, 2006. **273**(6): p. 1198-1209.
10. Eshaghi, S., et al., *Crystal structure of a divalent metal ion transporter CorA at 2.9 angstrom resolution*. Science, 2006. **313**(5785): p. 354-357.
11. Lunin, V.V., et al., *Crystal structure of the CorA Mg<sup>2+</sup> transporter*. Nature, 2006. **440**(7085): p. 833.
12. Matthies, D., et al., *Cryo-EM structures of the magnesium channel CorA reveal symmetry break upon gating*. Cell, 2016. **164**(4): p. 747-756.
13. Dalmas, O., et al., *Structural dynamics of the magnesium-bound conformation of CorA in a lipid bilayer*. Structure, 2010. **18**(7): p. 868-878.
14. Pfoh, R., et al., *Structural asymmetry in the magnesium channel CorA points to sequential allosteric regulation*. Proceedings of the National Academy of Sciences, 2012. **109**(46): p. 18809-18814.
15. Dalmas, O., et al., *Molecular mechanism of Mg<sup>2+</sup>-dependent gating in CorA*. 2014. **5**: p. 3590.
16. Barton, P.G. and F. Gunstone, *Hydrocarbon chain packing and molecular motion in phospholipid bilayers formed from unsaturated lecithins. Synthesis and properties of sixteen positional isomers of 1, 2-dioctadecenoyl-sn-glycero-3-phosphorylcholine*. Journal of Biological Chemistry, 1975. **250**(12): p. 4470-4476.
17. Borhani, D.W., et al., *Crystal structure of truncated human apolipoprotein AI suggests a lipid-bound conformation*. Proceedings of the National Academy of Sciences, 1997. **94**(23): p. 12291-12296.
18. Bayburt, T.H., Y.V. Grinkova, and S.G. Sligar, *Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins*. Nano Letters, 2002. **2**(8): p. 853-856.



19. Denisov, I.G., et al., *Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size*. Journal of the American Chemical Society, 2004. **126**(11): p. 3477-3487.
20. Denisov, I.G., et al., *Thermotropic phase transition in soluble nanoscale lipid bilayers*. The Journal of Physical Chemistry B, 2005. **109**(32): p. 15580-15588.
21. Shih, A.Y., et al., *Coarse grained protein– lipid model with application to lipoprotein particles*. The Journal of Physical Chemistry B, 2006. **110**(8): p. 3674-3684.
22. Shih, A.Y., et al., *Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations*. Journal of structural biology, 2007. **157**(3): p. 579-592.
23. Arkhipov, A., P.L. Freddolino, and K. Schulten, *Stability and dynamics of virus capsids described by coarse-grained modeling*. Structure, 2006. **14**(12): p. 1767-1777.
24. Arkhipov, A., Y. Yin, and K. Schulten, *Four-scale description of membrane sculpting by BAR domains*. Biophysical journal, 2008. **95**(6): p. 2806-2821.
25. de Jong, D.H., et al., *Improved parameters for the martini coarse-grained protein force field*. 2012. **9**(1): p. 687-697.
26. Marrink, S.J., et al., *The MARTINI force field: coarse grained model for biomolecular simulations*. The journal of physical chemistry B, 2007. **111**(27): p. 7812-7824.
27. Alessandri, R., F. Grünewald, and S.J. Marrink, *The Martini Model in Materials Science*. arXiv preprint arXiv:2012.07194, 2020.
28. Souza, P.C., et al., *Martini 3: a general purpose force field for coarse-grained molecular dynamics*. Nature methods, 2021. **18**(4): p. 382-388.
29. Bahar, I., A.R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*. Folding and Design, 1997. **2**(3): p. 173-181.
30. Boas, F.E. and P.B. Harbury, *Potential energy functions for protein design*. Current Opinion in Structural Biology, 2007. **17**(2): p. 199-204.
31. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. Nat Struct Biol, 2002. **9**(9): p. 646-52.
32. Biro, J., *Amino acid size, charge, hydrophathy indices and matrices for protein structure analysis*. Theoretical Biology and Medical Modelling, 2006. **3**(1): p. 15.
33. K., B., *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*. 1995, Oxford University Press.
34. Kalipatnapu, S. and A. Chattopadhyay, *Membrane protein solubilization: recent advances and challenges in solubilization of serotonin1A receptors*. IUBMB Life, 2005. **57**(7): p. 505-12.
35. Linke, D., *Detergents: an overview*, in *Methods in enzymology*. 2009, Elsevier. p. 603-617.
36. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs*. FEBS letters, 2010. **584**(9): p. 1721-1727.

37. Liang, B. and L.K. Tamm, *NMR as a Tool to Investigate Membrane Protein Structure, Dynamics and Function*. Nature structural & molecular biology, 2016. **23**(6): p. 468.
38. Borhani, D.W., et al., *Crystal structure of truncated human apolipoprotein AI suggests a lipid-bound conformation*. Proceedings of the National Academy of Sciences, USA, 1997. **94**(23): p. 12291-12296.
39. Denisov, I.G., et al., *Thermotropic phase transition in soluble nanoscale lipid bilayers*. Journal of Physical Chemistry B, 2005. **109**(32): p. 15580-15588.
40. Koppaka, V., et al., *The structure of human lipoprotein AI Evidence for the "belt" model*. Journal of Biological Chemistry, 1999. **274**(21): p. 14541-14544.
41. Ravula, T., N.Z. Hardin, and A. Ramamoorthy, *Polymer nanodiscs: Advantages and limitations*. Chemistry and physics of lipids, 2019. **219**: p. 45-49.
42. Dörr, J.M., et al., *Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K<sup>+</sup> channel: The power of native nanodiscs*. Proceedings of the National Academy of Sciences, 2014. **111**(52): p. 18607-18612.
43. Tall, A.R. and D.M. Small, *Solubilisation of phospholipid membranes by human plasma high density lipoproteins*. Nature, 1977. **265**(5590): p. 163.
44. Damiati, S., et al., *Albumin-bound nanodiscs as delivery vehicle candidates: Development and characterization*. Biophysical Chemistry, 2019. **251**: p. 106178.
45. Sahoo, B.R., et al., *Self-Assembly of Polymer-Encased Lipid Nanodiscs and Membrane Protein Reconstitution*. The Journal of Physical Chemistry B, 2019. **123**(21): p. 4562-4570.
46. Marrink, S.J., J. Risselada, and A.E. Mark, *Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model*. Chemistry and Physics of Lipids, 2005. **135**(2): p. 223-244.
47. Leekumjorn, S. and A.K. Sum, *Molecular characterization of gel and liquid-crystalline structures of fully hydrated POPC and POPE bilayers*. J Phys Chem B, 2007. **111**(21): p. 6026-33.
48. Kranenburg, M., M. Venturoli, and B. Smit, *Phase behavior and induced interdigitation in bilayers studied with dissipative particle dynamics*. Journal of Physical Chemistry B, 2003. **107**(41): p. 11491-11501.
49. Müller, S., et al., *Mixing behaviour of bilayer-forming phosphatidylcholines with single-chain alkyl-branched bolalipids: effect of lateral chain length*. Biophysical Chemistry, 2019. **244**: p. 1-10.
50. Kariyazono, H., et al., *Formation of stable nanodiscs by bihelical apolipoprotein A - I mimetic peptide*. Journal of Peptide Science, 2016. **22**(2): p. 116-122.
51. Shaw, A.W., M.A. McLean, and S.G. Sligar, *Phospholipid phase transitions in homogeneous nanometer scale bilayer discs*. Febs Letters, 2004. **556**(1-3): p. 260-264.
52. Jo, S., et al., *CHARMM-GUI PDB manipulator for advanced modeling and simulations of proteins containing nonstandard residues*, in *Advances in protein chemistry and structural biology*. 2014, Elsevier. p. 235-265.

53. Lee, J., et al., *CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field*. Journal of Chemical Theory and Computation, 2015. **12**(1): p. 405-413.
54. Qi, Y., et al., *CHARMM - GUI Nanodisc Builder for modeling and simulation of various nanodisc systems*. Journal of computational chemistry, 2019. **40**(7): p. 893-899.
55. Humphrey, W., A. Dalke, and K. Schulten, *VMD: Visual molecular dynamics*. Journal of Molecular Graphics, 1996. **14**(1): p. 33-38.
56. James C. Phillips, et al., *Scalable molecular dynamics with NAMD*. Journal of Computational Chemistry, 2005. **26**: p. 21.
57. Martyna, G.J., D.J. Tobias, and M.L. Klein, *Constant pressure molecular dynamics algorithms*. Journal of Chemical Physics, 1994. **101**(5): p. 4177-4189.
58. Feller, S.E., et al., *Constant pressure molecular dynamics simulation: the Langevin piston method*. Journal of Chemical Physics, 1995. **103**(11): p. 4613-4621.
59. Seeber, M., et al., *Wordom: a program for efficient analysis of molecular dynamics simulations*. Bioinformatics, 2007. **23**(19): p. 2625-2627.
60. Svetlovics, J.A., S.A. Wheaten, and P.F. Almeida, *Phase separation and fluctuations in mixtures of a saturated and an unsaturated phospholipid*. Biophys J, 2012. **102**(11): p. 2526-35.
61. Nath, A., W.M. Atkins, and S.G. Sligar, *Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins*. Biochemistry, 2007. **46**(8): p. 2059-2069.
62. Siuda, I. and D.P. Tieleman, *Molecular models of nanodiscs*. Journal of chemical theory and computation, 2015. **11**(10): p. 4923-4932.
63. Shih, A.Y., et al., *Molecular dynamics simulations of discoidal bilayers assembled from truncated human lipoproteins*. Biophysical journal, 2005. **88**(1): p. 548-556.
64. Furukawa, Y. and N. Nukina, *Functional diversity of protein fibrillar aggregates from physiology to RNA granules to neurodegenerative diseases*. Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2013. **1832**(8): p. 1271-1278.
65. Bai, Y., Q. Luo, and J. Liu, *Protein self-assembly via supramolecular strategies*. Chemical Society Reviews, 2016. **45**(10): p. 2756-2767.
66. McManus, J.J., et al., *The physics of protein self-assembly*. Current opinion in colloid & interface science, 2016. **22**: p. 73-79.
67. Sgarbossa, A., *Natural biomolecules and protein aggregation: emerging strategies against amyloidogenesis*. International journal of molecular sciences, 2012. **13**(12): p. 17121-17137.
68. Sun, H., et al., *Nanostructures based on protein self-assembly: From hierarchical construction to bioinspired materials*. Nano Today, 2017. **14**: p. 16-41.
69. Garcia-Seisdedos, H., et al., *Proteins evolve on the edge of supramolecular self-assembly*. Nature, 2017. **548**(7666): p. 244-247.

70. Pandey, R.B., B.L. Farmer, and B.S. Gerstman, *Self-assembly dynamics for the transition of a globular aggregate to a fibril network of lysozyme proteins via a coarse-grained Monte Carlo simulation*. AIP Advances, 2015. **5**(9): p. 092502.
71. Yang, L., et al., *Self - assembly of proteins: Towards supramolecular materials*. Chemistry - A European Journal, 2016. **22**(44): p. 15570-15582.
72. Hmiel, S., et al., *Magnesium transport in Salmonella typhimurium: genetic characterization and cloning of three magnesium transport loci*. Journal of bacteriology, 1989. **171**(9): p. 4742-4751.
73. Maguire, M.E., *MgtA and MgtB: prokaryotic P-type ATPases that mediate Mg<sup>2+</sup> influx*. Journal of bioenergetics and biomembranes, 1992. **24**(3): p. 319-328.
74. KEHRES, D.G., C.H. LAWYER, and M.E. MAGUIRE, *The CorA magnesium transporter gene family*. Microbial & comparative genomics, 1998. **3**(3): p. 151-169.
75. Payandeh, J., et al., *Probing structure-function relationships and gating mechanisms in the CorA Mg<sup>2+</sup> transport system*. Journal of Biological Chemistry, 2008. **283**(17): p. 11721-11733.
76. Payandeh, J. and E.F. Pai, *A structural basis for Mg<sup>2+</sup> homeostasis and the CorA translocation cycle*. The EMBO journal, 2006. **25**(16): p. 3762-3773.
77. Neale, C., et al., *Hydrophobic gating of ion permeation in magnesium channel CorA*. PLoS computational biology, 2015. **11**(7).
78. Kitjaruwankul, S., et al., *Inner and Outer Coordination Shells of Mg<sup>2+</sup> in CorA Selectivity Filter from Molecular Dynamics Simulations*. The Journal of Physical Chemistry B, 2016. **120**(3): p. 406-417.
79. Chakrabarti, N., et al., *An iris-like mechanism of pore dilation in the CorA magnesium transport system*. Biophysical journal, 2010. **98**(5): p. 784-792.
80. Nordin, N., et al., *Exploring the structure and function of Thermotoga maritima CorA reveals the mechanism of gating and ion selectivity in Co<sup>2+</sup>/Mg<sup>2+</sup> transport*. Biochemical Journal, 2013. **451**(3): p. 365-374.
81. Kitjaruwankul, S., et al., *Asymmetry in structural response of inner and outer transmembrane segments of CorA protein by a coarse-grain model*. Journal of Chemical Physics, 2016. **145**(13).
82. Kitjaruwankul, S., et al., *Thermal-induced folding and unfolding of a transmembrane protein (CorA)*. Physica A: Statistical Mechanics and its Applications, 2018. **506**: p. 987-992.
83. Pandey, R.B. and B.L. Farmer, *Aggregation and network formation in self-assembly of protein (H3.1) by a coarse-grained Monte Carlo simulation*. J Chem Phys, 2014. **141**(17): p. 175103.
84. Betancourt, M.R. and D. Thirumalai, *Pair potentials for protein folding: choice of reference states and sensitivity of predicted native states to variations in the interaction schemes*. Protein Sci, 1999. **8**(2): p. 361-9.
85. Miyazawa, S. and R.L. Jernigan, *Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation*. Macromolecules, 1985. **18**(3): p. 534-552.

86. Miyazawa, S. and R.L. Jernigan, *Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading*. J Mol Biol, 1996. **256**(3): p. 623-44.
87. Tanaka, S. and H.A. Scheraga, *Medium- and Long-Range Interaction Parameters between Amino Acids for Predicting Three-Dimensional Structures of Proteins*. Macromolecules, 1976. **9**(6): p. 945-950.
88. Godzik, A., *Knowledge-based potentials for protein folding: what can we learn from known protein structures?* Structure, 1996. **4**(4): p. 363-366.
89. Huang, S.Y. and X. Zou, *Statistical mechanics - based method to extract atomic distance - dependent potentials from protein structures*. Proteins: Structure, Function, and Bioinformatics, 2011. **79**(9): p. 2648-2661.
90. Sorokina, I. and A. Mushegian, *Modeling protein folding in vivo*. Biology direct, 2018. **13**(1): p. 13.
91. Duncan, A.L., et al., *Protein crowding and lipid complexity influence the nanoscale dynamic organization of ion channels in cell membranes*. Sci Rep, 2017. **7**(1): p. 16647.
92. Zeindlhofer, V. and C. Schroder, *Computational solvation analysis of biomolecules in aqueous ionic liquid mixtures : From large flexible proteins to small rigid drugs*. Biophys Rev, 2018. **10**(3): p. 825-840.
93. Kitjaruwankul, S., et al., *Thermal-induced folding and unfolding of a transmembrane protein (CorA)*. Physica A: Statistical Mechanics and its Applications, 2018. **506**: p. 987-992.
94. S. Kitjaruwankul, P.B., Sunita Subedi Paudel, Warin Jetsadawisut, Pornthep Sompornpisut, R.B. Pandey, *Conformational modulation of transmembrane segments of a protein (CorA) by effective media*. Int. J. Vir. Inf. Dis., 2017. **2**: p. 8-13.
95. Beveridge, D.L., *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, ed. K. Binder. 1995, New York: Oxford University Press.
96. R. B. Pandey, D.J.J., and B. L. Farmer, *Preferential binding effects on protein structure and dynamics revealed by coarse-grained Monte Carlo simulation*. J. Chem. Phys, 2017. **146**: p. 195101.
97. Warin Rangubpit, S.K., Panisak Boonamnaj, Pornthep Sompornpisut, Ras B. Pandey *Globular bundles and entangled network of proteins (CorA) by a coarse-grained Monte Carlo simulation*. AIMS Biophysics, 2019. **6**(2): p. 68-82.
98. Kersey, C.M., P.A. Agyemang, and C.K. Dumenyo, *CorA, the magnesium/nickel/cobalt transporter, affects virulence and extracellular enzyme production in the soft rot pathogen Pectobacterium carotovorum*. Molecular plant pathology, 2012. **13**(1): p. 58-71.
99. Maguire, M.E., *Magnesium transporters: properties, regulation and structure*. Front Biosci, 2006. **11**: p. 3149-3163.
100. Smith, R.L. and M.E. Maguire, *Distribution of the CorA Mg<sup>2+</sup> transport system in gram-negative bacteria*. Journal of bacteriology, 1995. **177**(6): p. 1638-1640.
101. Moomaw, A.S. and M.E. Maguire, *The unique nature of Mg<sup>2+</sup> channels*. Physiology, 2008. **23**(5): p. 275-285.

102. Schweyen, R.J. and E.M. Froschauer, *CorA-Mrs2-Alr1 Superfamily of Mg<sup>2+</sup> Channel Proteins*, in *New Perspectives in Magnesium Research*. 2007, Springer. p. 46-54.
103. Palombo, I., D.O. Daley, and M. Rapp, *Why is the GMN motif conserved in the CorA/Mrs2/Alr1 superfamily of magnesium transport proteins?* *Biochemistry*, 2013. **52**(28): p. 4842-4847.
104. Rostkowski, M., et al., *Graphical analysis of pH-dependent properties of proteins predicted using PROPKA*. *BMC structural biology*, 2011. **11**(1): p. 1-6.
105. Jo, S., T. Kim, and W. Im, *Automated builder and database of protein/membrane complexes for molecular dynamics simulations*. *PloS one*, 2007. **2**(9): p. e880.
106. Lamoureux, G., et al., *A polarizable model of water for molecular dynamics simulations of biomolecules*. *Chemical Physics Letters*, 2006. **418**(1-3): p. 245-249.
107. Huang, J. and A.D. MacKerell, *CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data*. *Journal of Computational Chemistry*, 2013. **34**(25): p. 2135-2145.
108. Martyna, G.J., D.J. Tobias, and M.L. Klein, *Constant-Pressure Molecular-Dynamics Algorithms*. *Journal of Chemical Physics*, 1994. **101**(5): p. 4177-4189.
109. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems*. *The Journal of chemical physics*, 1993. **98**(12): p. 10089-10092.
110. Miyamoto, S. and P.A. Kollman, *Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models*. *Journal of computational chemistry*, 1992. **13**(8): p. 952-962.
111. Kräutler, V., W.F. Van Gunsteren, and P.H. Hünenberger, *A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations*. *Journal of computational chemistry*, 2001. **22**(5): p. 501-508.
112. Wedekind, J.E., G.H. Reed, and I. Rayment, *Octahedral coordination at the high-affinity metal site in enolase: Crystallographic analysis of the MgII-enzyme complex from yeast at 1.9 Å resolution*. *Biochemistry*, 1995. **34**(13): p. 4325-4330.

## VITA

**NAME** Warin Rangubpit

**DATE OF BIRTH** 11 December 1991

**PLACE OF BIRTH** Bangkok

**INSTITUTIONS ATTENDED**

2017-2020:  
Chulalongkorn University, Thailand  
Aoyama Gakuin University, Japan (2019, June-July)  
University of Southern Mississippi, United State (2019, November - 2020, May)

2014-2016:  
Chulalongkorn University, Thailand

2009-2013:  
Chulalongkorn University, Thailand

2003-2008:  
Bodindecha (Sing Singhaseni) 2, Thailand

**HOME ADDRESS** Bangrakpattana Bangbuatong Nontaburee 11110

**PUBLICATION**

Rangubpit, W., Pandey R.B., Sompornpisut P., Thermal-induced unfolding-refolding of a nucleocapsid COVN protein. *AIMS Biophysics* 2021, 8(1), 103-110.

Rangubpit, W., Pandey R.B., Sompornpisut P., Thermally induced structural organization of nanodiscs by coarse-grained molecular dynamics simulations. *Biophysical Chemistry* 2020, 267, 106464.

Rangubpit, W., Pandey R.B., Sompornpisut P., Temperature Driven Shape Transformation of Nanodiscs by Coarse-Grained Molecular Dynamics Simulations. *Biophysical Journal* 2020, 118(3), 139a.

Rangubpit, W., Kitjaruwankul S., Sompornpisut P., Pandey R.B., Pinning the conformation of a protein (CorA) in a solute matrix with selective binding. *Physica A: Statistical Mechanics and its Applications* 2020, 124823.

W. Rangubpit, P. Sompornpisut, R. B. Pandey, Structure and dynamics of aquaporin-1, Academic Press, 2019.

P. Boonamnaj, S. S. Paudel, W. Jetsadawisut, S. Kitjaruwankul, P. Sompornpisut, R. B. Pandey, Thermal-response of a protein (hHv1) by a coarse-grained MC and all-atom MD computer simulations, *Physica A: Statistical Mechanics and its Applications*, 2019.

W. Rangubpit, S. Kitjaruwankul, P. Boonamnaj, P. Sompornpisut, R. B. Pandey, Globular bundles and entangled network of proteins (CorA) by a coarse-grained Monte Carlo simulation, *AIMS Biophysics*, 2019.

W. Worawalai, T. Doungwichitkul, W. Rangubpit, P. Taweachat, P. Sompornpisut, P. Phuwapraisirisan, Furofuran lignans as a new series of antidiabetic agents exerting  $\alpha$ -glucosidase inhibition and radical scavenging: Semisynthesis, kinetic study and molecular modeling, *Bioorganic chemistry*, 2019.

S. Kitjaruwankul, P. Boonamnaj, S. S. Paudel, W. Jetsadawisut, P. Sompornpisut, R. B. Pandey, Thermal-induced folding and unfolding of a transmembrane protein (CorA), *Physica A: Statistical Mechanics and its Applications*, 2018.

S. Kitjaruwankul, P. Boonamnaj, S. S. Paudel, W. Jetsadawisut, P. Sompornpisut, R. B. Pandey, Conformational Modulation of Transmembrane Segments of a Protein (CorA) By Effective Media, *Int J Virol Infect Dis*, 2017.

W. Jetsadawisut, B. Nutho, A. Meeprasert, T. Rungrotmongkol, N. Kungwan, P. Wolschann, S. Hannongbua, Susceptibility of inhibitors against 3C protease of coxsackievirus A16 and enterovirus A71 causing hand, foot and mouth disease: A molecular dynamics study, *Biophysical chemistry*, 2016.

**AWARD RECEIVED**

Note: W. rangubpit = W. Jetsadawisut  
- The Chulalongkorn University Dusadi Phipat Scholarship  
- Short-Term Internship Program in the College of Science & Engineering, Aoyama Gakuin University, Japan