Computational Study of Alpha-globin Constant Spring Structure and Binding Affinity to Alpha Haemoglobin Stabilising Protein at an Early Phase of Haemoglobin Formation

Nawanwat Chainuwong Pattaranggoon

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of
Master of Science in Biomedical Sciences
Prince of Songkla University
2016

Computational Study of Alpha-globin Constant Spring Structure and Binding Affinity to Alpha Haemoglobin Stabilising Protein at an Early Phase of Haemoglobin Formation

Nawanwat Chainuwong Pattaranggoon

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Biomedical Sciences
Prince of Songkla University
2016

Thesis Title	Computational Study of Alpha-globin Constant Spring Structure and Binding Affinity to Alpha Haemoglobin Stabilising Protein at an Early Phase of Haemoglobin Formation
Author	Mr Nawanwat Chainuwong Pattaranggoon
Major Programme	Biomedical Sciences

Major Advisor

.....
(Varomyalin Tipmanee, PhD)

Co-advisor:

.....
(Assoc.Prof.Paramee Thongsuksai, M.D.)

.....
(Thanyada Rungrotmongkol, Ph.D.)

Examining Committee:

.....Chairperson
(Assoc.Prof.Kiattawee Choowongkomon, Ph.D.)

.....
(Varomyalin Tipmanee, PhD)

.....
(Assoc.Prof.Paramee Thongsuksai, M.D.)

.....
(Thanyada Rungrotmongkol, Ph.D.)

.....
(Asst.Prof.Chamnong Nopparatana, Ph.D.)

The Graduate School, Prince of Songkla University, has approved this thesis as partial fulfilment of the requirements for the Master of Science Degree in Biomedical Sciences

.....
(Assoc.Prof.Teerapol Srichana, Ph.D.)
Dean of Graduate School

This is to certify that the work here submitted is the result of the candidates own investigations. Due acknowledgement has been made of any assistance received.

..... Signature
(Varomyalin Tipmanee, PhD)
Major Advisor

..... Signature
(Mr Nawanwat C Pattaranggoon)
Candidate

I hereby certify that this work has not been accepted in substance for any degree, and is not being currently submitted in candidature for any degree.

..... Signature
(Mr Nawanwat C Pattaranggoon)
Candidate

ชื่อวิทยานิพนธ์ การศึกษาเชิงคำนวณทางโครงสร้างของแอลฟาโกลบินคอนสแตนท์สปริงและสัมพรรคภาพการยึดจับกับโปรตีนช่วยรักษาเสถียรภาพของแอลฟาฮีโมโกลบินในระยะเริ่มแรกของการสร้างโมเลกุลฮีโมโกลบิน
ผู้เขียน นายนวันวัจน์ ไชยนุวงศ์ ภัทรางกูน
สาขาวิชา ชีวเวชศาสตร์
ปีการศึกษา 2558

บทคัดย่อ

บทนำ ความผิดปกติชนิดฮีโมโกลบินเอ็ชคอนสแตนท์สปริงเป็นโรคทางโลหิตที่ถ่ายทอดทางพันธุกรรม ความผิดปกตินี้สามารถพบความชุกได้ในประชากรไทยที่เป็นโรคโลหิตจางธาลัสซีเมียชนิดกลุ่มแอลฟาธาลัสซีเมีย สาเหตุของความผิดปกตินี้เกิดจากการกลายพันธุ์ที่รหัสหยุดบนยีนสร้างสายแอลฟาโกลบิน ทำให้มีการสังเคราะห์กรดอะมิโนในสายแอลฟาโกลบินเพิ่มขึ้น 31 กรดอะมิโน เรียกว่า หางคอนสแตนท์สปริง ส่งผลให้การสร้างฮีโมโกลบินไม่เสถียร โดยสมมติฐานของงานวิจัยนี้คือลำดับกรดอะมิโนที่สร้างเกินออกมาอาจมีผลต่อการเปลี่ยนแปลงของโครงสร้างและการเข้าจับกับโมเลกุลโปรตีนช่วยรักษาเสถียรภาพของสายแอลฟาฮีโมโกลบินในระยะเริ่มแรกของการสร้างโมเลกุลฮีโมโกลบิน อย่างไรก็ตามในปัจจุบันกลไกการสร้างฮีโมโกลบินเอจากกรณีการกลายพันธุ์ชนิดคอนสแตนท์สปริงบนสายแอลฟาโกลบินในแง่ของพลวัติเชิงโมเลกุล ยังไม่พบการศึกษาวิจัยและข้อสรุปอย่างเด่นชัด อีกทั้งโครงสร้างสามมิติของโปรตีนแอลฟาโกลบินคอนสแตนท์สปริงยังไม่มีการรายงานไว้ในฐานข้อมูลโครงสร้างสามมิติของโปรตีน ดังนั้นการวิจัยครั้งนี้จึงสนใจที่จะสร้างโครงสร้างโปรตีนชนิดนี้โดยวิธีการทำนายโครงสร้างสามมิติของโปรตีนด้วยเครื่องมือทางชีวสารสนเทศ ในการทำนายโครงสร้างสามมิติของโปรตีนแอลฟาโกลบินคอนสแตนท์สปริงได้ใช้โครงสร้างสามมิติของโปรตีนดีออกซีฮีโมโกลบินเอ ซึ่งมีการรายงานไว้แล้วในฐานข้อมูลโครงสร้างโปรตีนมาเป็นแม่แบบในการทำนายเพื่อจะศึกษาการยึดจับกับโปรตีนช่วยรักษาเสถียรภาพของแอลฟาฮีโมโกลบิน ในระยะเริ่มแรกของการสร้างโมเลกุลฮีโมโกลบินเอ
วัตถุประสงค์ ศึกษาการเปลี่ยนแปลงเชิงโครงสร้างจากผลกระทบของการกลายพันธุ์ชนิดคอนสแตนท์สปริงบนสายแอลฟาโกลบินเปรียบเทียบกับสายแอลฟาโกลบินปกติ ภายใต้สภาวะเทียบเคียงทางสรีรวิทยาของมนุษย์ และศึกษาสัมพรรคภาพการยึดจับของสายแอลฟาโกลบินคอนสแตนท์สปริงกับโปรตีนช่วยรักษาเสถียรภาพของแอลฟาฮีโมโกลบินในระยะเริ่มแรกของการสร้างโมเลกุลฮีโมโกลบินเอ
วิธีเชิงคำนวณ โครงสร้างสามมิติ ของโปรตีนแอลฟาโกลบินคอนสแตนท์สปริงได้ถูกทำนายด้วยเครื่องมือทางชีวสารสนเทศเพื่อสร้างโครงสร้างทุติยภูมิและสร้างโครงสร้างสามมิติขึ้น โดยใช้ PSIPRED NetSurfP และ CABS-fold ตามลำดับ จากนั้นโครงสร้างสามมิติของโปรตีนทุกโครงสร้างที่ศึกษาใช้โมดูล PMEMD ในการคำนวณเชิงพลวัติ จากโปรแกรม AMBER12 ที่ความ

เข้มข้นโซเดียมคลอไรด์ 0.15 โมลาร์ กำหนดอุณหภูมิคงที่ 310 แคลวิน และความดันคงที่ 1.031 ความดันบรรยากาศมาตรฐาน การวิเคราะห์การเปลี่ยนแปลงโครงสร้างใช้โมดูล PTRAJ CPPTRAJ และโปรแกรมที่ได้เขียนขึ้นเอง โปรแกรม VMD ได้ใช้ในการแสดงภาพโครงสร้างสามมิติของโปรตีนและวิเคราะห์สมบัติบางประการของโปรตีน ส่วนการคำนวณพลังงานเสรีการยึดจับสัมพัทธ์โดยใช้สคริปส์ MMPBSA จากโปรแกรม AMBER12

ผลการศึกษา โครงสร้างสามมิติของ 31 กรดอะมิโน ที่ได้จากการทำนายนั้นหลังจากทำการตรวจสอบความผิดพลาดของโครงสร้าง ทั้งการตรวจสอบความผิดพลาดของมุมทอร์ชันด้วยกราฟรามาชานดราน และการเปรียบเทียบความเป็นไปได้เชิงทฤษฎีของการม้วนพับของโครงสร้างสามมิติโปรตีน พบว่าสามารถยอมรับโครงสร้างสามมิตินี้ได้และนำหางคอนสแตนท์สปริงนี้ต่อกับโครงสร้างสามมิติหลักของสายแอลฟาโกลบินเพื่อสร้างโครงสร้างสามมิติสายแอลฟาโกลบินคอนสแตนท์สปริง แบบจำลองพลวัติเชิงโมเลกุลของสายแอลฟาโกลบินคอนสแตนท์สปริง มีค่า RMSD ใกล้เคียงกับโครงสร้างสามมิติอ้างอิงปกติ อีกทั้งความหนาแน่นประจุเชิงผิวบริเวณเข้าจับสายแอลฟาโกลบินปกติมีความคล้ายกันเมื่อเปรียบเทียบกับสายแอลฟาโกลบินคอนสแตนท์สปริง การเปรียบเทียบระยะทางจากทุก ๆ กรดอะมิโนในโครงสร้างถึงโมเลกุลฮีมของสายแอลฟาโกลบินคอนสแตนท์สปริง พบมีค่าไม่แตกต่างจากสายแอลฟาโกลบินปกติ และส่วนหางคอนสแตนท์สปริงที่ยาวออกมานั้นบ่งชี้ว่าไม่ส่งผลกระทบเชิงโครงสร้างต่อโครงสร้างโมโนเมอร์สามมิติหลักของสายแอลฟาโกลบินปกติและชนิดกลายพันธุ์ แต่อย่างไรก็ตามโครงสร้างสามมิติของสายแอลฟาโกลบินสแตนท์สปริงแสดงให้เห็นถึงความผิดปกติของการจัดเรียงตัวทุติยภูมิบริเวณเข้าจับขณะเกิดพลวัติ 10 นาโนวินาที สุดท้าย นอกจากนั้นผลจากการวิเคราะห์สัมพรรคภาพการยึดจับพบว่าพลังงานเสรีการยึดจับสัมพัทธ์ของสายแอลฟาโกลบินคอนสแตนท์สปริงกับโปรตีนช่วยรักษาเสถียรภาพแสดงค่าแย่กว่าการยึดจับของสายแอลฟาโกลบินปกติ

สรุป ส่วนของ 31 กรดอะมิโน ที่สร้างเพิ่มออกมานั้นไม่ส่งผลในเชิงโครงสร้างโมโนเมอร์สามมิติและความหนาแน่นประจุเชิงผิวบริเวณเข้าจับของโมเลกุลโมโนเมอร์ที่กลายพันธุ์แต่มีผลต่อการกีดกันการเข้าจับของโปรตีนช่วยรักษาเสถียรภาพของแอลฟาฮีโมโกลบินต่อบริเวณเข้าจับของสายแอลฟาโกลบินที่กลายพันธุ์เพื่อเป็นโมเลกุลไดเมอร์ ซึ่งอาจส่งผลต่อเสถียรภาพของโครงสร้างโมโนเมอร์แอลฟาโกลบินคอนสแตนท์สปริงและอาจทำให้เกิดการตกตะกอนบริเวณผนังเม็ดเลือดแดง ก่อให้เกิดอนุมูลอิสระทำลายเม็ดเลือดแดงได้ โดยสามารถตรวจพบจากสเมียร์เลือด ซึ่งจากบทสรุปนี้อาจอธิบายว่าหางคอนสแตนท์สปริงเกี่ยวข้องกับกลไกภาวะโลหิตจางธาลัสซีเมียชนิดฮีโมโกลบินเอ็ชคอนสแตนท์สปริง

Thesis Title	Computational Study of Alpha-globin Constant Spring Structure and Binding Affinity to Alpha Haemoglobin Stabilising Protein at an Early Phase of Haemoglobin Formation
Author	Mr Nawanwat Chainuwong Pattaranggoon
Major Programme	Biomedical Sciences
Academic Year	2015

ABSTRACT

Introduction: Haemoglobin H Constant Spring thalassaemia disease (HbH CS disease) is a hereditary haematologic disease that has a very high prevalence among α-type one found in Thai thalassaemia patients. The tetrameric HbCS-globin comes from the abnormal elongated α-globin, containing 31 amino acid extension in its structure so called a CS-tail. The mutant α^{CS}-globin structure may cause unstable Hb formation. Herein this thesis has focussed on early phase where mutant α^{CS}-globin binds to alpha haemoglobin stabilising protein (AHSP) within a hypothesis is the CS-tail may affect main α-globin conformation and interfere for binding site on α^{CS}-globin while α^{CS}·AHSP formation. However, the tetrameric HbA formation mechanism is underleased remains unclear and no report about these. In such a case 3-D structure of mutant α^{CS}-globin is necessary in order to insight to the mechanism. Unfortunately, a 3-D α^{CS}-globin structure has not yet been reported, so in this work we attempt to construct 3-D structures α^{CS}-globin by bioinformatics tools using α-globin from deoxyHbA as a structural template and study the protein-protein binding for mutant α^{CS}-globin at early phase of HbA formation
Objective: To study the conformations change of elongated C-terminal α^{CS}-globin subunit mutation comparing to wild-type α-globin in the dynamics physiological environment, and to study the structural effects associated with the dimer formation in case of dimeric α^{CS}·AHSP structure.
Computational method: Mutant structures were predicted the 2-D and 3-D of CS-tail using bioinformatics tools for PSIPRED, NetSurfP, and CABS-fold, respectively. All MD simulation was carried out with PMEMD module in AMBER12 package under a condition of 0.15 M NaCl and 310 K at standard pressure 1.031 atm. The protein conformations were eventually analysed using PTRAJ, CPPTRAJ module, and some manually written programmes. A structure visualisation was performed and analysed of some properties using packages on VMD programme. The relative binding free energy was computed by the MMPBSA script in AMBER12.
Results: 3-D CS-tail structures had been successfully predicted using various tools. The final chosen CS-tail structure had been justified for the 2-D and 3-D geomet-

rical error detection by Ramachandran plot. The result consistently showed with theory of protein folding. Later the CS-tail was docked to the C-terminus of main α^{WT}-globin for creating predicted α^{CS}-globin. From monomeric structure analysis results of trajectory analysis had the similarly RMSD to the reference structure, and found similar surface charge distribution on binding site for wild-type to the mutant. Moreover, the internal interaction distance amongst each amino acids residues to haem molecule also demonstrated that wild-type was not different from mutant. These are indicated that CS-tail may not strong effect to main conformation. Unfortunately the time evolution of secondary structure of mutant protein had a few changed locating on binding site in the last 10 ns trajectory. For dimeric binding mode found the relative binding free energy of mutant complex was shown poor binding than wild-type complex.

Conclusion: The CS-tail may neither effect monomeric main-globin-structure conformation nor surface charge distribution on the binding site. However, CS-tail seems to interfere with interactions on the mutant globin binding site to AHSP and influence mutant globin binding affinity to AHSP at an early phase of Hb formation. There is the presence of unstable momoneric α^{CS}-globin which may be bound to erythrocytic membrane and made the oxidative damage. Therefor, the oxidative damage in erythrocytic membrane was detected on a blood smear. The evidences reveal molecular effect was due to the CS-tail on anaemia occurrence via HbH CS diseases.

ACKNOWLEDGEMENTS

Though the following thesis is an individual work, I could never have reached the heights or explored the depths without the helps, supports, guidances and efforts of a lot of people. It is a pleasure to show my appreciation here.

First of all, my deepest gratitude goes to my both very beloved mother, Khunmae Sopa Chainuwong, and father, Khunphor Ugrid Pattaranggoon for their unflagging love and unconditional supports throughout my *life* and studies. Both of you have been making me live the most unique, magical, and carefree their son that has made me who am I now.

Foremost, I want to thank my supervisor Varomyalin Tipmanee, PhD. It has been an honour to be his first MSc student. He has taught me, both consciously and unconsciously, how good experiment of computational modelling is done. I appreciate all his contributions of time, ideas, and knowledge to make my thesis experience productive and stimulating. I am thankful to him for spending a lot of time to correct my English writing and prove my thesis. I am also thankful for the excellent example he has provided as a successful in molecular-dynamics-simulation scientist and protein modeller. Special thanks belong to Asst. Prof. Chamnong Nopparattana, Ph.D., who has given the best research question and guidance for molecular biology of thalassaemia disorders. I should like to show special courtesy to my co-supervisors, Assoc. Prof. Paramee Thongsuksai, M.D. and Thanyada Rungrotmongkol, Ph.D., who has given me necessary advices and supported until the research finish. Assoc. Prof. Paramee Thongsuksai, M.D. is also acknowledged again for giving my opportunity in MSc programme in BioMed, PSU and shapes me into a proper person. In addition, Assoc. Prof. Surasak Sangkhathat Na Aayudhaya (P'A), M.D., Ph.D. is my idol and inspires me to carry on the project as well as moral support and illumination during my period of study. Thank you to Asst. Prof. Kanyanatt Kanokwiroon, Ph.D., Kamonnut Singkhamanan, Ph.D., and especially to Raphatphorn Navakanitworakul, Ph.D., she is not only lecturer, but also sister in some time that I got it off my chest. For my thesis, I offer my sincerest gratitude to committee members exclusively Assoc .Prof. Kiattawee Choowongkomon, Ph.D. for their time, interest, and helpful comments. I should like to give an appreciations to all stiffs in the Department of Biomedical Sciences, Faculty of Medicine, PSU especially Khun Kanokon Khannui (P'Jai), Khun Panwasri Sanengsuwan (P'Not) and Khun Mayuree Boonrach (P'Aoy), who alway gives the recycle papers for me to print papers

I am also indebted to all lecturers and teachers in curricular instruction. I wound like to show the special politeness to mentor when I was undergraduate study, Manit Nuinoon, Ph.D. is acknowledged as the first important teacher illustrating

knowledge of thalassaemia disorders and others blood disorders. He is also a person inspire me to study in higher education level.

My time at BioMed, PSU is made enjoyable in large part due to the many friends and Varomyalin's groups that became a part of my life. I should never forget to my BioMed's friends, Surada Satthakarn, Ph.D., Somruedee Yorsin, (Ph.D.), Areerat Hnoonual, (Ph.D.), Siriphorn Chimplee, (Ph.D.), Rassanee Bissanum, Somrudee Nunun, Kesara Nittayaboon, Prerit U Aryal, and close friend Chonticha Romyasamit for their friendship. I should like to thankful to Korawit Kanjana, (Ph.D.) at Mahidol University who never feels annoyed and grumble during I need him help to download many papers importantly to my thesis. Thank you much to Nicha Khiaorit, a dental student who is loaned a textbook of Proteins: Structure and Function by Whitford D to help me read and understanded more via proteins structure.

Furthermore, I have offered my special thanks to my dearest friend, *KT* who always gives me hearty supports when I have depressed. A spiritual support from *KT* is also indispensable to uphold me to accomplish the thesis. I should never disremember all the chats and moments of happiness I got. And I should like to thanks much to Ugrid Sanggaraniti who has kept encouraging me when I bumped into difficulty during the writing this thesis.

Finally, in particular, I should also like to express my gratitude to Graduate School, PSU who gave me the support funding for MSc thesis.

Nawanwat C Pattaranggoon

TABLE OF CONTENTS

Abstract (Thai) **v**

Abstract (English) **vii**

Acknowledgements **ix**

Table of Contents **xi**

List of Tables **xiii**

List of Illustrations **xiv**

List of Abbreviations and symbols **xvi**

1 Introduction **1**
1.1 Background and motivation 1
1.2 Objectives 2

2 Literature reviews and theories **4**
2.1 Molecular dynamics (MD) simulation 4
2.1.1 Equations of motion for MD simulations 4
2.1.2 The numerical techniques 4
2.1.3 Leap-Frog algorithm 6
2.1.4 Periodic boundary condition (PBC) 6
2.1.5 Ensembles 7
2.1.6 Multi-scale modelling and simulation 8
2.2 Haemoglobin (Hb) 8
2.2.1 Genetics 9
2.2.2 Hb assembly 9
2.3 Alpha-globin 9
2.3.1 3-D structure 9
2.4 Alpha haemoglobin stabilising protein (AHSP) 10
2.4.1 Genetics and synthesis 11
2.4.2 ASHP function in Hb synthesis 11
2.5 Haemoglobin Constant Spring (HbCS) 11

3 Computational methodology **13**
3.1 Materials 13
3.1.1 Hardwares 13

3.1.2 3-D structures . . . 13
3.1.3 Bioinformatics tools . . . 13
3.1.4 Softwares . . . 14
3.2 Prediction of α-globin Constant Spring monomer (α^{CS}) . . . 15
3.3 Structural validation . . . 15
3.3.1 Ramachandran plot . . . 15
3.4 Preparation of α^{CS}·AHSP, and α^{WT}·AHSP dimer . . . 15
3.5 Simulation protocol . . . 17
3.6 Analysis of MD simulations . . . 19
3.7 Calculating binding energies . . . 20
3.8 Evolution of secondary structure analysis . . . 21
3.9 Data analysis . . . 22

4 Results and discussion **23**
4.1 3-D structure of CS-tail . . . 23
4.2 Mutagenesis of AHSP structure . . . 28
4.3 Trajectory analysis of MD simulations . . . 28
4.3.1 MD simulation of monomeric α^{CS}-globin . . . 29
4.4 Surface charge distribution of monomeric α^{CS}-globin . . . 36
4.5 Dimeric structure prediction of α^{CS}·AHSP and α^{WT}·AHSP . . . 39
4.6 Hydrogen bond analysis of dimeric α^{CS}·AHSP and α^{CS}·AHSP . . . 40
4.7 2-D RMSD . . . 46
4.8 Trajectory analysis of MD simulations of dimeric α^{CS}·AHSP . . . 49
4.9 Calculation of binding free energy . . . 49

5 Conclusion **52**

Bibliography **52**

Appendices **65**
A Proceeding . . . 65
B E-poster presentation . . . 70
C AMBER input files . . . 72

Vitae **84**

LIST OF TABLES

Table

3.1 Number of water molecules and counter ions used are applied for each protein simulated 18

4.1 Ramachandran plot values of predicted models using RAMPAGE . 23
4.2 Secondary structure predictions by NetSurfP v1.1 24
4.3 Backbone RMSD values of structures 29
4.4 Structural parameters of intermolecular hydrogen bonds in α^{WT}·AHSP dimer 43
4.5 Structural parameters of intermolecular hydrogen bonds in α^{CS}·AHSP dimer 43
4.6 Relative binding free energies MM/GBSA of binding modes of α^{WT}·AHSP and α^{CS}·AHSP dimer 50
4.7 Relative binding free energies MM/PBSA of binding modes of α^{WT}·AHSP and α^{CS}·AHSP 50
4.8 Entropy of binding modes of α^{WT}·AHSP and α^{CS}·AHSP 51

LIST OF ILLUSTRATIONS

Figure

1.1 Schematic flow chart of research methodology 3

2.1 Periodic boundary conditions in 2-D 7
2.2 Diagrammatic multi-scale modelling and simulation 8
2.3 DeoxyHbA and α-globin structure 10
2.4 Secondary structure of α-globin (PDB code: 2DN2 chain C) 10
2.5 Molecular chaperones and Hb synthesis 11
2.6 Predicted α^{CS}-globin structure 12

3.1 Amino acid sequences of α^{CS}-globin 14
3.2 Schematic flow chart of α^{CS}-globin prediction 16
3.3 Chemical structure of histidine 16
3.4 Schematic brief of MD simulation protocol 18

4.1 Schematic representation of secondary structure predictions using PSIPRED web-server of CS-tail 26
4.2 Secondary structure prediction of CS-tail by CASB-fold 26
4.3 Predicted CS-tail structure and Ramachandran plot of predicted-CS-tail (Model07) 27
4.4 Structural alignment of residue T137 to R141 between α^{WT}-globin and pre-CS-tail 28
4.5 Structural alignment of wild-type AHSP and mutant AHSP P30A . 28
4.6 Time evaluation of RMSD: α^{WT}, α^{CS} residue V1 to E172, α^{CS} residue V1 to R141, AHSP, α^{WT}·AHSP, and α^{CS}·AHSP 30
4.7 α^{WT}-globin snapshots from MD simulation 31
4.8 α^{CS}-globin snapshots from MD simulation 31
4.9 α^{WT}·AHSP dimer snapshots from MD simulation 32
4.10 α^{CS}·AHSP dimer snapshots from MD simulation 32
4.11 AHSP dimer snapshots from MD simulation 33
4.12 RMSF per residue number of α^{CS} structure 33
4.13 RMSF per residue of α^{WT}, α^{CS} residue V1 to E172, α^{CS} residue V1 to R141, α^{WT}·AHSP, and α^{CS}·AHSP 34
4.14 Distance analysis of α^{WT}, and α^{CS} 34
4.15 Time evolution of secondary structure of α^{WT}-globin and α^{CS}-globin 35
4.16 Electrostatic surface potential of monomeric α^{WT}-globin structure . 36
4.17 Electrostatic surface potential of monomeric α^{CS}-globin structure . 37
4.18 Electrostatic surface potential of monomeric AHSP structure 38

4.19 Diagrammatic of protein-protein interactions for reference α^{WT}·AHSP dimer (PDB code 1Z8U) 39
4.20 Geometrical arrangement of docked α^{WT}·AHSP and α^{CS}·AHSP structure 41
4.21 Diagrammatic of protein-protein interactions for α^{WT}·AHSP dimer and α^{CS}·AHSP 42
4.22 H-bond network of α^{WT}·AHSP and α^{CS}·AHSP 43
4.23 H-bond presence during the MD trajectory of α^{WT}·AHSP dimer . . 44
4.24 H-bond presence during the MD trajectory of α^{CS}·AHSP dimer . . 45
4.25 2-D RMSD plot in simulations of monomeric AHSP 46
4.26 2-D RMSD plot in simulations of monomeric α^{WT}-globin 47
4.27 2-D RMSD plot in simulations of monomeric α^{CS}-globin 47
4.28 2-D RMSD plot in simulations of dimeric α^{WT}·AHSP 48
4.29 2-D RMSD plot in simulations of dimeric α^{CS}·AHSP 48

LIST OF ABBREVIATIONS AND SYMBOLS

Acronyms

CS	Constant Spring
COM	Centre-of-mass
Hb	Haemoglobin molecule
HbA	Adult haemoglobin
MD	Molecular dynamics
Native	Wild-type of globin molecule
NPT	Isobaric-isothermal ensemble
NVT	Canonical ensemble
WT	Wild-type
RMSD	Root-mean-square deviation
RMSF	Root-mean-square fluctuation
VMD	Molecular visualiser: Visual Molecular Dynamics
AMBER	Simulation programme: Assisted Model Building with Energy Refinement
2-D	Two dimensional
3-D	Three dimensional
X-ray	X-ray crystallography
MM/PBSA	Molecular mechanics Poisson-Boltzmann Surface Area
MM/GBSA	Molecular mechanics Generalised Born Surface Area

Greek letters and units

α^{WT}	wild-type alpha-globin: α^{WT}-globin
α^{CS}	Alpha-globin Constant Spring: α^{CS}-globin
α_1	Alpha-globin subunit 1, translated from *HBA1*
α_2	Alpha-globin subunit 2, translated from *HBA2*
β_1	Beta-globin subunit 1
β_2	Beta-globin subunit 2
fs	Femtosecond
ns	Nanosecond
ps	Picosecond
K	Kelvin
k	Constant force springs
Å	Angstrom or 10^{-10} m

Amino acids

Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid (Aspartate)
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid (Glutamate)
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Ser	S	Serine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine
Asx	B	Aspartic acid or Asparagine
Glx	Z	Glutamine or Glutamic acid
Xaa	X	Any amino acid

CHAPTER 1

INTRODUCTION

1.1 Background and motivation

Nowadays, biological techniques are associated in various biomedical researchers. Understanding the molecular and cellular mechanisms of disease is one of the key goals of modern biomedical researches by using computational biology and Bioinformatics. It is critical to dissect the initiation, progression, and dissemination of human diseases, and vital for the identification of appropriate therapeutic targets. Structural Bioinformatics become common to analyse and predict the three-dimensional (3-D) structure of biological macromolecules such as proteins, RNA, and DNA. Proteins are molecular devices, in nanometre to micrometre scales, where biological function is work. Proteins are the building blocks of all cells in our bodies and all living things. The information crucial to proceed life is controlled by the DNA molecule, the dynamic process of life maintenance, defence, replication and reproduction are carried out by the proteins.

Molecular functions are described the burdens performed by individual proteins and can be broadly divided into twelve subcategories; namely cellular processes, metabolism, DNA modification/replication, cell-cell communication, intracellular signalling, protein folding/degradation, transport, multifunctional proteins, cytoskeletal/structural, defence and immunity, and miscellaneous functions. A remarkable fact is that all tasks they can perform are based on a common principle that a protein is formed by 20 amino acids. That is the reason why studying proteins, their composition, structure, dynamics and function, is so important. All protein functions are dependent on its structure, which, in turn, depends on physical and chemical parameters. Multidisciplinary subjects involve studying these molecules; physical, chemical, classical biological, mathematical and informatics sciences have been running together in a new area known as bioinformatics to allow a further level of knowledge about life organisation. This thesis is focussed on a structural biology, which is a branch of bioinformatics by using the molecular dynamics simulation method to study the interesting protein behaviours. An available bioinformatics tool for understanding function/structure applied in this thesis is RCSB Protein Data Bank (PDB) database of protein. As of October, 2015 there are 104,942 are proteins from 112,968 structures in the RCSB/PDB.

Herein, this work has concentrated on computational study regarding mechanism of molecular assembly in the early phase of alpha haemoglobin stabilising protein (AHSP) binds to alpha-globin (α-globin) subunit, in case of alpha-globin

Constant Spring (α^{CS}-globin) mutant. The α^{CS}-globin is a structural abnormal Hb which is generated 31 amino-acid-residues longer than wide-type α-globin, so called CS-tail. A structurally abnormal Hb is one of haemoglobinopathy disorders which causes anaemias and eventually leads to unstable α^{CS}-globin. The later precipitates and causes oxidative damage in developing red cells (causing dyserythropoiesis) and in mature red cells (causing haemolysis)[1]. Precipitated unstable α^{CS}-globin play a role in HbA formation via imbalanced quantity of globin protein. The abnormality is a quantitative defect in the biosynthesis of HbA for thalassaemia disorders. Recently, it has been shown that highly expressed protein called AHSP can act as a chaperon for free α-globin and prevent their precipitation[1,2]. Free α-globins are highly unstable in red blood cell, which formed Heinz bodies[3,4]. Heinz bodies are a clinical laboratory feature found in G6PD (glucose-6-phosphate dehydrogenase) deficiency disorders or chronic liver diseases. There are detected by staining with crystal violet. However, the mechanism of forming tetrameric Hb structure formation including abnormal α^{CS}-globin is still unclear[4]. To understand protein behavior as above mentioned, must understand n atomic level model of this proteins is necessary.

1.2 Objectives

- To study the conformations change of elongated C-terminal α^{CS}-globin subunit mutation comparing to wild-type α-globin in the dynamics physiological environment.

- To study the structural effects associated with the dimer formation in case of dimeric α^{CS}·AHSP structure.

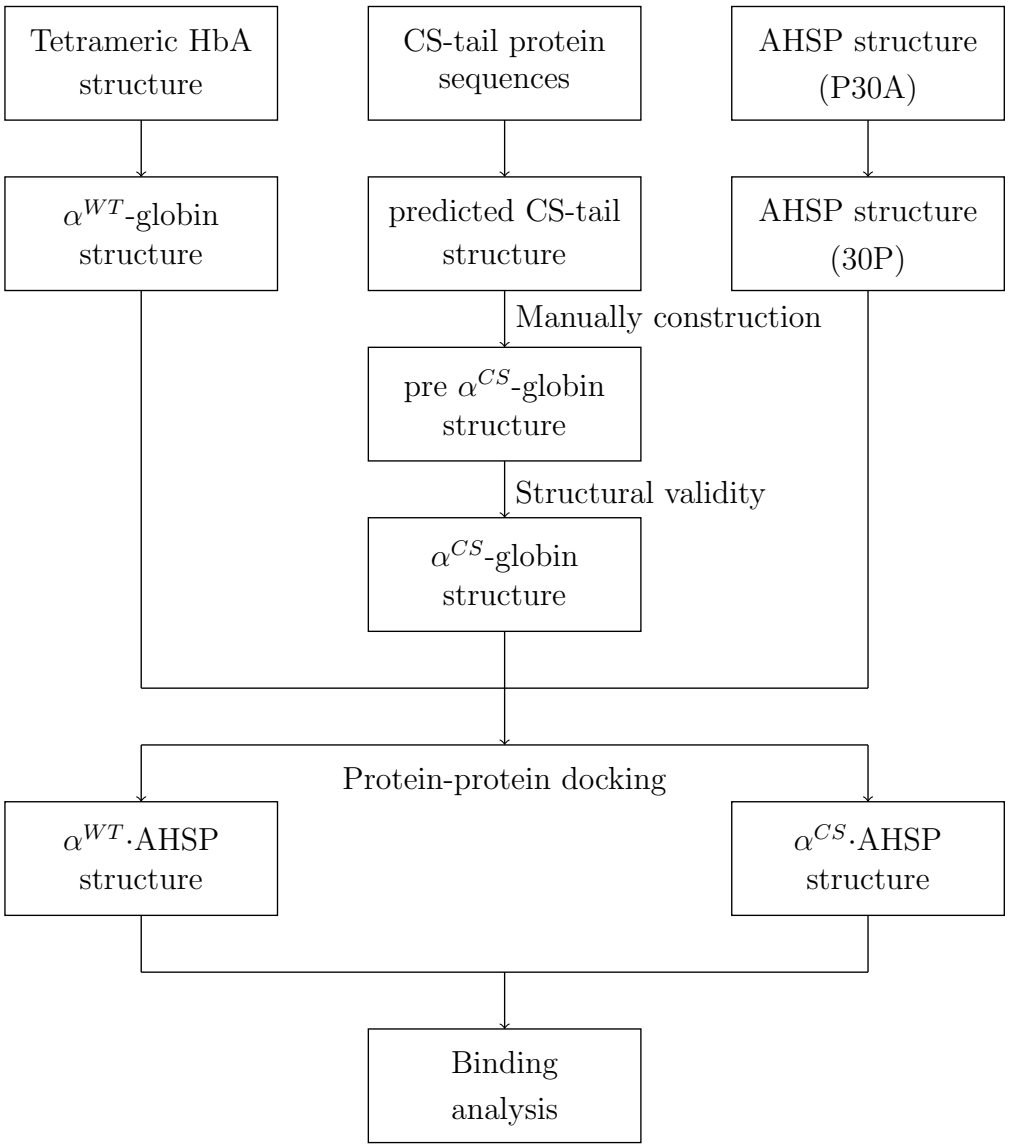

Figure 1.1 Schematic flow chart of research methodology

CHAPTER 2

LITERATURE REVIEWS AND THEORIES

To learn how such versatile molecular machines as proteins work, and why, this interest is reflected by the exponential growth of solved spatial structures of proteins deposited in the RCSB Protein Data Bank (PDB) [*].

2.1 Molecular dynamics (MD) simulation

2.1.1 Equations of motion for MD simulations

Molecular dynamics is a method in which the motion of each individual entity is computed according to Newton's equation of motion (**Equation 2.1**). It involves a large number of particles, to even several millions of particles. The method has been widely used in many areas from the simulation of liquids and solids to biological molecules by solving the set of Newton's equations.

$$m_i \frac{\mathrm{d}^2 \mathbf{r}_i(t)}{\mathrm{d}t^2} = \mathbf{F}_i(r_1, r_2, ..., r_N) \qquad \textbf{(2.1)}$$

Here the particle of interest is atoms. There are a total N atoms in the system. $\mathbf{r}_i$ are the position vectors and m_i is the mass of particle. $\mathbf{F}_i$ are the forces acting upon the i^{th} particles at time (t) in the system.

The forces derive from potential functions, $U(r_1,r_2,...,r_N)$, representing the potential energy of the system for the specific geometric arrangement of the particles,

$$-\nabla_{\mathbf{r}_i} U(r_1, r_2, ..., r_N) = \mathbf{F}_i(r_1, r_2, ..., r_N) \qquad \textbf{(2.2)}$$

The form implies the conservation of the total energy $E_t = E_k + E_p$, where E_t, E_k, and E_p is the total energy, instantaneous kinetic energy, and potential energy (E_p represented to U in **Equation 2.2**) respectively.

2.1.2 The numerical techniques

The most time-consuming component of MD calculation is force evaluation. Any method requiring more than one force evaluation per time step is inefficient. The classification of MD integrators are,

1. low-order methods, for example, leapfrog, Verlet, velocity Verlet algorithms, *etc.*

[*] http://www.rcsb.org/pdb/

2. predictor-corrector methods, which is high accuracy for large time-steps

Thus to obtain the dynamic behaviour of our system we must solve this second order differential equation for every particle in the system integrating with respect to time gives,

$$\frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t} = \frac{\mathbf{F}_i(r_1, r_2, ..., r_N)}{m_i} t + C_1 \qquad \textbf{(2.3)}$$

At time t=0 the first term vanishes and the velocity is given by the constant c_1, the initial velocity u_i. At time t equation written as,

$$\frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t} = a_i t + u_i \qquad \textbf{(2.4)}$$

This simple derivation produces and expression which corresponds to the truncated Taylor series for displacement.

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t}\Delta t + \frac{1}{2} \cdot \frac{\mathrm{d}^2\mathbf{r}_i(t)}{\mathrm{d}t^2}\Delta t^2 + \cdots \qquad \textbf{(2.5)}$$

therefore a small, persistent error is introduced into the calculation at every time step through the neglect of the higher order term. Note also this is based on the assumption that made the acceleration remains constant throughout the time step Δt . Unless infinitesimal steps are taken, this is another error including assumption. In practice, the time steps are used are of the order of 0.5–1 fs, with the restriction that the time difference must be smaller than that for the highest frequency vibration in the system (typically bond stretches). Using smaller time step would produce fewer errors but would require a complementary increase in computer time to allow the simulation of interesting phenomena. A number of algorithms have been developed to circumvent the problems associated with finite time steps and truncation errors[5].

The potential energy is a function of the atomic positions of all the atoms in the system. Due to the function complexity, there is no analytical solution to the equations of motion; they must be solved numerically. Finite difference techniques are used to generate molecular dynamics trajectories with continuous potential models. Numerous numerical algorithms using finite difference methods have been developed for integrating the equations of motion and commonly used in molecular dynamics calculations such as Verlet algorithm, Leap-frog algorithm, Velocity Verlet and Beeman's algorithm. AMBER12 applies the Leap-Frog algorithm for the integration[6], describes in details later. All algorithms assume that the positions and dynamic properties (velocities, accelerations, etc.) can be approximated as Taylor series expansion, as previously mentioned (**Equation 2.5**).

2.1.3 Leap-Frog algorithm

The first method used is the leapfrog algorithm, a modified version of the Verlet algorithm. The Verlet algorithm uses the positions and accelerations at the time t and the positions at the time $t+\Delta t$ where Δt is the a Taylor expansion of the 3^{rd} order integration step,

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \frac{\mathrm{d}^2\mathbf{r}_i}{\mathrm{d}t^2}(t+\Delta t^2) \tag{2.6}$$

The error in the atomic positions is of the order of Δt^4. The velocities are obtained from the basic definition of differentiation,

$$\frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t} = \frac{1}{2}\cdot\frac{\mathbf{r}_i(t+\Delta t) - \mathbf{r}_i(t-\Delta t)}{\Delta t} \tag{2.7}$$

with an error of the order of Δt^2. To obtain more accurate velocities, the leapfrog algorithm is used, using velocities at half time step,

$$\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}(t+\frac{\Delta t}{2}) = \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}(t-\frac{\Delta t}{2}) + \frac{\mathrm{d}^2\mathbf{r}_i(t)}{\mathrm{d}t^2}\Delta t \tag{2.8}$$

At time t, velocities can be computed by,

$$\frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t} = \frac{1}{2}\cdot\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}(t+\frac{\Delta t}{2}) + \frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t}(t-\frac{\Delta t}{2}) \tag{2.9}$$

This is advantageous when the kinetic energy is needed at time t, as in the case where velocity rescaling must be carried out (**Equation 2.10**). The atomic positions are then obtained by,

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \frac{\mathrm{d}\mathbf{r}_i(t)}{\mathrm{d}t}\Delta t \tag{2.10}$$

The leapfrog algorithm is computationally less expensive than the Predictor-Corrector approach, and requires less storage. In the case of large scale calculations, this could be an crucial advantage. Furthermore, the energy conservation is respected, even at large time steps. Therefore, computation time could be huge decreased. However, when more accurate velocities and positions are needed, another algorithm should be used for computing, as well as the Predictor-Corrector algorithm. To show in the following the effects of this algorithm on the results of the calculations.

2.1.4 Periodic boundary condition (PBC)

A PBC enable a simulation to be achieved using a relatively small number of particles in such a way which the particles experience forces as though they were in a bulk solution. The box containing the system is repeated infinitely in all

directions to give a periodic array. In term of the simulation, when an atom leaves the basic simulation box, attention has been switched to the incoming image. This is display in **Figure 2.1**, it is important to keep in mind the appointed artificial periodicity when in consideration of properties that are influenced by long-range correlations. The number of interacting particles is placed in a cubic cell which is surrounded by an infinite array of identical cubic cells is called a periodic boundary conditions (PBC). The minimum image of particle i, may be within the primitive cell shaded in lined pattern, or one of the surrounding image cells. The minimum image cell of particle (•) is surrounded by dashed square, where are a cubic box with box length (L) and the surface area of a sphere of radius cutoff (r_{cutoff})

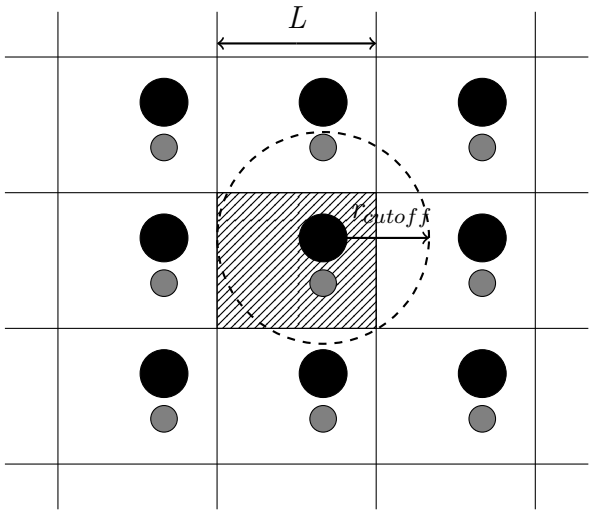

Figure 2.1 Periodic boundary conditions in 2-D

Special attention must be used to the case where the potential range is not short, for example, dipolar systems and charged. The shape of a periodic simulation cell must fill all space by translational operations of the central box in 3-D, a cube or rectangular prism is the most easy-to-use and common choice but can be computationally expensive due to dispensable amounts of solvent molecules in the corners, distant from the central macromolecules. A universal alternative that requires less volume is the truncated octahedron. Under PBC, inter-particle distances are measured by the minimum image convention and r_{cutoff} less than $^1/_2$ so particles do not interact with multiple images of neighbours.

2.1.5 Ensembles

Canonical ensemble (NVT) – in the canonical ensemble, number of particles (N), volume (V) and temperature (T) are conserved. It is also sometimes called constant temperature molecular dynamics (CTMD). In NVT, the energy of endothermic and exothermic processes is exchanged with a thermostat. A variety

of thermostat methods are available to add and remove energy from the boundaries of an MD system in a realistic way, approximating the canonical ensemble[7]. Isothermal-Isobaric ensemble (NPT) – in the isothermal-isobaric ensemble, number of particles (N), pressure (P) and temperature (T) are conserved. In addition to a thermostat, a barostat is needed. It corresponds most closely to laboratory conditions with a flask open to ambient temperature and pressure. In the simulation of biological membranes, isotropic pressure control is not appropriate[7].

2.1.6 Multi-scale modelling and simulation

Multi-scale structural biology modelling combines existing and emerging methods from diverse scientific disciplines to bridge the wide range of time and length scales that is inherent in a number of essential phenomena and processes in structural biology especially protein simulation. **Figure 2.2** illustrates the different scales and characteristic means of investigation by computer simulations (modified from Praprotnik *et al.* 2008)[8].

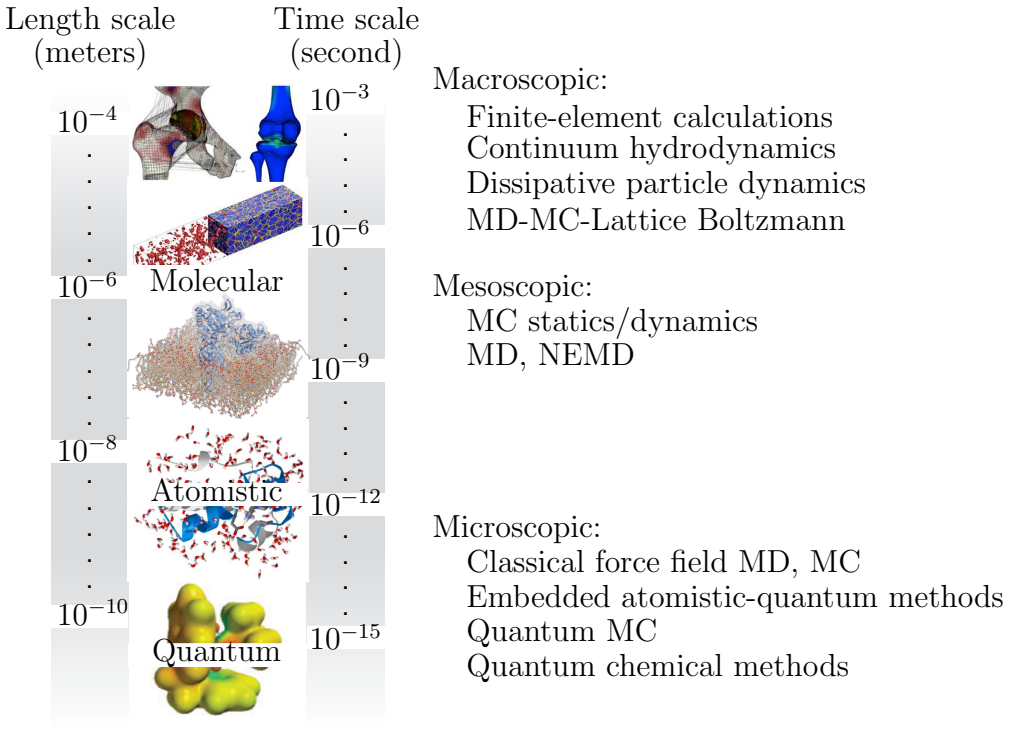

Figure 2.2 Diagrammatic multi-scale modelling and simulation

2.2 Haemoglobin (Hb)

Hb is found in the red blood cells of all vertebrates (excepting Family *Channichthyidae*) and it is an iron-containing oxygen-transport metalloprotein re-

sponsible for the O_2 delivery of oxygen from the lungs to tissues, and the transport of carbon dioxide from the tissues back to the lungs.

2.2.1 Genetics

The α_1-globin (*HBA1*) and α_2-globin (*HBA2*) genes are composed of three exons interrupted by two short introns[9]. The two mature α-globin mRNAs only diverge in structure in this 3'-untranslated region and therefore encode identical α-globin proteins. Subsequent to the structural characterisation of the two α-globin mRNAs, it was possible to devise a number of assays which could distinguish the two mRNA species bases upon the structural divergence in the 3' non-translated regions. These studies demonstrated that α_2-globin mRNA was present at approximately 2.6 folds higher level than α_1-mRNA in the reticulocytes of normal individual[10]. These and related reports further demonstrated the same excess of α_2-globin mRNA in bone marrow and early metal erythroid tissue indicating that the excess in α_2-globin gene expression is transcriptional in origin and which this ratio is not subject to developmental switching. The fact that α_2-globin gene has a dominant role in α-globin synthesis is accepted in an analysis of the α-globin mRNA and protein expression in eight individuals with distinct α-globin structural mutations at either the α_1- or α_2-globin locus. The results established that the α_2-gene encodes approximately 2.7 fold more protein than the α_1-gene, parallelling the 2.6 to 1 ratio of α_2 to α_1-mRNA levels[10]. The parallel between the α_2 to α_1 ratio of mRNA levels and protein expression is further authenticated by demonstrating of the two α-globin mRNAs are translated with equal efficiencies.

2.2.2 Hb assembly

The balancing component of α- and β-globins have been independently encoded on their respective polyribosomes occurring in pre-erythrocyte and haem, modified from mitochondria that have been inserted in each of them. Before $\alpha{\cdot}\beta$ formation, the alpha haemoglobin stabilising protein (AHSP) stabilises a nascent α-globin, subsequently, handle a possible stability of free α-globin[11]. Current evidences suggest that free α-globin is less stable and may form the precipitated inclusion bodies inside the erythrocyte[2,12–15]. AHSP binds specifically to α-globin and delivers newly α-globin to β-globin for dimeric globin formation[11,12].

2.3 Alpha-globin

2.3.1 3-D structure

Haem locates between helices H3 and H4, where H4 helix provides the proximal H87 to axially bind the haem iron atom.

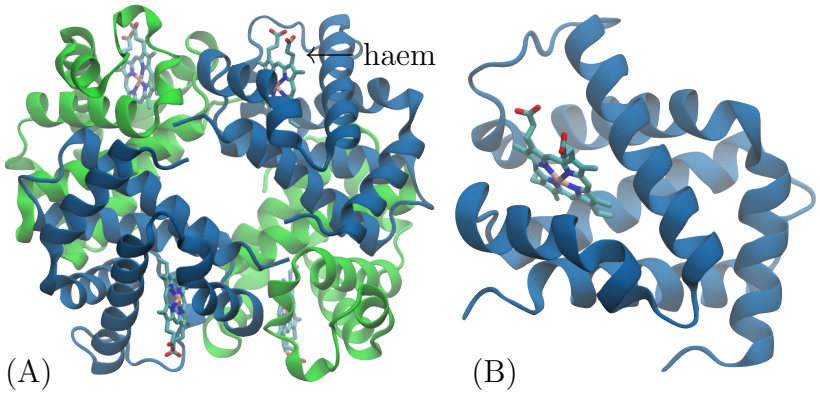

Figure 2.3 (A) tetrameric deoxyHbA structure and (B) monomeric α-globin structure (PDB code: 2DN2)

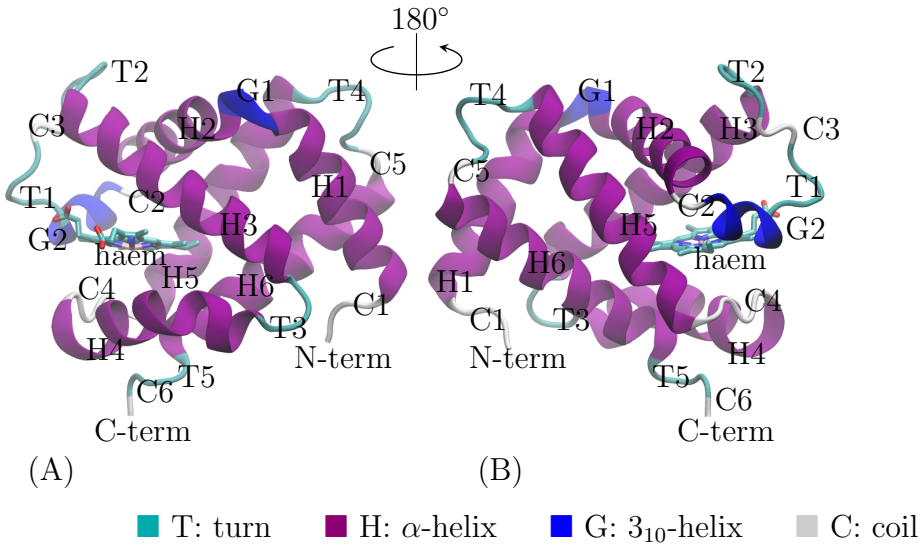

Figure 2.4 Secondary structure of α-globin (PDB code: 2DN2 chain C)

2.4 Alpha haemoglobin stabilising protein (AHSP)

AHSP has a function as an erythroid-specific molecular chaperone protein[2,16,17] to prevent the harmful aggregation of α-globin during normal erythroid cell development. Specifically protects free α-globin from precipitation. AHSP binding inhibits subunit reactions with oxidants such as hydrogen peroxide (H_2O_2)[18]. It is predicted to modulate pathological states of α-globin excess such as in thalassaemia. The AHSP protein binds to monomeric α-globin until it has been transferred to β-globin to form a heterodimer, which in turn binds to another heterodimer to form the stable tetrameric HbA.

2.4.1 Genetics and synthesis

ASHP is discovered by studying *GATA-1* gene[19] which is an essential erythroid transcription factor. The erythroid phenotype are known to be regulated by *GATA-1*. AHSP is a novel gene that is strongly induced by GATA-1[19]. AHSP encodes a small (102 amino acids), acidic protein with no recognisable signature motifs.

2.4.2 ASHP function in Hb synthesis

AHSP protects free α-globin from oxidative damage and participation in harmful redox reactions *in vivo*, it has also been shown to accelerate the rate of O_2 autoxidation to the ferric (MetHbA) state[20]. MetHbA, free met-α-globins, and free met-β-globins are unstable due to accelerated rates of haem loss, denaturation, aggregation, and precipitation[16]. Thus, it is puzzling why a molecular chaperone for α-globins would accelerate autoxidation. The AHSP stabilises a haemichrome folding intermediate and prevents its incorporation into HbA until the bound α-globins can be reduced to the ferrous form (see **Figure 2.5**).

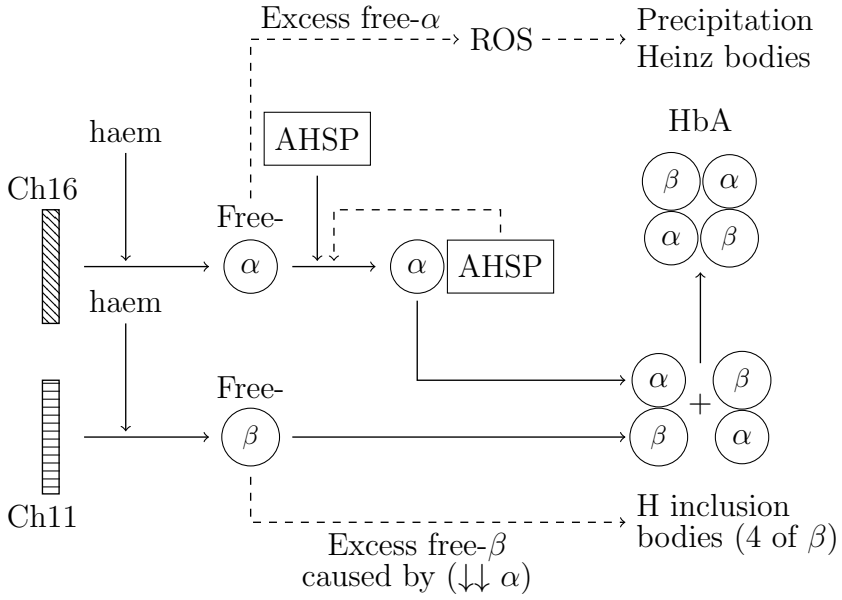

Figure 2.5 Molecular chaperones and Hb synthesis

2.5 Haemoglobin Constant Spring (HbCS)

HbCS is the most common non-deletional α-thalassaemic mutation (new sense mutation) and is an important cause of HbH-like disease in Southeast Asia. HbCS is caused by a mutation in the stop codon of the α_2 -globin gene that results

in poor output (1% of normal) of an α-globin with an additional 31 amino acids extension. HbH CS disease has a very high prevalence rate amongst α-thal group found in Thai thalassaemia patients, according to the report of Siriraj Hospital, Thailand[21]. Generally, clinical features of the disease can be growth retardation, jaundice, anaemia, gallstones, hepatomegaly, and splenomegaly. In addition the hypochromic condition and targeted cell of blood smear are commonly observed via laboratory investigation along with over 50% of HbH (abnormal Hb in a form of four-β; β_4)[10,22] inclusion bodies blood smear[10]. Severe anaemia is observed from when Hb level is about 9.5–11.0 g/dL in male and 7.5–9.0 g/dL in female[22,23]. Although, HbH CS disease comes from just only two alpha genes deletion, compared to HbH thalassaemia which has three alpha genes deletion, clinical manifestations of HbH CS disease is clearly worse than HbH thalassaemia[23].

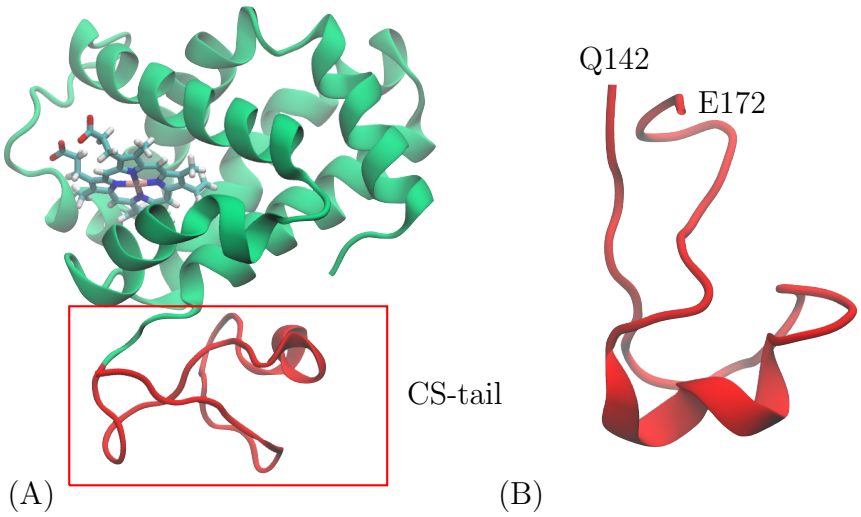

Figure 2.6 (A) predicted α^{CS}-globin structure at 80 ns of MD trajectory time, (B) predicted CS-tail structure (residue Q142 to E172)

CHAPTER 3

COMPUTATIONAL METHODOLOGY

This chapter lists materials and computational methodologies. Materials are including PDB file and amino acid sequence, web-based bioinformatics tools and resources, MD simulation, and visualising softwares.

3.1 Materials

3.1.1 Hardwares

- The Eclipse Cluster of the National Centre for Genetic Engineering and Biotechnology (BIOTEC) in the National Science and Technology Development Agency (NSTDA) located at 113 Thailand Science Park, Pathum Thani 12120, Thailand
- Personal computers, Dell^{TM} OptiPlex^{TM} 7010, Rock Cluster on CentOS® 7, Integrated Intel® HD Graphics 4000, Intel® Core^{TM} i7 (3^{rd} Gen) 3770/3.4 GHz, DDR3 SDRAM - non-ECC 8 GB/1600 MHz of RAM located at Department of Biomedical Sciences, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla 90112, Thailand

3.1.2 3-D structures

The PDB files consist of atomic coordinates of deoxyHbA and AHSP use as structural templates of the simulations containing the as structural templates of the amino acid sequence of CS-tail is artificially created. Two crystal structures and sequence are selected for the study presented in this thesis:

- A 1.25 Å resolution of human HbA in the deoxy form (PDB code 2DN2)[24].
- A 2.40 Å resolution of human oxidised α-globin bounded to AHSP (PDB code 1Z8U)[18].
- A CS-tail, 31-amino-acid-residues elongated in the α^{CS}-globin (**Figure 3.1**)[25].

3.1.3 Bioinformatics tools

- RCSB PDB – a database that provides an overview of some properties obtained from the analysis of 3-D coordinates of the crystal structures deposited in the PDB[26].

```
  1   VLSPADKTNV KAAWGKVGAH AGEYGAEALE RMFLSFPTTK
 41   TYFPHFDLSH GSAQVKGHGK KVADALTNAV AHVDDMPNAL
 81   SALSDLHAHK LRVDPVNFKL LSHCLLVTLA AHLPAEFTPA
121   VHASLDKFLA SVSTVLTSKY RQAGASVAVP PARWASQRAL
161   LPSLHRPFLV FE
```

Figure 3.1 Amino acid sequences of α^{CS}-globin. The CS-tail residues are represented from residue Q142 to E172 (underlined)

- PROPKA – a prediction method for prediction and rationalisation of pKa values. The server PDB2PQR v2.0.0 is used to predict pKa values of the ionisable residues[27].
- PSIPRED – a secondary structure prediction web-server bases on amino acid sequence.[28]
- NetSurfP – a web-server predicts the secondary structure and surface accessibility of amino acids from determined sequence[29].
- CABS-fold – a web-server provides tools for protein tertiary structure prediction from sequence only (*de novo* modelling) and also using alternative templates (consensus modelling)[30].
- RosettaBackrub – a web-server is based on *ab initio* modelling technique. This study used a Rosetta v3.1 to generate a mutagenesis[31].
- ClusPro – a Critical Assessment of Predicted Interactions which is the protein-protein Docking web-server that rapidly docks, filters, and ranks putative protein complexes within a short amount of time[32].

3.1.4 Softwares

- VMD – a Visual Molecular Dynamics, a molecular graphics software version 1.9.2[33]. The programme is used for the atomistic modelling, the analysis of MD simulations, and for rendering high resolution graphical images.
- UCSF Chimera – a programme for interactive visualisation and analysis of molecular structures and related data[34].
- AMBER12 – a molecular simulation programme[6]. An AMBER10 force field[35] with customised topology and parameter files is used for building and initial energy minimisation of the atomistic models.
- LigPlot$^+$ – a programme generates schematic 2-D representations of protein-ligand and protein-protein complexes from standard Protein Data Bank file input[36].

3.2 Prediction of α-globin Constant Spring monomer (α^{CS})

In the case of wide-type α-globin (αHbA), representing α^{WT}, chain C of deoxy tetrameric Hb structure is adopted (PDB code 2DN2) [24]. An additional CS-tail structure is constructed as showed in **Figure 3.1**. CS-tail, secondary structure (helix, sheet, and turn) is predicted using PSIPRED web-server * [28] and NetSurfP v1.1 † [29] and a 3-D structure is independently predicted using CABS-fold ‡ [30] according to the amino acid sequence. All above-mentioned data are then analysed and the tertiary structure of CS-tail is finally refined and obtained. The 3-D structure is determined using PDB2PQR v1.9.0 web-server § [27] to assign protonation states at pH 7.4 within a selection of AMBER force field to compute. H50, H72, and H103 of α-globin subunit are set as doubly protonated [37,38], the proximal histidine, H87 of α-globin subunit is set as singly protonated at the delta position (δ1-N) [37,38]. The position of the nitrogen atom is demonstrated in **Figure 3.3**.

3.3 Structural validation

3.3.1 Ramachandran plot

A Ramachandran plot was developed in 1963 by Ramachandran *et.al.* This is a structural validation method to visualise a backbone dihedral angles and use as geometrical error detection tool of amino acid residues in protein structure. A structural validation is carried out via theoretical values of dihedral angles of an amino acid residue in a protein and shown the empirical distribution of datapoint observed in a single structure. In the thesis, Ramachandran plot is constructed using RAMPAGE web-server¶ [39].

3.4 Preparation of α^{CS}·AHSP, and α^{WT}·AHSP dimer

The monomeric structure of AHSP (PDB code 1Z8U chain C) [18] is adopted from chain C. A mutagenesis is generated by RosettaBackrub web-server, in which change AHSP (P30A) mutant into wide-type AHSP (P30). The wide-type AHSP is set to AHSP monomer and submitted to ClusPro v2.0 web-server with α^{CS} for dimeric prediction. The ten best structures based on scoring selection are chosen and then clustered according to RMSD considerations. Each top five cluster is characterised by Ramachandran plot and the best structure is chosen to a starting wide-type dimeric structure for MD simulation. A dimeric α^{CS}·AHSP structure is

*http://bioinf.cs.ucl.ac.uk/psipred/
†http://www.cbs.dtu.dk/services/NetSurfP/
‡http://biocomp.chem.uw.edu.pl/CABSfold/
§http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/
¶http://mordred.bioc.cam.ac.uk/ rapper/rampage.php

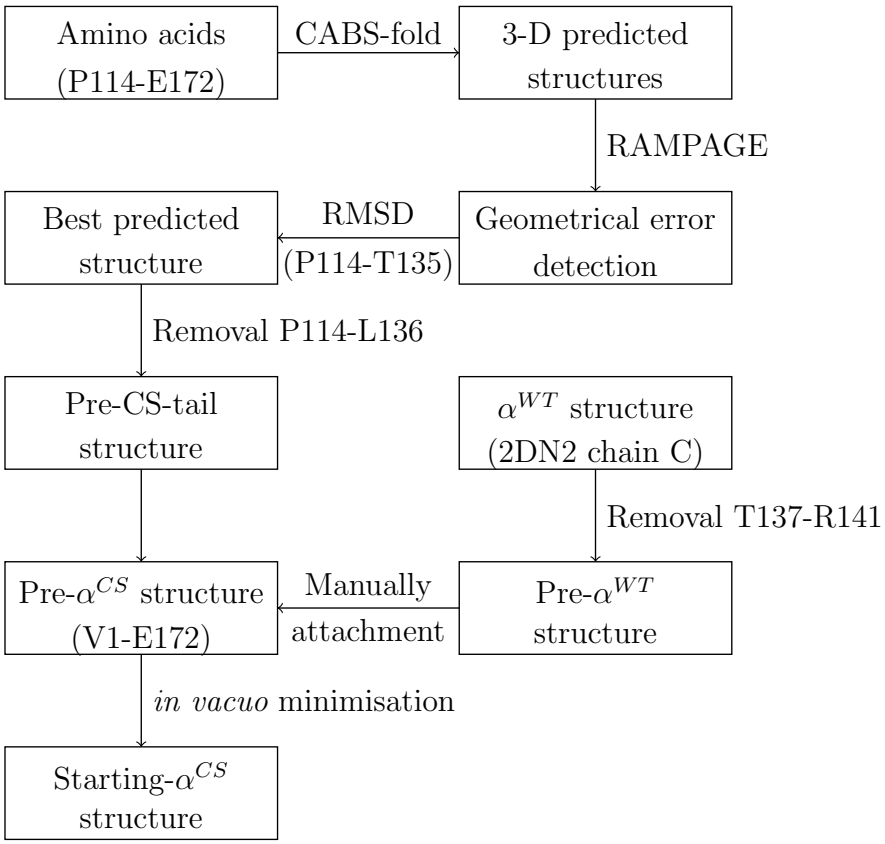

Figure 3.2 Schematic flow chart of α^{CS}-globin prediction

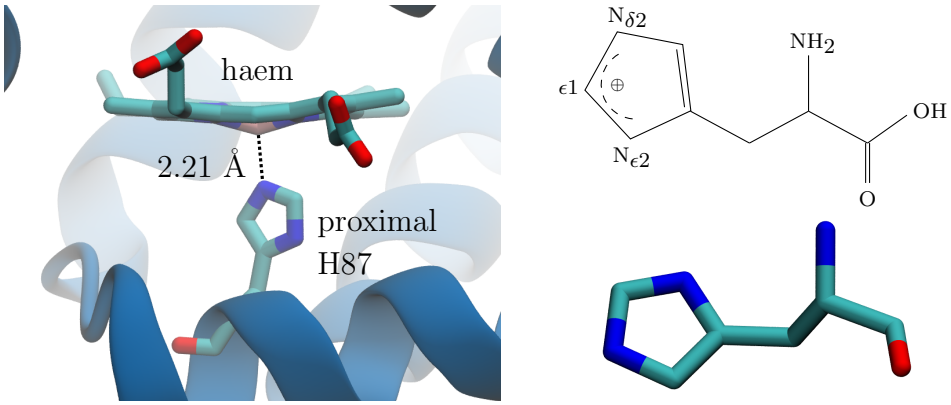

Figure 3.3 Chemical structure of histidine

constructed identical to α^{WT}·AHSP dimer, denoting that the α^{CS} monomer is taken from the **Step 3.2**. For the α^{WT}·AHSP dimer prediction uses the same docking method as the preparation of α^{CS}·AHSP.

- Haem molecular parameter – the parameters archived here are for use with

the AMBER force field modifications for all-atom haem parameters around the iron atom are appropriate for a six-coordinate haemoglobin/myoglobin. This force field parameters were adapted from[40] and finally N-terminus of H87 and Fe^{2+} atom in centre of haem is manually connected.

3.5 Simulation protocol

All structures, the hydrogen atoms are added using LEaP programme of AMBER12 package[6]. Then, the energy minimisation is presented 3,000 cycles using SANDER programme of AMBER12 package to remove bad interatomic contacts. MD simulation of all proteins, can be summarised as following:

- α^{WT} monomer
- α^{CS} monomer
- AHSP monomer
- α^{WT}·AHSP dimer
- α^{CS}·AHSP dimer.

The MD simulation starts with the knowledge of the potential energy of the system with respect to its position coordinates. There are multiple steps involved in simulation as summarised in **Figure 3.4**. Initial structures are prepared by LEaP programme of AMBER12 package to create topology and parameter files for each model. The AMBER10 force field[35] is adopted to analyse the proteins, haem parameters are taken from the study by Giammona[40].

All protonation states are set at pH 7.4 (see in **Step 3.2**). Each system is neutralised by either Cl^- or Na^+ and NaCl is added, corresponding to a physiological concentration of 0.15 M, equivalent to osmotic pressure in erythrocyte is shown on **Table 3.1**. The energy minimisation sets for 3,000 cycles of *in vacuo* using SANDER programme[41]. The systems are equilibrated in the canonical ensemble (NVT condition), the temperature is heated rescaling to 310 K and set the condition to dynamics. Van der Waals interactions are modelled using 6-12 Lennard-Jones potentials[42] and the electrostatic interactions to be performed using particle-mesh Ewald (PME) method[43], with cutoff is 12 Å. The structure is solvated in a cubic water box filled with TIP3P model[44] under the periodic boundary conditions at constant volume. Protein structures are equilibrated for 600 ps and harmonically restrained with force constants from 250, 150, 100, 50, 20, and 10 kcal·mol^{-1}·Å^{-2} every 100 ps, respectively. After NVT, the system is switched to to the isothermal-isobaric ensemble (NPT condition) at 310 K, the system is maintained constant pressure at 1.013 atm. Temperature and pressure are controlled using Berendsen

Table 3.1 Number of water molecules and counter ions used are applied for each protein simulated

System	Total	TIP3P	Box (Å^3)	Na^+	Cl^-
α^{WT}	28,270	8,667	73×68×69	24	27
α^{CS}	35,444	10,893	93×69×69	40	35
AHSP	22,590	7,682	79×63×57	25	21
α^{WT}·AHSP	49,088	15,097	99×79×82	43	42
α^{CS}·AHSP	45,150	13,625	82×73×76	38	39

weak-coupling algorithm[45], time step is set to 2 fs. The NPT simulation is performed for 80 ns of simulation time and the last 25 ns (2,500 snapshots), is taken for a configuration average.

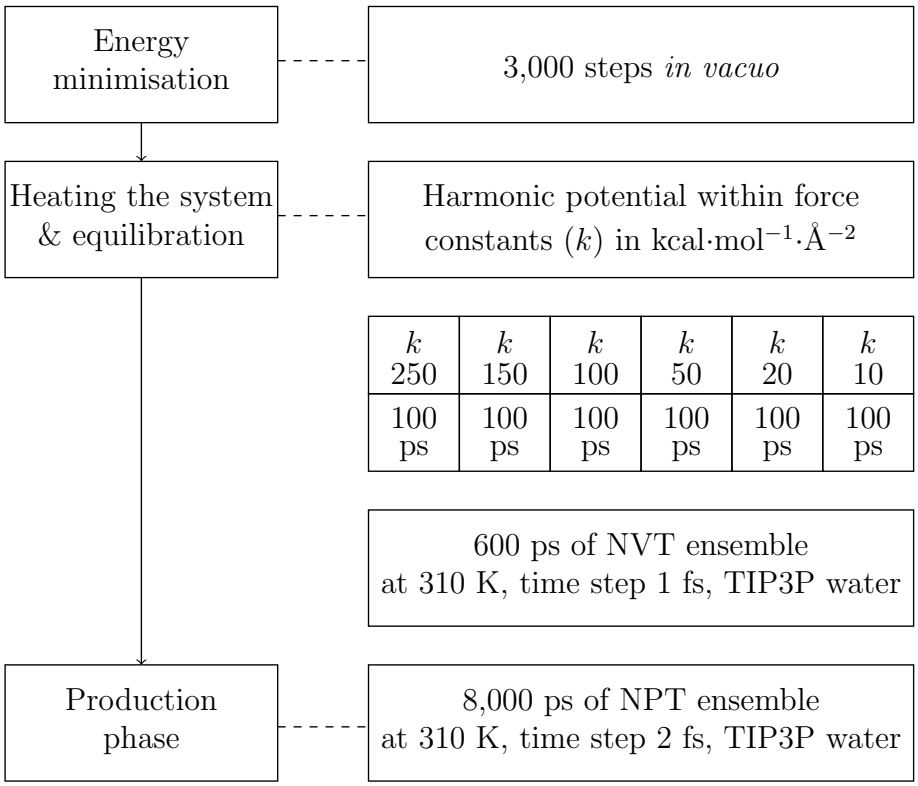

Figure 3.4 Schematic brief of MD simulation protocol

- Energy minimisation – commonly referred to as geometry optimisation which is usually performed to determine a stable conformation and remove the bad interatomic contact in the protein structure. The stable structure indicates the lowest potential energy of the system to the lowest possible point.
- Heating system and equilibration – the simulation is allowed to continue until

desired temperature is achieved. Force constraints on different subdomains of the simulation system are gradually removed as structural tensions dissipated by heating. In accurate solvent simulation, protein positions are fixed and solvents (water and ions) move accordingly. Once the solvent is equilibrated, the constraints on the protein can be removed and the whole system (protein in solvent) can derive in time.

- Production phase – this is a last step of the simulation methodology to remove constraints on protein and carry on a simulation in the isothermal-isobaric ensemble (NPT). The time scale can be varied from several hundred picoseconds to microseconds.

3.6 Analysis of MD simulations

All the proteins structure are visualised by VMD v1.9.2[33]. The trajectories are analysed using AMBER12 package. The Root-Mean-square Fluctuation (RMSF) and intramolecular distance pattern are computed using AMBER12 package. RMSD Trajectory Tools and VMD MultiSeq plugin on VMD programme is used to fit and align structures for computing Root-Mean-Square Deviation (RMSD). Surface charge distribution is analysed by a Coulombic Surface Colouring tool with UCSF Chimera programme. The molecular surfaces are coloured by the potential values (10 to -10 kcal·mol^{-1}·e^{-1}, gradient shading in blue to red colour).

- RMSD – a deviation of the structure with respect to a particular conformation is measured by RMSD. It indicates how structurally similar. The smaller RMSD value, the more convergent the structures may be; on the other hand, the structure is more similar to the reference structure. Acceptable simulation wild-type structures typically have an average RMSD of 1 Å to 3 Å, whereas an RMSD of 10 Å would be considered a poor fit for a small protein[46]. It is calculated for all trajectory frames, the RMSD for trajectory frame x is:

$$RMSD_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i(t_x) - r_{ref}(t_0)\right)^2} \qquad \textbf{(3.1)}$$

  where N is the number of residues in the molecule; r_{ref} is the reference structure position, starting structure in **Step 3.2** is at time $t = 0$, r_i is the position of the interested residue in trajectory frame x after superimposing on the reference frame, trajectory frame x is recorded at time t_x. The procedure is repeated for every trajectory frame in the simulation trajectory.

- RMSF – useful for characterising local changes along the protein chain and showing flexible regions of the protein flexibility. To interpret the flexibility

of structure, a large RMSF values are associated with increases in flexibility compares to the WT as the reference structure. The RMSF per residue i of C_α atom is computed as:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\langle (r_i(t) - r_{ref})^2 \rangle} \tag{3.2}$$

Where T is the trajectory time. The r_{ref} is the reference position of residue i and r_i is the position of residue $i$and the $\langle \cdots \rangle$ indicate that the average of the square distance is taken over the selection of atoms in the residue.

- Intramolecular distance pattern – a distance measurement between the centre-of-mass of haem residue and each interested residue is calculated with:

$$D_{avg} = \frac{1}{T}\sum_{i=1}^{N} |r_{HEM}^{com} - r_i^{com}| \tag{3.3}$$

Where r_{HEM}^{com} refers to centre-of-mass of haem residue and r_i^{com} is the position of centre-of-mass residue i.

- Surface charge distribution – a Coulombic Surface Colouring shades molecular surfaces by the potential values and can handle structures with or without explicit hydrogens. It can also generate a grid of potential values according to Coulomb's law in the scalar form:

$$|F| = k_e \frac{|q_i|}{r^2} \tag{3.4}$$

Where F is the potential which varies in space, q are the atomic partial i charges, the scalar r is the distance between the charges and k_e is Coulomb's constant ($k_e = (4\pi\epsilon_0)^{-1}$ is about 8.99×10^9 N$\cdot$ m$^2\cdot$C^{-2})[47].

3.7 Calculating binding energies

The MM/GBSA[48], MM/PBSA[49], and binding entropy calculation are methods used to calculate binding energy in this thesis. Both MM/GBSA and MM/PBSA have been used to estimate protein-protein binding affinities in wild-type and mutant complex systems. The binding affinities is used implicit (continuum) solvation and molecular mechanics (forcefields) solvation models[50].

The energy is relative value and not able to directly compare with the experiments. Mean gold is to compare ones between wild-type and mutant cases instead. In the MM/GB(PB)SA method, the free energy of the protein-protein binding, $\Delta G_{binding}$, is obtained from the difference between the free energies of

protein$_1$·protein$_2$ complex ($\langle G_{complex} \rangle$) and the unbound receptor, protein$_1$ ($\langle G_{receptor} \rangle$) and ligand, protein$_2$ ($\langle G_{ligand} \rangle$) as follows:

$$\Delta G_{binding} = \langle G_{complex} \rangle - (\langle G_{receptor} \rangle + \langle G_{ligand} \rangle) \quad \textbf{(3.5)}$$

Each free energy term in **Equation 3.5** is calculated with the absolute free energy of the species (protein$_1$, protein$_2$, and their complex) in gaseous phase (E_{gas}) as well as *in vacuo* of MD system. The free energy of each term in **Equation 3.5** is estimated as in **Equation 3.6**

$$G = \langle E_{gas} \rangle + \langle E_{solvation} \rangle - T\langle S \rangle \quad \textbf{(3.6)}$$

where $\langle E_{gas} \rangle$ is the molecular mechanics energy of the molecule expressed as the sum of the internal energy that are bonds, angles and torsions ($\langle E_{internal} \rangle$), Van der Waals term ($\langle E_{vdw} \rangle$), and electrostatic energy ($\langle E_{electrostatic} \rangle$):

$$\langle E_{gas} \rangle = \langle E_{internal} \rangle + \langle E_{vdw} \rangle + \langle E_{electrostatic} \rangle \quad \textbf{(3.7)}$$

The term $\langle E_{solvation} \rangle$ free energy is divided into a polar part of $\langle E_{GB} \rangle$ or $\langle E_{PB} \rangle$ and $\langle E_{surface} \rangle$ is a surface energy (a nonpolar solvation) follow:

$$\langle E_{solvation} \rangle = \langle E_{GB/PB} \rangle + \langle E_{surface} \rangle \quad \textbf{(3.8)}$$

After simplify and put energetic terms for protein$_1$ (monomeric α^{WT}, α^{CS}-globin), protein$_2$ (monomeric AHSP structure) and the complex **Equation 3.5** can be re-organisated and expressed as:

$$\Delta G_{binding} = \langle \Delta E_{gas} \rangle + \langle \Delta E_{solvation} \rangle \quad \textbf{(3.9)}$$

The calculation the average binding entropy ($\langle \Delta S \rangle$) using Normal Mode Analysis (normal mode) is performed in a vacuum, where the potential energy of a protein is a complex function. In **Equation 3.6**, the $T\langle S \rangle$ term represents the entropy loss of the flexibility upon binding in system. The entropies are calculated by a normal mode analysis of harmonic frequencies from minimised snapshots of MD simulations. This method also uses the MMPBSA script in AMBER12 for computation. The most common method to estimate the vibrational entropy in the MM/GBSA method is to use frequencies from a normal mode analysis performed at the molecular mechanics level.

3.8 Evolution of secondary structure analysis

To animate the secondary structure is defined by Timeline plugin in VMD v1.9.2 programme. The graphical display of residues and time-steps which are

scrolled and zoomed as necessary to see results for secondary structure assignment trajectories. Its result depends on colour code for magenta denotes α-helix, red denotes π-helix, cyan denotes turn, blue denotes 3_{10}-helix and white denotes coil.

3.9 Data analysis

The descriptive statistics are used to summarise and describe data. The t-Test for two independent samples is used for data competition between wild-type and mutant structure. True difference is considered statistically significant at the 95% confidence interval ($p \geq 0.05$) performs in R programme[51].

CHAPTER 4

RESULTS AND DISCUSSION

4.1 3-D structure of CS-tail

The 3-D structure prediction is suggested based on the results of different methods[52,53]. There are three methods for protein structure prediction namely homology modelling, fold recognition or threading, and *ab initio* or *de novo* method. The 3-D structure of CS-tail is predicted using CABS-fold server, depending on unique comparative modelling with Monte Carlo simulation for 3-D protein structure[30]. This method is the most efficient bioinformatics tool for *de novo* protein structure prediction in year 2013[30]. These ten possible structures are validated the geometrical error detection of dihedral angle of Ramachandran plot as showing in **Table 4.1**.

Table 4.1 Ramachandran plot values of predicted models using RAMPAGE

Structure	Favoured	Allowed	Outlier
Model01	55 (96.5%)	2 (3.5%)	-
Model02	54 (94.7%)	3 (5.3%)	-
Model03	52 (91.2%)	3 (5.3%)	2 (3.5%)
Model04	57 (100.0%)	-	-
Model05	50 (87.7%)	2 (3.5%)	5 (8.8%)
Model06	51 (89.5%)	3 (5.3%)	3 (5.3%)
Model07	53 (93.0%)	2 (3.5%)	2 (3.5%)
Model08	50 (87.7%)	4 (7.0%)	3 (5.3%)
Model09	49 (86.0%)	4 (7.0%)	4 (7.0%)
Model10	50 (87.7%)	6 (10.5%)	1 (1.8%)

To choose the best pre-structure of CS-tail, the secondary structure is assigned to amino acid sequence (P114 to E172) of CS-tail using the computational tools servers, which are NetSurfP v1.1 and PSIPRED web-server resulting in **Table 4.2** and **Figure 4.1**, respectively. The predicted secondary structure result of NetSurfP v1.1 indicates that from residue P114 to T118, A143 to A145, V149 to W154, and L161 to E172 are coil, residue P119 to R141, and A155 to L160 are α-helix and residue Q142 to A143, and S146 to A148 are β-strand. The method of protein structure prediction is the primary and secondary neural network[29] in term of homology modelling method. The predicted secondary structure of CS-tail

from PSIPRED web-server shows the similarity prediction to NetSurfP v1.1, residue P114 to T118, A143 to P151, and L161 to E172 are coil and residue P119 to Q142, and A152 to L160 are α-helix. Both predicted results have been compared to select the most possible 3-D structural model of CABS-fold server.

Table 4.2 Secondary structure predictions by NetSurfP v1.1

Amino acid	Residue no.	Probability		
		α-helix	β-strand	Coil
P	114	0.003	0.003	0.994
A	115	0.058	0.017	0.925
E	116	0.115	0.016	0.868
F	117	0.257	0.016	0.727
T	118	0.339	0.016	0.645
P	119	0.717	0.014	0.269
A	120	0.802	0.014	0.185
V	121	0.802	0.014	0.185
H	122	0.831	0.044	0.125
A	123	0.802	0.014	0.185
S	124	0.802	0.014	0.185
L	125	0.879	0.010	0.111
D	126	0.923	0.002	0.076
K	127	0.938	0.007	0.055
F	128	0.938	0.007	0.055
L	129	0.879	0.010	0.111
A	130	0.879	0.010	0.111
S	131	0.831	0.044	0.125
V	132	0.831	0.044	0.125
S	133	0.751	0.050	0.199
T	134	0.779	0.100	0.120
V	135	0.725	0.163	0.112
L	136	0.649	0.163	0.188
T	137	0.649	0.163	0.188
S	138	0.538	0.173	0.289
K	139	0.538	0.173	0.289
Y	140	0.428	0.171	0.402
R	141	0.307	0.165	0.527
Q	142	0.113	0.087	0.800
A	143	0.052	0.084	0.864
G	144	0.018	0.088	0.893
A	145	0.021	0.279	0.699

Table 4.2 Secondary structure predictions by NetSurfP v1.1 (cont.)

Amino acid	Residue no.	Probability		
		α-helix	β-strand	Coil
S	146	0.022	0.552	0.426
V	147	0.021	0.756	0.223
A	148	0.023	0.655	0.322
V	149	0.021	0.451	0.528
P	150	0.019	0.141	0.840
P	151	0.113	0.043	0.844
A	152	0.268	0.043	0.689
R	153	0.455	0.046	0.498
W	154	0.455	0.046	0.498
A	155	0.561	0.047	0.393
S	156	0.561	0.047	0.393
Q	157	0.660	0.049	0.291
R	158	0.561	0.047	0.393
A	159	0.561	0.047	0.393
L	160	0.455	0.046	0.498
L	161	0.339	0.016	0.645
P	162	0.339	0.016	0.645
S	163	0.455	0.046	0.498
L	164	0.354	0.048	0.598
H	165	0.268	0.043	0.689
R	166	0.115	0.016	0.868
P	167	0.053	0.043	0.903
F	168	0.066	0.296	0.638
L	169	0.021	0.451	0.528
V	170	0.021	0.451	0.528
F	171	0.022	0.359	0.619
E	172	0.003	0.003	0.994

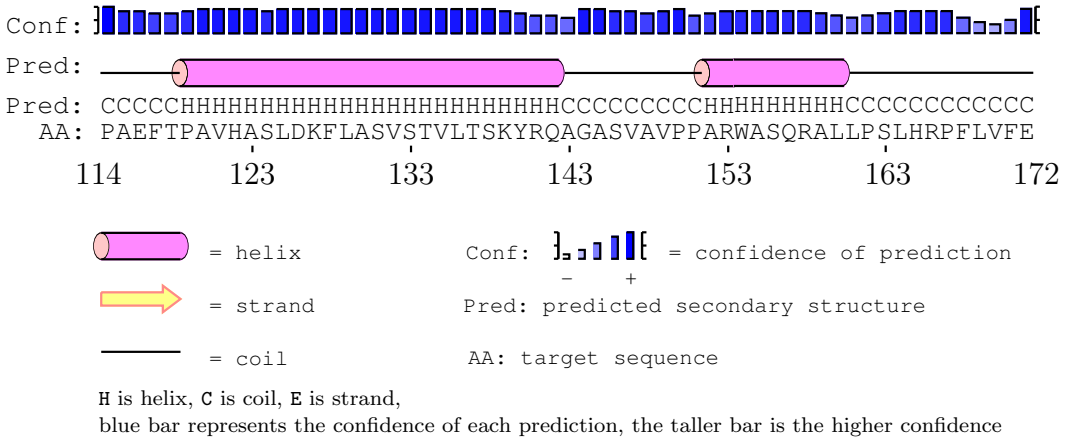

Figure 4.1 Schematic representation of secondary structure predictions using PSIPRED web-server of CS-tail

```
                 114      .        .        .        .        .      172
Predicted seq.   PAEFTPAVHASLDKFLASVSTVLTSKYRQAGASVAVPPARWASQRALLPSLHRPFLVFE
Validated seq.   CCCCCHHHHHHHHHHHHHHHHHHHCCCC----------------------------

NetSurfP         CCCCCHHHHHHHHHHHHHHHHHHHHHHHHBBCCBBBCCCCCCHHHHHHCCCCCCCCCCCC
PSIPRED          CCCCCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHCCCCCCCCCCC
Model01          CCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCHHHHHHHHHHHHHHHHHHHHC
Model02          CCHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHHHHHHHCCCCCCC
Model03          CCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCHHHHHHHHHHHHCCCCCCC
Model04          CCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCHHHHHHHHHHHHHHHHHHHHC
Model05          CCCCHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCHHHHHHHHHHHHCCCCCCCCC
Model06          CCCCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHHHHCCCCCCCC
Model07          CCHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCC
Model08          CCCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCHHHHHHCCCCCCCCCCCCCC
Model09          CCCCHHHHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCHHHHHHHHHHCCCCCCCCCCC
Model10          CCCCCCCCHHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
```

Figure 4.2 Secondary structure prediction of CS-tail by CASB-fold (Model1 to Model10)

A Model07 is chosen to be the predicted-CS-tail structure, which is visualised in **Figure 4.3A**. The predicted-CS-tail structure is *in vacuo* minimised with 3,000 steps (**Figure 4.3B**) then the residue P114 to L136 of pre-CS-tail is removed. The backbone residue T137 to R141 of α^{WT}-globin chain C (PDB code 2DN2) is aligned with residue P114 to L136 of pre-CS-tail, RMSD 0.28 Å by using VMD v1.9.2 (**Figure 4.4**). The alignment structure, residue T137 to R141 are also removed from α^{WT}-globin chain C. A residue T137 of pre-CS-tail is manually attached to residue L136 of α^{WT}-globin chain C, which becomes the pre-α^{CS}-globin by using AMBER12. The pre-α^{CS}-globin structure is *in vacuo* minimised with 3,000 steps by fixing up the positions of the atoms in order to remove any bad contacts where is manually bonded the CS-tail then the structure becomes the starting structure as predicted-α^{CS}-globin.

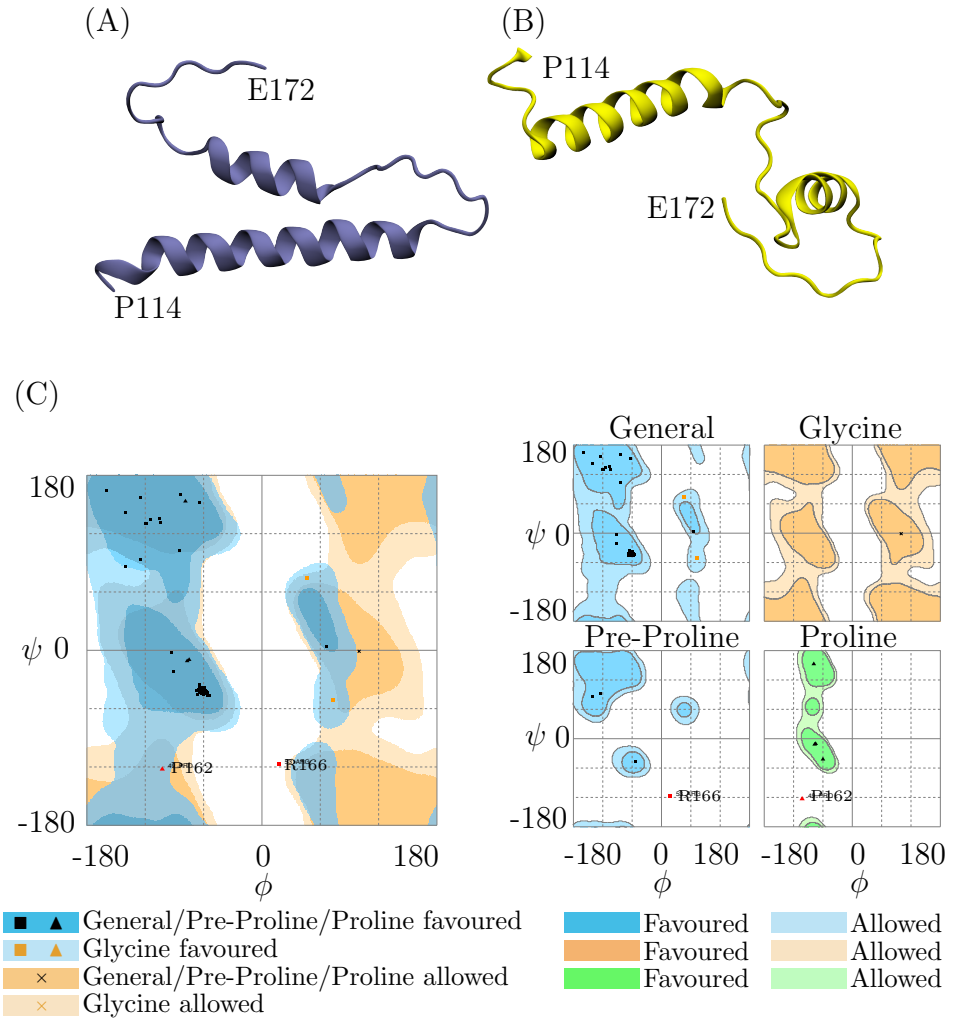

Figure 4.3 CS-tail structure for: (A) predicted-CS-tail structure (Model07); (B) pre-CS-tail structure; (C) the Ramachandran plot of predicted-CS-tail (Model07)

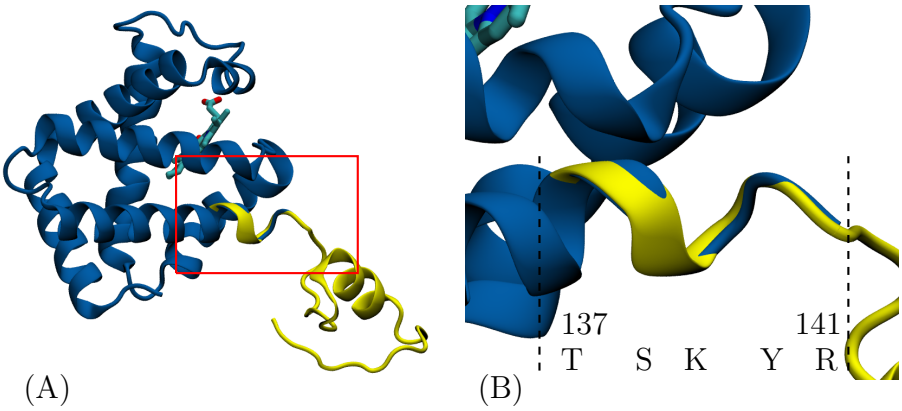

Figure 4.4 Structural alignment of residue T137 to R141 between α^{WT}-globin and pre-CS-tail, RMSD is 0.28 Å

4.2 Mutagenesis of AHSP structure

The alpha haemoglobin stabilising protein (AHSP) structure is taken from PDB code 1Z8U chain C. The structure is P30A mutation, so the structure is changed back to wild-type using point mutagenesis[54] of RosettaBackrub web-serv The predicted wild-type structure is *in vacuo* minimised with 3,000 steps and has RMSD 0.36 Å comparing to mutant and uses as the starting-AHSP structure in MD simulation.

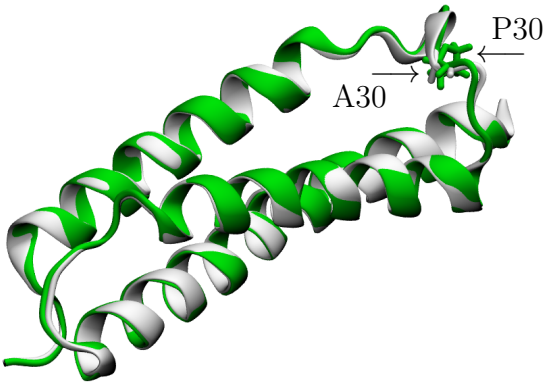

Figure 4.5 Structural alignment of wild-type AHSP (green) and AHSP P30A (grey), RMSD 0.36 Å

4.3 Trajectory analysis of MD simulations

For the MD simulations, the trajectories of the α^{WT}-globin, α^{CS}-globin, AHSP, α^{WT}·AHSP, and α^{CS}·AHSP in the explicit solvent are simulated. The backbone RMSD values for all structures during the production phase relative to the

starting structures are plotted in **Figure 4.6**. The RMSD values which are increased because the temperature of system is increased and the structure requires to stable conformation. The equilibrium phase is selected after 5,500 ps and 4,500 ps to the end of MD trajectory time for α^{WT}-globin, α^{CS}-globin, α^{WT}·AHSP, and α^{CS}·AHSP and AHSP, respectively. In the equilibrium phase, RMSD values change in a narrow range and there are nearly stable. Structures in the equilibrium phase are the representative sample for MD trajectory analysis.

Table 4.3 Backbone RMSD values of structures

Structure	RMSD ± SD (Å)	
	All simulation time	Equilibrium phase
α^{WT}-globin	1.44 ± 0.47	1.04 ± 0.11
α^{CS}-globin (V1-E172)	6.04 ± 0.67	6.27 ± 0.21
α^{CS}-globin (V1-R141)	1.16 ± 0.14	1.22 ± 0.17
AHSP	3.62 ± 0.94	4.46 ± 0.32
α^{WT}·AHSP	3.02 ± 0.48	3.46 ± 0.36
α^{CS}·AHSP	2.11 ± 0.35	1.82 ± 0.15

The MD simulations give stable equilibrium structures with backbone RMSD of 1.00 Å, 6.27 Å, 1.22 Å, 4.46 Å, 3.46 Å, and 1.82 Å for α^{WT}-globin, α^{CS}-globin residue V1 to E172 (whole mutant structure), α^{CS}-globin residue V1 to R141 (main mutant structure), AHSP, α^{WT}·AHSP, and α^{CS}·AHSP, respectively (MD simulation snapshot show in **Figure 4.7** to **4.11**).

4.3.1 MD simulation of monomeric α^{CS}-globin

The conformation of main α^{CS}-globin (residue V1 to R141) is not different from wild-type but the high RMSD value of whole α^{CS}-globin (residue V1 to E172) results from the flexibility of CS-tail movement. To support conformational unchanged on main α^{CS}-globin by the distances where they measure from centre of mass of each residue to haem molecule is the same pattern as wild-type (**Figure 4.14**). However, the result in **Figure 4.15** shows that the secondary structure of monomeric α^{CS}-globin during 62 ns to the end of simulation time has changed. These secondary structural change of monomeric α^{CS} has influenced residue T108 to P119, especially, residue L113 to E116 and residue F117 to P119 change from turn to α-helix and turn to coil, respectively starting at ≈67 ns to the end of simulation time. The positions from the secondary structural change are included hotspots binding site to AHSP molecule which are F117, and P119[55] which locate on flexible region, see **Figure 4.12** that may interfere the binding to AHSP.

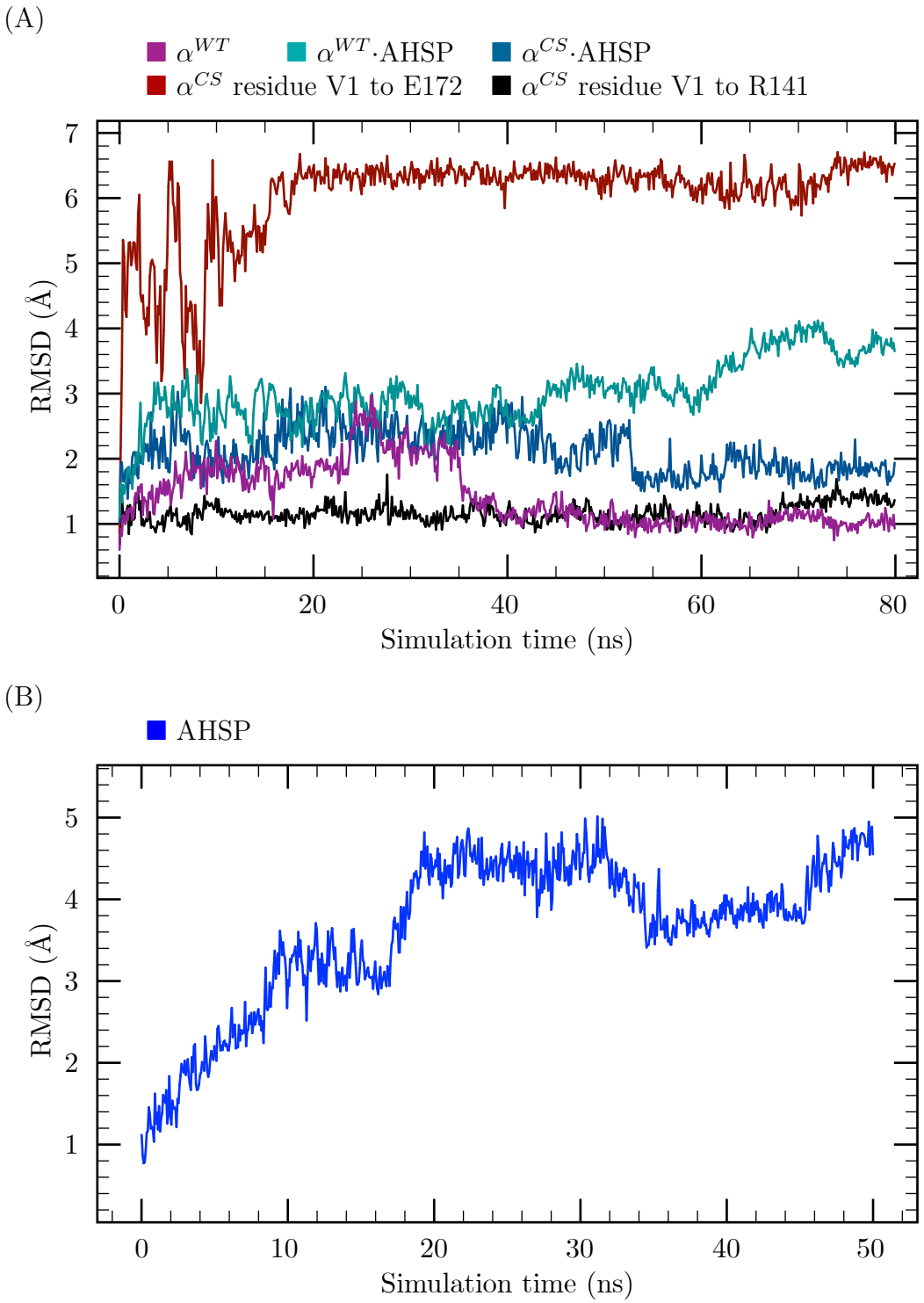

Figure 4.6 MD trajectories for: (A) the RMSD plots of α^{WT}, α^{CS} residue V1 to E172, α^{CS} residue V1 to R141, α^{WT}·AHSP, and α^{CS}·AHSP; (B) the RMSD plots of AHSP

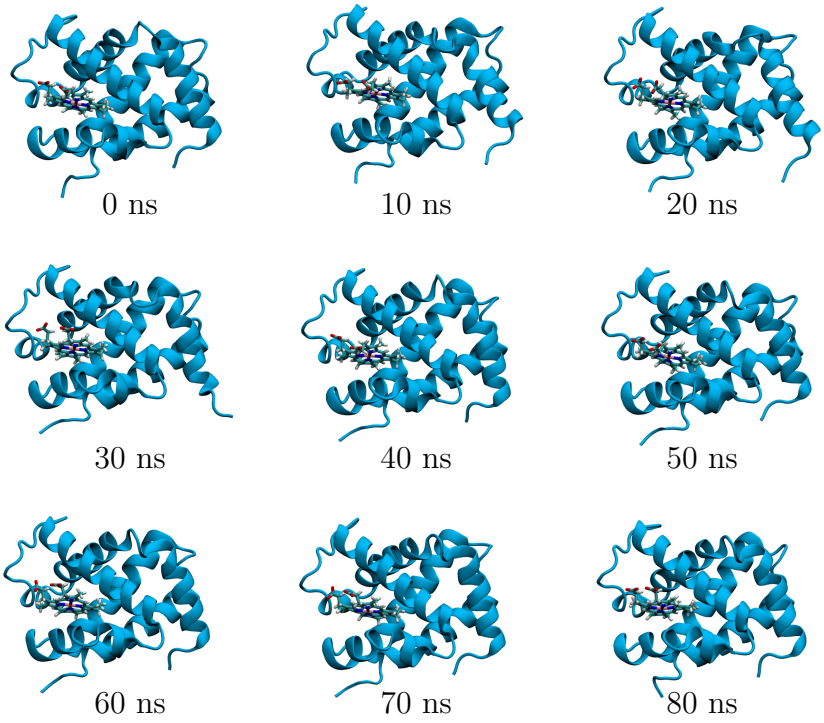

Figure 4.7 α^{WT}-globin snapshots from MD simulation

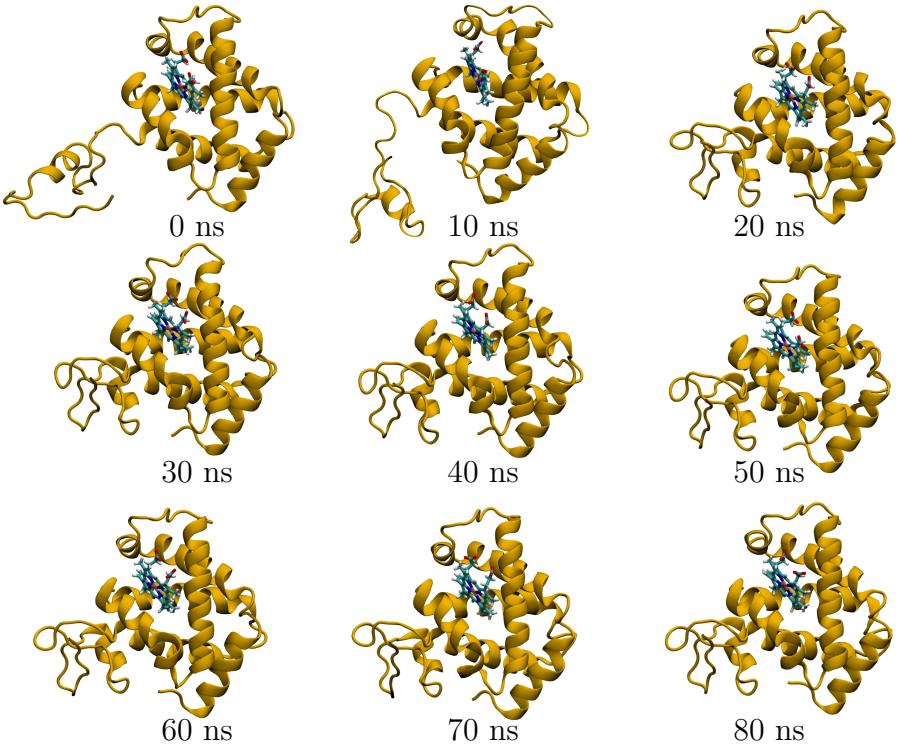

Figure 4.8 α^{CS}-globin snapshots from MD simulation

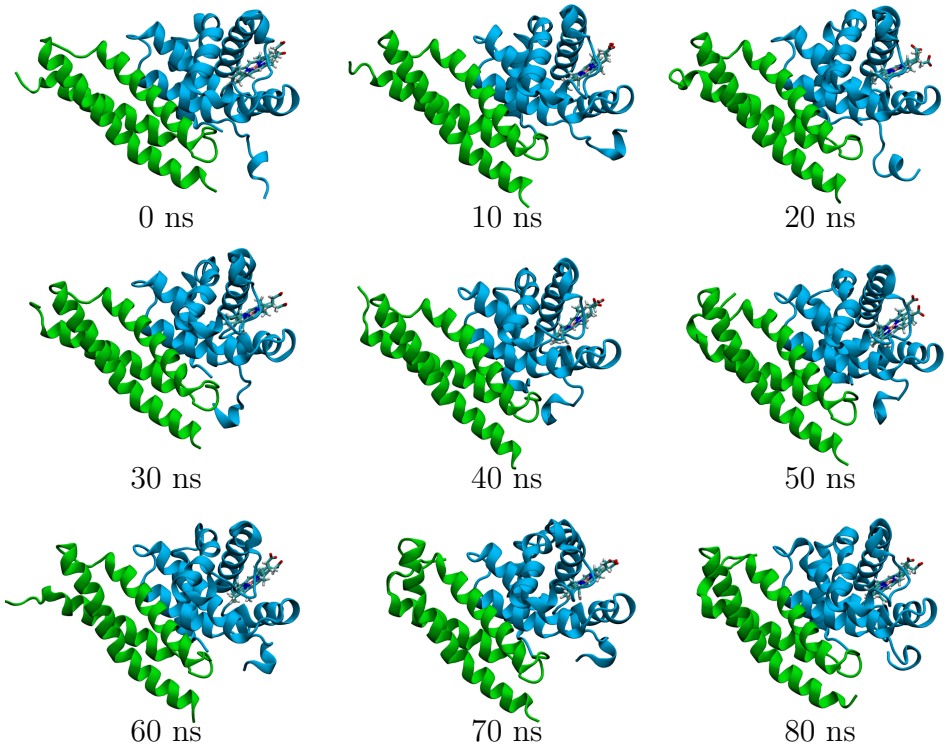

Figure 4.9 α^{WT}·AHSP dimer snapshots from MD simulation

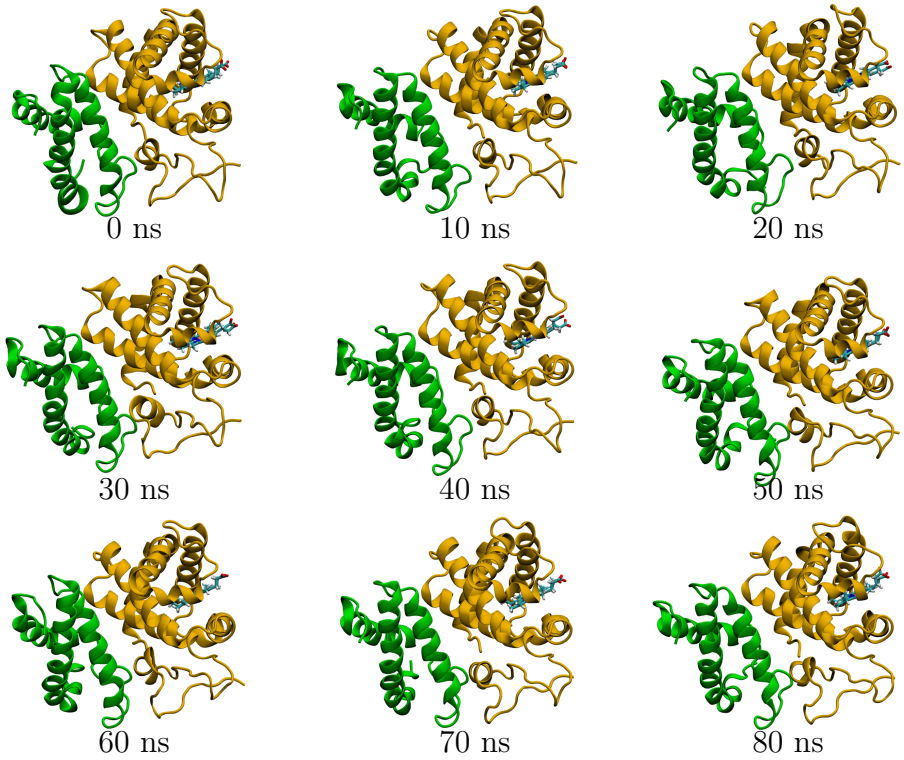

Figure 4.10 α^{CS}·AHSP dimer snapshots from MD simulation

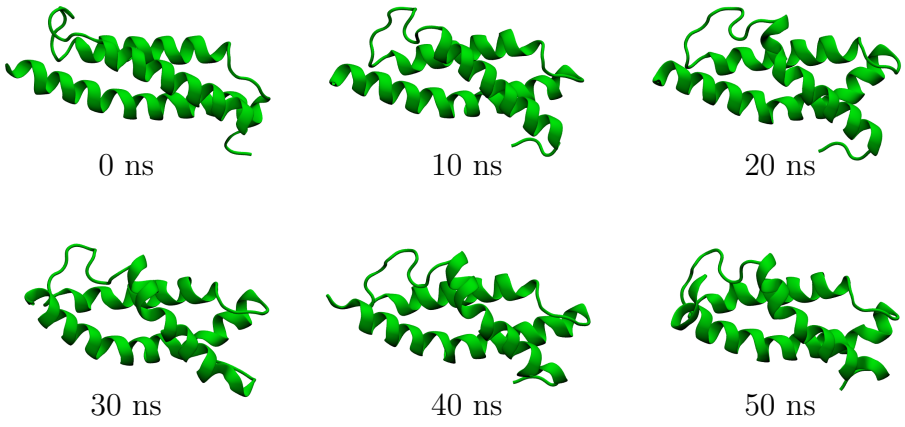

Figure 4.11 AHSP dimer snapshots from MD simulation

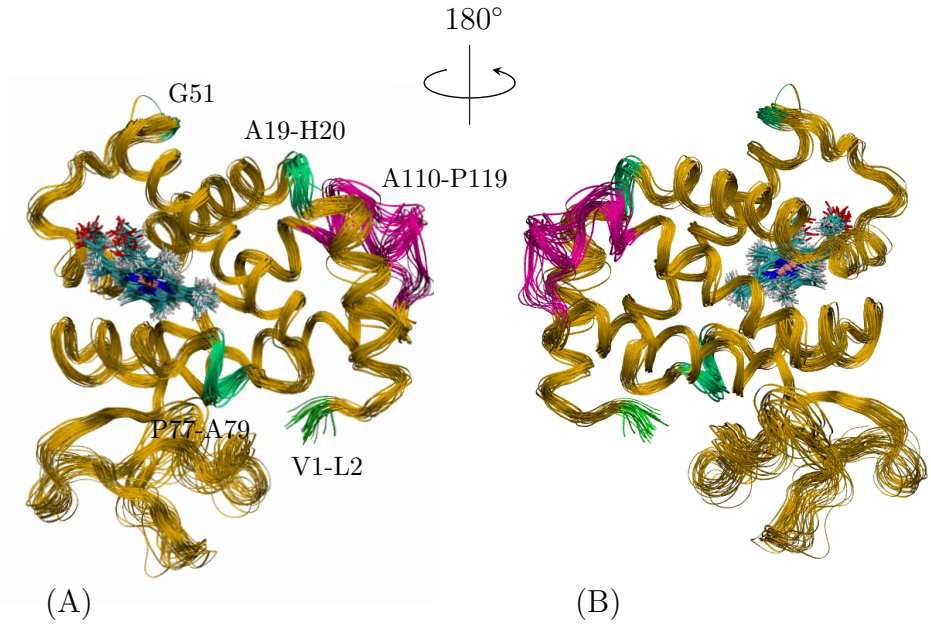

Figure 4.12 RMSF per residue number of α^{CS} structure: green and magenta colour are represented by RMSF $\geq$ 1.00 Å and RMSF $\geq$ 1.50 Å , respectively

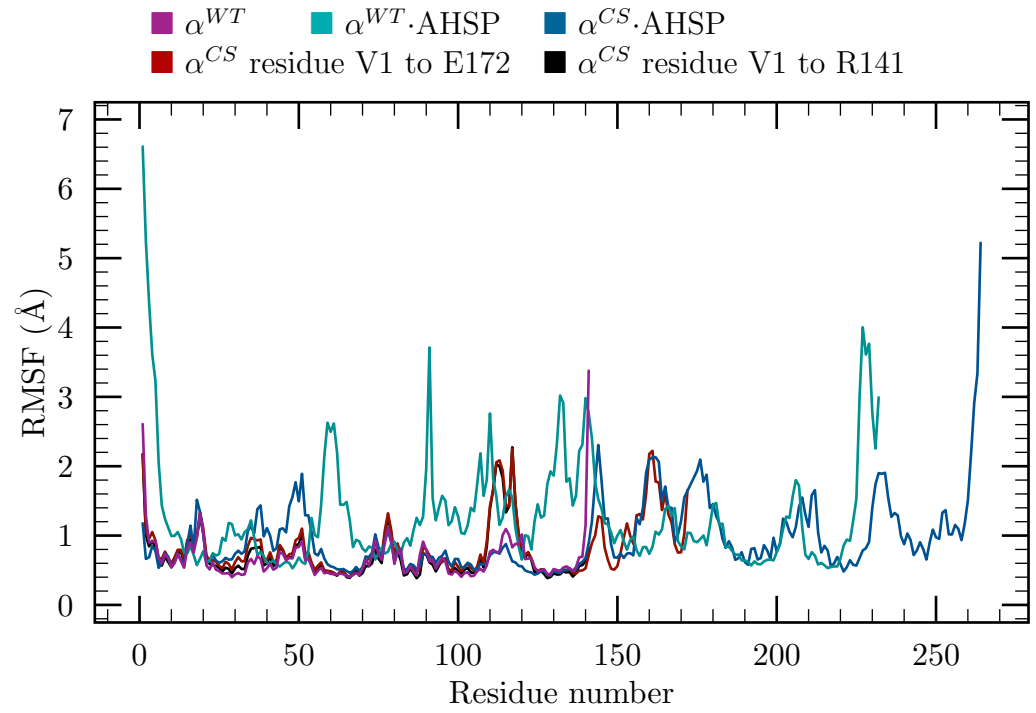

Figure 4.13 MD trajectories of RMSF values of atomic positions computed for the backbone atoms are shown as a function of residue number

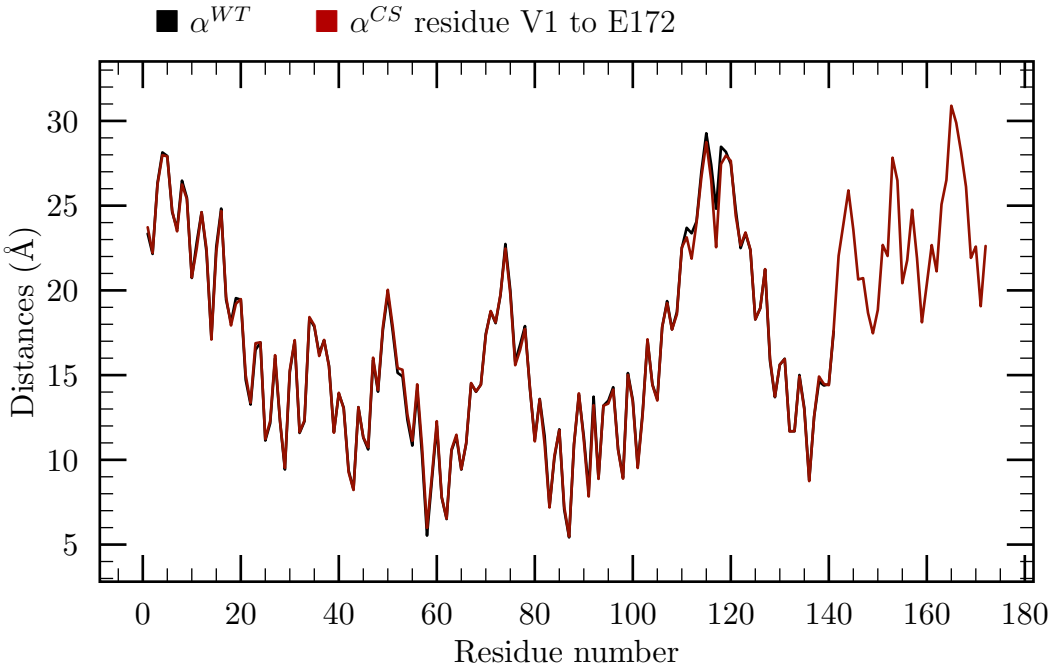

Figure 4.14 MD trajectories for: the distance between C_α atoms of any residue to centre of mass of haem

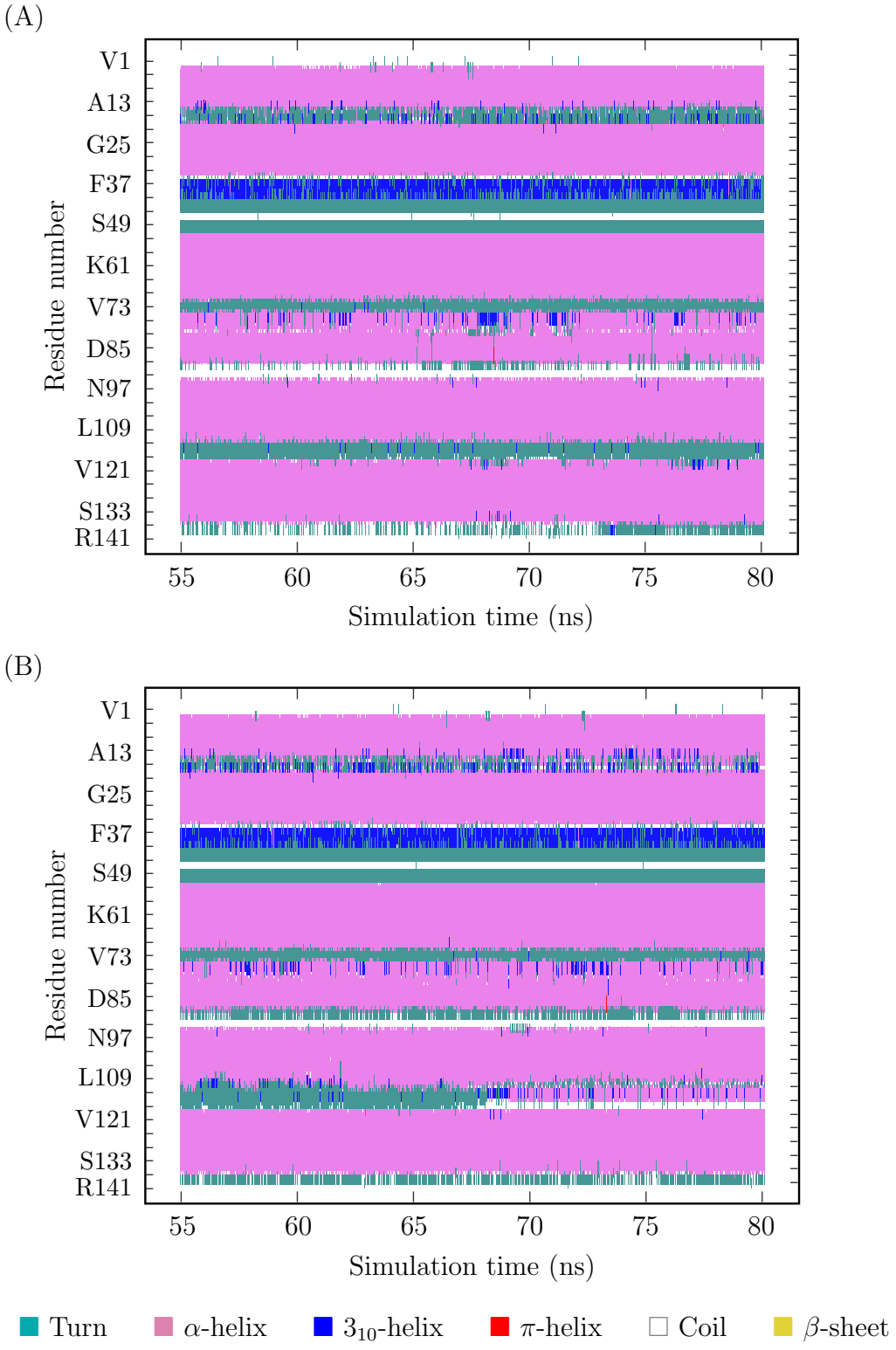

Figure 4.15 Time evolution of secondary structure of: (A) α^{WT}-globin monomer and (B) α^{CS}-globin monomer

4.4 Surface charge distribution of monomeric α^{CS}-globin

Surface charge distribution is computed by the Coulombic Surface Colouring tool on the UCSF Chimera programme. The colours of molecular surfaces represent the electrostatic surface potential values. These results are shown hardly any difference between mutant and wild-type structure where the binding site of AHSP molecule is located. Whereas most of binding surface that K40, K99, H100, and K127 contribute positively charged appearing in both mutant and wild-type.

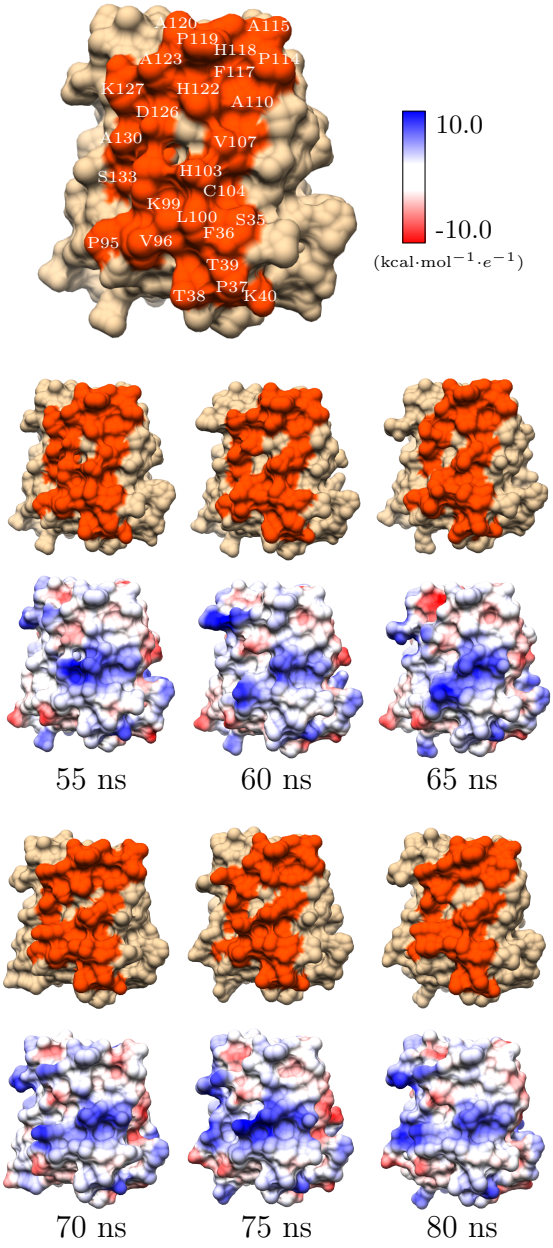

Figure 4.16 Electrostatic surface potential of monomeric α^{WT}-globin structure

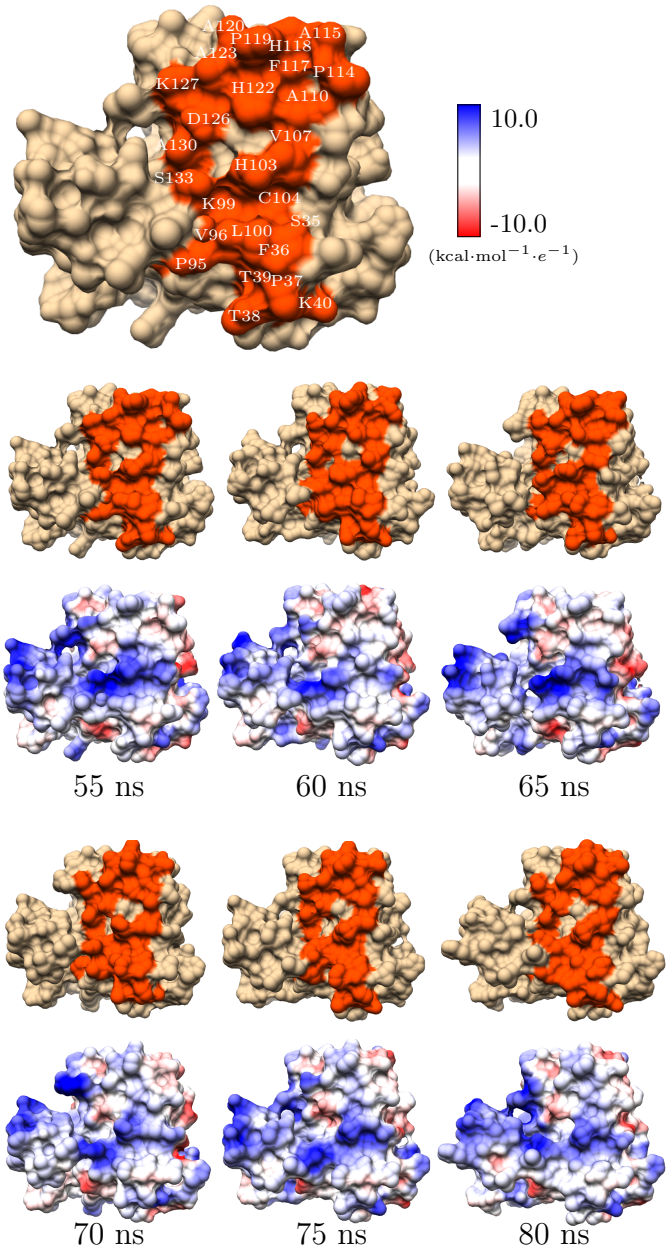

Figure 4.17 Electrostatic surface potential of monomeric α^{CS}-globin structure

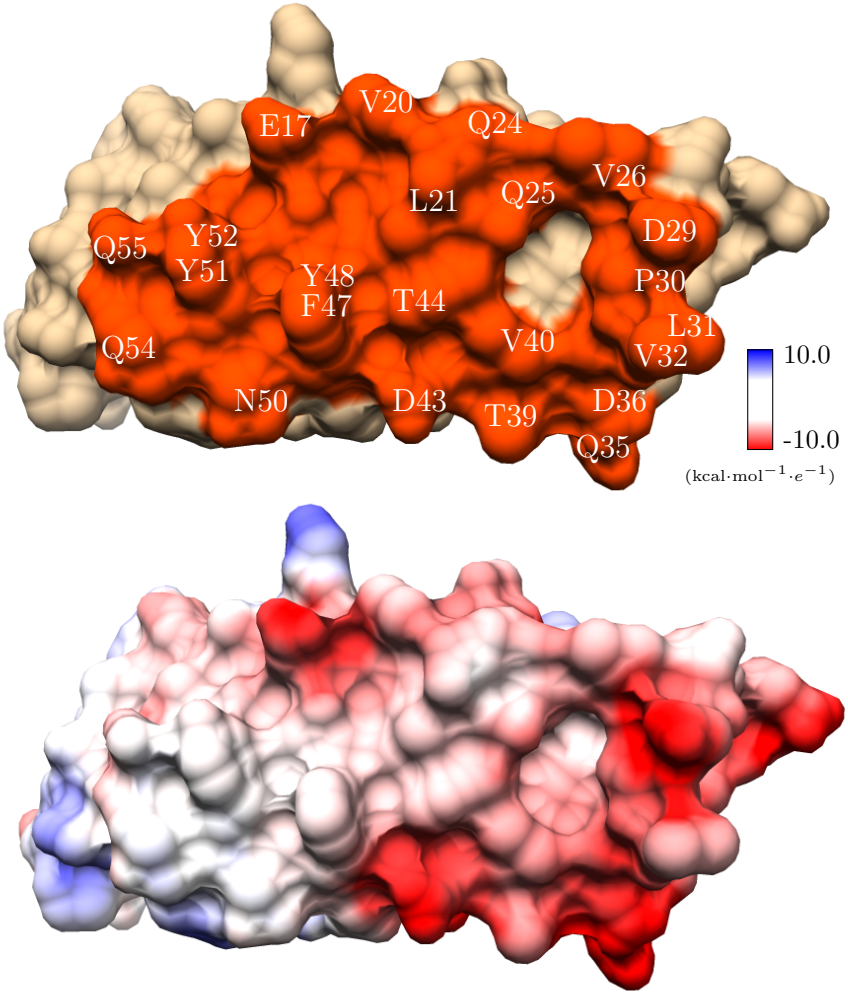

Figure 4.18 Electrostatic surface potential of monomeric AHSP structure

4.5 Dimeric structure prediction of α^{CS}·AHSP and α^{WT}·AHSP

The docked dimeric conformation is computed using ClusPro protein-protein docking tool. The binding residues on α^{CS}-globin to AHSP molecule are depended on the experimental analysis of residues interface interaction by Yu *et al.* (2009) [55] which are residue K99, H103, F117, P119, A123, and D126. The binding residues on AHSP molecule where interface with α^{WT} are predicted by LigPlot^{+} programme has display in **Figure 4.19**. There are obtainable H-bonds between residue H103 of α^{WT} and D43 of AHSP, residue A123 of α^{WT} and Q24 of AHSP, and residue D126 of α^{WT} and Q25 of AHSP, others are hydrophobic interaction. Interestingly, result is displayed aromatic interaction between residue F117 of α^{WT} and F47 of AHSP.

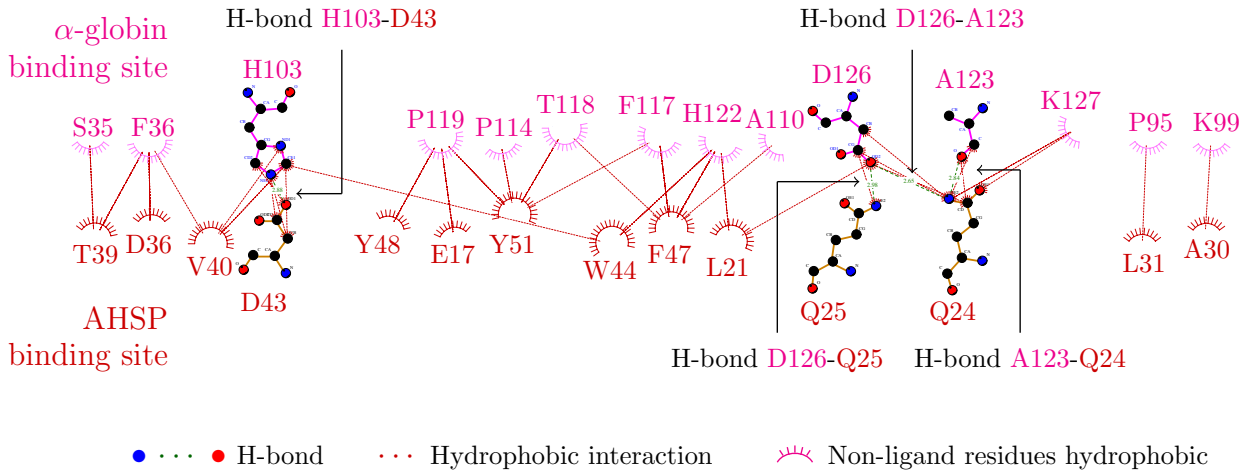

Figure 4.19 Diagrammatic of protein-protein interactions for reference α^{WT}·AHSP dimer (PDB code 1Z8U, chain C and D) analysing by LigPlot^{+}

The monomeric structure of α-globin mutant and AHSP are taken form **Step 4.1** and **4.2**, respectively and the monomeric structure of wild-type α-globin is adopted from chain C, PDB code 2DN2 and AHSP is also taken form **Step 4.2**. For the best α^{WT}·AHSP, conformation, there are obtainable H-bonds between residue T38 of α^{WT} and D36 of AHSP, residue K99 of α^{WT} and A30 of AHSP, residue H103 of α^{WT} and D42 of AHSP, and residue D126 of α^{WT} and Q25 of AHSP, others are hydrophobic interaction. Interestingly, result is displayed aromatic-aromatic interaction between residue F117 of α^{WT} and F47 of AHSP. The RMSD value is 3.46 $\pm$ 0.36 Å comparing to wild-type dimer chain C and D, PDB code 1Z8U and theoretical dihedral angle of structure by Ramachandran plot results that 94.3% of favoured region, 4.4% of allowed region, and 1.3% of outlier region. For the best α^{CS}·AHSP, conformation, there are obtainable H-bonds between residue V1 of α^{CS} and D43 of AHSP, residue D6 of α^{CS} and Q25 of AHSP, residue N9 of α^{CS} and

Q24 of AHSP, residue F117 of α^{CS} and Q54 of AHSP, residue H122 of α^{CS} and N50 of AHSP, residue K127 of α^{CS} and D43 of AHSP, residue R153 of α^{CS} and D36 of AHSP, and residue R153 of α^{CS} and T39 of AHSP, others are hydrophobic interaction. Interestingly, result is displayed aromatic-aromatic interaction between residue F117 of α^{CS} and F47 of AHSP. The RMSD value is 1.82 ± 0.15 Å and theoretical dihedral angle of structure by Ramachandran plot results that 93.4% of favoured region, 5.8% of allowed region, and 0.5% of outlier region.

From interface interactions of binding residue indicate that both of docked dimeric structures contain some H-bond, hydrophobic interaction, and aromatic-aromatic interaction when compare to binding residues on α^{CS}-globin to AHSP molecule according to experiment[55]. However, docked α^{CS}·AHSP structure has formed seven difference H-bond comparing to reference wild-type dimer while dimeric α^{WT}·AHSP has form three H-bonds that is are the same as reference wild-type dimer and another is difference. This procedure which does not use the α·AHSP(P30A) form chain C and D of PDB code 1Z8U to be a starting docking structure because α-globin in chain D lacks of the residue 137 to 141 and the residue 30 in AHSP chain C is a mutant to proline. The RMSD value of docked α·AHSP is 0.46 Å compared to crystal structure of α·AHSP(P30A) so, the docked α·AHSP structure is usable. Inspire the docked α^{CS}·AHSP is use to be a staring MD structure, it is RMSD 10.16 Å. Hight RMSD of mutant dimer is described by the CS-tail is quite jumbled near interface area where the AHSP performs binding to (**Figure 4.20**). This docked mutant structure is the lowest energy score and the lowest RMSD vale in docking results. There docked dimeric structures which become staring MD structures are the slightly different binding residues but the binding sites still occur in the same location, it means that CS-tail may interfere to interface interaction residues because it stays nearly to binding site of α^{CS}·AHSP.

4.6 Hydrogen bond analysis of dimeric α^{CS}·AHSP and α^{CS}·AHSP

H-bond analysis is determined of the distance between heavy atoms using a cutoff 3.00 Å and a cutoff 135° using for the angle between the acceptor, hydrogen, and donor atoms. A **Table 4.4** displays average H-bond distance and angle value of binding interface residues between α^{WT}·AHSP. The number of H-bond are observed at percentages higher than 45 during equilibrium phase of simulation time. There H103 of α^{WT} and D43 of AHSP, D126 of α^{WT} and Q25 of AHSP, and A123 of α^{WT} and Q24 of AHSP are still bonding in $^{3}/_{6}$ when compares to starting structure of dimeric α^{WT}·AHSP docking **Figure 4.21A**. Moreover, two more H-bond are found, D126 of α^{WT} and Q24 of AHSP, and R31 of α^{WT} and D43 of AHSP. Two H-bonds absent in equilibrium phase of simulation time are T38 of α^{WT} and D36 of AHSP, and K99 of α^{WT} and A30 of AHSP. A **Table 4.5** displays average H-bond distance

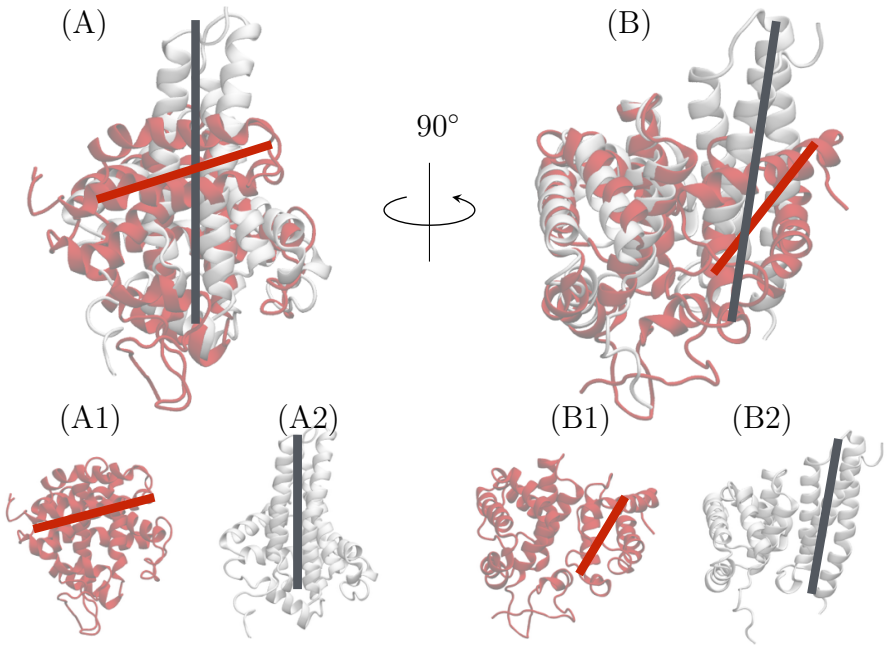

Figure 4.20 Geometrical arrangement of docked (A) and (B) structural superposition of α^{WT}·AHSP (grey colour) and structural superposition of α^{CS}·AHSP (red colour). The red line represents geometrical arrangement of AHSP in docked α^{CS}·AHSP structure and the grey line represents geometrical arrangement of AHSP in docked α^{WT}·AHSP structure

and angle value of binding interface residues between α^{CS}·AHSP. The number of H-bond are also observed at percentages higher than 45 during equilibrium phase of simulation time. There R153 of α^{CS} and D36 of AHSP, R153 of α^{CS} and T39 of AHSP, and D6 of α^{CS} and Q25 of AHSP are still $^{3}/_{7}$ bonding when compares to starting structure of dimeric α^{CS}·AHSP docking **Figure 4.21B**. Meanwhile, the absence H-bonds during equilibrium phase of simulation time are H122 of α^{CS} and N50 of AHSP, F117 of α^{CS} and Q54 of AHSP, N9 of α^{CS} and Q24 of AHSP, K127 of α^{CS} and D43 of AHSP, and V1 of α^{CS} and D43 of AHSP.

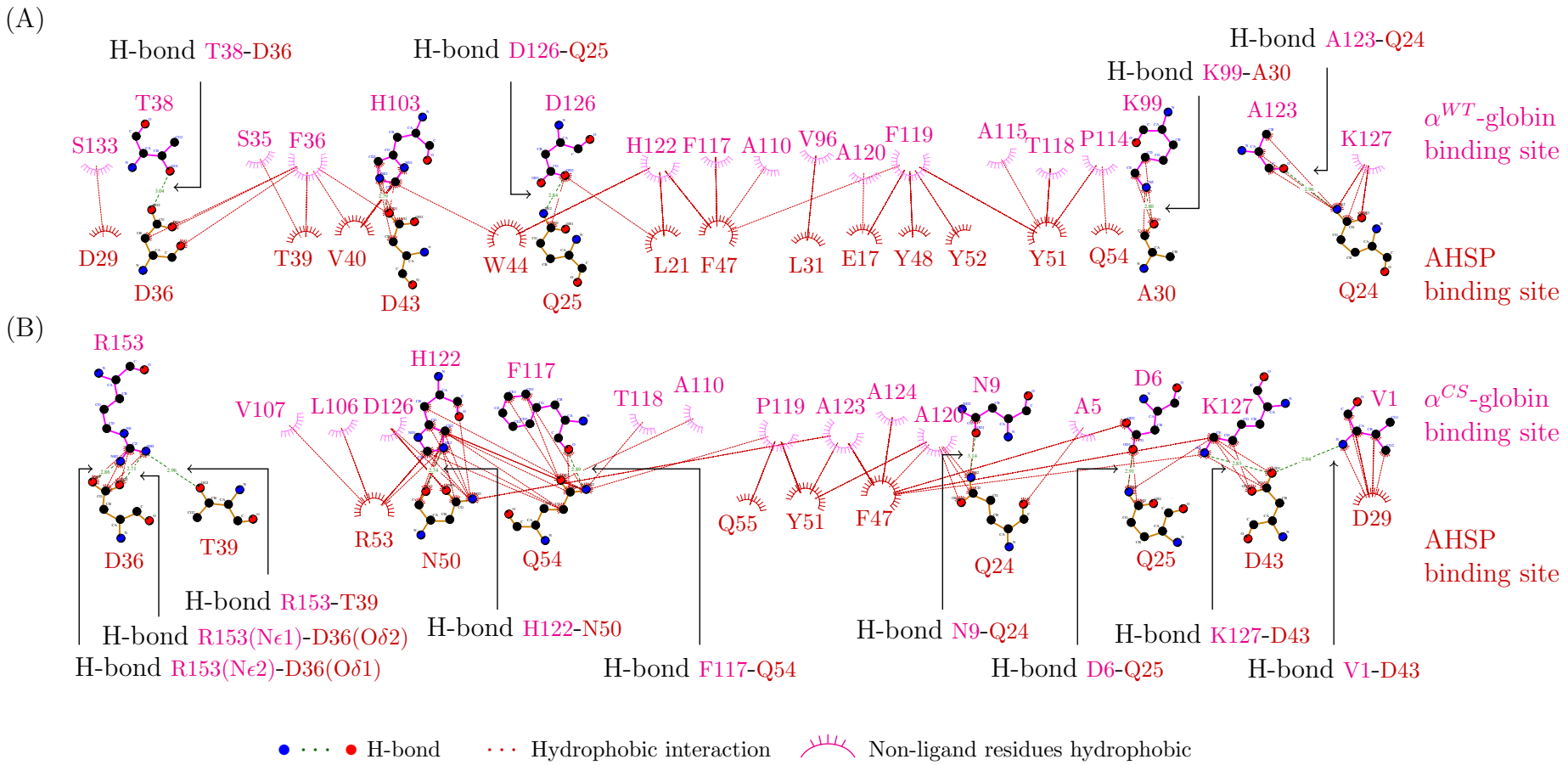

Figure 4.21 Diagrammatic of protein-protein interactions for docking starting-structures: (A) α^{WT}·AHSP dimer, (B) α^{CS}·AHSP dimer analysing by LigPlot^{+}

Table 4.4 Structural parameters of intermolecular hydrogen bonds in α^{WT}·AHSP dimer

Acceptor	DonorH	Donor	⟨Dist⟩ (Å)	⟨Ang⟩ (Å)	Found (%)
D43@OD2^{AHSP}	H103@HE2$^{\alpha^{WT}}$	H103@NE2^{AHSP}	2.72	152.94	89.04
D126@OD2$^{\alpha^{WT}}$	Q25@HE21^{AHSP}	Q25@NE2^{AHSP}	2.83	161.36	84.80
D43@OD1^{AHSP}	R31@HH21$^{\alpha^{WT}}$	R31@NH2$^{\alpha^{WT}}$	2.79	152.78	76.72
D126@OD2$^{\alpha^{WT}}$	Q24@HE22^{AHSP}	Q24@NE2^{AHSP}	2.84	155.79	70.40
A123@O$^{\alpha^{WT}}$	Q24@HE21^{AHSP}	Q24@NE2^{AHSP}	2.85	162.70	68.24
D43@OD1^{AHSP}	R31@HE$^{\alpha^{WT}}$	R31@HE$^{\alpha^{WT}}$	2.85	149.80	55.44

Table 4.5 Structural parameters of intermolecular hydrogen bonds in α^{CS}·AHSP dimer

Acceptor	DonorH	Donor	⟨Dist⟩ (Å)	⟨Ang⟩ (Å)	Found (%)
R153@O$^{\alpha^{CS}}$	T39@HG1^{AHSP}	T39@OG1^{AHSP}	2.72	160.89	57.76
T39@OG1^{AHSP}	R153@HH11$^{\alpha^{CS}}$	R153@NH1$^{\alpha^{CS}}$	2.86	157.73	54.56
D6@OD2$^{\alpha^{CS}}$	Q25@HE21^{AHSP}	Q25@NE2^{AHSP}	2.86	151.78	51.76
D36@OD2^{AHSP}	R153@HH22$^{\alpha^{CS}}$	R153@NH2$^{\alpha^{CS}}$	2.78	159.52	47.76

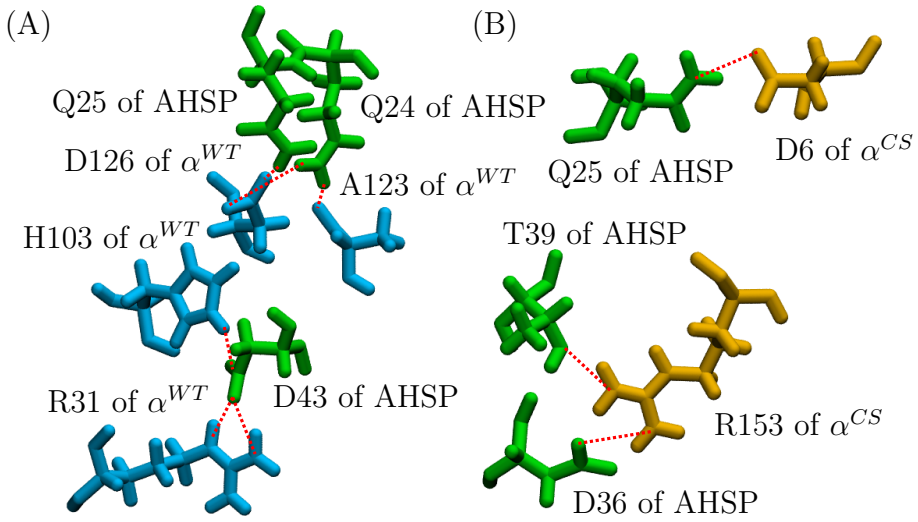

Figure 4.22 Representative simulation snapshot (at simulation time 80 ns) of H-bond network of (A) α^{WT}·AHSP and (B) α^{CS}·AHSP

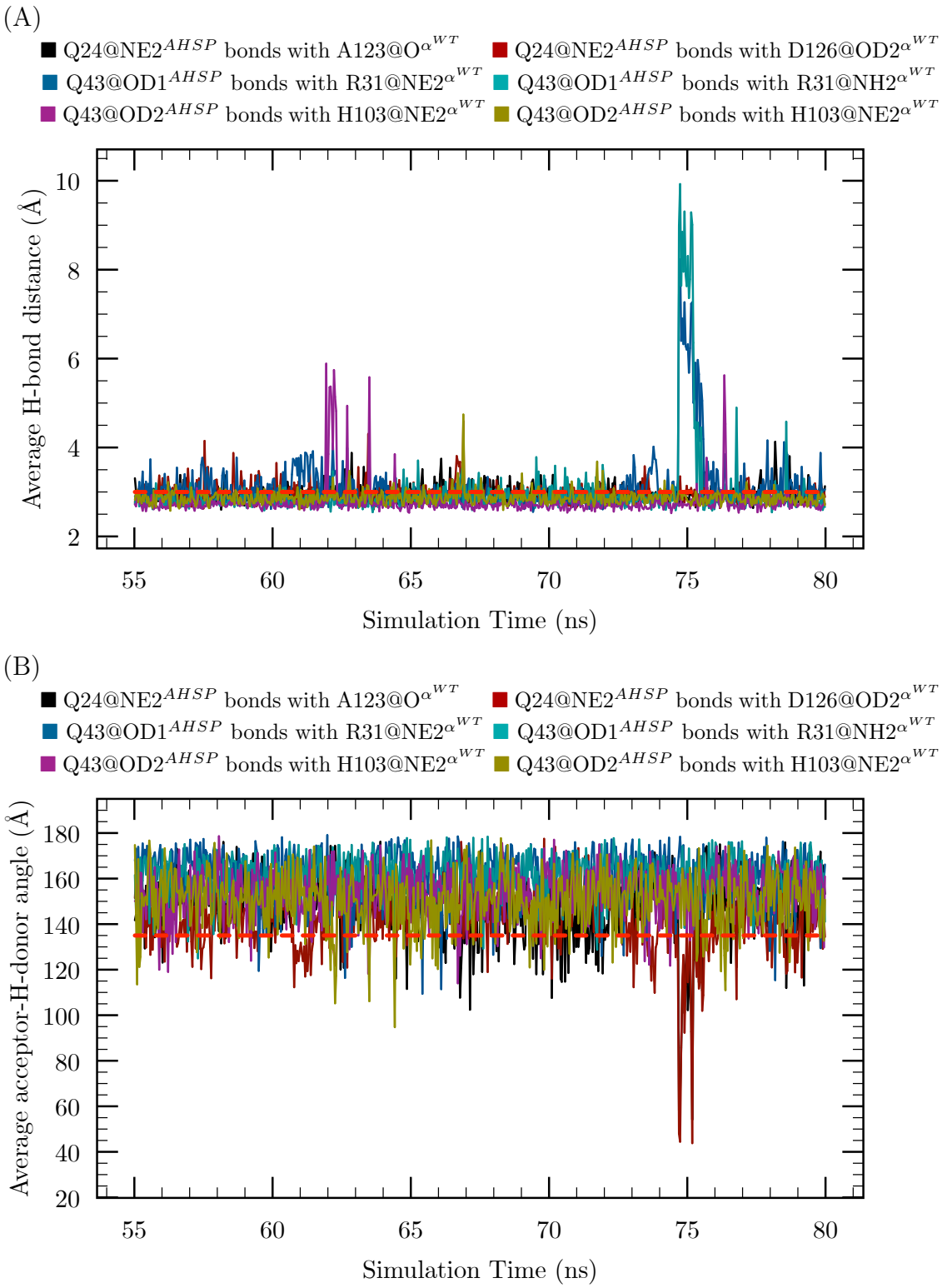

Figure 4.23 H-bond presence during the MD trajectory of α^{WT}·AHSP dimer: (A) a plot of average H-bond distance, H-bond length cutoff of 3.00 Å (red dashed line) and (B) a plot of the average donor and acceptor atoms (D–H···A) angle, cutoff of 135.00 Å (red dashed line)

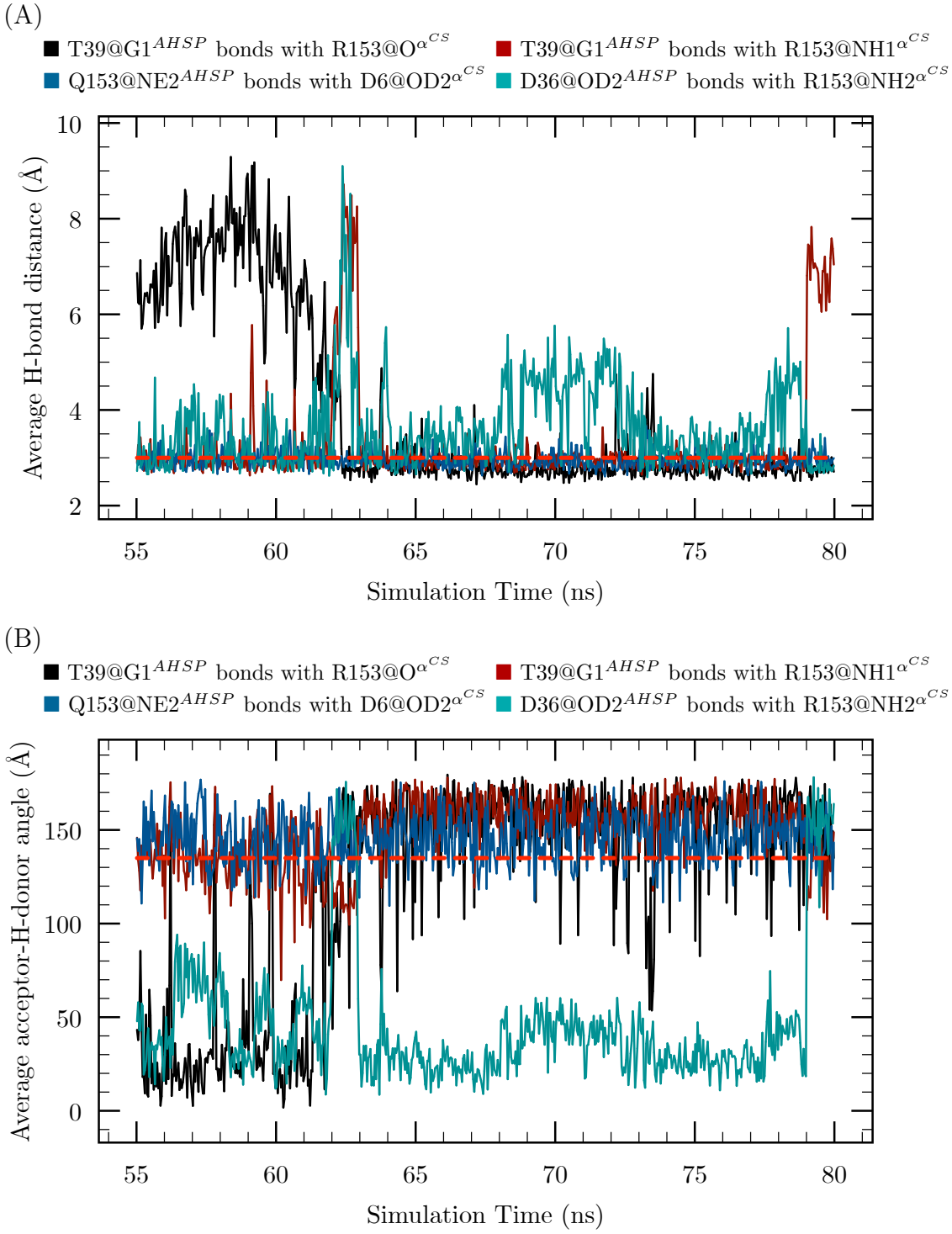

Figure 4.24 H-bond presence during the MD trajectory of α^{CS}·AHSP dimer: (A) a plot of average H-bond distance, H-bond length cutoff of 3.00 Å (red dashed line) and (B) a plot of the average donor and acceptor atoms (D–H···A) angle, angle cutoff of 135.00 Å (red dashed line)

4.7 2-D RMSD

The 2-D RMSD plots of all structures are shown for the respective molecular dynamics simulations. Each colour represents the RMSD value between Conformational during simulation time on the x-axis and the y-axis. For monomeric structures are illustrated in **4.25**, **Figure 4.26**, and **4.27** gradually move into a stable conformation in an equilibrium phase. In contrast, the α^{CS}·AHSP dimer (**Figure 4.29**) moves without reaching any stable conformation meanwhile the α^{WT}·AHSP dimer (**Figure 4.28**) moves quite into a stable conformation. It may indicate that the CS-tail may affect to the stability of mutant dimer.

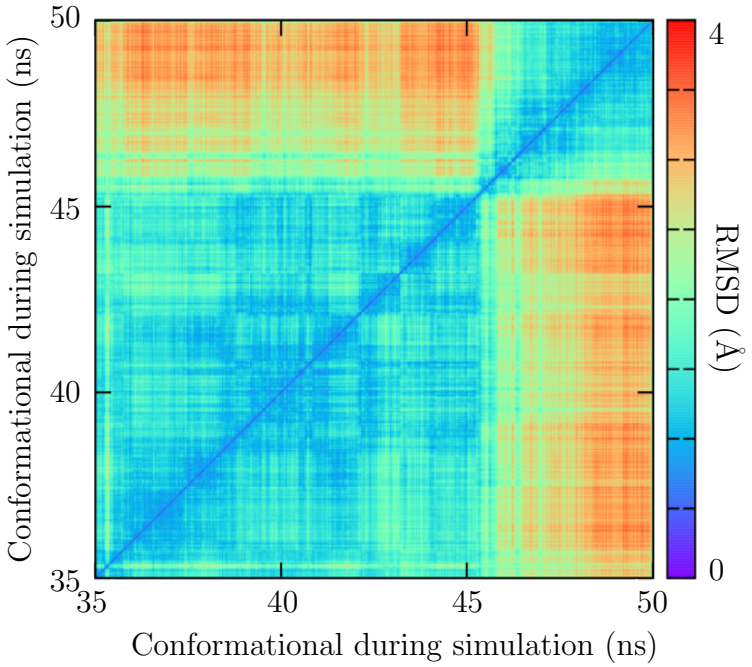

Figure 4.25 2-D RMSD plot in simulations of monomeric AHSP

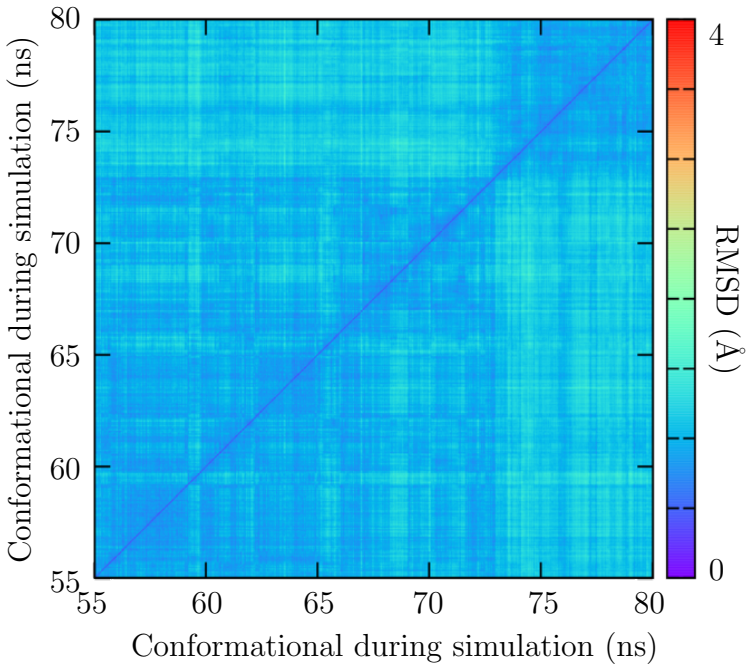

Figure 4.26 2-D RMSD plot in simulations of monomeric α^{WT}-globin

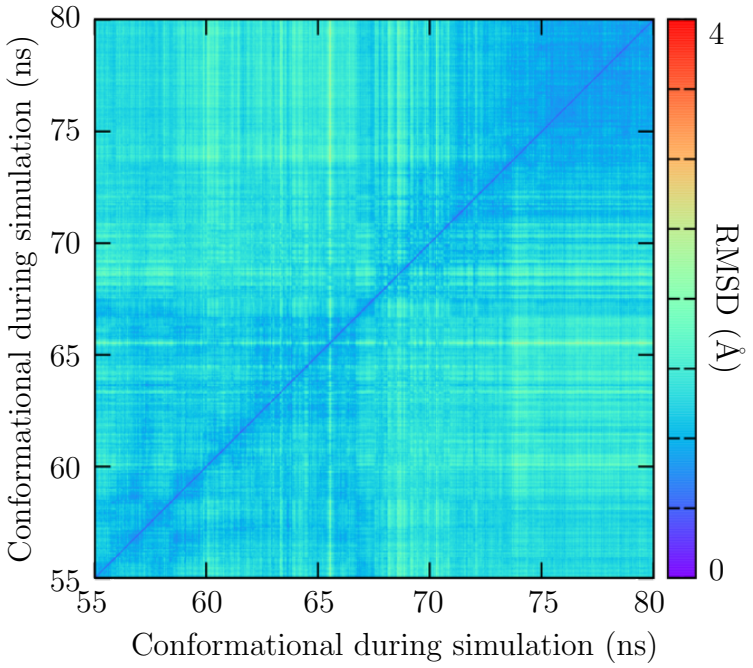

Figure 4.27 2-D RMSD plot in simulations of monomeric α^{CS}-globin

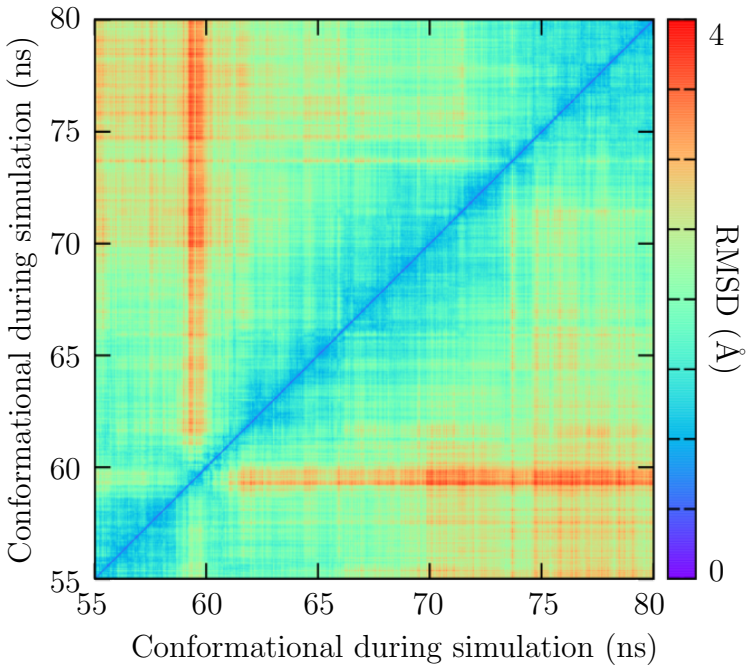

Figure 4.28 2-D RMSD plot in simulations of dimeric α^{WT}·AHSP

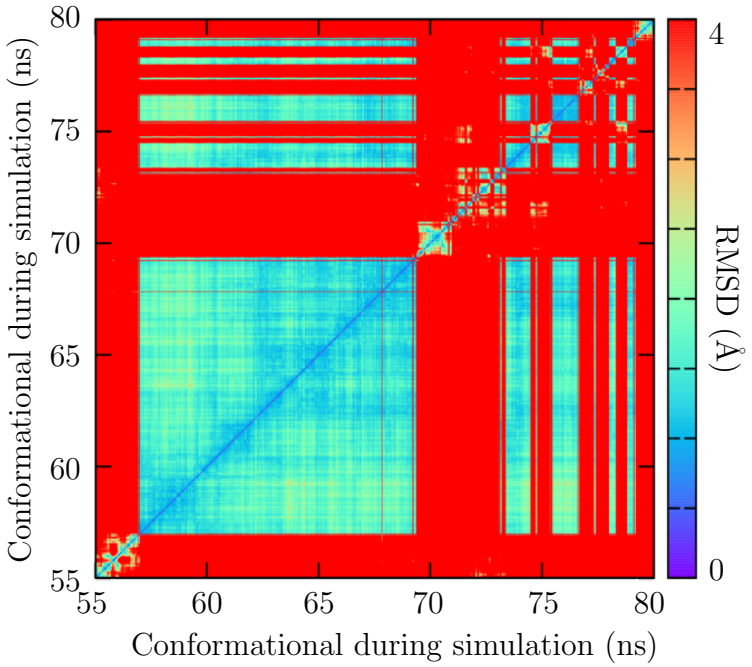

Figure 4.29 2-D RMSD plot in simulations of dimeric α^{CS}·AHSP

4.8 Trajectory analysis of MD simulations of dimeric α^{CS}·AHSP

In **Table 4.3**, RMSD values of α^{WT}·AHSP and α^{CS}·AHSP in equilibrium phase are displayed no statistically significant difference at the 95% confidence level.

4.9 Calculation of binding free energy

The solvent environment plays a significant role on molecular structures, dynamics, and energetics. Calculation of MD simulations of biological macromolecules in explicitly water molecules is very costly on the nanosecond time scale. However, an analytical binding free energy model efficiently describes electrostatics of molecules in a water environment. The binding free energy model of AMBER12 was adopted to introduce more flexibility of residue side chain for further refining of dimeric α^{WT}·AHSP and α^{CS}·AHSP. The MM/GB(PB)SA analysis allows us to separate the total binding free energy of binding into electrostatic and van der Waals solute-solute and solute-solvent interactions, thereby gaining additional insights into the physics of α^{WT}·AHSP and α^{CS}·AHSP association process. All binding free energies are calculated using 1,250 snapshots that are sampled with every 5 snapshots; these snapshots cover the last 25 ns of the trajectory. The total binding free energy is -49.95 ± 6.60 kcal/mol for α^{WT}·AHSP dimer occurring at an interface between α^{WT} and AHSP contribute favorably to the binding with the electrostatic contribution to the solvation free energy calculated by MM/GBSA. On the other hand, an unfavourable interaction is indicated to α^{CS}·AHSP dimer by the binding free energy is 7.08 ± 10.54 kcal/mol with the electrostatic contribution to the solvation free energy calculated by MM/GBSA. Moreover, the total binding free energy of α^{WT}·AHSP and α^{CS}·AHSP which is calculated with the electrostatic contribution to the solvation free energy calculated by MM/PBSA shows energetically unfavourable for binding. Values of total binding free energy are 22.83 ± 10.18 kcal/mol and 53.73 ± 12.59 kcal/mol for α^{WT}·AHSP and α^{CS}·AHSP, respectively. The difference of relative binding free energies (MM/GBSA and MM/PBSA) appearing the value α^{CS}·AHSP is higher than α^{WT}·AHSP can be emphasised that the AHSP molecule may be able to bind with wild-type better than mutant. The entropic contribution ($\langle \Delta S \rangle$) is also support for well binging of α^{WT}·AHSP dimer. The average entropic value calculated to be energetically favourable -48.32 ± 10.01 and -36.61 ± 7.82 for α^{WT}·AHSP and α^{CS}·AHSP, respectively at 298.15 K. These results is included to $T\langle S \rangle$ term that is necessary because the nonpolar solvation incorporates an estimate of the entropy changes implicitly but does not account for an entropy change upon α^{WT}·AHSP or α^{CS}·AHSP formation *in vacuo*. Denote that this results do not equal the real binding free energy and results have been not estimate the unfavourable entropy contribution to binding.

Table 4.6 Relative binding free energies MM/GBSA of binding modes of α^{WT}·AHSP and α^{CS}·AHSP dimer

Energy compartment	α^{WT}·AHSP dimer (mean ± SD)	α^{CS}·AHSP dimer (mean ± SD)
$\langle \Delta E_{vdw} \rangle$	-88.98 ± 6.23	-50.58 ± 8.93
$\langle \Delta E_{electrostatic} \rangle$	-487.03 ± 42.03	-801.16 ± 45.86
$\langle \Delta E_{GB} \rangle$	539.10 ± 40.80	841.71 ± 41.40
$\langle \Delta E_{surface} \rangle$	-13.04 ± 0.74	-9.76 ± 0.70
$\langle \Delta G_{gas} \rangle$	-576.01 ± 43.16	-824.86 ± 45.42
$\langle \Delta G_{solvation} \rangle$	526.06 ± 40.44	831.94 ± 41.15
$\langle \Delta G_{total} \rangle$	-49.95 ± 6.60	7.08 ± 10.54

Note: ΔE_{GB} represents $\Delta E_{polar\ solvation}$
$\Delta E_{surface}$ represents $\Delta E_{nonpolar\ solvation}$

Table 4.7 Relative binding free energies MM/PBSA of binding modes of α^{WT}·AHSP and α^{CS}·AHSP

Energy compartment	α^{WT}·AHSP dimer (mean ± SD)	α^{CS}·AHSP dimer (mean ± SD)
$\langle \Delta E_{vdw} \rangle$	-87.47 ± 6.16	-50.58 ± 8.93
$\langle \Delta E_{electrostatic} \rangle$	-486.47 ± 42.07	-801.17 ± 45.86
$\langle \Delta E_{PB} \rangle$	532.55 ± 39.02	827.26 ± 41.17
$\langle \Delta E_{surface} \rangle$	-70.41 ± 3.43	-50.03 ± 3.55
$\langle \Delta G_{gas} \rangle$	-575.94 ± 43.08	-824.86 ± 45.42
$\langle \Delta G_{solvation} \rangle$	598.77 ± 38.38	878.60 ± 41.74
$\langle \Delta G_{total} \rangle$	22.83 ± 10.18	53.73 ± 12.59

Note: ΔE_{PB} represents $\Delta E_{polar\ solvation}$
$\Delta E_{surface}$ represents $\Delta E_{nonpolar\ solvation}$

Table 4.8 Entropy of binding modes of α^{WT}·AHSP and α^{CS}·AHSP

Energy compartment	α^{WT}·AHSP dimer (mean ± SD)	α^{CS}·AHSP dimer (mean ± SD)
$\langle \Delta S \rangle$	-48.32 ± 10.01	-36.61 ± 7.82

CHAPTER 5

CONCLUSION

According to our simulation, two conclusions can be formulated as followings;

1. The additional CS-tail is suggested to neither effect on main-globin-structure conformation nor surface charge distribution on the binding site.

2. The CS-tails seems to interfere intermolecular interactions on the binding interface with AHSP and influence the haemoglobin binding affinity with the AHSP. Therefore the CS-tail may have an effect on an early phase of Hb formation.

These conclusions are also in good agreement with a previously reported experiment[56]. Furthermore it is supported by the report that presence of membrane-bound α^{CS}-globin chains may make for the oxidative damage found on the erythrocytic membrane by unstable α^{CS}-globin. More recently, the oxidative damage in erythrocytic membrane is detected on blood smear as hypochromic, codocytes, and Heinz bodies[3,4]. All aforementioned evidences reveal molecular effect due to CS-tail on anaemia occurrence via HbH CS diseases.

For future perspectives,

1. The MD simulation should be varied under different physiological condition such as hight temperature (greater than 310 K), low temperature (less than 310 K), blood acidosis (pH less than 7.2), or blood alkalosis (pH greater than 7.4). Various physiological conditions may be important to understand the mechanism of patient carried HbCS combining symptoms with that factors. The normal body strictly regulates temperature and pH to prevent their proteins[57] for keep working as its functions. In case of an abnormal condition, the amino acid sequence (the protein's primary structure) does not change, the α^{CS}'s shape may change. If the α^{CS} conformation is changed, it may allow AHSP for binding force close to wild-type.

2. The Hb formation within mutant CS-tail should be completely studied of $\alpha^{CS}\cdot\beta$ assembly, and tetrameter Hb formation.

3. For the most accurate result, this study should be compared with the result of an MD simulation using X-ray crystal α^{CS}-globin structure.

BIBLIOGRAPHY

1. Viprakasit V, Tanphaichitr VS, Chinchang W, Sangkla P, Weiss MJ, Higgs DR. Evaluation of alpha hemoglobin stabilizing protein (AHSP) as a genetic modifier in patients with β thalassemia. Blood 2004;103(9):3296–3299.

2. Mollan TL, Yu X, Weiss MJ, Olson JS. The role of alpha-hemoglobin stabilizing protein in redox chemistry, denaturation, and hemoglobin assembly. Antioxidants & redox signaling 2010;12(2):219–231.

3. Kihm AJ, Kong Y, Hong W, Russell JE, Rouda S, Adachi K, et al. An abundant erythroid protein that stabilizes free [alpha]-haemoglobin. Nature 2002;p 758–763.

4. Sugawara Y, Shigemasa Y, Hayashi Y, Abe Y, Ueno I, Ohgushiand E. New Mode (Molecular-Sensing) of Heinz Body Formation Mechanisms Inherent in Human Erythrocytes: Basis for Understanding of Clinical Aspects of Drug-Induced Hemolytic Anemia and the Like. Journal of Bioanalysis & Biomedicine 2013;.

5. Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. vol 1. Academic press; 2001.

6. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, et al.. AMBER 12. University of California, San Francisco; 2012.

7. Tipmanee V. Towards a quantitative prediction of reorganisation energy for intraprotein electron transfer. University of Cambridge; 2012.

8. Praprotnik M, Site LD, Kremer K. Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution. Annual Review of Physical Chemistry 2008;59(1):545–571.

9. Modell B, Darlisona M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ 2008;86(6):480–487.

10. Chui DHK, Fucharoen S, Chan V. Hemoglobin H disease: not necessarily a benign disorder. Blood 2003;101(3):791–800.

11. Luzzatto L, Notaro R. Physiology: Haemoglobin's chaperone. Nature 2002;417(6890):703–705.

12. Kiger L, Vasseur C, Domingues-Hamdi E, Truan G, Marden MC, Baudin-Creuza V. Dynamics of α-Hb chain binding to its chaperone AHSP depends

on heme coordination and redox state. Biochimica et Biophysica Acta (BBA)-General Subjects 2014;1840(1):277–287.

13. Thom CS, Dickson CF, Gell DA, Weiss MJ. Hemoglobin variants: biochemical properties and clinical correlates. Cold Spring Harbor perspectives in medicine 2013;3(3):a011858.

14. Liu J, Konermann L. Assembly of hemoglobin from denatured monomeric subunits: heme ligation effects and off-pathway intermediates studied by electrospray mass spectrometry. Biochemistry 2013;52(10):1717–1724.

15. Bunn HF. Subunit assembly of hemoglobin: an important determinant of hematologic phenotype. Blood 1987;69(1):1–6.

16. Mollan T. The Role of Alpha–Hemoglobin Stabilizing Protein in Human Hemoglobin Assembly. Biochemistry and Cell Biology, Rice University. Houston, Texas; 2011.

17. Weiss MJ, dos Santos CO. Chaperoning erythropoiesis. Blood 2009;113(10):2136–2144.

18. Feng L, Zhou S, Gu L, Gell DA, Mackay JP, Weiss MJ, et al. Structure of oxidized alpha-haemoglobin bound to AHSP reveals a protective mechanism for haem. Nature 2005;435(7042):697–701.

19. Weiss MJ, Zhou S, Feng L, Gell DA, Mackay JP, Shi Y, et al. Role of Alpha Hemoglobin-Stabilizing Protein in Normal Erythropoiesis and β-Thalassemia. Annals of the New York Academy of Sciences 2005;1054(1):103–117.

20. Zhou S, Olson JS, Fabian M, Weiss MJ, Gow AJ. Biochemical fates of α hemoglobin bound to α hemoglobin-stabilizing protein AHSP. Journal of Biological Chemistry 2006;281(43):32611–32618.

21. Korapin S. Serum Ferritin Levels in Various Status of Thalassemic Children [Diplomate, Thai Board of Pediatrics]. The Royal College of Pediatricians of Thailand; 2000.

22. Harteveld C, Higgs D. alpha-thalassaemia. Otphanet J Rare Dis 2010;5(1):13.

23. Fucharoen S, Viprakasit V. Hb H disease: clinical course and disease modifiers. Hematology Am Soc Hematol Educ Program 2009;2009(1):26–34.

24. Park SY, Yokoyama T, Shibayama N, Shiro Y, Tame JRH. 1.25 Å Resolution Crystal Structures of Human Haemoglobin in the Oxy, Deoxy and Carbonmonoxy Forms. J Mol Biol 2006;360(3):690–701.

25. Milner PF, Weatherall DJ, Clegg JB. Haemoglobin-H Disease Due to A Unique Haemoglobin Variant with An Elongated α-chain. Lancet 1971;297(7702):729–732.

26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235–242.

27. Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW, et al. Web servers and services for electrostatics calculations with APBS and PDB2PQR. J Comput Chem 2011;32(7):1488–1491.

28. Buchan DW, Minneci F, Nugent TC, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic acids research 2013;41(W1):W349–W357.

29. Petersen B, Petersen T, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Structural Biology 2009;9(1):51.

30. Blaszczyk M, Jamroz M, Kmiecik S, Kolinski A. CABS-fold: server for the de novo and consensus-based prediction of protein structure. Nucleic acids research 2013;41(W1):W406–W411.

31. Lauck F, Smith CA, Friedland GF, Humphris EL, Kortemme T. RosettaBackrub?a web server for flexible backbone protein structure modeling and design. Nucleic Acids Res 2010;38(suppl 2):W569–W575.

32. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for proteinprotein docking. Nucleic Acids Res 2004;32(suppl 2):W96–W99.

33. Humphrey W, Dalke A, Schulten K. VMD – Visual Molecular Dynamics. J Mol Graphics Modell 1996;14(1):33–38.

34. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera a visualization system for exploratory research and analysis. J Comput Chem 2004;25(13):1605–1612.

35. Chen AA, Pappu RV. Parameters of Monovalent Ions in the AMBER-99 Forcefield:? Assessment of Inaccuracies and Proposed Improvements. The Journal of Physical Chemistry B 2007;111(41):11884–11887.

36. Laskowski RA, Swindells MB. LigPlot+: Multiple LigandProtein Interaction Diagrams for Drug Discovery. J Chem Inf Model 2011;51(10):2778–2786.

37. Takayanagi M, Kurisaki I, Nagaoka M. Oxygen entry through multiple pathways in T-state human hemoglobin. J Phys Chem B 2013;117(20):6082–6091.

38. Chatake T, Shibayama N, Park SY, Kurihara K, Tamada T, Tanaka I, et al. Protonation States of Buried Histidine Residues in Human Deoxyhemoglobin Revealed by Neutron Crystallography. J Am Chem Soc 2007;129(48):14840–14841.

39. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, et al. Structure validation by C_*alpha* geometry: ϕ,ψ and C_*beta* deviation. Proteins: Structure, Function, and Bioinformatics 2003;50(3):437–450.

40. Giammona DA. An Examination of Conformational Flexibility in Porphyrins and Bulky-ligand Binding in Myoglobin. University of California, Davis; 1984.

41. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23(3):327–341.

42. Scott R. Math Comp 1991;57(195):442–444.

43. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 1993;98(12):10089–10092.

44. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79(2):926–935.

45. Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81(8):3684–3690.

46. Reva BA, Finkelstein AV, Skolnick J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? Folding and Design 1998;3(2):141–147.

47. Amis ES. Coulomb's law and the quantitative interpretation of reaction rates. J Chem Educ 1953;30(7):351.

48. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, et al. The Amber biomolecular simulation programs. J Comput Chem 2005;26(16):1668–1688.

49. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Accounts of chemical research 2000;33(12):889–897.

50. Hayes JM, Archontis G. MM-GB (PB) SA calculations of protein-ligand binding free energies. INTECH Open Access Publisher; 2012.

51. Chongsuvivatwong V. Analysis of Epidemiological Data Using R and Epicalc. Book Unit, Faculty of Medicine, Prince of Songkla University; 2008.

52. Zaki M, Bystroff C. Protein structure prediction. vol 413. Springer Science & Business Media; 2008.

53. Sourirajan J. Protein Structure Prediction. Tech. Rep; 2004.

54. Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. Journal of molecular biology 2008;380(4):742–756.

55. Yu X, Mollan TL, Butler A, Gow AJ, Olson JS, Weiss MJ. Analysis of human α globin gene mutations that impair binding to the α hemoglobin stabilizing protein. Blood 2009;113(23):5961–5969.

56. Turbpaiboon C. Alpha-globin termination mutants: intra-erythrocytic pathophysiology and DNA detection. Department of Biochemistry, Faculty of Science, Mahidol University. Sala Ya; 2006.

57. Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiology and Molecular Biology Reviews 2001;65(1):1–43.

58. Jeffrey GA, Saenger W. Hydrogen bonding in biological structures. Springer Science & Business Media; 2012.

59. Aruksakunwong O. Study of full range structure of p53 tetramer and binding mechanism with DNA using molecular dynamics simulation and X-ray crystal structure [Research report (MRG5180135)]; 2001.

60. Krissinel E, Henrick K. Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute. J Mol Biol 2007;372:774–797.

61. Achary MS, Reddy ABM, Chakrabarti S, Panicker SG, Mandal AK, Ahmed N, et al. Disease-Causing Mutations in Proteins: Structural Analysis of the CYP1b1 Mutations Causing Primary Congenital Glaucoma in Humans. Biophys J 2006;91(12):4329–4339.

62. Bergeron BP. Bioinformatics Computing. Prentice Hall; 2003.

63. Wu Z. Lecture Notes on Computational Structural Biology. California:World Scientific Publishing Company Pte Limited; 2008.

64. Sitkoff D, Lockhart DJ, Sharp KA, Honig B. Calculation of electrostatic effects at the amino terminus of an alpha helix. Biophys J 1994;67(6):2251–2260.

65. Lobanov MY, Bogatyreva NS, Galzitskaya OV. Radius of gyration as an indicator of protein structure compactness. Mol Biol 2008;42(4):623–628.

66. Patodia S, Bagaria A, Chopra D. Molecular Dynamics Simulation of Proteins: A Brief Overview. J Phys Chem Biophys 2014;4(6):1.

67. Terry CA, Fernndez MJ, Gude L, Lorente A, Grant KB. Physiologically Relevant Concentrations of NaCl and KCl Increase DNA Photocleavage by an N-Substituted 9-Aminomethylanthracene Dye. Biochemistry 2011;50(47):10375–10389.

68. Nelson DL, Cox MM. Lehninger Principles of Biochemistry, Fourth Edition. Fourth edition ed Freeman, editor; 2004.

69. Eaton WA, Henry ER, Hofrichter J, Mozzarelli A. Is cooperative oxygen binding by hemoglobin really understood? Nat Struct Mol Biol 1999;6(4):351–358.

70. Hardison RC. Evolution of Hemoglobin and Its Genes. Cold Spring Harbor Perspectives in Medicine 2012;2(12).

71. Freitas TAK, Hou S, Dioum EM, Saito JA, Newhouse J, Gonzalez G, et al. Ancestral hemoglobins in Archaea. Proceedings of the National Academy of Sciences of the United States of America 2004;101(17):6675–6680.

72. Levitt M, Warshel A. Computer simulation of protein folding. Nature 1975;253(5494):694–698.

73. Levitt M, Lifson S. Refinement of protein conformations using a macromolecular energy minimization procedure. J Mol Biol 1969;46(2):269–279.

74. Louise N Johnson DCP. Structure of some crystalline lysozyme–inhibitor complexes determined by X-ray analysis at 6 Angstrom resolution. Nature 1965;206(4986):761–763.

75. Campbell ID. The march of structural biology. Nature Reviews Molecular Cell Biology 2002;3(1):377–381.

76. Pray LA. Discovery of DNA Structure and Function: Watson and Crick. Nature 2008;1(1):100.

77. Eisenberg D. The discovery of the α-helix and β-sheet, the principal structural features of proteins. Proceedings of the National Academy of Sciences 2003;100(20):11207–11210.

78. Sadowski MI, Jones DT. The sequencestructure relationship and protein function prediction. Current Opinion in Structural Biology 2009;19(3):357 – 362.

79. Street TO, Courtemanche N, Barrick D. Protein Folding and Stability Using Denaturants. In: Correia DJJ, Dr H William Detrich I, editors. Biophysical Tools for Biologists, Volume One: In Vitro Techniques. vol 84 of Methods in Cell Biology Academic Press; 2008. p 295–325.

80. Ollikainen N, Kortemme T. Computational Protein Design Quantifies Structural Constraints on Amino Acid Covariation. PLoS Comput Biol 2013;9(11):e1003313.

81. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK server: interactive docking prediction of protein?protein complexes and symmetric multimers. Bioinformatics 2014;30(12):1771–1773.

82. Krawczyk K, Liu X, Baker T, Shi J, Deane CM. Improving B-cell epitope prediction and its application to global antibody-antigen docking. Bioinformatics 2014;30(16):2288–2294.

83. Shang YD, Zhang JL, Wang Y, Zhang HX, Zheng QC. Molecular simulation investigation on the interaction between barrier-to-autointegration factor dimer or its Gly25Glu mutant and LEM domain of emerin. Computational Biology and Chemistry 2014;53, Part B(0):184–190.

84. Zu X, Liu Y, Wang S, Jin R, Zhou Z, Liu H, et al. Peptide inhibitor of Japanese encephalitis virus infection targeting envelope protein domain III. Antiviral Research 2014;104(0):7–14.

85. Abernethy CD, Codd GM, Spicer MD, Taylor MK. A highly stable N-heterocyclic carbene complex of trichloro-oxo-vanadium(v) displaying novel Cl—C(carbene) bonding interactions. J Am Chem Soc 2003;125(5):1128–1129.

86. Arduengo AJ III, Gamper SF, Calabrese JC, Davidson F. Low-coordinate carbene complexes of nickel(0) and platinum(0) 1994;116(10):4391–4394.

87. Appelhans LN, Zuccaccia D, Kovacevic A, Chianese AR, Miecznikowski JR, Macchioni A, et al. An anion-dependent switch in selectivity results from a change of CH activation mechanism in the reaction of an imidazolium salt with IrH5 (PPh3) 2. Journal of the American Chemical Society 2005;127(46):16299–16311.

88. Communication from the European Commission to the European Council and the European Parliament: 20 20 by 2020: Europe's climate change opportunity. Brussels, Belgium; 2008.

89. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al.. Gaussian 03. Wallingford, CT:Gaussian, Inc.; 2004. Gaussian, Inc., Wallingford, CT, USA.

90. Abarca A, Gómez-Sal P, Martín A, Mena M, Poblet JM, Yélamos C. Ammonolysis of mono(pentamethylcyclopentadienyl) titanium(IV) derivatives. Inorg Chem 2000;39(4):642–651.

91. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford);bar009.

92. Santiveri CM, Pérez-Cañadillas JM, Vadivelu MK, Allen MD, Rutherford TJ, Watkins NA, et al. NMR Structure of the α-Hemoglobin Stabilizing Protein: insights into conformational heterogeneity and binding. Journal of Biological Chemistry 2004;279(33):34963–34970.

93. Feng L, Gell DA, Zhou S, Gu L, Kong Y, Li J, et al. Molecular Mechanism of AHSP-Mediated Stabilization of α-Hemoglobin. Cell 2004;119(5):629–640.

94. Gibson G, Muse SV. A Primer of Genome Science. Sinauer Associates; 2009.

95. Frishman D, Argos P. Seventy-five percent accuracy in protein secondary structure prediction. Proteins 1997;27(3):329–335.

96. Weatherall DJ, Clegg JB. Hemoglobin Constant Spring, and unusual alpha-chain variant involved in the etiology of hemoglobin H disease. Annu NY Acad Sci 1974;232(1):168–178.

97. Thévenet P, Shen Y, Maupetit J, Guyon F, Derreumaux P, Tufféry P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. Nucleic Acids Res 2012;40(W1):W288–W293.

98. Maupetit J, Derreumaux P, Tuffery P. PEP-FOLD: an online resource for de novo peptide structure prediction. Nucleic Acids Res 2009;37(suppl2):W498–W503.

99. Laboratories of Computer Science, Engineering and Biochemistry, Molecular Biology at the Pennsylvania State University. Hb Constant Spring (or CS);.

100. Alder BJ, Wainwright TE. Studies in Molecular Dynamics. I. General Method. J Chem Phys 1959;31(2):459–466.

101. Leach AR. Chapter 7. Molecular Dynamics Simulation Methods. In: 2nd, editor. Molecular Modelling: Principles and Applications Pearson Education. New Jersey:Prentice Hall; 2001. p 353.

102. The Nobel Foundation. The Nobel Prize in Chemistry 1962; 2013.

103. Kumar A, Kushwaha R, Gupta C, Singh U. An analytical study on peripheral blood smears in anemia and correlation with cell counter generated red cell parameters. J Appl Hematol 2013;4(4):137–144.

104. Horn BKP. Closed-form solution of absolute orientation using unit quaternions. J Opt Soc Am A 1987;4(4):629–642.

105. Thomas S, Tang X, Tapia L, Amato NM. Simulating Protein Motions with Rigidity Analysis. J Comp Biol 2007;14(6):839–855.

106. Shehu A, Kavraki LE, Clementi C. Multiscale characterization of protein conformational ensembles. Proteins: Structure, Function, and Bioinformatics 2009;76(4):837–851.

107. Givon D, Kupferman R, Stuart A. Extracting macroscopic dynamics: model problems and algorithms. Nonlinearity 2004;17:55–127.

108. Adcock SA, McCammon JA. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. Chem Rev 2006;5(106):1589–1615.

109. Garman EF. Developments in X-ray Crystallographic Structure Determination of Biological Macromolecules. Science 2014;343(6175):1102–1108.

110. Mittermaier A, Kay LE. New Tools Provide New Insights in NMR Studies of Protein Dynamics. Science 2006;312(5771):224–228.

111. Berman HM. Annual Report 2013. Rutgers, The State University of New Jersey, University of California, San Diego:Research Collaboratory for Structural Bioinformatics (RCSB); 2014. SDSC Report GA-A21224.

112. Whitford D. Proteins: Structure and Function. West sussex:Wiley & Sons, Incorporated, John; 2005.

113. Petsko GA, Ringe D. Protein Structure and Function. Primers in biology. London:New Science Press; 2004.

114. Chiang TH, Hsu D, Latombe JC. Markov dynamic models for long-timescale protein motion. Bioinformatics 2010;26(12):i269–i277.

115. Cowan JA. Metal Activation of Enzymes in Nucleic Acid Biochemistry. Chem Rev 1998;98(3):1067–1088.

116. Robinson RA, Stokes RH. Electrolyte solutions. 2nd ed Mineola:Dover Publications; 2002.

117. Nantasenamat C, Prachayasittikul V, Bulow L. Molecular Modeling of the Human Hemoglobin-Haptoglobin Complex Sheds Light on the Protective Mechanisms of Haptoglobin. PLoS ONE 2013;8(4):e62996 (1–11).

118. Baldwin J, Chothia C. Haemoglobin: The structural changes related to ligand binding and its allosteric mechanism. J Mol Biol 1979;129(2):175–220.

119. Olsson MHM, Sazndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. J Chem Theory Compu 2011;7(2):525–537.

120. Kelley L, Bennett-Lovsey R, Herbert A, Fleming K. Protein Homology/analogY Recognition Engine V 2.0; 2009. Available from: `http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index` [cited 2014-03-08].

121. Kavanaugh JS, Rogers PH, Arnone A, Hui HL, Wierzba A, DeYoung A, et al. Intersubunit interactions associated with Tyr42 alpha stabilize the quaternary-T tetramer but are not major quaternary constraints in deoxyhemoglobin. Biochemistry 2005;44(10):3806–3820.

122. Eaton WA, Henry ER, Hofrichter J, Bettati S, Viappiani C, Mozzarelli A. Evolution of allosteric models for hemoglobin. IUBMB Life 2007;59(8-9):586–599.

123. Borgstahl GEO, Rogers PH, Arnone A. The 1.9 Å structure of deoxy beta 4 hemoglobin. Analysis of the partitioning of quaternary-associated and ligand-induced changes in tertiary structure. J Mol Biol 1994;236(3):831–843.

124. Jenkins JD, Musayev FN, Danso-Danquah R, Abraham DJ, Safo MK. Structure of relaxed-state human hemoglobin: insight into ligand uptake, transport and release. Acta Crystallogr Sect D-Biol Crystallogr 2009;65(Pt1):41–48.

125. Hub JS, Kubitzki MB, de Groot BL. Spontaneous Quaternary and Tertiary T-R Transitions of Human Hemoglobin in Molecular Dynamics Simulation. PLOS Comput Biol 2010;6(5):e1000774 (1–11).

126. Weatherall DJ. Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias. Nat Rev Genet 2001;2(4):245–255.

127. Provan D, Singer CRJ, Baglin T, Dokal I. Oxford HandBOOK of Clinical Haematology. 3rd ed Oxford:Oxford university press (OUP); 2009.

128. Stevenson A. Oxford Dictionary of English. Oxford reference online premium. Oxford:Oxford university press (OUP); 2010.

129. Bain BJ. Chapter 1. Haemoglobin and the Genetics of Haemoglobin Synthesis. In: 2nd, editor. Haemoglobinopathy Diagnosis Oxford:Blackwell Publishing Ltd; 2007. p 1–25.

130. Pornprasert S, Panyasai S, Waneesorn J, Kongthai K, Singboottra P. Quantification of hemoglobin Constant Spring in heterozygote and homozygote by a capillary electrophoresis method. Int J Lab Hematol 2012;34(2):143–147.

131. Nguyena VH, Sanchaisuriya K, Wongprachum K, Nguyen MD, Phan TTH, Vo VT, et al. Hemoglobin Constant Spring is markedly high in women of an ethnic minority group in Vietnam: A community-based survey and hematologic features. Blood Cells Mol Dis 2014;52(4):161–165.

132. Fang J, Chen L, Zeng R, Tian Q, Jiang W, Li H, et al. The Hb H Disease Genotypes in Southern China. Hemoglobin 2014;38(1):76–78.

133. Winichagoon P, Fucharoen S. Haemoglobinopathies in Southeast Asia. Indian J Med Res;134(4):498.

134. Clegg JB, Weatherall DJ, Milner PF. Haemoglobin Constant Spring–A Chain Termination Mutant?. Nature 1971;234(5328):337–340.

135. Steinberg MH, Forget BG, Higgs DR, Weatherall DJ. Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. 2nd ed Cambridge medicine. Cambridge:Cambridge University Press; 2009.

APPENDICES

APPENDIX A

PROCEEDING

Computational Study of the Effect due to Chain Terminal Mutation in Human Alpha-haemoglobins

Nawanwat C Pattaranggoon[1], Varomyalin Tipmanee[1,a]
[1]Department of Biomedical Sciences, Faculty of Medicine, Prince of Songkla University, Thailand
[a] ***E-mail:*** *tvaromya@medicine.psu.ac.th;* ***Fax:*** *+ 66 7 442 9584;* ***Tel.*** *+ 66 7445 1180*

ABSTRACT

Introduction: Thalassaemia is an inherited autosomal recessive blood disorder which is concerning quantitative reduction in globin chain synthesis in thalassaemia syndromes, and can be differentiated from the structural changes seen in abnormal haemoglobin. The structurally abnormal haemoglobin are mostly produced in normal amounts. However, some abnormal haemoglobin are also associated with reduced globin chain production. Popular types of newsense mutations, for example Paksé (HbPS), Icaria (HbIC), and Seal Rock (Hb SR) as well as some frameshift mutation i.e. Wayne type I and Wayne type II, (HbW-I) and (HbW-II), respectively are included in this study, compared to previous study of Constant Spring (HbCS) case and wild type monomeric α-globins. Objective: To investigate the effect due to additional mutated structure on overall α-globin conformation under virtual human *in vivo* condition. Methodology: All mutated tertiary structures were predicted the elongated tail using bioinformatics tools. All MD simulation was carried out with PMEMD module in AMBER12 package under 0.15 M NaC, 310 K condition. The protein conformations were eventually analysed using PTRAJ module and some manually written programmes. A structure visualisation was done using VMD package. Conclusion: The study suggested that the additional protein structure at the C-terminus did not affect the protein conformation significantly. In other words, our work excluded the hypothesis that the mutation can lead to an absurdity in conformation, in addition, this work demonstrated a visualisation of how α-globins with an additional part due to the causative mutation behaved under *in vivo* condition, with a reference of native protein.

Keywords: Structure predictions, Molecular Dynamic Simulation, Haemoglobin, Thalassaemia

1. INTRODUCTION

Thalassaemia is a hereditary disease highly found in tropical country including Thailand. The disease comes from an abnormality in haemoglobin (Hb) protein, locating in an erythrocyte (red blood cell) and leading to various symptoms from mild anaemia to even a death. Two Hb types, α- and β-globins, are responsible to the disease, and the disease can therefore be categorised into two groups, α- and β-thalassemia, according to the abnormality in corresponding globin type. The genes responsible in coding these proteins are HBA and HBB respectively. Once a mutation occurs in these genes, the obtained globin protein will be affected. In order to understand the disease mechanism in molecular scale, a tertiary structure of abnormal Hb is necessary. Unfortunately, three dimensional structure of Hb mutant is rarely available in database. Herein we focus on α-thalassaemia 2 (α2-thal or α+-thal) in which the mutation in α-globin gene yields an additional protein chain from the C-terminus of α-globin. In other word, the globin structure contains extension part in the structure. The number of extended amino acids is varied in each mutation. According to above-mentioned unavailability regarding a protein structure, we applied some bioinformatics tools and molecular dynamics simulation to study a set of abnormal Hb able to cause many α-thal types. The current aim is to investigate an effect due to the mutation on globin conformation. Many popular mutations [1] are included in this study namely Paksé (HbPS), Koya Dora (HbKD), and Seal Rock (HbSR), respectively. These simulations are compared to previous study of Constant Spring (HbCS) case and wild-type monomeric α-globins.

2. THEORY AND RELATED WORKS

Computer simulation is proved to be a powerful apparatus to study the structural and functional behaviours of macromolecules such as enzymes or polynucleotides. One of the commonplace methods is molecular dynamics simulation, relied on Newtonian mechanics and treating the atom as a rigid sphere. The method can successfully reflect the dynamics system and reproduce an experimental results [10]. In 1959, the molecular structure of myoglobin (similar to Hb) was discovered by Max Ferdinand Perutz, the 1962 Nobel Prize in Chemistry. The role of oxygen binding cooperatively Hb in the blood was also elucidated in many reports [11,12,13]. Results form the studies of William A. Eaton et al. are summarised that in the T (deoxyHb) to R (oxyHb) state transition, the structure rotates ≈ 15 Angstroms. This conformational changes is that α1 interacts with β2, on the other hand, α2 interacts with β1. From studies of Barbara J. Bain and H. Franklin Bunn proposed that α·β dimer was normally formed by α-globin monomer, which prefers to pair with β-globin monomer. However, the crystal structure cannot represent the phenomenon under natural condition since the crystallisation media is far different from the erythrocytic entity. Furthermore, the thermal fluctuation because of temperature is neglected in the crystal since the structure is generally solved at cryogenic temperature.

Recently the MD studies of electron transfer in haem-containing-proteins [14,15] were performed and published results were in good agreement with experiments. Also The study concerning formation of dimer by protein-protein binding which resembles Hb dimer formation or Hb tetramer formation by the study of Chanin Nantasenamat et al. [16] was reported. Therefore we have already put these studies as a benchmark for our heroin simulations.

3. COMPUTATIONAL DETAILS

All additional mutated structures from C-terminus are constructed via secondary and tertiary structure prediction. The best predicted structure from each mutation will be manually attached to an experimental monomeric α- globin (X-ray structure, PDB code 2DN2) [2]. The secondary structure prediction was performed independently from PREDATOR [3] and NetSurfP version 1.1 [4] webserver tools. The tertiary structure prediction was carried out with CABS-fold webserver [5]. The results from these predictions were analysed together to obtain the best predicted structure. Molecular dynamics simulation was exploited in order to investigate the mutation effect on protein conformation under virtual human *in vivo* condition. Thus the protonation state of all ionisable amino acids in the mutant globin was modelled at pH 7.4 using PDB2PQR (formerly known as ProPKa) [6]. The protein was then solvated by TIP3P water and NaCl, and simulated under 0.15 M NaCl solution at 310 K and 1 atm, equivalent to a human erythrocytic environment, with AMBER10 force field [7]. For a charge model of haem in haemoglobin, reported RESP charges and bonded parameters were adopted [8]. The simulation consists of 400ps-NVT ensemble as a pre-equilibrated phase, and NPT simulation as a production run. In NVT simulation, a temperature of 310 K was regulated using Langevin Dynamics, while a similar temperature and a pressure of 1 atm (1.013 bar) were controlled by a weak coupling (Berendsen) algorithm. All Van der Waals interactions and electrostatics forces were handled with a cutoff of 12 Angstroms, which the latter were computed via Particle Mesh Ewald (PME) calculation. All MD simulation was carried out with PMEMD module in AMBER12 package. The protein conformations were eventually analysed using PTRAJ module and some manually written programmes. A structure visualisation was performed using Visual Molecular Dynamics package [9].

4. RESULTS AND DISCUSSION

All mutant α-globins structures of all were successfully constructed through the bioinformatics tools. The predicted tertiary structure was corresponding to the results from the secondary structure predictions. Merging the predicted protein tail to the wild-type structure to create an initial coordinate for MD simulation, the energy-minimisation was performed, yielding the low root-mean-square displacement (RMSD), less than 0.5 Angstroms compared to an experiment (2DN2.pdb). For MD simulation, energetics parameters such as energy terms,

temperature, pressure and density become stable and converged to the set values. Furthermore all mutant protein gives a steady RMSD of less than 1.5 Angstroms, Figure 1A.

Besides, after structural analysis for each amino acid residue using the relative distance from a haem group to an interested residue, we found that the dynamic trajectories from all mutant proteins show no significant structural difference in residue 1-141 from wild-type dynamic structure, Figure 1B. This indicates a structural similarity amongst all globin proteins. The study suggests that the additional protein structure at the C-terminus does not affect the protein conformation significantly. In other words, our work excludes the hypothesis that the mutation can lead to an absurdity in conformation, in addition, this work demonstrate a visualisation of how α-globins with an additional part due to the causative mutation behaves under *in vivo* condition, with a reference of native protein. In short, the mutation of α-thal type II which provides C-terminus elongation in α-globins, seems to have no effect on a protein conformation in monomeric form. Instead this elongated component may play a role in further functional stages, such as dimeric or tetrameric formations between two globin types (α- and β-types). A further study needs to be accomplished to verify how causative mutation contributes to a disease occurrence.

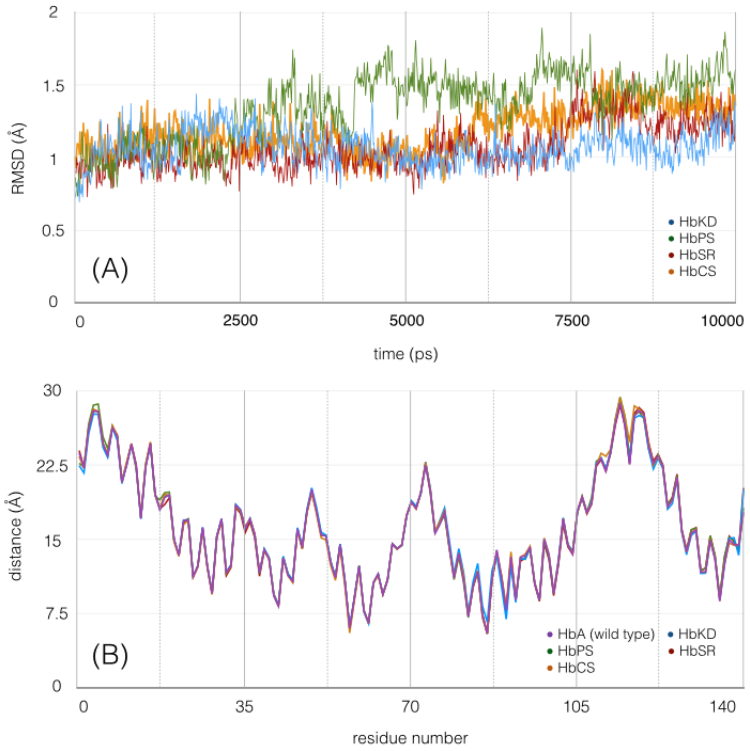

Figure 1. (A), RMSD of all MD simulations with respect to backbone atoms in the experimental structure PDB code 2DN2. (B), Distance pattern of residue 1 to 141 of all MD simulations. A similar pattern indicates the structural similarity between MD trajectories.

5. CONCLUSION

In our current study, we have applied bioinformatic apparatus to predict the Hb mutant composed of an additional elongated terminus related to α-thal. We have later carried out Md simulations in order to elucidate information at atomistic scale under in vivo circumstance. Our study provides molecular insight of conformational change with a presence of interested mutation. The results indicate no significant impact due to the mutation on Hb native core domain. An assumption associated with structural disturbance in alpha Hb monomeric structure is discarded, and the focus on disease origin has subsequently moved into other key steps in tetrameric formation instead.

REFERENCES

1. Steinberg M.H., Forget B.G., Higgs D.R., and Weatherall D.J., *Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management*, 2nd, Cambridge University Press, New York, 2009, 255-6.
2. Park S.Y., Yokoyama T., Shibayama N., Shiro Y., andTame J.R.H., *J Mol Bio,* 2006, 360 (3): 690-701.
3. Frishman D., and Argos P., *Proteins*, 1997, 27(3): 329-35.
4. Petersen B., Petersen T., Andersen P., Nielsen M., and Lundegaard C., *BMC Structural Biology,* 2009, 9(1): 51.
5. Blaszczyk M., Jamroz M., Kmiecik S., Kolinski A., *Nucleic Acids Res*, 2013, W406-11.
6. Unni S., Huang Y., Hanson R.M., Tobias M., Krishnan S., Li W.W., Nielsen J.E., Baker N.A., *J Comput Chem*, 2011, 32(7): 1488-91.
7. Chen A.A., and Pappu R.V., *J Phy Chem B*, 2007, 111(41): 11884-7.
8. Giammona D.A., Ph.D. thesis, University of California, Davis, 1984.
9. Humphrey W., Dalke A., and Schulten K., *J Mol Graph*, 1996, 14(1): 33-8.
10. Adcock S.A., and McCammon J.K., *Chem Rev,* 2006, 106(5): 1589−615.
11. Jenkins, J.D., Musayev F.N., Danso-Danquah R., Abraham, D. J., and Safo, M.K., (2009), *Acta Cryst,* D65, 41-8.
12. Hub J.S., Kubitzki M.B., de Groot B.L., (2010), *PLoS Comput Biol* 6(5): e1000774.
13. Takayanagi M., Kurisaki I., and Nagaoka M., (2013), *J Phys Chem B*, 117(20): 6082-91.
14. Tipmanee V., Oberhofer H., Park M., Kim K.S., and Blumberger J., *J Am Chem Soc*, (2010), 132(47): 17032-40.
15. Tipmanee V., Blumberger J., (2012), *J Phys Chem B*, 116(6): 1876-83.
16. Chanin N., Prachayasittikul V., and Bulow L., (2013), *PLoS One*, 8(4): e62996.

ACKNOWLEDGEMENTS

N.C.P. and V.T. would like to express a gratitude to Department of Biomedical Sciences, Faculty of Medicine, Prince of Songkla University for financial support and computing resources. V.T. also thanks Asst. Prof. Dr. Chamnong Nopparatana and Assoc. Prof. Paramee Thongsuksai, M.D. for a research initiative and productive guidance.

APPENDIX B

E-POSTER PRESENTATION

TRACKS TRENDS in HEALTHCARE
การประชุมวิชาการประจำปี คณะแพทยศาสตร์ ครั้งที่ 30

204

The homology structure Ppediction of alpha-globin Constant spring using bioinformatics tools

Nawanwat Chainuwong, Varomyalin Tipmanee
Department of Biomedical Sciences, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand.

Introduction: Haemoglobin H Constant Spring thalassaemia disease (HbH CS disease) is a hereditary haematologic disease that has a very high prevalence among alpha-type one found in Thai thalassaemia patients. The disease comes from the abnormal elongated alpha haemoglobin (alpha-Hb), containing 31 amino acid extension in its structure. The abnormal structure can affect the functional form of haemoglobin protein, a tetramer (alpha2beta2) consisting of 2 alpha- and 2 beta- subunits. However, the tetrameric formation mechanism remains unclear. In such a case 3-D structure of Hb is necessary in order to elucidate and understand the mechanism. Unfortunately a 3-D HbCS structure has not yet been reported, so in this work we attempt to construct 3-D structures HbCS by bioinformatics tools using deoxy alpha Hb as a structural template.

Objective: To construct and predict homology structure of HbCS by bioinformatics tools

Materials and Methods: The human deoxyHbA crystal structure (PDB code 2DN2) will take from the Protein Data Bank. A secondary structure of CS tail in HbCS has been predicted by NetSurfP (1.1) and PREDATOR (2.1.2). For alpha-globin 3-D monomeric structure, both alpha amino acid residues from HbA and HbCS amino acid residues will be predicted for the homology structure of HbCS using PHYRE2. All structures HbA will be determined at pH 7.4 using the PROPKA bioinformatics tool. The rigid-body protein-protein docking programme ZDOCK (3.0.2) is used for the elucidation of dimeric and tetrameric structures.

Results and conclusion: 3-D CS tail structures were predicted using various tools. The finalised chosen structure was justified using the 2-D structure prediction data. Later the CS tail was docked to the C-terminus of human alpha-HbA. The obtained structure was then energy minimised along with hydrogen addition. The protonation state of the alpha-CS-globin protein was considered at pH 7.4 using the ProPkA web-based tool. Finally this energy minimised structure will be used as the starting structure for investigating the interaction and simulated “in solution” structure by molecular dynamics simulation method.

Keywords: haemoglobin constant spring, alpha-globin, homology structure prediction, bioinformatics

APPENDIX C

AMBER INPUT FILES

C.1 Input files for MD simulation of α^{WT}-globin

The residue 1 to 141 is represented α^{WT}-globin, and the residue 142 is represented haem molecule.

C.1.1 Energy minimisation *in vacuo*

```
#energy_minimisation.in
&cntrl
imin = 1,
maxcyc = 3000,
ncyc = 500,
cut = 12.0,
igb = 0,
ntb = 0,
ntb = 500,
/
```

C.1.2 Heating and equilibration (k=250 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_250.in
&cntrl
imin = 1,
irest = 0, ntx = 1,
ntb = 1, cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
250.0
RES 1 142
END
END
```

C.1.3 Heating and equilibration (k=150 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_150.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
150.0
RES 1 142
END
END
```

C.1.4 Heating and equilibration (k=100 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_100.in
&cntrl
imin = 1,
```

irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
100.0
RES 1 142
END
END

C.1.5 Heating and equilibration (*k*=50 kcal·mol^{-1}·Å^{-2})

#heat_nvt_50.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
50.0
RES 1 142
END
END

C.1.6 Heating and equilibration (*k*=20 kcal·mol^{-1}·Å^{-2})

#heat_nvt_20.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
20.0
RES 1 142
END
END

C.1.7 Heating and equilibration (*k*=10 kcal·mol^{-1}·Å^{-2})

#heat_nvt_10.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein

10.0
RES 1 142
END
END

C.1.8 Production for 80 ns

#production_npt.in
&cntrl
imin = 0,
irest = 1, ntx = 5,
ntb = 2,
cut = 12.0,
ntr = 0, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 1,
tautp = 1.0,
nstlim = 4000000, dt = 0.002,
ntpr = 1000,
ntwx = 5000, ntwr = 5000,
nscm = 1000,
iwrap = 1,
/

C.2 Input files for MD simulation of α^{CS}-globin

The residue 1 to 172 is represented α^{CS}-globin, and the residue 173 is represented haem molecule.

C.2.1 Energy minimisation *in vacuo*

#energy_minimisation.in
&cntrl
imin = 1,
maxcyc = 3000,
ncyc = 500,
cut = 12.0,
igb = 0,
ntb = 0,
ntb = 500,
/

C.2.2 Heating and equilibration (k=250 kcal·mol^{-1}·Å^{-2})

#heat_nvt_250.in
&cntrl
imin = 1,
irest = 0, ntx = 1,
ntb = 1, cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
250.0
RES 1 173
END
END

C.2.3 Heating and equilibration (k=150 kcal·mol^{-1}·Å^{-2})

#heat_nvt_150.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,

ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
150.0
RES 1 173
END
END

C.2.4 Heating and equilibration (*k*=100 kcal·mol^{-1}·Å^{-2})

#heat_nvt_100.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
100.0
RES 1 173
END
END

C.2.5 Heating and equilibration (*k*=50 kcal·mol^{-1}·Å^{-2})

#heat_nvt_50.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
50.0
RES 1 173
END
END

C.2.6 Heating and equilibration (*k*=20 kcal·mol^{-1}·Å^{-2})

#heat_nvt_20.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
20.0
RES 1 173
END
END

C.2.7 Heating and equilibration (*k*=10 kcal·mol^{-1}·Å^{-2})

#heat_nvt_10.in

```
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
10.0
RES 1 173
END
END
```

C.2.8 Production for 80 ns

```
#production_npt.in
&cntrl
imin = 0,
irest = 1, ntx = 5,
ntb = 2,
cut = 12.0,
ntr = 0, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 1,
tautp = 1.0,
nstlim = 4000000, dt = 0.002,
ntpr = 1000,
ntwx = 5000, ntwr = 5000,
nscm = 1000,
iwrap = 1,
/
```

C.3 Input files for MD simulation of AHSP

The residue 1 to 91 is represented AHSP.

C.3.1 Energy minimisation

in vacuo

```
#energy_minimisation.in
&cntrl
imin = 1,
maxcyc = 3000,
ncyc = 500,
cut = 12.0,
igb = 0,
ntb = 0,
ntb = 500,
/
```

C.3.2 Heating and equilibration

(k=250 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_250.in
&cntrl
imin = 1,
irest = 0, ntx = 1,
ntb = 1, cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2, tempi = 10.0,
temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
250.0
RES 1 91
END
```

```
END
```

C.3.3 Heating and equilibration (k=150 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_150.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
150.0
RES 1 91
END
END
```

C.3.4 Heating and equilibration (k=100 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_100.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
100.0
RES 1 91
END
END
```

C.3.5 Heating and equilibration (k=50 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_50.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
50.0
RES 1 91
END
END
```

C.3.6 Heating and equilibration (k=20 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_20.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
```

nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
20.0
RES 1 91
END
END

C.3.7 Heating and equilibration (*k*=10 kcal·mol^{-1}·Å^{-2})

#heat_nvt_10.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
10.0
RES 1 91
END
END

C.3.8 Production for 50 ns

#production_npt.in
&cntrl
imin = 0,
irest = 1, ntx = 5,
ntb = 2,
cut = 12.0,
ntr = 0, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 1,
tautp = 1.0,
nstlim = 4000000, dt = 0.002,
ntpr = 1000,
ntwx = 5000, ntwr = 5000,
nscm = 1000,
iwrap = 1,
/

C.4 Input files for MD simulation of α^{WT}·AHSP

The residue 1 to 141 is represented α^{WT}-globin, the residue 142 to 232 is represented AHSP, the residue 233 and haem molecule.

C.4.1 Energy minimisation *in vacuo*

#energy_minimisation.in
&cntrl
imin = 1,
maxcyc = 3000,
ncyc = 500,
cut = 12.0,
igb = 0,
ntb = 0,
ntb = 500,
/

C.4.2 Heating and equilibration (*k*=250 kcal·mol^{-1}·Å^{-2})

#heat_nvt_250.in
&cntrl

```
imin = 1,
irest = 0, ntx = 1,
ntb = 1, cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
250.0
RES 1 233
END
END
```

C.4.3 Heating and equilibration (k=150 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_150.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
150.0
RES 1 233
END
END
```

C.4.4 Heating and equilibration (k=100 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_100.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
100.0
RES 1 233
END
END
```

C.4.5 Heating and equilibration (k=50 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_50.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
```

fix protein
50.0
RES 1 233
END
END

C.4.6 Heating and equilibration (k=20 kcal·mol^{-1}·Å^{-2})

#heat_nvt_20.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2, tempi = 10.0,
temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
20.0
RES 1 233
END
END

C.4.7 Heating and equilibration (k=10 kcal·mol^{-1}·Å^{-2})

#heat_nvt_10.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
10.0
RES 1 233
END
END

C.4.8 Production for 80 ns

#production_npt.in
&cntrl
imin = 0,
irest = 1, ntx = 5,
ntb = 2,
cut = 12.0,
ntr = 0, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 1,
tautp = 1.0,
nstlim = 4000000, dt = 0.002,
ntpr = 1000,
ntwx = 5000, ntwr = 5000,
nscm = 1000,
iwrap = 1,
/

C.5 Input files for MD simulation of α^{CS}·AHSP

The residue 1 to 172 is represented α^{CS}-globin, the residue 173 to 263 is represented AHSP, the residue 264 and haem molecule.

C.5.1 Energy minimisation *in vacuo*

```
#energy_minimisation.in
&cntrl
imin = 1,
maxcyc = 3000,
ncyc = 500,
cut = 12.0,
igb = 0,
ntb = 0,
ntb = 500,
/
```

C.5.2 Heating and equilibration (k=250 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_250.in
&cntrl
imin = 1,
irest = 0, ntx = 1,
ntb = 1, cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
250.0
RES 1 263
END
END
```

C.5.3 Heating and equilibration (k=150 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_150.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
150.0
RES 1 263
END
END
```

C.5.4 Heating and equilibration (k=100 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_100.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
100.0
RES 1 263
END
END
```

C.5.5 Heating and equilibration (k=50 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_50.in
```

```
&cntrl
imin = 1,
irest = 1, ntx = 5, ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
50.0
RES 1 263
END
END
```

C.5.6 Heating and equilibration (k=20 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_20.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
20.0
RES 1 263
END
END
```

C.5.7 Heating and equilibration (k=10 kcal·mol^{-1}·Å^{-2})

```
#heat_nvt_10.in
&cntrl
imin = 1,
irest = 1, ntx = 5,
ntb = 1,
cut = 12.0,
ntr = 1, ntc = 2,
ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 3, gamma_ln = 1.0,
tautp = 0.1,
nstlim = 100000, dt = 0.001,
ntpr = 1000,
ntwx = 1000, ntwr = 1000,
/
fix protein
10.0
RES 1 263
END
END
```

C.5.8 Production for 80 ns

```
#production_npt.in
&cntrl
imin = 0,
irest = 1, ntx = 5,
ntb = 2,
cut = 12.0,
ntr = 0, ntc = 2, ntf = 2,
tempi = 10.0, temp0 = 310.0,
ntt = 1,
tautp = 1.0,
nstlim = 4000000, dt = 0.002,
ntpr = 1000,
ntwx = 5000, ntwr = 5000,
nscm = 1000,
iwrap = 1,
/
```

C.6 Input files for calculating free binding energies

C.6.1 MMBGSA_entopy

```
#energy_MMGB(PB)SA.in
&general
verbose=1,
interval=1,
/
#MMGBSA
&gb
igb=2,
saltcon=0.15,
/
#MMPBSA
&pb
inp=1,
saltcon=0.15,
/
#delta_S_entropy
&nmode
nmode_igb=1,
nmode_istrng=0.15,
/
```

VITAE

Name Mr Nawanwat Chainuwong Pattarnggoon

Student ID 5610320012

Educational Attainment

Degree	Name of Institution	Year of Graduation
B.Sc. (Medical Technology)	Walailak University	2013

Scholarship Awards during Enrollment

1. Graduate School Research Support Funding for Thesis from Graduate School, Prince of Songkla University

List of Publications and Proceedings

1. **Chainuwong N**, Yimtiang T. The Study of Prevalence and Haematological Parameters of G6PD Deficiency Patient: Case Report from Trang Hospital. Journal of the Medical Technologist Association of Thailand. 2014 Aug 1;42(2).
2. **Chainuwong N** and Tipmanee V. The homology structure Prediction of alpha-globin Constant spring using bioinformatics tools. 30th Tracks and Trends in Healthcare; 2014. Aug 6-8; Annual Medical conference, Faculty of Medicine, Prince of Songkla University, Thailand. (E-poster presentation)
3. **Pattaranggoon NC** and Tipmanee V. Computational Study of the Effect due to Chain Terminal Mutation in Human Alpha haemoglobins. The 19th International Annual Symposium on Computational Science and Engineering, At Ubon Ratchathani University, Ubon Ratchathani, Thailand. (Oral presentation)