



## รายงานฉบับสมบูรณ์

โครงการ “การออกแบบเชิงโมเลกุลและการทำนายโครงสร้างแอนติบอดีและแอนติเจน”

โดย ดร. วรณจันทร์ แสงหิรัญ ลี

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

พฤษภาคม 2554

สัญญาเลขที่ RMU5180021

รายงานฉบับสมบูรณ์

โครงการ “การออกแบบเชิงโมเลกุลและการทำนายโครงสร้างแอนติบอดีและแอนติเจน”

ดร. วรณจันท์ แสงหิรัญ ลิ  
ภาควิชาเคมี คณะวิทยาศาสตร์  
มหาวิทยาลัยเชียงใหม่

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว.ไม่จำเป็นต้องเห็นด้วยเสมอไป)

## กิตติกรรมประกาศ

งานวิจัยนี้สำเร็จลงได้ด้วยทุนสนับสนุนจากสำนักกองทุนสนับสนุนการวิจัย และสำนักงานคณะกรรมการการอุดมศึกษา นอกจากนี้ทีมงานวิจัยใคร่ขอขอบคุณ รศ.ดร. ชัชชัย ตะยาภิวัฒนา และทีมวิจัย จากภาควิชาภูมิคุ้มกันวิทยาคลินิก คณะเทคนิคการแพทย์ มหาวิทยาลัยเชียงใหม่ โดยมีความความร่วมมือในส่วนของโครงการโคลนนิ่ง และวิเคราะห์ลำดับกรดอะมิโนแอนติบอดีสายเดี่ยว เพื่อใช้ยืนยันผลการคำนวณ ขอขอบคุณห้องปฏิบัติการออกแบบโมเดล และการจำลองโดยวิธีทางคอมพิวเตอร์ (Computational Simulation and Modeling Laboratory: CSML) ภาควิชาเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่ให้ความอนุเคราะห์ สถานที่อุปกรณ์ในการทำวิจัย ขอขอบคุณศูนย์นาโนเทคโนโลยีแห่งชาติที่เอื้อเฟื้อซอฟต์แวร์ในการคำนวณ ตลอดจนผู้ที่มีส่วนร่วมในการดำเนินการ การประเมิน และให้ความคิดเห็นในการพัฒนา งานวิจัยนี้ทุกท่าน

วรรณจันทร์ แสงหิรัญ ลี

พฤษภาคม 2554

**Abstract (บทคัดย่อ)**

**Project Code :** RMU5180021

**Project Title :** Molecular Design and Prediction of Antibody-Antigen Structure

**(ชื่อโครงการ):** การออกแบบเชิงโมเลกุลและการทำนายโครงสร้างแอนติบอดีและแอนติเจน

**Investigator :** Dr. Vannajan Sanghiran Lee

Department of Chemistry, Faculty of Science, Chiang Mai University

**(ชื่อนักวิจัย):** ดร. วรณจันทร์ แสงหิรัญ ลี

ภาควิชาเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

**E-mail Address :** vannajan@gmail.com

**Project Period :** May 16, 2008 – May 15, 2011

**(ระยะเวลาโครงการ)** 16 พฤษภาคม 2551 - 15 พฤษภาคม 2554

## Abstract

Computational assisted modeling was carried out for several systems such as scFv anti-p17, scFv G46 (scFv binds with HIV epitope p17 which is still attached to p24), and antibody-CD147. The models were built by homology modeling and evaluate with Ramachandran plot and PROCHECK. Molecular docking has been performed to obtain the complex structure of antibody and antigen. In this study, scFv against HIV-1 epitope at the C-terminal on p17 (scFv anti-p17) was mainly used as a candidate molecule for evaluating the method. The wild-type p17 and its nine natural mutants were docked with scFv anti-p17. Potential mean force (PMF) scores predicted the most favorable binding interaction, and the correlation agreed well with the corresponding activity data from the peptide based on competitive ELISA. In the interaction with solvent molecules, the 3D structures of scFv anti-p17 and selected peptide epitopes were further investigated by molecular dynamics (MDs) simulation with the AMBER 9 program. Post-processing of the snapshot at equilibrium was performed to evaluate the binding free energy and pairwise decomposition or residue-based energy calculation of complexes in solution using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) protocol. Our results demonstrated that the specific residues located in the complementary determining regions (CDRs) of scFv anti-p17, MET100, LYS101, ASN169, HIS228, and LEU229, play a crucial role in the effective binding interaction with the absolute relative decomposed energy more than 2.00 kcal/mol in comparison to the original substrate. A computational alanine scanning mutagenesis was submitted to quest for the residues in the energetic contribution of the mutation sites in scFv anti-p17 binding free energy with p17 c-terminal peptide alternatives. The result demonstrated several influenced amino acid residues in the scFv binding pocket i.e. ASP31, ASN35, TRP50, MET100, LYS101, SER103, TYR104, SER162, ASN169, LEU185, LYS188, ASP190, TRP224, GLN225, and LEU229. A candidate point mutation, M100R and M100G, were found to improve remarkably improved the binding activity scFv anti-p17 of by the indirect ELISA. However, some concerns on the correlation between the theoretical and experimental results of competitive and indirect ELISA have been raised in several cases.

Besides, intracellularly expressed antibody fragments (intrabodies) have been used as powerful tools for clinical applications and for the functional analysis of proteins inside the cell. Among several types of intrabodies, single chain fragment variables (scFv) composed of only the variable regions (VH or VL) of antibodies are the smallest and thus the easiest to design. However, normal antibody fragments do not form disulfide bonds in the cytoplasm and usually are unable to achieve a stable native fold in the absence of the disulfide bonds. Recently, the crystal structure of anti-RAS VH and VL fragments without disulfide bonds after substitution of cysteine residues and the wild type VH and VL with intact disulfide bonds showed no structural differences between the two types of the VH and VL. Another system of interest is the soluble scFv in cytoplasm. There is great interest in engineering antibody fragments that will fold and are stable under reducing conditions, and that could serve as framework to which other specificities could be grafted. We have undertaken molecular dynamics simulations to investigate the two types of such systems with and without disulfide bonds. The structural analysis in term of distance geometry analysis, hydrogen bond analysis, residue interactions, the disulfide strain energy and binding affinity between the domain, were observed in order to explain the stability of the antibody scFv fragments. The experimental study is still ongoing for comparison with the calculation results.

**Keywords :** Molecular design, Antibody-antigen, HIV-1, Single chain Fv, Docking,

Molecular dynamics

### บทคัดย่อ

การโมเดลโดยใช้คอมพิวเตอร์ในหลายระบบเช่น scFv anti-p17, scFv G46 ที่ยึดกับ HIV epitope บน p17 ที่ไม่จำเป็นต้องถูก cleave ออกจาก p24 และ แอนติบอดี-CD147 ด้วยการทำให้โมเลกุลโมเดลลิงและประเมินโครงสร้างจากกราฟามาซานดราและโปรเซค ใช้ด็อกกิงเชิงโมเลกุลสร้างโครงสร้างเชิงซ้อนระหว่างแอนติบอดีและแอนติเจน ในงานวิจัยนี้ศึกษา scFv ต่อ HIV-1 อีพิโทปบริเวณปลาย C-terminal ที่ p17 (scFv anti-p17) โดยด็อกเปปไทด์ wild-type p17 และมิวแทนท์อีก 8 ตัวที่พบในธรรมชาติกับ scFv anti-p17 ใช้ Potential mean force (PMF) ซึ่งเป็นพลังงานที่สอดคล้องกับข้อมูลการทดลอง competitive ELISA สำหรับอันตรกิริยาน้ำไขการจำลองพลศาสตร์ของโครงสร้างเชิงซ้อนในน้ำด้วยโปรแกรม AMBER 9 ใช้โครงสร้างที่สมดุลในการคำนวณพลังงานยึดจับอิสระ พลังงานย่อยที่ละคู่ หรือ พลังงานแต่ละหมู่อะมิโนโดยใช้ Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) พบว่าหมู่อะมิโนบริเวณ CDRs ที่สำคัญและมีค่าพลังงานยึดจับสมบูรณ์สัมพันธ์เทียบกับสารแอนติเจนดั้งเดิมมากกว่า 2.00 kcal/mol ได้แก่ MET100, LYS101, ASN169, HIS228, และ LEU229 เมื่อทำอะลานีนสแกนนิ่งโดยคอมพิวเตอร์พบว่าหมู่อะมิโนที่มีผลกระทบได้แก่ ASP31, ASN35, TRP50, MET100, LYS101, SER103, TYR104, SER162, ASN169, LEU185, LYS188, ASP190, TRP224, GLN225, และ LEU229 เมื่อเปลี่ยนแปลงหมู่อะมิโนของ scFv จาก M100R และ M100G พบว่าประสิทธิภาพการยึดจับของ scFv anti-p17 ดีขึ้นกับเปปไทด์ที่พบในธรรมชาติ สอดคล้องกับการทดลอง indirect ELISA อย่างไรก็ตามควรเปรียบเทียบผลการคำนวณ และใช้การทดลองทั้ง competitive และ indirect ELISA อย่างระมัดระวัง

นอกจากนี้การใช้อินทรานาอติในทางคลินิกในการนำโปรตีนเข้าสู่เซลล์ โดยการใช้ส่วนแอนติบอดีสายเดี่ยวที่ประกอบด้วย VH หรือ VL ซึ่งเป็นส่วนขนาดเล็กที่สุดในการออกแบบพบว่า มีปัญหาตรงที่พันธะไดซัลไฟด์ไม่เสถียร และถูกทำลายได้เมื่อเข้าไปอยู่ใน cytoplasm จากข้อมูลโครงสร้างผลึกของ scFv anti-RAS พบว่าเมื่อแทนที่หมู่ซิสเทอีน ด้วยหมู่อื่นปรากฏว่าโครงสร้างยังเสถียรอยู่ได้ อีกระบบที่น่าสนใจคือระบบ scFv ที่ละลายได้ในไซโตพลาสซึม ซึ่งจากโครงสร้างที่พบทำให้สามารถทำวิศวกรรมส่วนของแอนติบอดีที่สามารถม้วนและเสถียรในสภาวะรีดิวซิงได้โดยการกราฟติง โดยศึกษาการจำลองพลศาสตร์ของระบบ scFv ที่มีและไม่มีพันธะไดซัลไฟด์ วิเคราะห์โครงสร้าง พันธะไฮโดรเจน อันตรกิริยาระหว่างหมู่อะมิโน พลังงานความเครียดในหมู่ไดซัลไฟด์ เพื่ออธิบายความเสถียรของ scFv ซึ่งการศึกษาทางห้องปฏิบัติการยังอยู่ในการดำเนินการเพื่อเปรียบเทียบกับผลการคำนวณ

**คำหลัก:** การออกแบบเชิงโมเลกุล, แอนติบอดี-แอนติเจน, เอชไอวี-1, Fv สายเดี่ยว, ด็อกกิง, การจำลองพลศาสตร์

## สารบัญเรื่อง

กิตติกรรมประกาศ	ก
บทคัดย่อ	ข
สารบัญ	จ
1. บทนำ	1
2. วิธีดำเนินการวิจัย	9
3. ผลงานวิจัย และวิจารณ์ผล	30
4. บทสรุป	62
5. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.	63
ภาคผนวก	68
ก. Reprint Molecular Immunology 2009	
ข. Manuscript "Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning"	
ค. Manuscript "Stabilization of Anti-Ras scFv fragments with/without disulfide bond through theoretical modeling and molecular dynamics simulations investigation"	
ง. Manuscript "Modeling of CD147 binding sites against ScFv: a guide for epitope mapping"	

## 1. บทนำ

### 1. ความเป็นมาและความสำคัญของเรื่อง

การติดเชื้อไวรัสเอชไอวีเป็นปัญหาสำคัญทางสาธารณสุขอันดับต้นของประเทศ รัฐบาลต้องจ่ายค่ารักษาพยาบาลดูแลผู้ติดเชื้อและผู้ป่วยเอดส์ เฉพาะยาต้านไวรัสชั้นพื้นฐานและโรคติดเชื้อฉวยโอกาสต่างๆ 8.5 หมื่นบาทต่อคนต่อปี ถ้าต้องดูแลรักษาตลอดชีวิตเฉลี่ยคนละ 2 ล้านบาท หากเชื้อดื้อยาก็ต้องจ่ายเพิ่มเป็น 4 ล้านบาท ในปี 2549 กระทรวงสาธารณสุขตั้งเป้าว่าจะลดผู้ติดเชื้อรายใหม่ไม่เกิน 1.6 หมื่นคน และในปี 2551 จะต้องมิผู้ติดเชื้อรายใหม่ไม่เกิน 7,500 คน การพัฒนาวิธีการรักษารูปแบบใหม่ที่มีประสิทธิภาพจะสามารถลดการสูญเสียทางเศรษฐกิจ และเพิ่มคุณภาพชีวิตของผู้ติดเชื้อได้ แนวทางหนึ่งในการรักษาได้แก่การให้วัคซีน วัคซีนเอดส์ทดลองที่ได้รับการพัฒนาขึ้นมาในขณะนี้ แบ่งเป็น 2 ชนิดคือ วัคซีนป้องกันและวัคซีนรักษา วัคซีนป้องกันนั้นเพื่อป้องกันคนที่ยังไม่ได้ติดเชื้อเพื่อไม่ให้ติดเชื้อ ในขณะที่วัคซีนรักษานั้นเพื่อกระตุ้นภูมิให้คนที่ติดเชื้อแล้วสามารถควบคุมเชื้อให้อยู่ในระดับต่ำและมีการน้อยที่สุดเท่าที่จะน้อยได้ โดยทำร่วมไปกับการให้ยาต้านเชื้อไวรัส โดยวัคซีนที่มีการศึกษาส่วนมากผลิตจากชิ้นส่วนโปรตีนของเชื้อเอชไอวี เรียกว่า subunit protein vaccine เป็นการตัดเอาบางส่วนจากโปรตีนจากเชื้อที่สามารถกระตุ้นภูมิคุ้มกันได้ โดยชิ้นโปรตีนดังกล่าวไม่เพียงพอที่จะทำให้เกิดการติดเชื้อได้ในคน ตัวอย่างโปรตีนดังกล่าวคือ จีพี 120 (gp 120) และจีพี 160 (gp160) ซึ่งเป็นโปรตีนที่ตัดมาจากส่วนหนามที่มีอยู่รอบตัวเชื้อ เมื่อนำวัคซีนมาทดสอบกับมนุษย์ พบว่าอุปสรรคสำคัญในการพัฒนาวัคซีนเอดส์คือ เปลือกที่ห่อหุ้มไวรัสนั้นมีวิวัฒนาการในการหลบหลีกแอนติบอดีที่มีอยู่ในร่างกายมนุษย์ ถึงแม้ว่าวัคซีนสามารถเป็นตัวกระตุ้นแอนติบอดีได้ก็ตาม แต่ไวรัสยังสามารถหลบหลีก T cell ที่ต่อต้านมันได้ เพราะไวรัสมีการกลายพันธุ์ และเปลี่ยนแปลงอย่างต่อเนื่อง อีกแนวทางหนึ่งที่ทางผู้วิจัยในความสนใจคือการพัฒนาแอนติบอดีบริเวณ cleavage sites ต่างๆ ของเอชไอวี-1 โปรตีนเอส เพื่อผลิตเป็นวัคซีน หรือใช้แอนติบอดีในการตรวจวิเคราะห์เนื่องจากบริเวณ cleavage site ดังกล่าวมีการเกิด mutation ค่อนข้างน้อยเมื่อเทียบกับบริเวณอื่น ทำให้มีโอกาสที่จะพัฒนาเทคโนโลยีนี้เพื่อประยุกต์ใช้ในการรักษา อีกทั้งการทำงานของ scFv เกิดขึ้นภายในเซลล์ดังนั้นจึงไม่กระตุ้นให้ภูมิคุ้มกันของมนุษย์ตอบสนองต่อ scFv

ปัจจุบันความรู้เกี่ยวกับโครงสร้างของโปรตีนมีบทบาทสำคัญในทั้งวงการวิจัย และอุตสาหกรรม ความก้าวหน้าทางด้านชีววิทยาระดับโมเลกุล ทำให้สามารถพบลำดับของยีนส์ใหม่ๆ ได้หลายพันยีนส์จากสิ่งมีชีวิตที่แตกต่างกัน [1] ความเข้าใจโครงสร้างของโปรตีนจะทำให้เข้าใจหน้าที่ของโปรตีนเหล่านี้ โดยอาศัยการวิเคราะห์โครงสร้างของโปรตีน [2-4] จากการวิเคราะห์ทางรังสีด้วยเทคนิค NMR (NMR spectroscopy) และการศึกษา X-ray crystallography ของโปรตีน [4, 5] อย่างไรก็ตามโปรตีนบางชนิดก็ไม่สามารถวิเคราะห์ได้จากการทดลองดังกล่าว จึงต้องอาศัยการทำนายโครงสร้างของโปรตีนที่แม่นยำ เช่น การโมเดลแบบเปรียบเทียบ (comparative modeling หรือ homology modeling) [6-11] ซึ่งความแม่นยำของเทคนิคนี้ขึ้นอยู่กับความเหมือนและคล้ายคลึงระหว่างโปรตีนเป้าหมาย และโปรตีนที่ทราบโครงสร้างที่ใช้ในการสร้างโมเดล [10] หากมีลำดับที่เหมือนกันน้อย ความแม่นยำ ถูกต้องของโมเดลที่ได้ก็จะต่ำลง นอกจากนี้การโมเดลในส่วนของบริษัทที่มีความยืดหยุ่นสูง เช่น บริเวณ variable loops ที่พบในแอนติบอดี ก็ค่อนข้างที่จะทำนายได้ถูกต้อง โครงสร้างแอนติบอดีแม้ว่าจะมีส่วนที่เป็น conserved structural framework หากแต่มีบริเวณที่แตกต่างกันอย่างมากคือในบริเวณ surface loop ที่เรียกว่า complementarity - determining regions (CDRs) มีถึง 6 hypervariable loops ซึ่งจะจับกับบอกลึงตำแหน่งการยึดจับ (binding site) ความแตกต่างกันในโครงสร้างของ loops ดังกล่าวมีผลต่อความสามารถของแอนติบอดีในการจดจำโมเลกุลต่างๆ [12] แม้ว่า



การจัดเรียง (alignment) กับส่วนของ framework จากการม้วนของอิมมูโนโกลบิน (immunoglobulin) จะค่อนข้างถูกต้อง แต่การโมเดลบริเวณยี่ดจับที่ค่อนข้างยืดหยุ่นดังกล่าวยังเป็นส่วนที่ต้องพัฒนาต่อไป นอกจากนี้ปัญหาโครงสร้างบริเวณยี่ดจับยังซับซ้อนขึ้นอีก หากโครงสร้างบริเวณยี่ดจับเกิดจากการรวมของสองโดเมน (domains; heavy chain ( $V_H$ ) และ light chain ( $V_L$ ) ซึ่งแต่ละแอนติบอดีจะมีบริเวณนี้แตกต่างกันไป [12]

ความสำคัญของการศึกษาแอนติบอดีมีบทบาทมากในทางการแพทย์ เช่น ในการรักษามะเร็ง, การทำ biosensors และ catalysis ซึ่งความก้าวหน้าในการประยุกต์ดังกล่าว เป็นผลมาจาก library-based in vitro production of monoclonal antibodies การทำนายโครงสร้างด้วยคอมพิวเตอร์ โครงสร้างสามมิติของแอนติบอดีนอกเหนือจากโครงสร้างที่พบในฐานข้อมูลโปรตีน และ โครงสร้างเชิงซ้อนแอนติบอดี – แอนติเจน จะเป็นประโยชน์ในการศึกษา และการประยุกต์ได้หลายแนวทาง เช่น การทำวิศวกรรมบริเวณยี่ดจับของแอนติบอดีใหม่เพื่อให้ทำหน้าที่อื่นที่สำคัญ การออกแบบแอนติบอดีใหม่ที่ยี่ดจับได้ดีขึ้นสำหรับแอนติเจนเป้าหมาย การบ่งชี้ (identification) แอนติเจนที่สามารถจับกับแอนติบอดีได้ และ การสร้างความเข้าใจอันตรกิริยาระหว่างแอนติบอดี-แอนติเจน และกลไกของแอนติบอดีในการจดจำ (recognition) โมเลกุลเป้าหมาย ปัจจุบันส่วนของแอนติบอดี ที่เรียกว่า single chain Fv (scFv) สามารถประยุกต์ใช้ใน gene therapy โดยการใช้เทคนิค intracellular antibodies (intrabodies) strategy โดยที่ scFv สามารถไปยังเป้าหมายที่จำเพาะทั้ง cellular cytoplasm, endoplasmic reticulum (ER), nucleus และ mitochondria แต่ส่วนใหญ่ประสบความสำเร็จที่บริเวณ ER จากการศึกษาพบว่า intrachain disulfide bond ใน scFv ที่มีส่วนต่อการม้วนตัวและความเสถียร โดยปกติแล้วไม่เสถียรใน reducing environment ที่พบในไซโตพลาสซึม ทำให้ไม่สามารถทำหน้าที่ได้ ดังนั้นการพัฒนา scFv ที่สามารถใช้งานได้ใน ไซโตพลาสซึมจึงเป็นโจทย์วิจัยที่สำคัญ จากการศึกษาการทดลองมีผู้พบ scFv ที่ไม่มีไดซัลไฟด์ และยังคงทำงานได้ในไซโตพลาสซึม และได้ศึกษาการ graft CDRs ที่ใช้จับกับแอนติเจนไปบน framework regions ของ SS-bond free scFvs แต่ส่วนใหญ่แล้วมักไม่ประสบความสำเร็จ การจำลองโดยวิธีคอมพิวเตอร์ในการ grafting และการวิเคราะห์พลศาสตร์ที่มีผลกระทบต่อพันธะไดซัลไฟด์ น่าจะมีส่วนช่วยประเมินความเสถียรของโครงสร้างโมเดลที่สร้างขึ้นได้

## 2. ความรู้พื้นฐานเกี่ยวกับเชื้อเอชไอวี

โรคติดเชื้อไวรัสเอชไอวี Human Immunodeficiency Virus (HIV) และโรคเอดส์เป็นปัญหาสาธารณสุขที่สำคัญของประเทศไทย โดยเป็นสาเหตุสำคัญของการป่วยและการตายของประชากรอายุ 20 ถึง 50 ปี ซึ่งเป็นประชากรที่อยู่ในวัยทำงานและเป็นการกำลังสำคัญของประเทศ เชื้อไวรัสเอชไอวี ตรวจพบครั้งแรกในเลือดของผู้ป่วยชาวคองโกในปี 1959 ปัจจุบันพบว่า เชื้อไวรัสเอชไอวีที่สำคัญ มีเพียงสองชนิด คือ HIV type 1 (HIV-1) และ HIV type 2 (HIV-2) จากการศึกษาจีโนมของไวรัสเอชไอวีทั้งสองชนิด ย้อนหลังไปหลายสิบปี พบว่า HIV-1 เริ่มติดต่อสู่คนครั้งแรกเมื่อประมาณปี ค.ศ. 1930 ส่วน HIV-2 เริ่มติดต่อสู่คนครั้งแรกในช่วงทศวรรษ 1940-1950 เชื้อไวรัสเอชไอวีทั้งสองชนิด มีลักษณะทางพันธุกรรมที่ใกล้เคียงกับไวรัสก่อโรคในลิง ที่มีชื่อเรียกว่าไวรัสเอสไอวี (SIV) ย่อมาจาก simian immunodeficiency virus; simian หมายถึงสัตว์จำพวกลิง (=ape, monkey) เชื้อไวรัสเอชไอวีทั้งสองชนิด มีที่มาแตกต่างกัน โดยเชื้อไวรัส HIV-1 มาจากลิง chimpanzee ที่อาศัยอยู่ในแถบแอฟริกากลาง มีชื่อวิทยาศาสตร์ว่า Pan troglodytes ปรากฏหลักฐานชัดเจนว่าเชื้อไวรัส HIV-1 กลายพันธุ์มาจากเชื้อไวรัสเอสไอวีในลิง ที่เรียกว่า SIVcpz ส่วนไวรัส HIV-2 มาจากลิงคนละชนิดกัน โดยเป็นลิงท้องถิ่นที่เรียกว่า sooty mangabey ซึ่งอาศัยอยู่ตามชายฝั่งแอฟริกาตะวันตกจาก Senegal ถึง Ivory Coast ต่อมาพบว่าเชื้อ HIV-2 กลายพันธุ์มาจากเชื้อไวรัส SIVsm สำหรับลิง sooty

mangabey มีชื่อวิทยาศาสตร์ว่า *Cercocebus atys* ทั้งนี้และทั้งนั้นพบว่าเชื้อไวรัส HIV-1 และ HIV-2 มีชิ้นส่วนดีเอ็นเอเหมือนกันร้อยละ 40-60

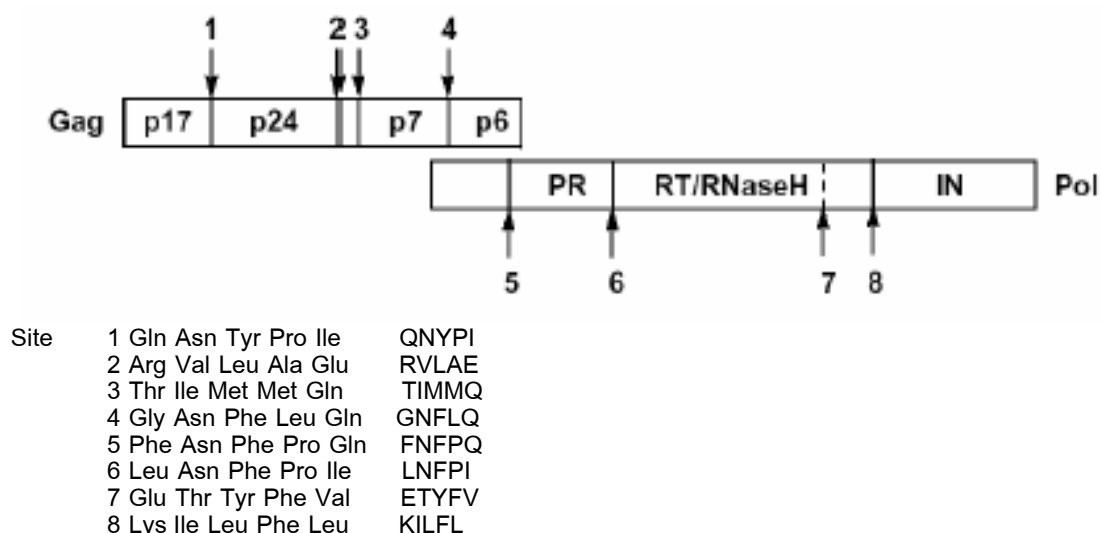
ลักษณะทางพันธุกรรมของเชื้อไวรัสเอชไอวี เชื้อ HIV-1 เป็นอาร์เอ็นเอไวรัส ชนิดสายเดี่ยว จีโนมมีความยาว 9 กิโลเบส ประกอบไปด้วยยีน 9 ชนิดควบคุมการสร้างโปรตีน โปรตีนที่สำคัญ ของเชื้อไวรัสเอชไอวี ได้แก่ โปรตีนโครงสร้าง หรือที่เรียกว่า structural proteins ได้แก่ Gag, Pol, และ Env โปรตีนควบคุมการทำงานของไวรัส เรียกว่า regulatory proteins ได้แก่ Tat, Rev โปรตีนที่ทำหน้าที่เสริม หรือ accessory proteins ได้แก่ Vpu, Vpr, Vif, และ Nef ชนิดหลักๆ ของเชื้อ HIV-1 ได้แก่ M (main), N (new), และ O (outlier) ในกลุ่ม M group ซึ่งพบได้มากถึงกว่าร้อยละ 90 ทั่วโลก แบ่งเป็น 9 ชนิดย่อย เรียกว่า clades กำหนดให้เรียกเป็นลำดับตัวอักษร A-D, F-H, J, และ K ในสหรัฐอเมริกาและทางยุโรปตะวันตก พบชนิด Clade B มากที่สุด แตกต่างจากที่พบในเอเชียและแอฟริกา ความหนาแน่นของไวรัสพบมากที่สุดในทวีปแอฟริกา

เชื้อเอชไอวีชนิด clade B เมื่อนำยาด้านไวรัสไปใช้ในบริเวณอื่นๆ การตอบสนองต่อการรักษาอาจแตกต่างกันไปบ้าง รวมทั้งกลไกการดื้อยาที่เกิดจากมิวเตชัน ความแตกต่างและความหลากหลายของสายพันธุ์เชื้อไวรัสเอชไอวี มีความสำคัญในการพัฒนาวัคซีนเป็นอย่างมากแอนติบอดีที่ร่างกายสร้างขึ้นเมื่อต่อต้านไวรัส และปฏิกิริยาจากเซลล์เม็ดเลือดขาวชนิด T-lymphocyte จะมีความจำเพาะเจาะจงสูง

เชื้อ Human Immunodeficiency Virus (HIV) เป็นเชื้อไวรัสที่ทำให้เกิดโรคเอดส์ เมื่อเชื้อเข้าสู่ร่างกายเชื้อจะแบ่งตัวอย่างมากและมีการเกิดโรคที่อวัยวะต่างๆ เช่นสมอง หัวใจ ไตและที่สำคัญคือจะทำลายระบบภูมิคุ้มกันของร่างกาย ระบบภูมิคุ้มกันนี้จะทำหน้าที่สร้างภูมิเพื่อต่อต้านการติดเชื้อและมะเร็งบางชนิด ในการสร้างภูมิจะต้องอาศัยเซลล์หลายชนิดที่สำคัญได้แก่เซลล์ CD4+ lymphocytes ซึ่งเป็นเซลล์ที่จับกับเชื้อ HIV เมื่อเซลล์ CD4+ lymphocytes ถูกทำลายโดยเชื้อมากจะทำให้ภูมิของร่างกายอ่อนแอ ดังนั้นปัญหาที่สำคัญของคนติดเชื้อ HIV คือปัญหาของโรคที่เกิดจากภูมิที่อ่อนแอลงเช่นโรคติดเชื้อฉวยโอกาส opportunistic infections เช่นโรคปอดบวมและโรคเยื่อหุ้มสมองอักเสบ และมะเร็งบางชนิด การติดเชื้อ HIV ไวรัสใช้ CD4 เป็นตัวรับเชื้อ envelope glycoprotein p120 ซึ่งมีขนาด 120 kd จะเชื่อมติดกับ CD4 molecule ที่บริเวณผิวเซลล์ของ host จุดที่เชื่อมติดในตอนแรกจะอยู่ในบริเวณ carboxy terminal ของ gp120 ซึ่งจะเชื่อมติดกับ residues 37-53 ณ amino-terminus ของ CD4 จะพยายามต่อต้านการเชื่อมติดและพยายามป้องกันการเข้าสู่เซลล์ อย่างไรก็ตามเมื่อมีการเชื่อมติดเกิดขึ้น envelope ของไวรัสจะสลายตัวรวมกับเยื่อหุ้มเซลล์และเข้าสู่ภายในเซลล์ CD4 ได้ ส่วนของ amino acid-terminus domain ซึ่งไม่ชอบน้ำ (hydrophobic) ที่อยู่บน envelope protein gp 41 มีความสำคัญอย่างยิ่งในการเชื่อม (fuse) envelope ให้เป็นเนื้อเดียวกับเยื่อหุ้มเซลล์ เมื่ออยู่ในสภาพนี้เท่ากับว่าไวรัสอยู่ในสภาพที่ไม่มีอะไรห่อหุ้ม สารพันธุกรรม (genome) ของไวรัสซึ่งอยู่ใน RNA ถูกถ่ายทอด (transcribe) ไปยัง deoxyribonucleic acid (DNA) โดยเอนไซม์ reverse transcriptase ซึ่งเป็น heterodimer ที่มีน้ำหนักของหน่วยย่อย (subunit) ขนาด 61 และ 52 kd หน่วยย่อยนี้เกิดขึ้นจากการแตกตัวของ polyproteins ซึ่งผลิตขึ้นมาจาก pol gene ของ HIV อย่างไรก็ตามเอนไซม์ตัวนี้ไม่ค่อยมีความแน่นอนทั้งในด้านปริมาณและการทำงาน อย่างไรก็ตามตราบเท่าที่ยังไม่มีความผิดพลาดการสังเคราะห์ DNA ก็จะดำเนินต่อไปอย่างไม่มีที่สิ้นสุด ความผิดพลาดดังกล่าวของเอนไซม์ reverse transcriptase เป็นสิ่งสำคัญทางด้านพันธุกรรมของเชื้อ HIV strains ต่างๆ ซึ่งสามารถแยกและเลี้ยงเชื้อได้ DNA ของไวรัสจะเข้าสู่กระบวนการขยายจำนวนใน DNA ของโครโมโซมของเซลล์ host โดยไม่เกี่ยวกับเอนไซม์ integrase หรือ viral endonuclease [13]

เชื้อเอชไอวีมีการกลายพันธุ์สูงเป็น ผลจากการเกิด recombination ระหว่างไวรัสต่างสายพันธุ์ ขณะนี้ยังไม่ทราบถึงภูมิคุ้มกันที่จำเพาะต่อการป้องกันการติดเชื้อหรือเกิดโรคเอดส์ แต่การศึกษาในกลุ่มผู้ที่สัมผัสเชื้อแต่ไม่ติดเชื้อ HIV-exposed uninfected จะตรวจพบ T cell proliferation responses เซลล์ CTL responses และแอนติบอดีเฉพาะที่ในเยื่อ จากการศึกษานิสัตว์ทดลอง SIV/macaque model ที่ได้รับวัคซีน SIV ชนิดเชื้อมีชีวิตอ่อนฤทธิ์ และภูมิคุ้มกันที่เกิดขึ้นสามารถป้องกันการติดเชื้อ SIV ที่ให้ทางเลือด และเยื่อ ภูมิคุ้มกันที่ตรวจพบ คือแอนติบอดีจำเพาะที่สามารถ neutralize เชื้อ SIV ชนิด primary isolates ได้ และพบเซลล์ CTL ปริมาณสูง เป็นหลักฐานแสดงว่าทั้ง neutralizing antibody ที่จำเพาะต่อ primary isolates และ CTL จำเป็นต่อการป้องกันการติดเชื้อไวรัสในลิง

นอกจากนี้งานวิจัยล่าสุดโดย Tewari และคณะ [14] ได้ใช้เทคนิคอินทราบอดีจากส่วนของแอนติบอดีสายเดี่ยว (scFv) ของ hybridoma-secreting anti-p17 antibody ในเซลล์ที่ติดเชื้อ HIV-1 พบว่ามีผลต่อการขยายพันธุ์ของไวรัส(viral replication process) โดยแอนติบอดีดังกล่าวจับเฉพาะที่ C-terminal epitope (DTGHSSQVSQNY) ของ p17 domain (รูปที่ 1.1) โดย scFv platform สามารถแสดงออกได้ทั้งในระบบ prokaryotic และ eukaryotic systems อีกทั้งสามารถวิศวกรรมเพื่อพัฒนาฟังก์ชันการยึดจับและความเสถียรให้ดีขึ้นได้ [15-17] การวิจัยประยุกต์ในด้านนี้เป็นไปอย่างแพร่หลายและต่อเนื่อง ในการใช้ scFv ในการพัฒนาชุดตรวจวิเคราะห์และ gene therapy [18-22]



รูปที่ 1.1 Gag p55 polyprotein

### 3. ความรู้พื้นฐานเกี่ยวกับแอนติบอดี และแอนติเจน

แอนติเจน คือสิ่งแปลกปลอมหรือสารที่ไม่มีอยู่ในร่างกายเมื่อเข้าสู่ร่างกายสามารถกระตุ้นให้เกิดการตอบสนองทางภูมิคุ้มกันแบบจำเพาะได้ (specific immune response) โดยทำให้เกิดการสร้างแอนติบอดี (antibody, Ab) หรือ T lymphocyte ที่จำเพาะ (specific sensitized T lymphocyte)

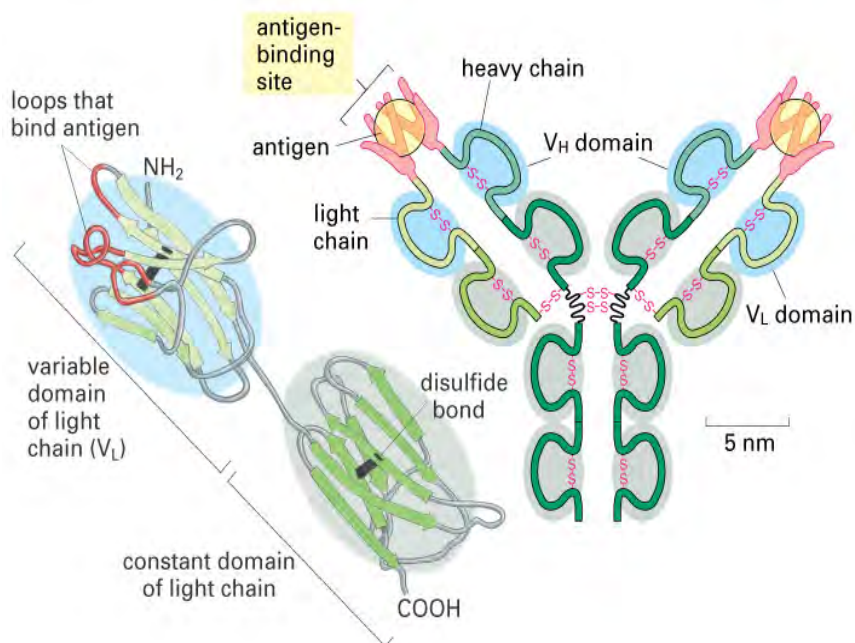
แอนติบอดี (Antibody, Ab) เป็นสารที่สร้างและหลั่งจาก plasma cell ซึ่งสามารถจับกับ antigenic determinant ของแอนติเจน (antigen, Ag) ที่เป็นตัวกระตุ้นได้อย่างจำเพาะ แอนติบอดีหรืออิมมูโนโกลบูลินพบได้ในซีรัมส่วนน้ำอื่นๆของร่างกายและเนื้อเยื่อ เช่น น้ำไขสันหลัง น้ำนม น้ำลาย น้ำตา ต่อมเหงื่อ ม้ามและบ่นผิวของ ลิมโฟไซท์บางชนิด อิมมูโนโกลบูลินเป็นสารพวก glycoprotein ประกอบด้วย polypeptide 82-96% และ

carbohydrate 4-18% เมื่อมีแอนติเจนเข้าสู่ร่างกาย B lymphocyte ที่มีความจำเพาะต่อแอนติเจนนั้น ๆ จะมีปฏิกิริยาตอบสนองโดยการแบ่งตัว (proliferation) และเปลี่ยนแปลง (differentiation) เกิดเป็นกลุ่ม (clone) ของเซลล์ที่ทำหน้าที่ผลิตแอนติบอดีที่จำเพาะต่อแอนติเจนนั้น

แอนติบอดีมีโครงสร้างพื้นฐานประกอบด้วย polypeptides 4 สาย (รูปที่ 1.2) คือสายที่ยาวและมีน้ำหนักโมเลกุลมาก เรียกว่า Heavy chain (H) ซึ่งมี 2 สายที่เหมือนกัน ส่วนอีก 2 สายที่สั้นและมีน้ำหนักโมเลกุลน้อยกว่า เรียกว่า Light chain (L) ซึ่งมีลักษณะเหมือนกันทั้ง 2 สายเช่นเดียวกัน Heavy chain และ Light chain ทั้ง 4 สายจะเชื่อมต่อกันด้วย interchain disulfide bonds ซึ่งสามารถแยกแรงยึดนี้ออกจากกันได้ด้วย mercaptoethanol ปลายข้างหนึ่งของแต่ละสายจะเป็น -N หรือ  $\text{-NH}_2$  หรือ amino terminal ส่วนปลายอีกข้างหนึ่งจะเป็น -C หรือ  $\text{-COOH}$  หรือ carboxy terminal ทั้ง 4 สายจะหันปลายข้าง  $\text{-NH}_2$  หรือ  $\text{-COOH}$  ไปทางเดียวกัน โมเลกุลของอิมมูโนโกลบูลินมีรูปร่างคล้าย ๆ กับตัว T หรือ Y โดยมี hinge region อยู่ตรงกลางของ H-chain ซึ่ง hinge region จะเป็นบริเวณที่มีการยืดหยุ่น (flexible) ได้มากทำให้แขนทั้ง 2 ข้าง ยึดห่างออกจากกันเพื่อจับกับแอนติเจนได้ดียิ่งขึ้น ส่วนของพันธะไดซัลไฟด์ ทั้ง intra และ inter-domain disulfide ที่มีส่วนทำให้โครงสร้างเสถียร แสดงได้ดังรูป

Variable region (V-region) คือบริเวณที่มีการดะมิโนเรียงตัวไม่คงที่ จะอยู่ทางด้าน N-terminal และในบริเวณ V-region นี้เองก็ยังมีบริเวณที่มีการดะมิโนผันแปรไปได้มาก ๆ เรียกว่า hypervariable region จากการวิเคราะห์การเรียงตัวของกรดอะมิโนพบว่า hypervariable segment สั้น ๆ อยู่ 3 ช่วง แต่ละช่วงจะมีการดะมิโน 3-11 ตัว เรียกส่วน hypervariable region นี้ว่า complementarily determining region (CDR) ซึ่งเป็นบริเวณสำคัญที่ใช้จับกับแอนติเจน โดยประกอบด้วยทั้ง variable heavy and light chain domains ( $V_H$ ,  $V_L$ ) บริเวณส่วนปลายสายเดี่ยว (single chain Fv fragment (scFv)) ที่เชื่อมด้วยเปปไทด์ที่ยืดหยุ่นสายสั้น (flexible peptide linker)

Constant region (C-region) คือบริเวณที่มีการดะมิโนเรียงตัวคงที่ ใน L-chain มี 1 segment คือ  $C_L$  ส่วน H-chain มี 3 หรือ 4 segments คือ  $C_{H1}$ ,  $C_{H2}$ ,  $C_{H3}$  (และ  $C_{H4}$ )



รูปที่ 1.2 โครงสร้างแอนติบอดี [23, 24]

#### 4. วัตถุประสงค์ของโครงการและผลที่คาดว่าจะได้รับ

1) สร้างโปรโตคอลการออกแบบโมเดลโครงสร้างแอนติบอดี แอนติเจน และโครงสร้างสารเชิงซ้อนแอนติบอดี – แอนติเจน

2) หาลำดับอะมิโนของแอนติเจนที่เหมาะสมที่มีส่วนสำคัญในโครงสร้างยึดจับกับแอนติบอดีจากการคำนวณ เปรียบเทียบกับการทำ epitope mapping และ ผลจากห้องปฏิบัติการ เพื่อประโยชน์ในการผลิตแอนติบอดี

3) วิศวกรรมลำดับอะมิโนของแอนติเจนที่มีความจำเพาะสูงกับแอนติบอดี

โดยศึกษาวิธีทาง in silico ในสองระบบ เพื่อออกแบบ intracellular antibody ที่เสถียรได้แก่

ระบบที่ 1 แอนติบอดีต่อเอชไอวี-1 โปรตีนเอสบีเวิล HIV-1 Gag proteins (p17, p24 and p17-p24 complex)

โดยโปรตีนของไวรัส Gag precursor เป็นโปรตีนที่ทำหน้าที่ได้ (functional proteins) ได้แก่ matrix (MA or p17), capsid (CA or p24), nucleocapsid (NC) and p6 domains พบว่า p17 เกี่ยวข้องกับวงจรชีวิตของไวรัสในหลายขั้นตอน โดยเฉพาะอย่างยิ่ง assembly and maturation steps

ระบบที่ 2 แอนติบอดีต่อ human leukocyte surface molecule CD147

ในการศึกษาครั้งนี้ผู้วิจัยมีเป้าหมายในการศึกษาการเข้าจับกันระหว่าง scFv กับ HIV-1 p17 ที่ตำแหน่ง c-terminal epitope โดยใช้เทคนิคเคมีคำนวณทางทฤษฎี เช่น การคำนวณควอนตัม กลศาสตร์เชิงโมเลกุล การ docking และการจำลองพลศาสตร์ ซึ่งจะช่วยให้ทราบรายละเอียดต่างๆ โดยเฉพาะอย่างยิ่งพลังงานในการยึดจับกัน ซึ่งแสดงถึงความสามารถของแอนติบอดีในการยึดจับกับ peptide epitope โดยได้ข้อมูลลำดับอะมิโนของ scfv ที่จำเพาะต่อ HIV-1 c-terminal epitope บน p17 จากค้นหาจากวารสาร และข้อมูลจากทางห้องปฏิบัติการโดยการโคลน immunoglobulin (Ig) gene variable fragments (single chain Fv, scFv) จาก hybridoma cells ภาควิชาภูมิคุ้มกันวิทยาคลินิก คณะเทคนิคการแพทย์ มหาวิทยาลัยเชียงใหม่ เพื่อยืนยันผลการคำนวณ การปรับปรุงแอนติบอดี และเป็นแนวทางการพัฒนาเพื่อผลิตเป็นวัคซีน หรือใช้แอนติบอดีในการตรวจวิเคราะห์หัดต่อไป โดย scFv จะทำหน้าที่ในการเข้าจับกับ HIV-1 ในบริเวณ c-terminal epitope บน p17 เพื่อขัดขวางการเข้าจับของ HIV-1 protease ในการย่อย substrate ส่งผลให้การสร้างไวรัสใหม่น้อยลง เนื่องจากบริเวณ cleavage sites ต่างๆของเอชไอวีมีการเกิด mutation น้อยมาก ทำให้โอกาสที่จะพัฒนาเทคโนโลยีนี้เพื่อประยุกต์ใช้ในการรักษา เนื่องจากการทำงานของ scFv เกิดขึ้นภายในเซลล์ดังนั้นจึงไม่กระตุ้นให้ภูมิคุ้มกันของมนุษย์ตอบสนองต่อ scFv

#### เอกสารอ้างอิง

1. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., Walters, L., et al. 1998 New goals for the U.S. human genome project: 1998-2003, *Science* 282, 682-689.
2. Burley, S. K., Almo, S. C., Bonanno, J. B., Capel, M., Chance, M. R., Gaasterland, T., et al. 1999 Structural genomics: beyond the human genome project, *Nat. Genet.* 23, 151-157.
3. Maggio, E. T., and Ramnarayan, K. 2001 Recent developments in computational proteomics, *Trends Biotechnol.* 19, 266-272.
4. Teichmann, S. A., Murzin, A. G., and Chothia, C. 2001 Determination of protein function, evolution and interactions by structural genomics, *Curr. Opin. Struct. Biol.* 11, 354-363.
5. Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., et al. 2000 Structural proteomics of an archaeon, *Nat. Struct. Biol.* 7, 903-909.

6. Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E., and Thornton, J. M. 1987 Knowledge-based prediction of protein structures and the design of novel molecules, *Nature* 326, 347-352.
7. Blundell, T., Carney, D., Gardner, S., Hayes, F., Howlin, B., Hubbard, T., et al. 1988 Knowledge-based protein modeling and design, *Eur. J. Biochem.* 172, 513-520.
8. Sali, A., Overington, J. P., Johnson, M. S., and Blundell, T. L. 1990 From comparisons of protein sequences and structures to protein modelling and design, *Trends Biochem. Sci.* 15, 235-240.
9. Sanchez, R., and Sali, A. 1997 Advances in comparative protein structure modelling, *Curr. Opin. Struct. Biol.* 7, 206-214.
10. Sanchez, R., Pieper, U., Melo, F., Eswar, N., Marty-Renom, M., Madhusudhan, M. S., et al. 2000 Protein structure modeling for structural genomics, *Nat. Struct. Biol.* 7 (Suppl.), 986-990.
11. Sanchez, R., and Sali, A. 1999 Comparative protein structure modeling in genomics, *J. Comput. Phys.* 151, 388-401.
12. Padlan, E. A. 1994 Anatomy of the antibody molecule, *Mol. Immunol.* 31, 169-217.
13. McMichael, A. J. 2006 HIV vaccines. *Annu. Rev. Immunol.* 24, 227-55.
14. Tewari, D., Goldstein, S.L., Notkins, A.L., Zhou, P. 1998. cDNA encoding a single-chain antibody to HIV p17 with cytoplasmic or nuclear retention signals inhibits HIV-1 replication. *J. Immunol.* 161, 2642–2647.
15. Quintero-Hernandez, V., Juarez-Gonzalez, V.R., Ortiz-Leon, M., Sanchez, R., Possani, L.D., Becerril, B. 2007. The change of the scFv into the Fab format improves the stability and in vivo toxin neutralization capacity of recombinant antibodies. *Mol. Immunol.* 44, 1307–1315.
16. Pavoni, E., Flego, M., Dupuis, M.L., Barca, S., Petronzelli, F., Anastasi, A.M., D'Alessio, V., Pelliccia, A., Vaccaro, P., Monteriu, G., Ascione, A., De Santis, R., Felici, F., Cianfriglia, M., Minenkova, O. 2006. Selection, affinity maturation, and characterization of a human scFv antibody against CEA protein. *BMC Cancer* 6, 41.
17. Park, S.G., Jung, Y.J., Lee, Y.Y., Yang, C.M., Kim, I.J., Chung, J.H., Kim, I.S., Lee, Y.J., Park, S.J., Lee, J.N., Seo, S.K., Park, Y.H., Choi, I.H. 2006. Improvement of neutralizing activity of human scFv antibodies against hepatitis B virus binding using CDR3 V(H) mutant library. *Viral. Immunol.* 19, 115–123.
18. Inui, H., Takehara, A., Doi, F., Nishi, K., Takai, M., Miyake, S., Ohkawa, H. 2009. A scFv antibody-based immunoaffinity chromatography column for clean-up of bisphenol A-contaminated water samples. *J. Agric. Food. Chem.* 57, 353–358.
19. Shen, Z., Yan, H., Zhang, Y., Mernaugh, R.L., Zeng, X. 2008. Engineering peptide linkers for scFv immunosensors. *Anal. Chem.* 80, 1910–1917.
20. Wang, G.P., Qi, Z.H., Chen, F.P. 2008. Treatment of acute myeloid leukemia by directly targeting both leukemia stem cells and oncogenic molecule with specific scFvimmunolipoplexes as a deliverer. *Med. Hypotheses* 70, 122–127.

21. Depetris, M., Casalis, P., Kratje, R., Etcheverrigaray, M., Oggero, M. 2008. A scFv antibody fragment as a therapeutic candidate to neutralize a broad diversity of human IFN-alpha subtypes. *J. Immunol. Methods* 334, 104–113.
22. Stocks, M. 2005. Intrabodies as drug discovery tools and therapeutics. *Curr. Opin. Chem. Biol.* 9, 359–365.
23. Burton, D.R. 1985 Immunoglobulin G; Functional sites. *Mol.Immunol.* 22, 161-206.
24. Burton, D.R. and Woof, J.M. 1992 Human antibody effector function. *Adv. Immunol.* 51, 1-84.

## 2. วิธีในการออกแบบโมเดลแอนติบอดี และแอนติเจน

ดังที่ได้กล่าวมาแล้วว่าโครงสร้างบริเวณ complementarity - determining regions (CDRS) มีส่วนสำคัญในการยึดจับกับแอนติเจน จำนวนกรดอะมิโนใน CDRS ซึ่งมีประมาณ 70 เรสิดิว (residues) จากจำนวนส่วนของ Fv แอนติบอดีทั้งหมดประมาณ 230 residues ปัจจุบันพบว่าการใช้ Single-chain antibody fragment (scFv) ซึ่งประกอบด้วยส่วนของ  $V_H$  และ  $V_L$  ที่ต่อกันด้วย polypeptide linker สายสั้นที่มีความยืดหยุ่นสูง สามารถนำไปประยุกต์ใช้ได้หลายทาง โดยยังมีความจำเพาะกับแอนติเจนสูงแม้ว่าจะจำเพาะน้อยกว่าโครงสร้างแอนติบอดีจริง เช่น การผลิต scFv ได้ปริมาณมากในแบคทีเรียโดยใช้ต้นทุนต่ำ การ fusion กับ protein drug ไปยัง target and kill pathogens หรือ การ fusion กับ marker molecules เพื่อใช้ในการตรวจวิเคราะห์ [1, 2]

จากโครงสร้างเชิงซ้อนแอนติบอดี-แอนติเจนที่พบครั้งแรกในฐานข้อมูลโปรตีน [3-6] ซึ่งเพิ่มจำนวนโครงสร้างเชิงซ้อนดังกล่าวขึ้นตามลำดับ ทำให้ทราบถึงรูปแบบการจดจำ (recognition) ของแอนติบอดี และมีการวิจัยหลายกลุ่มศึกษาอันตรกิริยาระหว่างแอนติบอดี-แอนติเจน [7-8] และศึกษาการทำนายโครงสร้างของโปรตีนซึ่งปัจจุบันวิธีที่นิยมใช้หลักๆ มี 3 ขอบเขตที่สำคัญคือ การทำนายโดยวิธี *ab initio* (*ab initio* prediction) การจดจำลักษณะการม้วนพับ (fold recognition) และการโมเดลโดยวิธีโฮโมโลยี (homology modeling) ซึ่งทั้งสามวิธีต่างกันขึ้นกับข้อมูลลำดับและโครงสร้างในฐานข้อมูลที่จะใช้สร้างโมเดล กรณีการทำนายโดยวิธี *ab initio* จะไม่ใช้ฐานข้อมูลใดๆ เลย [9] โครงสร้างที่ทำนายได้ทำนายโดยใช้กฎทางฟิสิกส์และเคมีเท่านั้น คำว่า *ab initio* หรือ *de novo* ใช้ในการทำนายโครงสร้างของโปรตีนที่ไม่พบโครงสร้างที่คล้ายคลึงกันในฐานข้อมูลโปรตีน แต่พบลำดับอะมิโน และความสัมพันธ์ระหว่างโครงสร้างในส่วนโปรตีนสั้นๆ และใช้การในการทำนายโครงสร้างทุติยภูมิ (secondary structure) [10] ส่วนกรณีการจดจำ การม้วนพับ หรือ นิยมเรียกว่า “threading” ใช้ในกรณีที่ไม่มีหลายโครงสร้างแม่แบบที่คล้ายคลึงกับโครงสร้างที่ต้องการสร้าง และต้องเลือกแม่แบบที่ดีที่สุด ซึ่งเป็นไปได้ยากที่จะได้โมเดลที่ถูกต้องแม่นยำ ปัจจุบันวิธีที่แม่นยำสร้างได้เมื่อทราบแม่แบบจากฐานข้อมูลโปรตีน ที่มีลำดับอะมิโนที่มีความคล้ายคลึงกับโปรตีนเป้าหมาย การสร้างโปรตีนจากแม่แบบดังกล่าวเรียกว่าการโมเดลโดยวิธีโฮโมโลยี (homology modeling) หรือ การโมเดลด้วยการเทียบแบบ (comparative modeling) ซึ่งในงานนี้จะขอเสนอการใช้ homology modeling ซึ่งเป็นวิธีที่น่าจะมีโอกาสได้โมเดลที่ถูกต้อง ในการสร้างโมเดลแอนติบอดี และแอนติเจนเริ่มต้น และทำการปรับปรุงโมเดลโดยการทำ refinement ด้วยกลศาสตร์เชิงโมเลกุล (Molecular Mechanics, MM) และ พลศาสตร์เชิงโมเลกุล (Molecular Dynamics, MD) และใช้วิธีการล็อกกิ้ง (docking method) เพื่อทำนายโครงสร้างเชิงซ้อนแอนติบอดี-แอนติเจน คำนวณพลังงานยึดจับอิสระในน้ำ และหาลำดับหมู่อะมิโนที่สำคัญที่ยึดจับระหว่างแอนติบอดีและแอนติเจน เพื่อแสดงบริเวณเฉพาะบนโครงสร้างของแอนติบอดีที่จับกับแอนติเจน ซึ่งหากเป็นบริเวณสำคัญที่เกี่ยวข้องกับหน้าที่ของโมเลกุล เช่น ควบคุมกลไกการเกิดมะเร็ง หรือ การส่งสัญญาณเข้าสู่เซลล์เพื่อควบคุมการแบ่งตัวของเซลล์ในระบบภูมิคุ้มกัน หรือ จับกับเชื้อก่อโรคและสามารถยับยั้งการเพิ่มจำนวนของตัวเชื้อได้ ก็จะส่งผลให้เกิดองค์ความรู้ในการพัฒนาแอนติบอดีที่มีประสิทธิภาพสูงขึ้นโดยทำ protein engineering ซึ่งอาศัยข้อมูลทางชีววิทยาที่ได้จากการทำนายเชิงโมเลกุล รวมทั้งการสร้างแอนติบอดีที่มีประสิทธิภาพสูง ในการจับกับบริเวณจำเพาะบนแอนติเจนได้จากการรู้โครงสร้างเชิงซ้อนของโมเลกุล

ในการศึกษาครั้งนี้ผู้วิจัยมีเป้าหมายในการศึกษาการเข้าจับกันระหว่าง scFv กับ HIV-1 p 17 ที่ตำแหน่ง c-terminal epitope โดยใช้เทคนิคทาง theoretical chemistry เช่น docking ซึ่งจะช่วยให้ทราบรายละเอียดต่างๆ อีกมาก โดยเฉพาะอย่างยิ่งพลังงานในการยึดจับกันซึ่งแสดงถึงความสามารถของแอนติบอดีในการยึดจับกับ peptide epitope โดยโครงสร้าง 3 มิติของโมโนโคลนอลแอนติบอดีที่จำเพาะต่อ HIV-1 c-terminal epitope บน p 17 นั้นได้ข้อมูลลำดับอะมิโนของ scfv จากข้อมูลทางห้องปฏิบัติการโดยการโคลน immunoglobulin (Ig) gene variable fragments (single chain Fv, scFv) จาก hybridoma cells



ข้อมูลในห้องปฏิบัติการโดยความร่วมมือในส่วนการโคลนนิ่ง และวิเคราะห์ลำดับกรดอะมิโนแอนติบอดีสายเดี่ยวที่มีความจำเพาะต่อ HIV-1 c-terminal epitope บน p 17 และการวิเคราะห์ epitope mapping โดย รศ.ดร. ชัชชัย ตะยาภักพัฒนา ภาควิชาภูมิคุ้มกันวิทยาคลินิก คณะเทคนิคการแพทย์ มหาวิทยาลัยเชียงใหม่ เพื่อยืนยันผลการคำนวณ และเป็นแนวทางการพัฒนาเพื่อผลิตเป็นวัคซีน หรือใช้แอนติบอดีในการตรวจวิเคราะห์ต่อไป โดย scFv จะทำหน้าที่ในการเข้าจับกับ HIV-1 ในบริเวณ c-terminal epitope บน p 17 เพื่อขัดขวางการเข้าจับของ HIV-1 protease ในการย่อย substrate ส่งผลให้การสร้างไวรัสใหม่ลดลง เนื่องจากบริเวณ cleavage sites ต่างๆของเอชไอวีมีการเกิด mutation น้อยมาก ทำให้โอกาสที่จะพัฒนาเทคโนโลยีนี้เพื่อประยุกต์ใช้ในการรักษา เนื่องจากการทำงานของ scFv เกิดขึ้นภายในเซลล์ดังนั้นจึงไม่กระตุ้นให้ภูมิคุ้มกันของมนุษย์ตอบสนองต่อ scFv

ในการศึกษานี้จะเลือกศึกษาโครงสร้างแอนติบอดี-แอนติเจน 2 ระบบโดยมีเครือข่ายความร่วมมือกับนักวิจัยในส่วนที่ทำการสังเคราะห์แอนติบอดีคือ

ระบบที่ 1 แอนติบอดีต่อเอชไอวี-1 โปรตีเอส โดยใช้แอนติเจนที่เป็นโครงสร้างของยา และ substrate

ปัจจุบันโรคเอดส์ยังคงเป็นปัญหาที่สำคัญ พบการแพร่ระบาดมากขึ้น และ ยังไม่มีการป้องกันที่ได้ผลอย่างแท้จริง [13-15] แม้ว่าความพยายามในการพัฒนาวัคซีนจะยังไม่ประสบความสำเร็จมากนักตลอดระยะเวลา 20 ปีที่ผ่านมา ยังมีความพยายามออกแบบวัคซีนรูปแบบใหม่ๆ เพื่อกระตุ้นภูมิคุ้มกันของผู้ป่วยเอดส์อย่างต่อเนื่อง วัคซีนที่มีการศึกษาส่วนมากผลิตจากชิ้นส่วนโปรตีนของเชื้อเอชไอวี เรียกว่า subunit protein vaccine เป็นการตัดเอาบางส่วนจากโปรตีนจากเชื้อที่สามารถกระตุ้นภูมิคุ้มกันได้ ตัวอย่างโปรตีนดังกล่าวคือ gp 120 และ gp160 โดยหวังว่าจะได้แอนติบอดีเพื่อยับยั้งการจับของเอชไอวีกับผิวเซลล์เป้าหมาย เช่น  $CD4^+$  T-lymphocyte ซึ่งเป็นแหล่งเพาะเชื้อ เนื่องจากการเตรียมโปรตีนของไวรัสเอชไอวีในปริมาณมากพอเพื่อใช้ในระดับอุตสาหกรรมผลิตวัคซีนมีความยุ่งยากซับซ้อน ซึ่งเปปไทด์ดังกล่าวมีโครงสร้างที่ไม่เสถียร ไม่เป็นเนื้อเดียว (nonhomogeneous) ตกตะกอน และไม่แสดงสมบัติของอิมูโนโพรตีนได้อย่างที่พบในธรรมชาติ [16-17] อีกวิธีหนึ่งคือ การให้ แอนติบอดีต่อเอชไอวี-1 จดจำและเลือกเปปไทด์ที่เหมาะสมจากการคัดกรองจากการใช้ phage epitope libraries [18-20] ซึ่งเป็นวิธีการทดลองที่ต้องมีค่าใช้จ่ายค่อนข้างสูง ดังนั้นทางผู้วิจัยเห็นความสำคัญของการพัฒนาวิธีที่มีประสิทธิภาพจะสามารถลดการสูญเสียทางเศรษฐกิจ และเพิ่มคุณภาพชีวิตของผู้ติดเชื้อได้

ในการศึกษานี้ผู้วิจัยในส่วนที่ทำการสังเคราะห์มีเป้าหมายในการสร้างโมโนโคลนอลแอนติบอดีจำเพาะต่อ HIV protease substrate ในบริเวณ cleavage sites เช่น ระหว่าง p17-p24 เพื่อใช้ในการโคลน immunoglobulin (Ig) gene variable fragments (single chain Fv, scFv) จาก hybridoma cells และนำไปทำการสอดเข้าสู่ adenovirus vector เพื่อใช้ในการนำส่ง scFv gene เข้าสู่เซลล์เป้าหมาย เรียกว่า intrabody technique ในการศึกษานี้ BSA-conjugated indinavir จะถูกเตรียมขึ้นและเลือกใช้เป็น immunogen เนื่องจาก indinavir เป็นยาต้านเอชไอวีในกลุ่มยาต้านโปรตีเอส ที่มีการเลียนแบบโครงสร้างของ substrate บริเวณ cleavage sites เพื่อ block binding groove ของ HIV protease ทำให้การกระบวนการ replication ของไวรัสไม่สมบูรณ์ โดยตั้งสมมติฐานว่าหากสามารถสร้างโมโนโคลนอลแอนติบอดีที่จำเพาะต่อ indinavir โมโนโคลนอลแอนติบอดีที่ได้ย่อมสามารถจับกับบริเวณ cleavage sites ได้ทุกบริเวณ ซึ่งจะมีผลดีกว่าการสร้างโมโนโคลนอลแอนติบอดีต่อ protease substrate cleavage site โดยตรง เนื่องจากโมโนโคลนอลที่ได้อาจจดจำลำดับกรดอะมิโนเฉพาะบริเวณนั้น ทำให้ไม่สามารถเข้าจับกับ cleavage sites ในตำแหน่งอื่นๆได้

การศึกษาโครงสร้างสามมิติของ scFv ที่ผลิตขึ้นว่าบริเวณใดในโครงสร้างของ indinavir ที่มีความสำคัญในการเลียนแบบโครงสร้างของ cleavage sites บน HIV substrate ที่ scFv เข้าจับอย่างจำเพาะ โดยเปรียบเทียบกับโครงสร้างสามมิติของ scFv ที่สามารถจับกับโครงสร้างของ indinavir แต่ไม่สามารถจับกับ substrate ได้ กระบวนการศึกษาในส่วนนี้จะใช้เทคนิคทาง theoretical chemistry เช่น โมเลกุลลาร์ด็อกกิง (molecular docking) กลศาสตร์ควอนตัม (Quantum Mechanics, QM) กลศาสตร์เชิงโมเลกุล (Molecular Mechanics, MM) และ

พลศาสตร์เชิงโมเลกุล (Molecular Dynamics, MD) ในอนาคตองค์ความรู้ใหม่ที่ได้จะเอื้อประโยชน์ต่อการทำ molecular drug design ของยาต้านโปรตีนเอส โดยใช้ scFv anti-cleavage sites เป็นแม่แบบ ในการทดสอบยาใหม่ ว่ามีค่าการยึดจับว่าสูงหรือต่ำเพียงไร

ระบบที่ 2 แอนติบอดีต่อ human leukocyte surface molecule CD147 โดยมีการคำนวณเพื่อทำนายอิพิโทป จากอันตรกิริยาระหว่างโปรตีน-โปรตีนดังกล่าว โดยใช้ single-chain antibody fragment (scFv) ประกอบด้วยส่วนของ VH และ VL ที่ต่อกันด้วย polypeptide linker สายสั้นที่มีความยืดหยุ่นสูง ซึ่ง CD147 มีหน้าที่เป็นตัวรับ สัญญาณ (signaling receptor) สำหรับ extracellular cyclophilins และยังพบว่า CD147 มีอันตรกิริยากับเซลล์ภูมิ ต้านทาน (immune cells) อีกด้วย

สำหรับประโยชน์ของการใช้ scFv ในการหาแอนติเจน คือมีความจำเพาะกับแอนติเจนสูงแม้ว่าจะจำเพาะน้อยกว่า โครงสร้างแอนติบอดีจริง และสามารถผลิต scFv ได้ปริมาณมากในแบคทีเรียโดยใช้ต้นทุนต่ำ นอกจากนี้ยังสามารถประยุกต์ได้หลายแนวทาง เช่น การ fusion กับ protein drug ไปยัง target and kill pathogens หรือ กับ marker molecules สำหรับการวิเคราะห์

## 8. ระเบียบวิธีวิจัย

วิธีที่ใช้ วิธีโฮโมโลยีโมเดลลิง (homology modelling) โมเลกุลลาร์ต็อกกิง (molecular docking) กลศาสตร์ควอนตัม (Quantum Mechanics, QM) กลศาสตร์เชิงโมเลกุล (Molecular Mechanics, MM) พลศาสตร์เชิงโมเลกุล (Molecular Dynamics, MD) และเครือข่ายนิวรอน (neuron network)

### 8.1 การสร้างโครงสร้างแอนติบอดี; scFv

8.1.1 หาโครงสร้างแม่แบบของแอนติบอดีที่เหมาะสมจากฐานข้อมูลโปรตีน (Protein Data Bank, PDB)

8.1.2 Homology modeling โครงสร้าง scFv จากลำดับกรดอะมิโน

8.1.3 Refinement โครงสร้าง scFv ด้วยกลศาสตร์เชิงโมเลกุล (Molecular Mechanics, MM) และ พลศาสตร์เชิง โมเลกุล (Molecular Dynamics, MD)

8.1.4 Evaluate model แอนติบอดีที่สร้างได้

### 8.2 การสร้างโครงสร้างแอนติเจน; indinavir, HIV-1 protease substrate, CD147

8.2.1 หาโครงสร้างแม่แบบของแอนติเจนที่เหมาะสมจาก Protein Data Bank (PDB)

8.2.2 Homology modeling โครงสร้าง CD147

8.2.3 Refinement โครงสร้าง CD147 ด้วยกลศาสตร์เชิงโมเลกุล (Molecular Mechanics, MM) และ พลศาสตร์เชิง โมเลกุล (Molecular Dynamics, MD)

8.2.4 Evaluate model CD147 ที่สร้างได้

### 8.3 การทำนายโครงสร้างแอนติบอดี – แอนติเจน; scFv-indinavir, scFv-HIV-1 protease substrate, scFv-CD147

8.3.1 ศึกษาโครงสร้างระหว่างแอนติบอดี - แอนติเจนที่พบปัจจุบันในฐานข้อมูลโปรตีน (Protein Data Bank, PDB)

8.3.2 ทำนายบริเวณยึดจับของแอนติบอดี และแอนติเจน

8.3.3 Molecular Docking ทำนายโครงสร้างยึดจับระหว่างแอนติบอดี – แอนติเจน

8.3.4 หาลำดับอะมิโนของแอนติเจนที่เหมาะสมที่มีส่วนสำคัญในโครงสร้างยึดจับกับแอนติบอดีจากโครงสร้างที่ ทำนายได้จาก Molecular Docking

8.3.5 การทำนายอิพิโทป ด้วยการใช้เครือข่ายนิวรอน กับฐานข้อมูลอิพิโทปเปปไทด์ และเปรียบเทียบผลกับวิธีดี ออกกิง

### 8.4 การตรวจสอบทำนายโครงสร้างแอนติบอดี - แอนติเจน, scFv-CD147

8.4.1 ศึกษา epitope mapping หาโครงสร้างแอนติเจนที่เหมาะสมจากโครงสร้างระหว่างแอนติบอดี - แอนติเจนที่ทำนายได้ในขั้นต้น

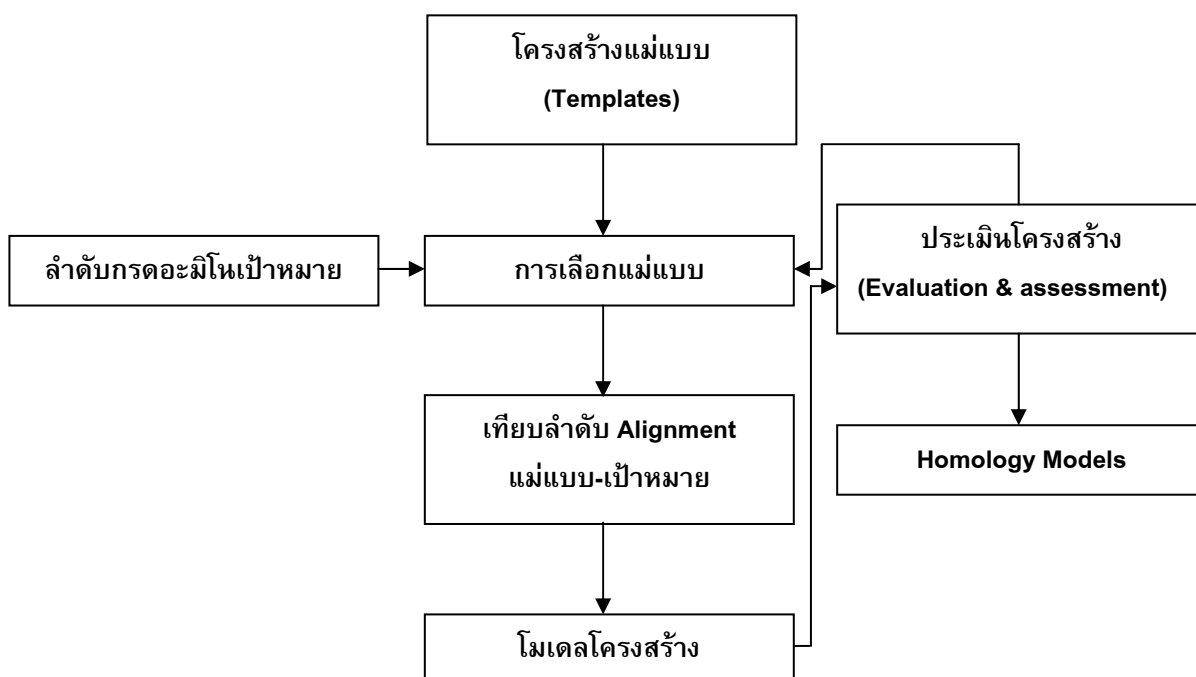
8.4.2 เปรียบเทียบผลจากห้องปฏิบัติการ

โดยจะขอล่าถึงหลักการอย่างคร่าวๆ ในแต่ละหัวข้อและสามารถอ่านรายละเอียดเพิ่มเติมได้จาก manuscript ในภาคผนวก ดังนี้

1. การจำลองโมเดลแอนติบอดี โดยใช้ Homology modeling และจาก Website ออกแบบโมเดลสามมิติแอนติบอดี
2. การศึกษาการยึดจับจาก molecular docking และ
3. การคำนวณพลังงาน และการจำลอง โดยใช้กลศาสตร์ควอนตัม กลศาสตร์เชิงโมเลกุล และ พลศาสตร์เชิงโมเลกุล
4. การทดลองในการศึกษา antibody-antigen affinity

#### 1. การจำลองโมเดลแอนติบอดี

a) **Homology modeling** ซึ่งในการศึกษาครั้งนี้ได้ใช้วิธีนี้เป็นส่วนใหญ่ โดยสร้างโมเดลแอนติบอดีจากข้อมูลโครงสร้างสามมิติของแอนติบอดีที่คล้ายคลึงกัน จากฐานข้อมูลโปรตีน (protein data bank) [11] โดยข้อมูลโครงสร้างได้มาจากการเลี้ยวเบนรังสีเอกซ์เรย์ ข้อมูล NMR หรือเทคนิคทางสเปกโตรสโกปีอื่นๆ ขั้นตอนหลักในการทำโฮโมโลยีโมเดลประกอบด้วย 4 ขั้นตอนตามแผนผังในรูปที่ 2.1



รูปที่ 2.1 ขั้นตอนหลักในการสร้าง homology model.

ขั้นแรกเป็นการเลือกโปรตีนต้นแบบจาก protein data bank (PDB) [11] โดยใช้ลำดับของโปรตีนเป้าหมาย [12] ในการค้นหา หากมีลำดับเหมือนกันสูงถึง >30% จะใช้การเปรียบเทียบลำดับเป็นคู่ (pairwise sequence

comparison) เช่นใช้โปรแกรม BLAST [13] และ FASTA [14] หาก <30% มักใช้วิธี multiple sequence alignment เช่น profile analysis [15, 16] ขั้นที่สองเป็นขั้นการเทียบลำดับ (align) เป้าหมายและแม่แบบ ตัวอย่างโปรแกรมที่ใช้ align แม่แบบที่เหมาะสมเช่น t-coffee [17] หากลำดับเหมือนมากกว่า 50% จะได้ alignment ที่ถูกต้องค่อนข้างมาก หากลำดับเหมือนน้อยกว่า 35% ที่เรียกว่า "twilight zone" ความถูกต้องจะลดลงไปอย่างมาก [18] เนื่องจากการเพิ่มจำนวนช่องว่างและทำให้การ align ผิดพลาดและต้องตรวจสอบอย่างระมัดระวัง เมื่อได้แม่แบบและเทียบลำดับแล้ว โมเดลของเป้าหมายจะถูกสร้างขึ้น โดยหลายวิธี เช่นวิธีที่ได้จาก rigid fragment assembly เช่น SWISS-MODEL [19, 20] หรือ Composer [21] และวิธี spatial restraint methods ใน Modeller [22] การออกแบบโดยใช้ rigid fragment assembly ใช้หลักการหาส่วนของโครงสร้าง conserved core sections ภายในตระกูลโปรตีน แต่ในส่วน loops และ side-chains เป็นส่วนที่เปลี่ยนและต้องโมเดลด้วยวิธีอื่น การออกแบบโดยใช้ spatial restraints จะสร้างส่วนคงที่ หรือ ส่วนไม่คงที่ (constraints or restraints) จำนวนมากบนลำดับเป้าหมาย จากระยะพันธะ มุม ทอร์ชัน บริเวณที่ไม่มีพันธะจากสนามแรงกลศาสตร์เชิงโมเลกุล sidechains มักออกแบบโดยใช้ rotamer libraries [23, 24] การสร้างส่วน loop ทำโดยวิธี ab initio และสร้าง conformations ที่เป็นไปได้จำนวน และทำ scoring [25, 26] หรือจาก database methods หาโครงสร้างโปรตีนที่คล้ายกับส่วน loop [27, 28].

ความผิดพลาดของการสร้างด้วยวิธีโฮโมโลยี เป็นไปได้ใน 5 ลักษณะด้วยกัน [29] ได้แก่ errors in side-chain prediction, alignment errors, errors ในบริเวณที่ไม่มีแม่แบบ, errors due to misalignments และจากการใช้แม่แบบที่ผิด โดยมีโปรแกรมใน webserver ที่ใช้ประเมินโมเดล PROCHECK [30] และ WHATIF [31] ใช้ตรวจสอบ stereochemistry และ structural packing quality ส่วน VERIFY3D [32] ใช้วิเคราะห์ compatibility ของโมเดลระดับอะตอมภายในลำดับอะมิโนเอง ขณะที่ ANOLEA [33] ใช้ตรวจ fitness ของลำดับกับโครงสร้างและประเมิน nonlocal environment แต่ละอะตอมหนักในโมเลกุล บางครั้งในการออกแบบจำเป็นต้องเลือกแม่แบบ หรือ align ใหม่เพื่อปรับปรุงคุณภาพของโครงสร้างสามมิติให้ดีขึ้น

**b) Website ออกแบบโมเดลสามมิติแอนติบอดี** ได้รวบรวมแหล่ง website ในการวิเคราะห์และวิศวกรรมแอนติบอดี ดังนี้

- The Kabat Database (<http://www.kabatdatabase.com>)

Created by E.A. Kabat and T.T. Wu in 1966, the Kabat database publishes aligned sequences of antibodies, T-cell receptors, major histocompatibility complex (MHC) class I and II molecules, and other proteins of immunological interest. A searchable interface is provided by the SeqHuntII tool, and a range of utilities is available for sequence alignment, sequence subgroup classification, and the generation of variability plot.

- Antibodies-structure and Sequence (<http://www.bioinf.org.uk/abs>)

This page summarizes useful information on antibody structure and sequence. It provides a query interface to the Kabat antibody sequence data, general information on antibodies, crystal structures, and links to other antibody-related information. It also distributes an automated summary of all antibody structures deposited in the Protein Databank (PDB). Of particular interest is a thorough description and comparison of the various numbering schemes for antibody variable regions.

- WAM-Web Antibody Modeling (<http://antibody.bath.ac.uk>)

Hosted by the Centre for Protein Analysis and Design at the University of Bath, United Kingdom. Based on the AbM package (formerly marketed by Oxford Molecular) to construct 3D models of antibody Fv sequences using a combination of established theoretical methods, this site also includes the latest antibody structural information. It is free for academic use.

- Mike's Immunoglobulin Structure/Function Page

(<http://www.path.cam.ac.uk/~mrc7/mikeimages.html>)

These page provide educational materials on immunoglobulin structure and function, and are illustrated by many color images, models, and animation. Additional information is available on antibody humanization and Mike Clark's Therapeutic Antibody Human Homology Project, which aims to correlate clinical efficacy and anti-immunoglobulin responses with variable region sequences of therapeutic antibodies.

- The Recombinant Antibody Pages (<http://www.mgen.uni-heidelberg.de/SD/SDScFvSite.html>)

This is large collection of links and information on recombinant antibody technology and general immunology that provides links to companies that exploit antibody technology

- Humanization by Design (<http://people.cryst.bbk.ac.uk/~ubcg07s>)

The resource provides an overview on antibody humanization technology. The most useful feature is a searchable database (by sequence and text) of more than 40 published humanized antibodies including information on design issues, framework choice, framework back-mutations, and binding affinity of the humanized constructs.

## 2. การศึกษาการยึดจับจาก molecular docking

ประสิทธิภาพของยาสามารถประเมินได้จากการศึกษาการยึดจับระหว่างโมเลกุล (ลิแกนด์) ในบริเวณ pocket หรือ ตำแหน่งยึดจับของโมเลกุลขนาดใหญ่กว่า (โปรตีน) เรียกว่า การด็อกกิง (docking) เช่น ยา กับ เอนไซม์ ในที่นี้ แอนติบอดีกับแอนติเจน โดยนำโมเลกุลที่กำลังออกแบบมาลองจับเข้ากับโมเลกุลเป้าหมาย โดยสามารถยับยั้งเปลี่ยนทิศทางในสามมิติ เพื่อหาตำแหน่งที่เหมาะสม เข้ากันได้ดี (fit) ในการจับกันของโมเลกุลทั้งสอง และได้รูปร่างโครงสร้างเชิงซ้อนที่เหมาะสมที่สุดแรงผลักระหว่างโมเลกุลที่คำนวณได้ โดยเทคนิคเดิมในห้องปฏิบัติการในการหาข้อมูลโครงสร้าง protein-ligand complexes มักเป็นไปอย่างล่าช้า ทางเลือกโดยการประเมินทางคอมพิวเตอร์เพื่อหาโครงสร้างยึดจับจึงมีส่วนช่วยสำคัญ ในการศึกษาครั้งนี้ได้ใช้วิธี CHARMM-based DOCKER (CDOCKER) ใช้ grid-based molecular dynamics (MD) docking algorithm โดยใช้ Monte Carlo simulated-annealing ด้วยสมมติฐานที่ว่าโปรตีนอยู่นิ่ง (rigid) และลิแกนด์ขยับได้ (flexible) แม้ว่าวิธีจำลองพลศาสตร์แอนนิลลิ่งปกติ (MD-based simulated annealing) จะใช้เวลาค่อนข้างนาน เนื่องจากการหา conformation ของลิแกนด์จำนวนมากทั้ง external และ internal rotational degrees of freedom ในบริเวณ active site ของโปรตีนเป้าหมาย แต่อัลกอริทึมใน CDOCKER โดยการใช้ grid-based tri-linear interpolation technique ได้ช่วยลดเวลาลงไปได้มาก อีกทั้งมีความแม่นยำมากขึ้นโดยการใช้ full force field minimization approach และการทำ refine the docked poses ในขั้นตอนสุดท้าย

โปรแกรมด็อกกิงอีกหนึ่งโปรแกรมที่เลือกใช้คือ CAChe [34-36] ซึ่งใช้สำหรับวินโดวส์มี โปรแกรมนี้สามารถใช้วาดรูปโมเดลโมเลกุลและทำการคำนวณบนโมเลกุลเพื่อค้นหาสมบัติและค่าพลังงานของได้ทั้งกลศาสตร์ดั้งเดิม (classical mechanics) และกลศาสตร์ควอนตัม (quantum mechanics) การศึกษาการเข้าจับของโมเลกุลที่ได้จากโปรแกรม CAChe ทำโดยใช้ grid-base energy evaluation คล้ายกับที่ใช้ใน AutoDock ส่วนที่แตกต่างกันของทั้งสองโปรแกรม คือ สกอร์ฟังก์ชันที่ใช้ในซอฟต์แวร์เท่านั้น โดย CAChe 6.1 สกอร์ฟังก์ชันจาก Potential Mean Force (PMF) ข้อมูลพลังงานศักย์ระหว่างคู่อะตอมจะถูกรวบรวมจากโครงสร้าง X-ray ในฐานะข้อมูลโปรตีน

$$A_{ij}(r) = -k_B T \ln \left[ f^{i \text{ vol-corr}}(r) \frac{p^{ij}_{seg}(r)}{p^{ij}_{bulk}(r)} \right]$$

$k_B$  คือ ค่าคงที่ Boltzmann

$T$  คือ อุณหภูมิ

$R$  คือ ระยะห่างคู่อะตอม

$f^{i \text{ vol-corr}}(r)$  คือ แฟคเตอร์ทั้งหมดของลิแกนด์

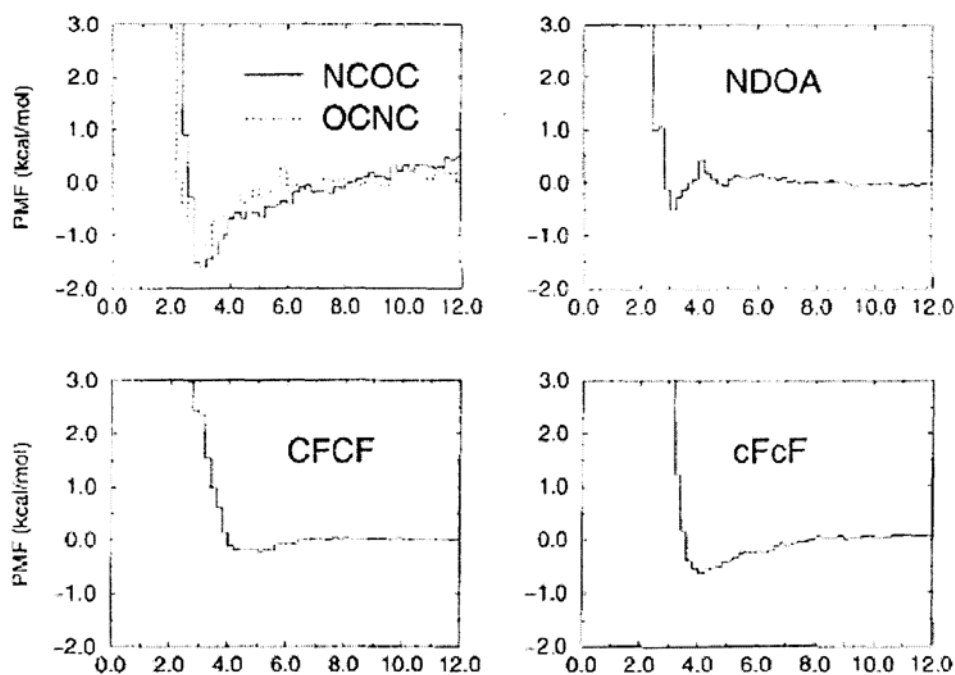
$p^{ij}_{seg}(r)$  คือ จำนวนความหนาแน่นของ  $ij$  ในโครงสร้างพื้นฐานที่ปรากฏ  
ในช่อง รัศมี กำหนดโดย “seg”

$p^{ij}_{bulk}(r)$  คือ จำนวนความหนาแน่นของ  $ij$  ใน bulk

ผลหารของ  $p^{ij}_{seg}(r)/p^{ij}_{bulk}(r)$  จะเป็นฟังก์ชันความสัมพันธ์ของอะตอมโปรตีน type  $i$  รวมกับอะตอมลิแกนด์ type  $j$  ในโครงสร้างพื้นฐานของเชิงซ้อนลิแกนด์โปรตีน และได้สคออร์ฟังก์ชันเป็นตัวกำหนดอันตรกิริยาทั้งหมดของเชิงซ้อนลิแกนด์โปรตีน

$$\text{PMF - score} = \sum_{kl, r < r^{ij}_{cutoff}} A_{ij}(r)$$

$r^{ij}_{cutoff}$  คือ การตัดรัศมีสำหรับอะตอม type  $ij$  และอะตอม  $kl$  ของลิแกนด์โปรตีนทั้งหมดในข้อมูลพื้นฐานเมื่อการคำนวณเสร็จสิ้นจะได้ curves ของศักย์ไฟฟ้าที่ชัดเจนและรูปร่างคล้ายคลึงกับ “Morse Curve” ดังตัวอย่างในรูปที่ 2.2 โดยแกน x ของ curves คือ ระยะห่างของคู่อะตอม และแกน y คือ สคออร์ของการคำนวณ



รูปที่ 2.2 ตัวอย่างของกราฟ PMF ตัวอักษร 4 ตัว หมายถึง ชนิดของคู่อะตอมที่มีอันตรกิริยาต่อกัน ตัวอักษร 2 ตัวแรกจากโปรตีน ตัวอักษร 2 ตัวหลังจาก [36]

### 3. การคำนวณพลังงาน และการจำลอง โดยใช้กลศาสตร์ควอนตัม กลศาสตร์เชิงโมเลกุล และ พลศาสตร์เชิงโมเลกุล

#### a) กลศาสตร์ควอนตัม (Quantum Mechanics)

วิธีการคำนวณทางเคมีควอนตัมนั้นคำนวณหาพลังงานของระบบอะตอมและโมเลกุลจากการแก้สมการคลื่น (Schrödinger equation) โดยพิจารณาว่าระบบล้นเป็นคลื่น ไม่นิยมใช้สำหรับระบบที่ใหญ่ เช่น ระบบชีวโมเลกุล เนื่องจากการแก้สมการดังกล่าวใช้เวลาค่อนข้างนาน และต้องการคอมพิวเตอร์สมรรถนะสูง วิธีการคำนวณทางควอนตัม มีระเบียบวิธีคิดที่หลากหลายซึ่งให้ผลที่มีความถูกต้อง (accuracy) แตกต่างกัน และอาศัยความรู้ทางทฤษฎีของผู้ใช้เป็นอย่างมาก ข้อมูลที่ได้จากวิธีการคำนวณทางเคมีควอนตัมนั้นส่วนใหญ่เป็นการประยุกต์เอาความรู้ทางพลังงานและสมบัติที่เป็นผลพลอยได้อื่นๆ เช่น ค่าประจุและแรงกระทำ เป็นต้น ไปใช้ร่วมกับวิธีการอื่น ตัวอย่างเช่น ค่าพลังงานสามารถนำไปสร้างพื้นผิวพลังงานศักย์ (potential-energy surface) เพื่อนำไปใช้กับการจำลองด้วยคอมพิวเตอร์

#### b) กลศาสตร์เชิงโมเลกุล

นิยมใช้คำนวณระบบชีวโมเลกุล แนวคิดหลักของวิธีการออกแบบเชิงโมเลกุลที่ถูกเรียกและนำมาใช้เพื่อแสดงแบบการเคลื่อนไหว (mode of motions) หรือสมบัติเชิงกล (mechanical properties) ของโมเลกุล โดยสมมติให้อะตอมเป็นก้อนมวลที่ยึดติดกันด้วยสปริง แรงกระทำระหว่างกันของอะตอมต่างๆ นั้นคำนวณจากสมการชุดหนึ่งซึ่งจะกำหนดรูปแบบการเคลื่อนที่ของอะตอมต่างๆ ในโมเลกุล เช่น การยืด-หดของพันธะ (stretching) การงอเข้า-ออกของมุม (bending) และการหมุนของพันธะ (torsional angle) แล้วอาศัยหลักการฟิสิกส์แบบเดิม (classical physics) เช่น กฎของฮุก (Hook's law) ร่วมกับพารามิเตอร์ที่มีความจำเพาะ เช่น ค่าคงที่แรงกระทำ (force constant) เพื่อสร้างฟังก์ชันซิมเพิลฮาร์โมนิก แล้วคำนวณพลังงานศักย์ โดยมีค่าพารามิเตอร์แต่ละชุดสำหรับเซตของอะตอม ที่เกี่ยวข้องในแบบของการเคลื่อนที่นั้น สมการตลอดจนค่าพารามิเตอร์เหล่านั้นถูกเรียกว่า force-field parameters ซึ่งได้มาจากการรวบรวมและวิเคราะห์ข้อมูลจากการทดลอง และจากการคำนวณทางทฤษฎีจำนวนมาก ข้อดีของวิธีนี้คือสามารถประยุกต์ใช้ศึกษาระบบใหญ่ๆ ที่มีหลายๆ อะตอมได้ โดยเฉพาะในระบบทางเคมีเชิงฟิสิกส์

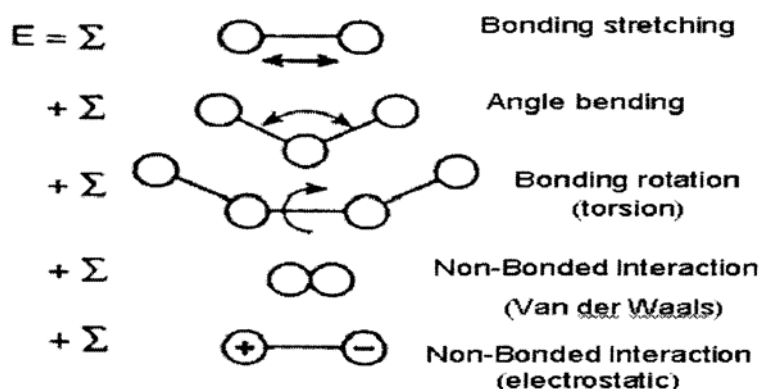
การคำนวณพลังงานศักย์พื้นผิว (potential energy surface) ของมหโมเลกุล (macromolecules) เช่น โปรตีน ใช้พารามิเตอร์ต่างๆ ที่กำหนดไว้ในศักย์ฟังก์ชันของสนามแรงกระทำ ตามหลักการของกลศาสตร์เชิงโมเลกุล พลังงานศักย์ของระบบเป็นฟังก์ชันของพิกัดของอะตอม พลังงานศักย์ของโมเลกุลสัมพันธ์กับโครงสร้างของโมเลกุลซึ่งสามารถคำนวณได้โดยอาศัยการเทียบข้อมูลอ้างอิง เช่น ความยาวพันธะ มุมระหว่างพันธะ เป็นต้น พลังงานศักย์ที่คำนวณมาจากสองส่วน คือ ส่วนแรกเป็นผลรวมของฟังก์ชันพลังงานศักย์ที่มีอันตรกิริยาจากพันธะโคเวเลนต์ เรียกว่า bonded interaction อันตรกิริยาในส่วนนี้ ได้แก่ อันตรกิริยาที่ได้จากการยืด-หดของพันธะ (bond stretching), จากการงอของมุมระหว่างพันธะ (bond bending), จากการหมุนของพันธะหรือการเปลี่ยนมุมไดฮีดรัล (bond rotation หรือ dihedral angle) และส่วนที่สองอันตรกิริยาเป็นผลรวมของฟังก์ชันพลังงานศักย์ในส่วนที่ไม่ได้สร้างพันธะโคเวเลนต์ต่อกัน หรือ non-bonded interaction อันตรกิริยาในส่วนนี้ เช่น แรงแวนเดอร์วาลส์ (van der Waal force) และแรงดึงดูดระหว่างประจุอิเล็กตรอน (electrostatic force) ตามหลักการกลศาสตร์เชิงโมเลกุลแล้ว พลังงานศักย์รวม ( $E_{tot}$ ) ของระบบสามารถคำนวณได้จากสมการ

$$E_{tot} = E_{bd} + E_{nb}$$

$$E_{bd} = E_{bond} + E_{angle} + E_{dihedral}$$

$$E_{nb} = E_{vdw} + E_{coulomb}$$

โดยที่	$E_{bd}$	คือ พลังงานศักย์ของการยืด-หดของพันธะ
	$E_{angle}$	คือ พลังงานศักย์ของการงอของมุมระหว่างพันธะ
	$E_{dihedral}$	คือ พลังงานศักย์ของการเปลี่ยนมุมไดฮีดรัล
	$E_{bd}$	คือ พลังงานศักย์ของอันตรกิริยาของอะตอมที่มีพันธะต่อกัน (bonded interaction)
	$E_{nb}$	คือ พลังงานศักย์ของอันตรกิริยาที่ไม่มีพันธะระหว่างกัน (non-bonded interaction)
	$E_{vdw}$	คือ พลังงานศักย์แวนเดอร์วาลส์
	$E_{coulomb}$	คือ พลังงานศักย์คูลอมบ์



รูปที่ 2.3 พลังงานศักย์เทอมต่างๆ

โดยทั่วไป ฟังก์ชันศักย์ (potential function) ของ  $E_{bond}$  และ  $E_{angle}$  อาจเป็นฟังก์ชันฮาร์มอนิก (harmonic function) หรือฟังก์ชันศักย์ของมอร์ส (Morse potential) ซึ่งมีลักษณะฟังก์ชันแอนฮาร์มอนิก (an harmonic function) (รูปที่ 2.2) ส่วน  $E_{dihedral}$  สอดคล้องกับฟังก์ชันของโคไซน์ (cosine) และสำหรับ  $E_{nb}$  ในเทอมที่แสดง แรเงวแวนเดอร์วาลส์ ใช้ฟังก์ชันของตัวแปร  $r^{-12}$  และ  $r^{-6}$  ซึ่งเรียกฟังก์ชันนี้ว่าฟังก์ชันเลนเนาร์ด-โจนส์ (Lennard-Jones function) และ เทอมที่แสดงแรงดึงดูดระหว่างประจุอิเล็กตรอน (electrostatic force) ใช้ฟังก์ชันคูลอมบ์ (coulomb function) ฟังก์ชันศักย์ สามารถแสดงตามสมการดังต่อไปนี้

$$\begin{aligned}
 E_{bond} &= \sum_{bond} K_r (r - r_0)^2 \\
 E_{angle} &= \sum_{angle} K_\theta (\theta - \theta_0)^2 \\
 E_{dihedral} &= \sum_{torsion} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\
 E_{nb} &= \sum_{i=1}^N \sum_{j=i+1}^N (4\epsilon_{ij}) \left[ \frac{A_{ij}}{r^{12}_{ij}} - \frac{B_{ij}}{r^6_{ij}} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}
 \end{aligned}$$

ระบบๆ หนึ่งอาจมีพลังงานที่ต่ำที่สุดของแต่ละพื้นผิวพลังงานศักย์อยู่จำนวนมากมาย ซึ่งเรียกว่า local energy minimum และค่าพลังงานที่น้อยที่สุดในพื้นผิวพลังงานศักย์กลุ่มนี้ global energy minimum โมเดลที่สร้างขึ้นจาก



โปรแกรมคอมพิวเตอร์ในตอนเริ่มต้น มักจะไม่สามารถนำมาใช้แทนโครงสร้างของโมเลกุลได้ทันที ดังนั้นจะต้องผ่านกระบวนการปรับโครงสร้างเพื่อให้มีพลังงานลดต่ำลงให้น้อยที่สุด ขั้นตอนนี้เรียกว่า energy minimization ตัวอย่าง อัลกอริทึมสำหรับใช้ในวิธีการลดพลังงานลงให้น้อยที่สุด เช่น สตรีเฟเพสท์ดีเซนท์ (steepest descent method) คอนจูเกต เกรเดียนท์ (conjugate gradients) และของนิวตันและราฟสัน (Newton-Raphson method)

สำหรับการวิจัยสำหรับแอนติบอดี และแอนติเจนนี้ใช้ AMBER force field ซึ่งเป็นสนามแรงที่ปรับให้เหมาะสมสำหรับระบบกรดอะมิโน และโปรตีน ข้อควรระวังเนื่องจากโครงสร้างแอนติบอดีประกอบด้วยพันธะไดซัลไฟด์ภายในโดเมน (intra disulfide) สนามแรงแอมเบอร์สามารถแยกระหว่าง cysteine (CYS) และหมู่ที่คล้ายกันแต่เกิดสะพานไดซัลไฟด์ (disulfide bridges, cystine (CYX)) โดย Cystine เป็น dimeric amino acid ที่สร้างโดยการออกซิเดชันของ cysteine สองหมู่เชื่อมกันด้วยพันธะโควาเลนต์เพื่อสร้างพันธะไดซัลไฟด์ โดยต้องเปลี่ยนส่วนของ disulfide bridge ที่โปรแกรมทำอัตโนมัติให้จาก CYS ให้เป็น CYX ให้หมดเพื่อให้ได้พารามิเตอร์ที่ถูกต้องสำหรับพันธะไดซัลไฟด์

### c) พลศาสตร์เชิงโมเลกุล

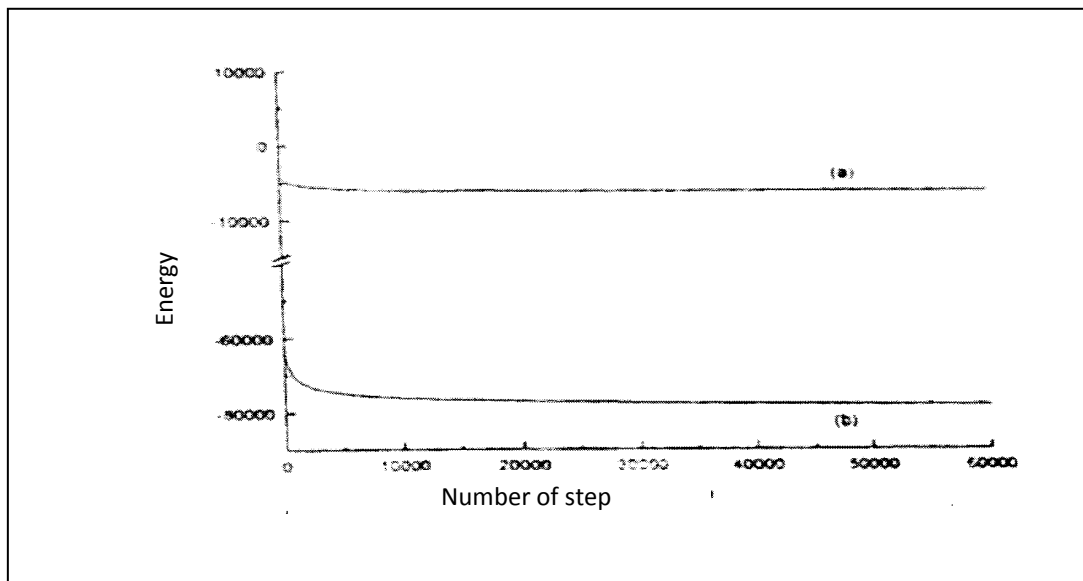
ในสภาพที่เป็นจริงนั้น โมเลกุลต่างๆ ของระบบมีการเคลื่อนไหวตลอดเวลา หรือ มีสมบัติทางพลวัต (dynamics properties) ไม่ว่าจะเป็นอยู่ในสถานะของแข็ง ของเหลว หรือ ก๊าซ ปฏิสัมพันธ์ของโมเลกุลต่างๆ เหล่านี้ยังขึ้นอยู่กับอุณหภูมิอีกด้วย สมบัติทางอุณหภูมิจศาสตร์ (thermodynamics properties) ของระบบจึงเป็นมาตรวัดชนิดหนึ่งที่สะท้อนให้เห็นสมบัติโดยรวมของระบบ ดังนั้นถ้าหากการศึกษานิยมสมบัติมหภาค (macroscopic properties) ของระบบแล้ว โครงสร้างของโมเลกุลในระดับพลังงานต่ำสุดจะไม่ได้ให้ความหมายเท่าใดนัก แต่การเปลี่ยนแปลงหรือสภาพของโมเลกุลต่างๆ ณ ระดับพลังงาน แม้ว่าจะมีพลังงานสูงกว่า แต่ภายใต้สภาวะสมดุลทางอุณหพลศาสตร์ เป็นสิ่งที่มีความหมายมากกว่าและถ้าหากสามารถสำรวจทราบสมบัติทางอุณหพลศาสตร์ของระบบ จะช่วยให้เข้าใจพฤติกรรมทำนายแนวโน้มการเปลี่ยนแปลง และอธิบายสมบัติมหภาคของระบบโดยเฉลี่ยนั้นๆ ได้ดียิ่งขึ้น อย่างไรก็ตามปฏิสัมพันธ์ในระดับโมเลกุลเป็นสมบัติทางจุลภาค (microscopic properties) ชนิดหนึ่ง แต่สมบัติทางอุณหพลศาสตร์เป็นสมบัติทางมหภาค การเชื่อมต่อระหว่างทั้งสองทฤษฎี อาศัยหลักการทางกลศาสตร์สถิติ (statistical mechanics) ซึ่งสามารถเชื่อมต่อระหว่าง สมบัติมหภาคของระบบกับสมบัติทางจุลภาคภายในระบบ การจำลองพลวัตจะสร้างชุดข้อมูลจำนวนมาก ซึ่งข้อมูลแต่ละชุดใช้แทนสมบัติต่างๆ เช่น พลังงาน จลน์ พลังงานศักย์ อุณหพลศาสตร์ โครงสร้าง ณ สถานะภาพหนึ่งๆ เนื่องจากเทคนิคการจำลองทางพลวัตเชิงโมเลกุลจะให้ระบบมีการเปลี่ยนแปลงทั้งโครงสร้างและโมเมนต์ภายใต้สมดุลทางอุณหพลศาสตร์ ถ้าหากเก็บชุดข้อมูลที่ได้จากการทดลองในสภาวะสมดุลได้อย่างเพียงพอและครอบคลุมทุกๆ สถานะภาพ แล้วนำมาวิเคราะห์ตามหลักการทางกลศาสตร์สถิติ การสำรวจเพื่อศึกษามหาภาคจากสมบัติทางจุลภาคก็สามารถกระทำได้

โมเลกุลของโปรตีนก็เหมือนกับโมเลกุลขนาดเล็กอื่นๆ คือมีพลศาสตร์ การเปลี่ยนแปลงคอนฟอร์เมชันของโปรตีนเกิดขึ้นได้มากมาย หลายๆ บริเวณ และค่อนข้างซับซ้อน ยากต่อการตรวจวัดในห้องปฏิบัติการด้วยเทคนิคใดเทคนิคหนึ่ง ในการศึกษาการทำงานของโปรตีน มักศึกษาโปรตีนในสภาวะสารละลายซึ่งถือได้ว่าเป็นสภาวะที่ใกล้เคียงกับในร่างกายมากที่สุด โครงสร้างของโปรตีนในสภาวะผลึกเดี่ยว (single crystal) ที่ได้จากเทคนิครังสีเอ็กซ์ขาดข้อมูลทางพลศาสตร์ ในขณะที่การศึกษาสมบัติทางพลศาสตร์ด้วยเทคนิคเอ็นเอ็มอาร์ หลายระบบยังพบว่า มีอุปสรรคอยู่มาก เทคนิคการจำลองด้วยพลวัตเชิงโมเลกุลของโปรตีนจึงเป็นเทคนิคหนึ่งในการสร้างสมบัติทางพลวัตของโปรตีน ให้ข้อมูลสนับสนุนและขยายความรู้ ความเข้าใจเกี่ยวกับกลไกการทำงานของโปรตีน หรืออธิบายพฤติกรรมของการเปลี่ยนแปลงของโปรตีนได้ ในการจำลองพลศาสตร์มีเทอมต่างๆ ที่ควรทราบดังนี้

#### เงื่อนไขขอบเขตแบบคาบ (periodic boundary condition)

การประมาณโดยการกำหนดขอบเขต หรือบาวนด์รีของระบบเป็นสิ่งจำเป็นในวิธีการจำลองเพราะการเชื่อมโยงสมบัติทางมหภาคกับสมบัติจุลภาค ใช้จำนวนอนุภาคไม่มากนักแต่เพียงพอที่จะใช้แทนทั้งระบบได้ periodic boundary condition ช่วยให้การจำลองมีสภาวะใกล้เคียงความจริงมากขึ้นโดยไม่จำเป็นต้องเพิ่มจำนวนของอนุภาคให้มากมายมหาศาลใกล้เคียงกับระบบจริง การจำลองนี้ทำได้โดยให้ระบบอยู่ในภาชนะลูกบาศก์ ศูนย์กลาง ซึ่งเรียกว่า กล่องจินตนาการ (image boxes) เมื่อให้แต่ละกล่องมีจำนวนอนุภาคให้เหมือนกัน วางเรียงกันไว้รอบๆ ทุกทิศทางในลักษณะคล้ายอาร์เรย์ (arrays) ถ้าเป็นสองมิติ แต่ละกล่องจินตนาการจะมีกล่องจินตนาการข้างเคียงที่เหมือนกันอยู่รอบๆ จำนวน 8 กล่องจินตนาการ แต่ถ้าเป็นสามมิติจะได้ทั้งหมด 26 กล่องจินตนาการ การเคลื่อนย้ายของอนุภาคออกจากกล่องจินตนาการ ระหว่างการจำลองจะขัดเขยด้วยอนุภาคในกล่องจินตนาการ ที่อยู่ตรงกันข้ามเคลื่อนย้ายเข้ามาแทน ทำให้จำนวนอนุภาคของระบบคงที่เสมอ

โปรโตคอลการจำลอง (simulation protocol) ระบบประกอบด้วยโมเลกุลตัวถูกละลายซึ่งอาจจะเป็นโปรตีน หรือโมเลกุลขนาดเล็ก และโมเลกุลตัวทำละลายที่ห้อมล้อมโมเลกุลตัวถูกละลายและกระจายอยู่ทั่วภาชนะซึ่งมีรูปทรงคล้ายกล่องลูกบาศก์ กล่องลูกบาศก์นี้อาจมีความกว้าง ความยาว และความสูง ไม่เท่ากันก็ได้ การใช้วิธีปริมาตรบาวนด์รี คอนดิชันจะช่วยลดข้อผิดพลาดที่เกิดบริเวณรอยต่อ หรือสันของกล่องซึ่งมีค่าให้พลังงานศักย์ของแรงกระทำไม่ถูกต้อง บางโปรแกรม สามารถเลือกการจัดวางโมเลกุลของระบบอาจมีลักษณะเป็นทรงกลม หรือ ให้ล้อมรอบโมเลกุลของตัวถูกละลายเท่านั้น ก็ได้ ก่อนเริ่มการจำลองแบบพลวัตเชิงโมเลกุล ควรปรับพลังงานของระบบให้มีค่าต่ำที่สุด (energy minimization) เพื่อลดหรือกำจัดแรงผลักเนื่องจากอิทธิพลของ steric ระหว่างโมเลกุลในบริเวณบางบริเวณของระบบเสียก่อน ดังแสดงการลดพลังงานในรูปแบบที่ 2.4



รูปที่ 2.4 Energy minimization ของเอนไซม์ (a) ในสุญญากาศ (b) ในน้ำ

ในบางครั้ง เนื่องจากการจำลองด้วยวิธีพลวัตเชิงโมเลกุลของมหโมเลกุลมักจะถูกจำกัดด้วยเวลาของการใช้หน่วยประมวลผลกลาง (CPU) และประสิทธิภาพของฮาร์ดแวร์ การจำลองระบบในสถานะก๊าซ หรือการจำลองระบบให้เสมือนมีโมเลกุลของตัวทำละลายอยู่ในระบบ (implicit solvent) ก็สามารถทำได้ โดยการกำหนดค่าคงที่ไดอิเล็กตริกที่เหมาะสม

เมื่อได้ปรับระบบที่มีพลังงานต่ำที่สุดแล้ว กระบวนการแรก คือ การให้ความร้อน หรือเพิ่มอุณหภูมิให้กับระบบ จนถึงอุณหภูมิที่ต้องการ อุณหภูมิของระบบคำนวณจากผลรวมของพลังงานจลน์ของทุกอะตอมภายในระบบ ดังสมการ

$$\frac{1}{2} \sum m_i v_i^2 = \frac{3}{2} n k_B T$$

อย่างไรก็ตาม ณ  $t = 0$  เราไม่ทราบความเร็ว ดังนั้นความเร็วเริ่มต้นจึงได้จากการกำหนดความเร็วให้กับอะตอมโดยไม่เลือกว่าจะเป็นอะตอมใด แต่จะขึ้นอยู่กับค่าการกระจายของแมกซ์เวลล์ตามอุณหภูมิ เมื่อทราบความเร็วเริ่มต้นแล้ว การจำลองพลวัตเชิงโมเลกุลก็จะสามารถดำเนินได้ด้วยตัวเอง และต่อไปได้อย่างต่อเนื่อง เมื่อระบบถึงอุณหภูมิที่ต้องการแล้ว ควรดำเนินการจำลองต่อไปสักระยะหนึ่งเพื่อให้ระบบเข้าสู่สมดุลและเสถียร กล่าวคือระบบกระจายพลังงานไปทั่วบริเวณ ผลก็คือ มีการเปลี่ยนแปลงของพลังงาน หรือ พารามิเตอร์ทางอุณหพลศาสตร์คงที่ ที่สภาวะสมดุลนี้ พลังงานรวม พลังงานจลน์ พลังงานศักย์ อุณหภูมิ ความดัน ปริมาตร ก่อนข้างคงที่ หรือมีการเปลี่ยนแปลงน้อยมาก ขึ้นถัดไป สิ่งสำคัญในการจำลองคือ ต้องตรวจสอบระบบเสียก่อนว่าอยู่ในสภาวะสมดุลทางพลวัตหรือไม่ และระบบมีความเสถียรมากน้อยเพียงไร เมื่อระบบมีความเสถียรแล้ว จึงจะผลิตข้อมูลทางพลวัต ข้อมูลโคออร์ดิเนตของตำแหน่งอะตอมต่างๆ การผลิตจำนวนข้อมูลจะขึ้นอยู่กับช่วงชั้นเวลาและไทม์สเกล (time scale) ดังที่ได้กล่าวมาแล้วข้างต้น ช่วงชั้นเวลาจะบอกการเปลี่ยนแปลงตำแหน่งในแต่ละขั้นมักใช้ในช่วงเฟมโตวินาที ซึ่งเป็นการสั้นที่เร็วที่สุดที่พบในโครงสร้างของโมเลกุล ช่วงชั้นเวลามีค่าคงที่ตลอดการจำลองพลวัต ส่วนไทม์สเกล บอกถึงเวลาทั้งหมดในการจำลองพลวัต หากเอาค่าไทม์สเกลหารด้วยค่าช่วงชั้นเวลา จะสามารถบอกได้ว่าระบบมีการเปลี่ยนแปลงตำแหน่งหรือมีการผลิตข้อมูลพลวัตไปกี่ครั้ง การเก็บรวบรวมข้อมูลพลวัต เป็นการบันทึกที่ระบบที่เวลาต่างๆ ที่กำหนดไว้ เช่น ข้อมูลจะถูกบันทึกทุกๆ หนึ่งครั้งหลังจากที่จำลองไปครบหนึ่งร้อยช่วงชั้นเวลา การบันทึกในลักษณะนี้ จะได้ข้อมูลเป็นชุดๆ (อาจเรียกว่า configuration) โดยแต่ละชุดจะห่างกันเป็น

ระยะเวลาที่เท่ากัน จำนวนชุดทั้งหมดที่บันทึกได้ เรียกว่า ไดนามิกส์ทราเจกทอรี (dynamics trajectory) โดยทั่วไปแล้ว ชุดข้อมูลทางพลวัต คือ ชุดของโคออร์ดิเนตของอะตอมทุกอะตอมที่เป็นองค์ประกอบของระบบ ซึ่งสามารถนำไปใช้คำนวณคุณสมบัติต่างๆ ของระบบที่สมดุลโดยรวมต่อไป ในการวิจัยนี้ได้ใช้ข้อมูล dynamics trajectory ในสภาวะสมดุล ศึกษาการคำนวณพลังงานยึดจับอิสระในน้ำด้วยวิธี MM-PBSA และ MM-GBSA และศึกษาผลของ side chain จากการทำ alanine scanning โดยมีรายละเอียดโดยคร่าวดังนี้

- **Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA)**

The mutual molecular recognition of proteins with their ligands is the beginning of most of biological processes. Therefore one of the main objectives of structure-based drug design is to be able to reliably and accurately predict the binding affinity of compounds that bind to the target protein. The "lock-and-key" analogy to substrates binding to enzymes like a key fits into a lock was first introduced by Emil Fischer in 1894 [37]. Although this is still a valid analogy since a certain shape complementarity has to be present between the substrate and the protein, it is also clear that there are many other factors that play a part in the ligand binding process. In this work we will only focus on non-covalent binding, which can be viewed as the association of Protein (P) with Ligand (L) to form a Protein-Ligand complex (PL). Figure 2.5 shows a schematic picture of non-covalent ligand binding to a protein, and some of the factors that contribute to protein-ligand binding.

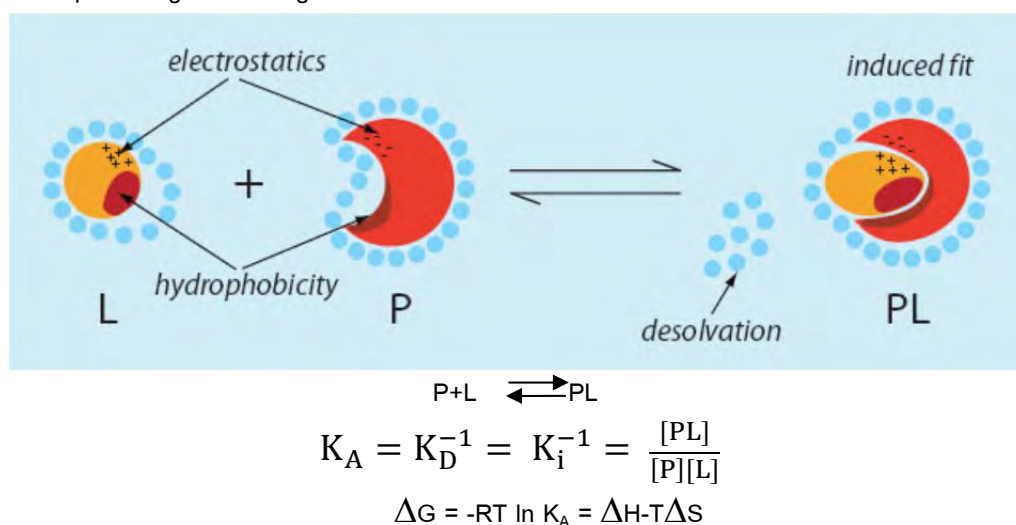


Figure 2.5 A schematic overview of the main factors of protein-ligand binding and how to derive binding energies from them.  $K_A$  is the association constant, and  $K_D$ ,  $K_i$  the dissociation and inhibition constants.  $\Delta G$  is the Gibb's free energy.

It has been suggested that the electrostatic interactions mainly determines molecular recognition and noncovalent binding [38], but this is by no means a general rule. There is equal evidence to the importance of shape complementarity [39]. Molecular recognition can be therefore attributed to contributions of electrostatic and van der Waals interactions, solvation/desolvation and flexibility of ligand and protein. These major contributors to molecular recognition are described briefly below

### *Noncovalent Interactions*

Noncovalent interactions are a number of relatively weak chemical interactions that stabilize the conformations and the interactions between molecules. Hydrogen bonds result from electrostatic attraction between an electronegative atom and hydrogen that is connected to an electronegative atom, which is usually oxygen or nitrogen and less frequently fluorine, or from a *p-p*-interactions, or stacking. This also implies that it is very important for theoretical calculations to have the protonation states of arginine, lysine, aspartic and glutamic acids, as well as histidine correctly determined for an accurate description of electrostatic interactions. Distances of hydrogen bonds are normally 2.5-3.2 Å and angles of 130°-180° are typically found [40]. The strength of a hydrogen bond depends on its directionality and its surroundings. The hydrogen bonds in the interior of proteins are stronger than the ones in the solvent-exposed regions [41]. In addition ionic bonds are very important to ligand-protein binding, but their strength is considerably reduced in water due to shielding.

Water plays an important role in the hydrophobic forces, by forcing hydrophobic groups together it abolishes disruptive effects on the hydrogen bonded network in water, which is also known as the hydrophobic effect. The van der Waals interactions or London dispersion forces are used to model the attractive and repulsive forces between molecules. If two atoms are too close to each other they will repel each other, which makes it possible to define a fixed radius for the "size" of each atom (van der Waals radius). The contact distance between two atoms is then the sum of their van der Waals radii. Van der Waals interactions can be very important when two surfaces of molecules fit well together.

### *Solvation and Desolvation*

Water plays an important role in the formation of protein-ligand complexes. Before a protein-ligand complex is formed, the individual partners that are not a part of hydrophobic surface are involved in hydrogen bonds with the surrounding water. Once the complex is formed, these hydrogen bonds are replaced with hydrogen bonds between the ligand and the protein. The contribution of hydrophobic interactions to protein-ligand binding is normally regarded to be proportional to the size of the hydrophobic surface buried during complex formation [42-43]. Hydrophobic interactions are also regarded to be the main driving force of conformational change of the receptor upon ligand binding [38].

### *Entropy*

The change in the degrees of freedom of the ligand and protein upon binding results in a change of the entropy [44]. It can be viewed as the ligand and protein both losing three degrees of translational and rotational freedom, while six new vibrational degrees of freedom are created for the complex [45].

Estimating binding free energies accurately is a very time-consuming process. The most accurate results can be obtained at a lower computational cost with methods such as MM-PBSA and MM-GBSA. These methods are however still considered too computationally intensive to be of much use in virtual screening approaches.

In MM-PBSA, the binding free energy,  $\Delta G_{\text{bind}}$ , is written as the sum of the gas phase contribution,  $\Delta H_{\text{bind}}^{\text{gas}}$ , the desolvation free energy of the system upon binding,  $\Delta G_{\text{desolv}}$ , and an entropic contribution,  $-T\Delta S$  [46-47] (Figure 2.6):

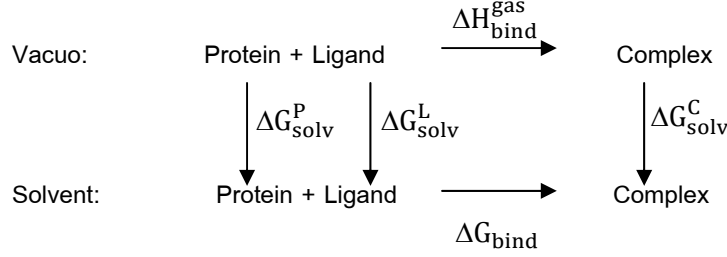


Figure 2.6 The thermodynamic cycle used for MM-PBSA/GBSA calculations

The term  $\Delta H_{\text{bind}}^{\text{gas}}$  contains the van der Waals ( $\Delta H_{\text{vdW}}$ ) and electrostatic ( $\Delta H_{\text{elec}}$ ) interaction energies between the two partners in the complex, and the internal energy variation (including bond, angle, and torsional angle energies) between the complex and the isolated molecules,  $\Delta H_{\text{intra}}$ :

$$\Delta H_{\text{bind}}^{\text{gas}} = \Delta H_{\text{elec}} + \Delta H_{\text{vdW}} + \Delta H_{\text{intra}}$$

$\Delta G_{\text{desolv}}$  is the difference between the solvation free energy,  $\Delta G_{\text{solv}}$ , of the complex and that of the isolated parts.  $\Delta G_{\text{solv}}$  is divided into the electrostatic,  $\Delta G_{\text{elec,solv}}$ , and the nonpolar,  $\Delta G_{\text{np,solv}}$ , contributions,

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec,solv}} + \Delta G_{\text{np,solv}}$$

In MM-PBSA,  $\Delta G_{\text{elec,solv}}$  is calculated by solving the Poisson or the Poisson-Boltzmann (PB) equation [48-49], depending whether the salt concentration is zero or nonzero.

- **Molecular Mechanics Gibbs-Boltzmann Surface Area (MM-GBSA)**

Recently, an approach related to MM-PBSA, where  $\Delta G_{\text{elec,solv}}$  is determined using a Generalized Born (GB) [50] model, has been introduced under the name Molecular Mechanics-Generalized Born Surface Area [51] (MM-GBSA). Despite its approximations, the GB model variant is attractive since it is much faster than PB. Recent advances of GB models, like GB-MV2 [52-53], reproduce more accurately the PB solvation energies of macromolecules as well as desolvation energies upon binding and support the use of GB in this context [54]. The term  $\Delta G_{\text{np,solv}}$ , which can be considered as the sum of a cavity term and a solute-solvent van der Waals term, is assumed to be proportional to the solvent accessible surface area (SASA),

$$\Delta G_{\text{np,solv}} = \gamma \text{SASA} + b$$

This well-known and often used approximation comes from the fact that the  $\Delta G_{\text{solv}}$  of saturated nonpolar hydrocarbons is linearly related to the SASA [55]. Recently, an alternative model for  $\Delta G_{\text{np,solv}}$  using a cavity solvation free energy term plus an explicit solute-solvent van der Waals interactions energy term has been tested. This model led to better results in estimating  $\Delta G_{\text{bind}}$  for the Ras-Raf association, although the transferability of the results was questioned [56]. The entropy term, due to the loss of

degrees of freedom upon association, is decomposed into translational,  $S_{\text{trans}}$ , rotational,  $S_{\text{rot}}$ , and vibrational,  $S_{\text{vib}}$ , contributions. These terms are calculated using standard equations of statistical mechanics [57]. The vibrational entropy term is calculated with the quantum formula from a normal mode analysis (NMA) [57]. A quasiharmonic analysis of the MD simulations is also possible. However, it has been found that it does not always yield convergent values, even using very long MD simulation trajectories, and also led to large deviations from the results obtained with NMA, giving an overall unreasonable entropic contribution [56].

- Computational Alanine-Scanning

MM-PB(GB)SA has been found to perform well at determining the effect of mutations on association processes, and identifying the hot-spots of protein-protein complexes [51, 56, 58-61]. Two main approaches exist. First, it is possible to perform the so-called computational alanine scanning (CAS) [58-59] in which the absolute  $\Delta G_{\text{bind}}$  is calculated for the wild type system, as well as for several mutants in which one residue has been replaced by an alanine. The alanine mutation is introduced by modifying the frames extracted from the MD simulation of the wild-type system. The difference in  $\Delta G_{\text{bind}}$  between the wild-type system and the mutants may be compared directly to the results of an experimental alanine scanning (AS) [58-59]. The second possibility is to perform a binding free energy decomposition (BFED) [51] for the wild-type system. This process aims at calculating the contributions to  $\Delta G_{\text{bind}}$  arising from each atom or groups of atoms (typically side chains). Like CAS, the BFED also identifies the nature of the energy change in terms of interaction and solvation energies, or entropic contributions. The detailed description of the BFED process can be found elsewhere [51, 60] and is summarized in the appendix of the present Review Article. The MM-GBSA variant is attractive for BFED, not only because it is much faster than MM-PBSA, but also because the pair-wise nature of the GB equation allows the decomposition of  $\Delta G_{\text{elec,solv}}$  into atomic contributions in a straightforward manner [51]. It is however interesting to note that the decomposition of a PB calculated  $\Delta G_{\text{elec,solv}}$  can also be performed [61], though it is more computationally demanding. Although its results cannot be compared directly to an experimental AS, the BFED offers a faster alternative to the CAS, since it only requires one binding free energy calculation. Also, it allows studying the contributions from nonmutable groups of atoms, such as backbone atoms. In addition, contrary to CAS, the BFED is a nonperturbing approach that does not require introducing a mutation in the system. A comparison between CAS and BFED results can be found in reference [60]. Obviously, these methods cannot be expected to provide results exactly comparable to values obtained from an experimental AS, since they both neglect the effect of the mutations on the protein conformation. However, fair agreements between the experimental and theoretical results have been found in several studies and open the way to rational protein engineering [51, 58, 61].

#### 4. การทดลองในการศึกษา antibody-antigen affinity

วิธีที่ใช้ในห้องปฏิบัติการภาควิชาภูมิคุ้มกันวิทยาคลินิก คณะเทคนิคการแพทย์ มหาวิทยาลัยเชียงใหม่ใช้เทคนิค Enzyme-linked immunosorbent assay (ELISA) ซึ่งแตกต่างกันสองวิธีคือ แบบ Indirect ELISA และแบบ Competitive ELISA โดยขอก้าวโดยคร่าว และแสดงภาพการทดลองอย่างง่าย ดังรูปดังนี้

### a) Indirect ELISA

In direct ELISA, an antigen is adsorbed to a plastic plate, then an excess of other protein (using bovine serum albumin) is added to block all the other binding sites. Then an enzyme is linked to an antibody in a separate reaction. The enzyme-antibody conjugate is allowed to adsorb to the antigen, then excess enzyme-antibody conjugate is washed off, leaving enzyme-antibody bound to antigen. The enzyme is detected by adding the enzyme's substrate. In an indirect ELISA, the steps are similar, but with important differences and an additional step. After the antigen is adsorbed to the plate (and after the BSA step), the next antibody to be added is the antibody that recognizes the antigen (this antibody does not have the enzyme attached to it). Then, an enzyme-antibody conjugate is prepared, which is added to the plate and detects the antibody that is adsorbed to the antigen (in a direct ELISA, the enzyme-antibody conjugate directly adsorbs to the antigen), then the substrate is added which detects the presence of the enzyme and thus the antigen.

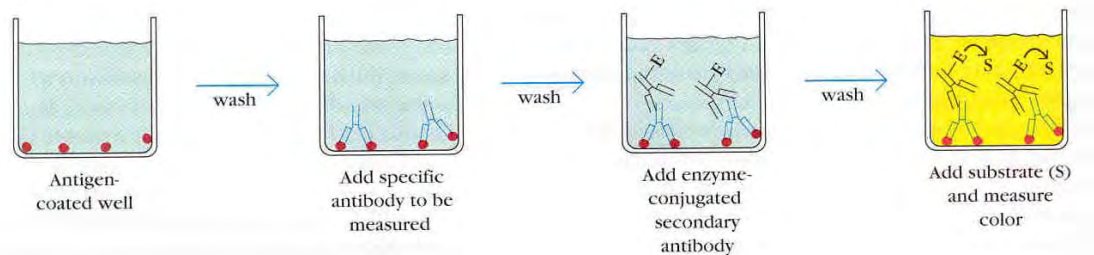


Figure 2.6 Indirect ELISA

### b) Competitive ELISA

The use of ELISA is through competitive binding with the following steps which are different than the indirect ELISA:

1. Unlabeled antibody is incubated in the presence of its antigen (Sample).
2. These bound antibody/antigen complexes are then added to an antigen-coated well.
3. The plate is washed, so that unbound antibody is removed. (The more antigen in the sample, the less antibody will be able to bind to the antigen in the well, hence "competition.")
4. The secondary antibody, specific to the primary antibody is added. This second antibody is coupled to the enzyme.
5. A substrate is added, and remaining enzymes elicit a chromogenic or fluorescent signal.

For competitive ELISA, the higher the sample antigen concentration, the weaker the eventual signal. The major advantage of a competitive ELISA is the ability to use crude or impure samples and still selectively bind any antigen that may be present.



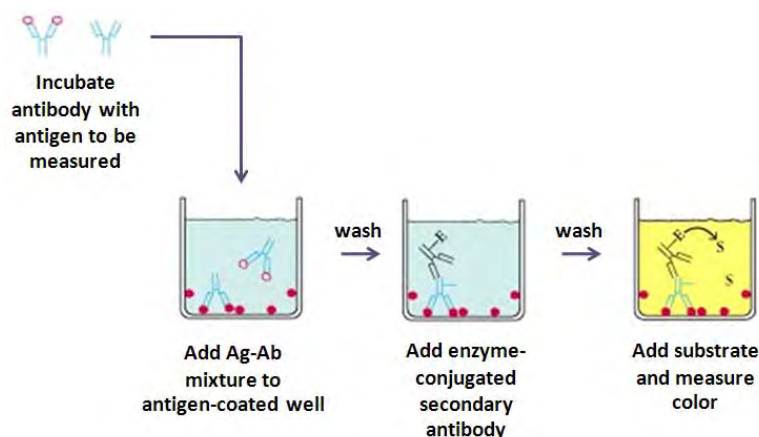


Figure 2.7 Competitive ELISA

## เอกสารอ้างอิง

1. Winter, G. A. D., Griffiths, R. E., Hoogenboom, H. R. (1994) *Annu. Rev. Immunol.* 12, 433-55.
2. Vaughan, T. J., Williams, K., Pritchard, J. K., Osbourn, A. R., Pope, J. C., Earnshaw, J., McCafferty, R. A., Hodits, J., Johnson, K. S. (1996) *Nat. Biotechnol.*, 14, 309-14.
3. Rees, A. R., Staunton, D., Webster, D. M., Searle, S. J., Henry, A. H., Pedersen, J. T. (1994). Antibody design: beyond the natural limits. *Trends Biotechnol.* 12, 188-206.
4. Webster, D. M., Henry, A. H. & Rees, A. R. (1994). Antibody-antigen interactions. *Curr. Opin. Struct. Biol.* 4, 123-129.
5. Wilson, I. A., Stanfield, R. L. (1993). Antibody-antigen interactions. *Curr. Opin. Struct. Biol.* 3, 113-118.
6. Wison, I. A., Stanfield, R. L. (1994). Antibody-antigen interactions – new structures and new formational-changes. . *Curr. Opin. Struct. Biol.* 4, 857-867.
7. Arevalo, J. H., Hassig, C. A., Stura, E. A., Sims, M. J., Taussig, M. J., Wilson, I. A. (1994). Structural-analysis of antibody specificity – detailed comparison of 5 Fab'-steroid complexes. *J. Mol. Biol.* 241, 663-690.
8. Wong, Y. W., Kussie, P. H., Parhamiseren, B., Margolies, M. N. (1995). Modulation of antibody-affinity by an engineered amino-acid substitution. *J. Immunol.* 154, 3351-3358.
9. Nancias, M., Chinchio, M., Oldziej, S., Czaplewski, C., Scheraga, H.A. (2005). Protein structure prediction with the UNRES force-field using Replica-Exchange Monte Carlo-with-Minimization; comparison with MCM, CSA, and CFMC. *J. Comput. Chem.* 26, 1472–1486.
10. Bradley, P., Misura, K.M.S., Baker, D. (2005). Toward high-resolution de novo structure prediction for small proteins. *Science* 309, 1868–1871.
11. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The protein data bank. *Nucleic Acids Res* 28(1), 235–42 (2000).
12. Tramontano, A. and Morea, V. Assessment of homology-based predictions in casp5. *Proteins* 53 Suppl 6, 352–68 (2003).

13. Pearson, W. R. and Lipman, D. J. Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85(8), 2444–8 (1988).
14. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res* 25(17), 3389–402 (1997).
15. Barton, G. J. Protein multiple sequence alignment and flexible pattern matching. *Methods Enzymol* 183, 403–28 (1990).
16. Suyama, M., Matsuo, Y., and Nishikawa, K. Comparison of protein structures using 3d profile alignment. *J Mol Evol* 44 Suppl 1, S163–73 (1997).
17. Notredame, C., Higgins, D. G., and Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302(1), 205–17 (2000).
18. Rost, B. Twilight zone of protein sequence alignments. *Protein Eng* 12(2), 85–94 (1999).
19. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. Swiss-model: An automated protein homology-modeling server. *Nucleic Acids Res* 31(13), 3381–5 (2003).
20. Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. The swiss-model workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22(2), 195–201 (2006).
21. Johnson, M. S., Srinivasan, N., Sowdhamini, R., and Blundell, T. L. Knowledgebased protein modeling. *Crit Rev Biochem Mol Biol* 29(1), 1–68 (1994).
22. Sali, A. and Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3), 779–815 (1993).
23. Ponder, J. W. and Richards, F. M. Tertiary templates for proteins. use of packing criteria in the enumeration of allowed sequences for different structural classes. *J Mol Biol* 193(4), 775–91 (1987).
24. Dunbrack, R. L., J. and Karplus, M. Backbone-dependent rotamer library for proteins. application to side-chain prediction. *J Mol Biol* 230(2), 543–74 (1993).
25. Fiser, A., Do, R. K., and Sali, A. Modeling of loops in protein structures. *Protein Sci* 9(9), 1753–73 (2000).
26. Rohl, C. A., Strauss, C. E., Chivian, D., and Baker, D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins* 55(3), 656–77 (2004).
27. Wojcik, J., Mornon, J. P., and Chomilier, J. New efficient statistical sequencedependent structure prediction of short to medium-sized protein loops based on an exhaustive loop classification. *J Mol Biol* 289(5), 1469–90 (1999).
28. Chothia, C. and Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196(4), 901–17 (1987).
29. Sanchez, R. and Sali, A. Evaluation of comparative protein structure modeling by modeller-3. *Proteins Suppl* 1, 50–8 (1997).
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283–291 (1993).
31. Vriend, G. What if: a molecular modeling and drug design program. *J Mol Graph* 8(1), 52–6, 29 (1990).

32. Eisenberg, D., Luthy, R., and Bowie, J. U. Verify3d: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277, 396–404 (1997).
33. Melo, F. and Feytmans, E. Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol* 277(5), 1141–52 (1998).
34. BioMedCACHe, *CACHe Release Notes, Release 6.1*, CACHe for Window2000/Xp, Inc: Fujitsu Limited.
35. CACHe group, “Getting started with BioMedCACHe: molecular modeling in drug design”, Fujitsu Limited, 2004.
36. Muegge, I.; Martin, Y. C., A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 1999, 42, (5), 791-804.
37. Fischer, E. Ber. Dtsch. Chem. Ges 27, 2985–2993 (1894).
38. Gohlke, H. and Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie-International Edition* 41(15), 2645–2676 (2002).
39. Homans, S. W. Water, water everywhere—except where it matters? *Drug discovery today* 12(13-14), 534–9 (2007).
40. Jeffrey, G. and Saenger, W. Hydrogen bonding in biological structures. Springer, Berlin, (1991).
41. Stahl, M. and Bohm, H. J. Development of filter functions for protein-ligand docking. *Journal of Molecular Graphics & Modelling* 16 (3), 121–132 (1998).
42. Chothia, C. Hydrophobic bonding and accessible surface area in proteins. *Nature* 248(446), 338–9 (1974).
43. Wang, J. M., Morin, P., Wang, W., and Kollman, P. A. Use of mm-pbsa in reproducing the binding free energies to hiv-1 rt of tibo derivatives and predicting the binding mode to hiv-1 rt of efavirenz by docking and mm-pbsa. *Journal of the American Chemical Society* 123(22), 5221–5230 (2001).
44. Brady, G. P. and Sharp, K. A. Entropy in protein folding and in protein-protein interactions. *Curr Opin Struct Biol* 7(2), 215–21 (1997).
45. Page, M. I. and Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci USA* 68(8), 1678–83 (1971).
46. Srinivasan J, Miller J, Kollman PA, Case DA. 1998. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 16: 671–682.
47. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham T Er. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33: 889–897.
48. Gilson MK, Honig B. 1988a. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* 4: 7–18.
49. Gilson MK, Honig BH. 1988b. Energetics of charge-charge interactions in proteins. *Proteins* 3: 32–52.
50. Still WC, Tempczyk A, Hawley RC, Hendrickson T. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112: 6127–6129.

51. Gohlke H, Kiel C, Case DA. 2003. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* 330: 891–913.
52. Lee MS, Salsbury FR, Jr Brooks CL, III. 2002. Novel generalized Born methods. *J. Chem. Phys.* 116: 10606–10614.
53. Lee MS, Feig M, Salsbury FR, Jr Brooks CL, III. 2003. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *J. Comput. Chem.* 24: 1348–1356.
54. Zoete V, Meuwly M, Karplus M. 2005. Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition. *Proteins* 61: 79–93.
55. Amidon GL, Yalkowsky SH, Anik ST, Valvani SC. 1975. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach. *J. Phys. Chem.* 79: 2239–2246.
56. Gohlke H, Kuhn LA, Case DA. 2004. Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* 56: 322–337.
57. Tidor B, Karplus M. 1994. The contribution of vibrational entropy to molecular association. *J. Mol. Biol.* 238: 405–414.
58. Massova I, Kollman PA. 1999. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc.* 121: 8133–8143.
59. Huo S, Massova I, Kollman PA. 2002. Computational alanine scanning of the 1: 1 human growth hormone-receptor complex. *J. Comp. Chem.* 23: 15–27.
60. Zoete V, Michielin O. 2007. Comparison between computational alanine scanning and per-residue binding free energy decomposition for protein-protein association using MM-GBSA: application to the TCRp-MHC complex. *Proteins* 67: 1026–1047.
61. Lafont V, Schaefer M, Stote RH, Altschuh D, Dejaegere A. 2007. Protein-protein recognition and interaction hot spots in an antigen-antibody complex: free energy decomposition identifies “efficient amino acids.” *Proteins* 67: 418–434.

### 3. ผลงานวิจัย และวิจารณ์ผล

ประกอบด้วยงานในส่วนต่าง ๆ ดังนี้

- 3.1 Homology modeling ของ scFv anti-p17  
รายละเอียดเพิ่มเติมใน paper ที่ตีพิมพ์ใน Molecular Immunology 2009
- 3.2 Molecular docking nine peptide epitopes into the binding site of scFv anti-p17  
รายละเอียดเพิ่มเติมใน paper ที่ตีพิมพ์ใน Molecular Immunology 2009
- 3.3 Molecular dynamics simulations (MD) of scFv anti-p17 - peptide epitopes complexes: Binding free energies and decomposition energy  
รายละเอียดเพิ่มเติมใน paper ที่ตีพิมพ์ใน Molecular Immunology 2009
- 3.4 Molecular dynamics simulations (MD) of scFv anti-p17 - peptide epitopes complexes: Computational Alanine-Scanning of an amino acids in the complementary determining regions (CDRs) of scFv anti-p17  
รายละเอียดเพิ่มเติมใน Manuscript "Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning"
- 3.5 Amino acids mutation of scFv anti-p17 for improving the specificity of binding affinity  
รายละเอียดเพิ่มเติมใน Manuscript "Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning"
- 3.6 Competitive ELISA and Indirect ELISA  
รายละเอียดเพิ่มเติมใน Manuscript "Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning"
- 3.7 การปรับปรุง scFv p17 ด้วยการ grafting บน SS-bond free scFvs  
รายละเอียดเพิ่มเติมใน Manuscript "Stabilization of Anti-Ras scFv fragments with/without disulfide bond through theoretical modeling and molecular dynamics simulations investigation"
- 3.8 การปรับปรุง scFv anti-p17 ด้วยการ grafting CDRs ของ scFv anti-p17 บน scFv ที่ละลายได้ดี (scFv13-R4)
- 3.9 การสร้างโครงสร้าง scFv G46 และศึกษาการยึดจับกับ gag-pol sequence
- 3.10 การศึกษาอันตรกิริยาระหว่าง antibody-CD147  
รายละเอียดเพิ่มเติมใน Manuscript "Modeling of CD147 binding sites against ScFv: a guide for epitope mapping"

#### 3.1 Homology modeling ของ scFv anti-p17

การสร้างโครงสร้าง scFv anti-p17 เริ่มต้นได้ใช้โปรแกรม modeller ใน Discovery studio 2.0 ในการสร้างโครงสร้าง antibody และใช้วิธี molecular docking ด้วย Cdocker ในการสร้างโครงสร้างเชิงซ้อน antibody-antigen และศึกษาการยึดจับในน้ำ โดยลำดับกรดอะมิโนเริ่มต้นได้มาจาก Tewari et al., 1998 พบว่าส่วน VH คล้ายกับ VH of idiotype-anti-idiotype Fab complex (pdb id:1iai), sharing 82% sequence identity ในขณะที่ส่วน VL เหมือนกับ VL of the Fab fragment of neutralizing antibody directed against an epitope of gp41 from HIV-1 (pdb id:1nld), sharing 96% of sequence identity และการวางตัวของ VH และ VL ใช้วิธี superposition กับ a crystal structure of the anti-DNA binding antibody (pdb id:2gki) จากนั้นทำการ minimize โครงสร้าง 500

steps of steepest descent ตามด้วย 500 steps of conjugate gradient จนได้ convergence criterion of 0.05 kcal/mol/Å ใช้ AMBER03 force field (Case et al., 2006) ตรวจสอบโครงสร้างด้วย PROCHECK พบว่าหมู่อะมิโนมากกว่า 92% อยู่ในบริเวณที่เหมาะสมของ Ramachandran

Plot โครงสร้าง scFv anti-p17 ประกอบด้วย six unique hypervariable units หรือ complementary determining regions (CDRs) ที่ใช้เป็น binding groove สำหรับยึดจับกับแอนติเจน ได้แก่ loops ในส่วน heavy chain H1(31–35), H2(50–66) และ H3(99–104) และอีก 3 loops ในส่วน light chain คือ L1(154–169), L2(185–190), and L3(224–232) ดังรูปที่ 3.1



รูปที่ 3.1 โมเดล scFv anti-p17. (a) 3D structural model for scFv anti-p17. VH and VL domains are colored white. The hypervariable binding loops are colored green (H1), pink (H2), cyan (H3), yellow (L1), purple (L2), and orange (L3). (b) Amino acid sequences of the hypervariable binding VH (1–114) and VL (131–241) domains of scFv anti-p17 and an artificial linker shown in bold (GGGGS)<sub>3</sub>; the loops in the heavy chain (CDRH1: 31–35, CDRH2: 50–66, and CDRH3: 99–104), and the loops in the light chain (CDRL1: 154–169, CDRL2: 185–190, and CDRL3: 224–232) of the scFv antibody.

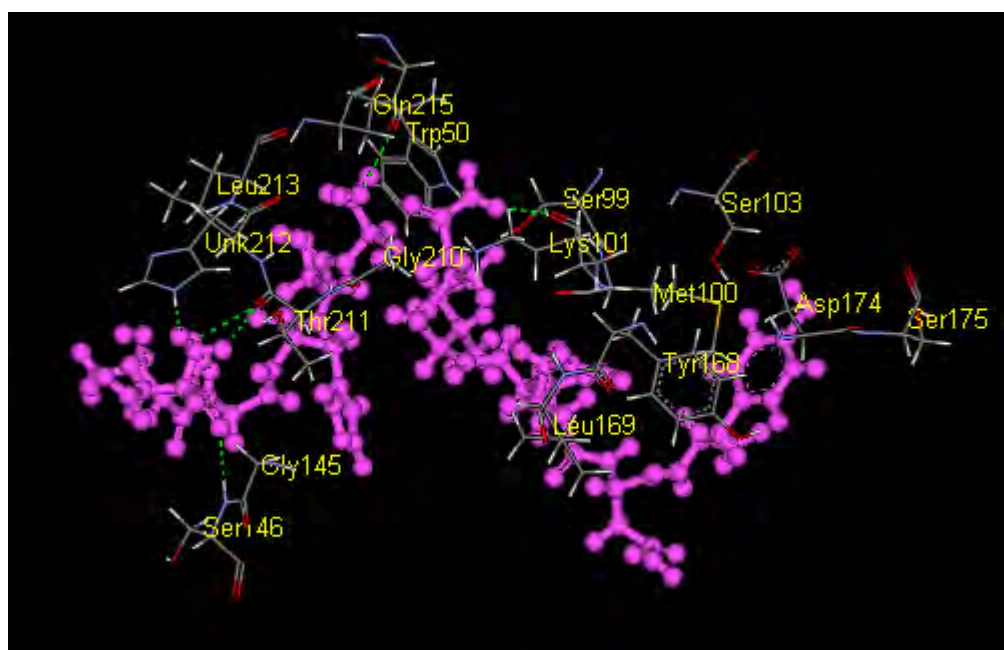
### 3.2 Molecular docking nine peptide epitopes into the binding site of scFv anti-p17

การออกแบบโครงสร้าง antibody-antigen และการศึกษาอันตรกิริยาเฉพาะที่ระหว่างโมเลกุล โดยประเมนหมู่อะมิโนที่มีส่วนสำคัญในการยึดจับของ single-chain antibody fragment (scFv) กับ HIV-1 p 17 โดยการสร้างโครงสร้างสารเชิงซ้อนของ scfv anti-p17 กับ c-terminal HIV-p17 wild type และมิวแทนท์อีก 8 ชนิดที่พบในธรรมชาติ สร้างโครงสร้างเปปไทด์บางส่วนจากโครงสร้างเอกซเรย์ pdb id:1kj4 และสร้างโครงสร้างยึดจับด้วยการทำ molecular docking ด้วยโปรแกรม BioMedCaChe 2.0 (Fujitsu, Inc.) program โดยใช้ CDR loops เป็นบริเวณยึดจับ ทั้งเปปไทด์และบริเวณ CDR loops สามารถยับยั้งได้ขณะจำลองการยึดจับ และประเมินพลังงานของโครงสร้างเชิงซ้อนด้วย potential of mean force (PMF) และใช้ genetic algorithm with a population size of 50, crossover rate of 0.80, elitism of 5, mutation rate of 0.2, and the maximum cycle generation was set to be 40,000 ขนาดของ grid box = 30 x 30 x 30 Å โดยได้ผลดังตารางที่ 1

ตารางที่ 3.1 ลำดับเปปไทด์, PMF score และหมู่อะมิโนบริเวณรอบอพิโทป

Peptide names	Peptide sequences	PMF score (kcal/mol)	Amino Acid in 4.5 Å from peptide epitope	
			Nonpolar hydrophobic	Polar hydrophilic
p17.1	<sup>121</sup> DTGHSSQVSQNY <sup>132</sup>	-902.11	GLY33 (H1), TRP50 (H2), MET100 (H3), GLY226 (L3), LEU229 (L3)	SER99 (H3), SER103 (H3), ASP163 (L1), ASP190 (L2), THR227 (L3)
p17.2	<sup>121</sup> DTGH <u>SN</u> QVSQNY <sup>132</sup>	-899.18	TRP50 (H2), MET100 (H3), GLY161 (L1), IEU185 (L2), GLY226 (L3), LEU229 (L3)	THR59 (H2), SER99 (H3), LYS101 (H3), SER103 (H3), SER162 (L1), ASP163 (L1), THR227 (L3), HIS228 (L3)
p17.3	<sup>121</sup> DTGHSSQVSQNY <sup>132</sup>	-882.65	TRP50 (H2), MET100 (H3), PHE167 (L1), GLY226 (L3), LEU229 (L3)	SER99 (H3), LYS101 (H3), ASP163 (L1), TYR184, ASP190 (L2), SER191, THR227 (L3), GLN231 (L3)
p17.4	<sup>121</sup> DTGH <u>NS</u> QVSQNY <sup>132</sup>	-898.71	GLY33 (H1), MET100 (H3), GLY226 (L3), LEU229 (L3)	ASP31 (H1), TYR32 (H1), ASN52 (H2), SER99 (H3), LYS101 (H3), HIS228 (L3), GLN231 (L3)
p17.5	<sup>121</sup> <u>NT</u> GHSSQVSQNY <sup>132</sup>	-843.51	TRP50 (H2), MET100 (H3), PHE167 (L1), IEU185 (L2), LEU229 (L3)	ASN35 (H1), SER99 (H3), LYS101 (H3), ASN169 (L1), TYR184, ASP190 (L2), SER191, THR227 (L3), HIS228 (L3)
p17.6	<sup>121</sup> DTG <u>NS</u> SQVSQNY <sup>132</sup>	-846.12	GLY33 (H1), TRP50 (H2), MET100 (H3), LEU229 (L3)	SER99 (H3), LYS101 (H3), LYS165 (L1), TYR184, SER191
p17.7	<sup>121</sup> DTGHSSQ <u>AS</u> QNY <sup>132</sup>	-829.94	TRP50 (H2), MET100 (H3), PHE167 (L1), GLY226 (L3)	THR59 (H2), LYS101 (H3), LYS165 (L1), LYS188 (L2), THR227 (L3), HIS228 (L3)
p17.8	<sup>121</sup> DTGH <u>SK</u> QVSQNY <sup>132</sup>	-926.79	TRP50 (H2), MET100 (H3), PHE167 (L1), GLY226 (L3), LEU229 (L3)	THR59 (H2), SER99 (H3), LYS101 (H3), SER103 (H3), SER162 (L1), ASP163 (L1), THR227 (L3)
p17.9	<sup>121</sup> DTG <u>NN</u> SQVSQNY <sup>132</sup>	-841.02	TRP50 (H2), MET100 (H3), PHE167 (L1), IEU185, LEU229 (L3)	THR59 (H2), SER99 (H3), LYS101 (H3), LYS165, ASN169 (L1), TYR184, ASP190 (L2), SER191, HIS228 (L3)

The underlined letters are the mutated residue in each sequence compared to the wild type.



รูปที่ 3.2 กรดอะมิโนในบริเวณยึดจับของ scfv ใน 4.5 Å จาก HIV epitope original substrate (DTGHSSQVSQNY)

● ผลของค่าพลังงานการยึดจับของแต่ละ CDR loops

การคำนวณ Residue interaction โดยไม่คิดอิทธิพลของน้ำ และใช้โครงสร้างเชิงซ้อน scfv anti-p17 กับ c-terminal HIV-p17 จาก BioMedCACHe และ PMF scoring และศึกษาอันตรกิริยาระหว่างโมเลกุล ระหว่างหมู่อะมิโน Residue interaction เป็นการคำนวณพลังงานที่ได้จากการคำนวณควอนตัม หรือกลศาสตร์เชิงโมเลกุล และเป็นการคำนวณในสถานะแก๊ส คำนวณได้จาก AB-Antigen เป็นโครงสร้างเชิงซ้อนในบริเวณยึดจับที่สนใจ เช่น หมู่อะมิโนใน complementary region โดยคำนวณ interaction ของหมู่อะมิโนที่ละตัวในบริเวณยึดจับ คำนวณได้จาก

$$E_{\text{res1-Antigen}} = E_{\text{res1-Antigen}} - E_{\text{res1}} - E_{\text{Antigen}}$$

Res1= หมู่อะมิโนเฉพาะที่สนใจในบริเวณยึดจับ

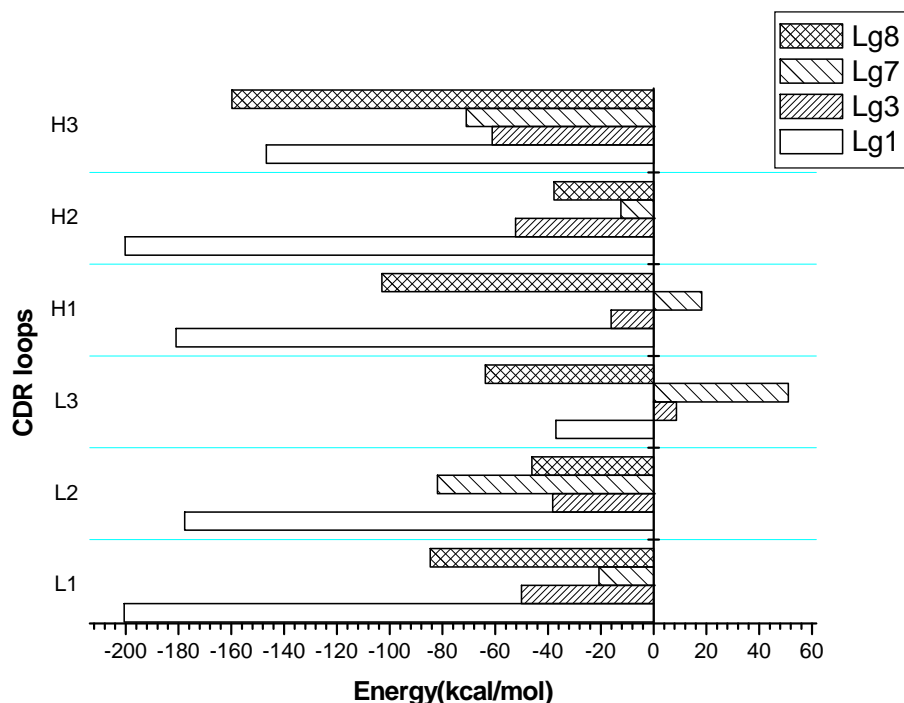
ในการศึกษาค่าพลังงานการยึดจับของ CDR loops ได้ทำการศึกษาเฉพาะ CDR loops ที่ยึดจับกับลิแกนด์ p17.1 17.3 17.7 p17.8 โดย p17.1 เป็นลิแกนด์ตั้งต้น ลิแกนด์ที่ 3 มีค่าพลังงานการยึดจับใกล้เคียงกับ ลิแกนด์ตั้งต้น ลิแกนด์ที่ 7 มีค่าพลังงานการยึดจับน้อยที่สุด และลิแกนด์ที่ 8 มีค่าพลังงานการยึดจับมากที่สุด ซึ่งผลการทดลองพบว่า ลิแกนด์แต่ละชนิดมีการยึดจับกับแต่ละ CDR loops ต่างกับของ wild type p17.1ค่อนข้างมาก โดยบริเวณที่ยึดจับไม่ดีในกรณีของ p17.7 น่าจะเป็นผลมาจากการยึดจับที่ไม่ดีกับ CDR loops L3 และ H1 โดยแสดงค่าการยึดจับเป็นบวกดังรูป และตาราง อย่างไรก็ตามการคำนวณพลังงานยึดจับในน้ำน่าจะให้ค่าอันตรกิริยาที่ถูกต้องมากกว่า

ตารางที่ 3.2 ค่าพลังงานการยึดจับของ CDR loops ที่ยึดจับกับลิแกนด์ที่ 1 3 7 และ 8

Ligand	CDR loops	$E_{\text{CDX}}$	$E_{\text{CDR loop}}$	$E_{\text{CDR loop ที่เหลือ + Lg}}$	$\Delta E$ kcal/mol	PMF score
Lg1	L1	-1520.9452	-128.92753	-1191.4655	-200.55214	-902.11
	L2	-1520.9452	-464.44398	-878.7993	-177.70191	
	L3	-1520.9452	-57.62006	-1426.246	-37.07913	
	H1	-1520.9452	-371.67571	-968.25535	-181.01413	
	H2	-1520.9452	-99.35067	-1221.3816	-200.21292	
	H3	-1520.9452	-103.69674	-1270.5459	-146.70252	
Lg 3	L1	-1581.7045	-128.92753	-1402.7213	-50.05568	-899.18
	L2	-1581.7045	-464.44398	-1079.0834	-38.17713	
	L3	-1581.7045	-57.62006	-1532.8045	8.72	
	H1	-1581.7045	-371.67571	-1193.9575	-16.07129	
	H2	-1581.7045	-99.35067	-1430.029	-52.32484	
	H3	-1581.7045	-103.69674	-1416.8688	-61.13896	
Lg7	L1	-1463.0242	-128.92753	-1313.4605	-20.63616	-829.94
	L2	-1463.0242	-464.44398	-916.66548	-81.9147	
	L3	-1463.0242	-57.62006	-1456.4604	51.05631	
	H1	-1463.0242	-371.67571	-1109.6534	18.30498	
	H2	-1463.0242	-99.35067	-1351.3159	-12.35759	
	H3	-1463.0242	-103.69674	-1288.3425	-70.98497	
Lg8	L1	-1705.3973	-128.92753	-1491.8571	-84.61269	
	L2	-1705.3973	-464.44398	-1194.7615	-46.19187	



L3	-1705.3973	-57.62006	-1584.003	-63.77433	-926.793
H1	-1705.3973	-371.67571	-1230.7455	-102.97611	
H2	-1705.3973	-99.35067	-1568.3251	-37.72156	
H3	-1705.3973	-103.69674	-1441.9818	-159.71884	



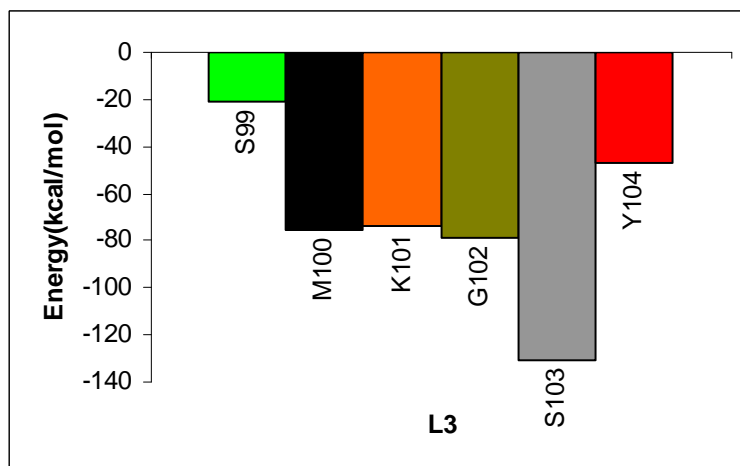
รูปที่ 3.3 กราฟแสดงค่าพลังงานการยึดจับของ CDR loops ต่างๆ

- ผลของค่าพลังงานการยึดจับของกรดอะมิโนแต่ละตัวใน CDR loops

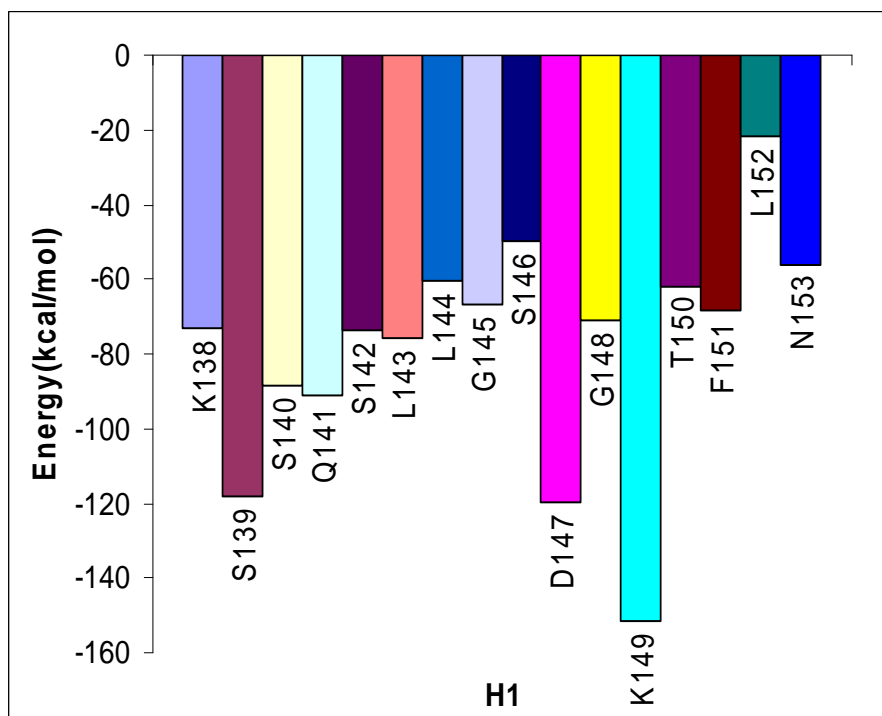
จากนั้นทำการศึกษาค่าพลังงานการยึดจับของกรดอะมิโนใน CDR loops กรดอะมิโนที่ทำการศึกษาเป็นกรดอะมิโนใน CDR loop L3 และ H1 ที่ยึดจับกับลิแกนด์ที่ 7 จากการศึกษาค่าพลังงานการยึดจับของ CDR loops จะเห็นว่า CDR loops L3 และ H1 มีค่าพลังงานการยึดจับกับลิแกนด์ที่ 7 ไม่ดี ดังนั้นจึงได้ทำการศึกษาค่าพลังงานการยึดจับของกรดอะมิโนเพื่อปรับปรุงการยึดจับของกรดอะมิโนใน CDR loops ดังกล่าว ให้สามารถยึดจับกับลิแกนด์ที่ 7 ได้ดีขึ้น ในการหาค่าพลังงานการยึดจับของกรดอะมิโนใน CDR loops L3 และ H1 ที่ยึดจับลิแกนด์ที่ 7 เมื่อเปรียบเทียบกับหมู่อะมิโนอื่นพบว่า Ser ตัวที่ 99 ใน CDR loops L3 มีการยึดจับไม่ดี มีค่าพลังงานประมาณ -21.34 kcal/mol ดังตาราง

ตารางที่ 3.3 ค่าพลังงานการยึดจับของกรดอะมิโนใน CDR loops L3 และ H1 ที่ยึดจับลิแกนด์ที่ 7

CDR loops	Amino acid of CDR loop	$E_{\text{cpx}}$	$E_{\text{amino acid}}$	$E_{\text{amino acid ที่เหลือ}} + Lg$	$\Delta E$ kcal/mol
L3	S99	-225.72707	24.42592	-228.80918	-21.34381
	M100	-225.72707	20.67997	-171.32588	-75.08116
	K101	-225.72707	70.9755	-222.59417	-74.1084
	G102	-225.72707	14.18032	-161.37302	-78.53437
	S103	-225.72707	28.81289	-123.66832	-130.87164
	Y104	-225.72707	10.57596	-189.10717	-47.19586
H1	K138	-562.67268	65.42807	-554.96922	-73.13153
	S139	-562.67268	30.65557	-474.95398	-118.37427
	S140	-562.67268	25.45731	-499.69269	-88.4373
	Q141	-562.67268	-12.55039	-459.09992	-91.02237
	S142	-562.67268	23.02384	-511.85764	-73.83888
	L143	-562.67268	22.26286	-509.11123	-75.82431
	L144	-562.67268	16.08759	-518.2999	-60.46037
	G145	-562.67268	9.12748	-505.254	-66.54616
	S146	-562.67268	33.48619	-546.60939	-49.54948
	D147	-562.67268	-30.76463	-412.30372	-119.60433
	G148	-562.67268	12.87716	-504.4024	-71.14744
	K149	-562.67268	54.70396	-465.66376	-151.71288
	T150	-562.67268	24.69047	-525.32765	-62.0355
	F151	-562.67268	17.43465	-511.99579	-68.11154
	L152	-562.67268	21.75446	-515.58566	-68.84148
	N153	-562.67268	32.74438	-539.08052	-56.33654



รูปที่ 3.4 ค่าพลังงานการยึดจับของกรดอะมิโนแต่ละตัวใน CDR loops L3



รูปที่ 3.5 ค่าพลังงานการยึดจับของกรดอะมิโนแต่ละตัวใน CDR loops H1

### 3.3 Molecular dynamics simulations (MD) of scFv anti-p17 - peptide epitopes complexes Binding free energies and decomposition energy

การจำลองพลวัตโดยใช้กรอบทั้งโครงสร้างสารเชิงซ้อน 20 Å และหา binding free energy โดยใช้เทคนิค MM-PBSA และคำนวณ decomposition energy และศึกษาอันตรกิริยาที่เกิดขึ้นในสถานะที่มีน้ำ ความเป็นกรดต่างเป็นกลาง จากการจำลองพลศาสตร์ในน้ำเป็นเวลา 2 ns และการคำนวณ binding free energy โดยใช้ Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) protocol คำนวณ pairwise decomposition or residue-based energy ได้ผลดังตารางที่ 3.4

การคำนวณพลังงานการยึดจับระหว่างโครงสร้างเชิงซ้อนของ scFv anti-p17 กับ peptide p17.1, p17.3, p17.7 และ p17.8 จะวิเคราะห์โดยใช้เทคนิค MMPBSA (Molecular Mechanics Poisson Boltzmann Surface Area) และเมื่อเปรียบเทียบประสิทธิภาพการยึดเหนี่ยวจากอันตรกิริยาที่เกิดขึ้นของโมเลกุลเชิงซ้อนระหว่าง scFv anti-p17 และ peptide epitopes ในตัวทำละลายของข้อมูลจากการตรวจด้วยวิธี ELISA และ Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) พบว่า ถ้ามี ค่า  $PB_{TOT}$  ติดลบมากๆ แสดงว่ามีประสิทธิภาพการยึดเหนี่ยวมาก โดยเรียงลำดับค่า  $PB_{TOT}$  ได้ดังนี้  $p17.1 > p17.8 > p17.3 > p17.7$  ซึ่งมีค่าเท่ากับ -29.98, -26.38, -8.82 และ -8.77 kcal/mol ตามลำดับ ซึ่งสอดคล้องกับค่า PMF ที่ได้จากวิธีการ Molecular Docking จากข้อมูลข้างต้นสามารถแบ่งแอนติเจนที่เกิดการกลายพันธุ์เป็นสองกลุ่ม คือ กลุ่มที่มีประสิทธิภาพการยึดเหนี่ยวสูงกับ scFv anti-p17 ได้แก่ peptides p17.1, p17.2 และ p17.8 และกลุ่มที่มีประสิทธิภาพการยึดเหนี่ยวต่ำกับ scFv anti-p17 ได้แก่ peptides p17.3, p17.4, p17.5, p17.6, p17.7 และ p17.9 จากการคำนวณพลังงานการยึดจับระหว่างโครงสร้างเชิงซ้อนของ scFv anti-p17 กับ peptide p17.1, p17.3, p17.7 และ p17.8 พบว่า พลังงานการยึดจับของลำดับกรดอะมิโนสำคัญบริเวณ complementary determining regions (CDRs) ของโมเลกุล scFv anti-p17 ที่มีอันตรกิริยาสูงระหว่างโมเลกุลเชิงซ้อนระหว่าง scFv anti-p17 กับแอนติเจนต่างๆ ประกอบด้วย MET100, LYS101, ASN169, HIS228 และ LEU229 โดยการเปรียบเทียบค่า absolute relative decomposed energy ที่มีค่ามากกว่า 2.00 kcal/mole กับแอนติเจนสายพันธุ์เดิม (peptide p17.1) และมีแรงยึดเหนี่ยวระหว่างโมเลกุลของโครงสร้างที่ได้จากการจำลองแบบพลศาสตร์เชิงโมเลกุล (Molecular Dynamics Simulations) ในช่วง 500 ps สุดท้ายด้วยพันธะไฮโดรเจนมากกว่า 95% occupancy ของ R181 ใน CDRL2 กับ D121 และ Q127

ตารางที่ 3.4 ความสัมพันธ์ระหว่างแอดติวิตีการยึดจับจาก competitive ELISA และ MM-PBSA ที่อุณหภูมิ

310 K

Method	Contribution	Peptide Names				CD147
		p17.1	p17.3	p17.7	p17.8	
MM	$\Delta(L)$	-227.32	-174.34	-169.29	-91.34	
	$\Delta VDW$	-72.28	-55.46	-48.01	-68.95	
	$\Delta GAS$	-299.60	-229.81	-217.30	-160.29	
PBSA	$\Delta PB_{SOL}$	-11.02	-9.73	-8.54	-10.56	
	$\Delta PB_{CAL}$	280.63	230.71	217.07	144.47	
	$\Delta PB_{SCL}$	269.61	220.99	208.53	133.91	
	$\Delta PB_{ELE}$	53.31	56.37	47.78	53.13	
	$\Delta PB_{TOT}$	-29.98	-8.82	-8.77	-26.38	
GBSA	$\Delta GB_{SOL}$	-11.02	-9.73	-8.54	-10.56	
	$\Delta GB_{CAL}$	273.07	220.02	207.41	136.89	
	$\Delta GB_{SCL}$	262.05	210.29	198.87	126.33	
	$\Delta GB_{ELE}$	45.75	45.68	38.12	45.55	
	$\Delta GB_{TOT}$	-37.55	-19.52	-18.43	-33.97	
	Experimental value <sup>a</sup>	75.94	55.25	44.60	79.46	7.94

ELE, electrostatic interactions; VDW, van der Waals interactions between the fragments; GAS, addition ELE + VDW + INT being the binding enthalpic contributions in vacuo;  $PB_{SOL}$ , nonpolar contribution to solvation;  $PB_{CAL}$ , polar contribution of solvation;  $PB_{SCL}$ , the  $PB_{SOL}$  +  $PB_{CAL}$ ;  $PB_{ELE}$ ,  $PB_{CAL}$  + ELE addition;  $PB_{TOT}$ , total binding free energy calculated by the MM-PBSA method.

<sup>a</sup> Competitive ELISA: PI (%)

ตารางที่ 3.5 Common interaction residues found among the natural peptide sequences

Residue	Loop	Decomposed energy (kcal/mol)			
		p17.1	p17.3	p17.7	p17.8
ASP31	H1	0.26	0.35 (0.09)	0.28 (0.02)	0.24 (-0.02)
TYR32	H1	0.04	-0.05 (-0.09)	0.03 (-0.01)	0.01 (-0.03)
GLY33	H1	0.01	-0.08 (-0.09)	-0.02 (-0.03)	-0.11 (-0.12)
ASN35	H1	0.14	-0.14 (-0.28)	-0.09 (-0.23)	-0.34 (-0.48)
TRP50	H2	-1.73	<b>-3.76 (-2.03)</b>	-2.05 (-0.32)	<b>-4.90 (-3.17)</b>
ASN52	H2	0.03	-0.25 (-0.28)	-0.05 (-0.08)	-0.84 (-0.87)
THR59	H2	0.04	-0.56 (-0.60)	-0.57 (-0.61)	-0.61 (-0.65)
SER99	H3	-0.18	-0.17 (0.01)	0.09 (0.27)	-0.62 (-0.44)
MET100	H3	-3.80	<b>-1.02 (2.78)</b>	-3.27 (0.53)	-3.59 (0.21)
LYS101	H3	-2.91	<b>-0.56 (2.35)</b>	-1.10 (1.81)	<b>4.24 (7.15)</b>
SER103	H3	-0.11	0.06 (0.17)	0.03 (0.14)	-1.31 (-1.2)
GLY161	L1	0.01	-0.01 (-0.02)	-0.07 (-0.08)	-0.02 (-0.03)
SER162	L1	0.11	0.03 (-0.08)	-0.07 (-0.18)	0.20 (0.09)
ASP163	L1	0.40	0.28 (-0.12)	0.39 (-0.01)	-0.27 (-0.67)
LYS165	L1	-0.50	-0.16 (0.34)	-0.26 (0.24)	0.46 (0.96)
PHE167	L1	-2.26	-3.14 (-0.88)	-1.59 (0.67)	<b>-4.87 (-2.61)</b>
ASN169	L1	-2.66	<b>0.00 (2.66)</b>	<b>0.09 (2.75)</b>	<b>-0.22 (2.44)</b>
TYR184	-	-2.41	-2.06 (0.35)	-1.44 (0.97)	-1.79 (0.62)
LEU185	L2	-0.07	-1.24 (-1.17)	-1.13 (-1.06)	-2.06 (-1.99)
LYS188	L2	-0.14	-0.07 (0.07)	-0.30 (-0.16)	0.02 (0.16)
ASP190	L2	0.26	-0.37 (-0.63)	0.20 (-0.06)	-0.10 (-0.36)
SER191	-	-0.03	0.00 (0.03)	0.00 (0.03)	0.01 (0.04)
GLY226	L3	-0.18	-0.57 (-0.39)	-0.07 (0.11)	-1.29 (-1.11)
THR227	L3	-0.30	-0.32 (-0.02)	-1.85 (-1.55)	<b>-2.90 (-2.6)</b>
HIS228	L3	-3.93	<b>0.14 (4.07)</b>	<b>-0.42 (3.51)</b>	<b>-1.12 (2.81)</b>
LEU229	L3	-4.53	<b>-0.70 (3.83)</b>	<b>-1.15 (3.38)</b>	<b>-2.07 (2.46)</b>
GLN231	L3	-0.52	<b>-2.75 (-2.23)</b>	-1.20 (-0.68)	0.14 (0.66)

Relative energy to p17.1 is in parenthesis. The amino acids which have significant contribution with the absolute relative energy larger than 2 Kcal/mol are indicated in bold letters.

### 3.4 Molecular dynamics simulations (MD) of scFv anti-p17 - peptide epitopes complexes: Computational Alanine-Scanning of an amino acids in the complementary determining regions (CDRs) of scFv anti-p17

การจำลองพลวัตโดยใช้น้ำรอบทั้งโครงสร้างสารเชิงซ้อน และศึกษาการทำ Alanine scanning ศึกษาอิทธิพลของ side chain โดยการนำ alanine scanning เปลี่ยนหมู่อะมิโนบริเวณ CDRs ที่ละตัวให้เป็น alanine แล้วคำนวณอันตรกิริยา จากผลการจำลอง และการคำนวณพบว่า scFv anti-p17 ที่ใช้มีความสามารถยึดจับกับ c-terminal HIV-p17 wild type และมิวแทนที่ได้แตกต่างกัน และสามารถระบุหมู่อะมิโนที่มีส่วนสำคัญต่อการยึดจับเพื่อใช้เป็นแนวทางในการปรับปรุง โครงสร้าง scFv anti-p17 ต่อไปได้ ผลการทดลองเป็นไปดังตารางที่ 3.6

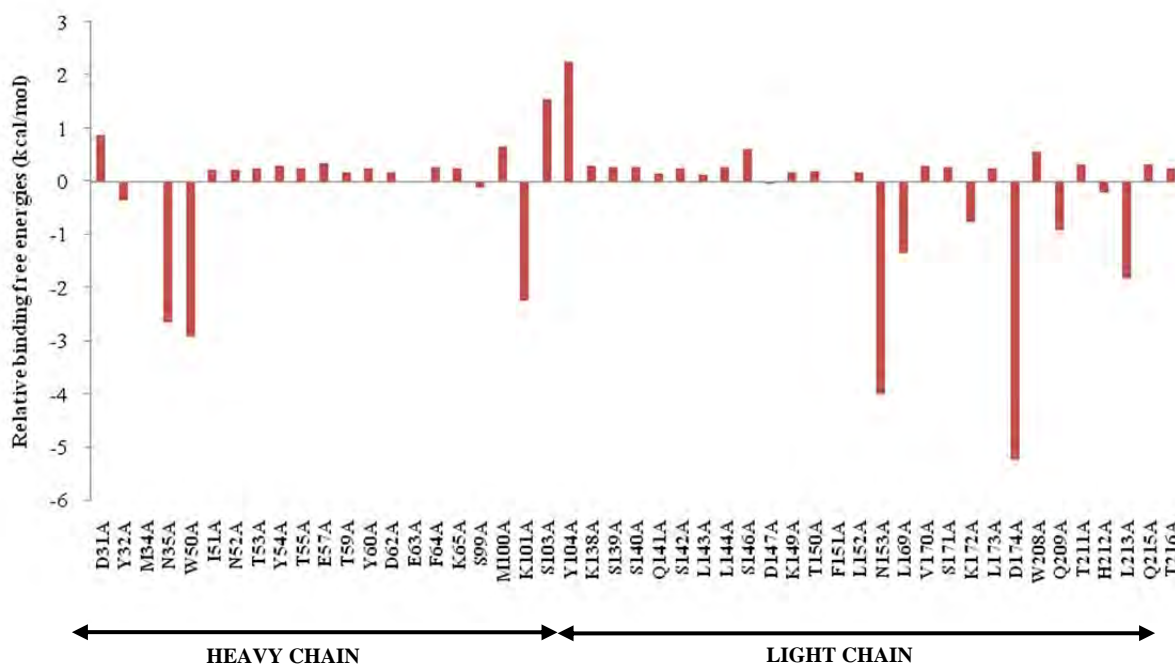
อะลานีนสแกนนิ่งเป็นฟังก์ชันพลังงานอิสระที่ใช้คำนวณพลังงานยึดเหนี่ยวระหว่างโครงสร้างของ scFv anti-p17 และเปปไทด์ เมื่อเปลี่ยนแปลงกรดอะมิโนบริเวณ complementary determining regions (CDRs) ของโมเลกุล scFv anti-p17 เป็นอะลานีน โดยจะคำนวณพลังงานยึดเหนี่ยวด้วยเทคนิค MM-PBSA และพลังงานการยึดจับของแต่ละลำดับกรดอะมิโนด้วยเทคนิค MM-GBSA ซึ่ง input สำหรับการทำอะลานีนสแกนนิ่งประกอบด้วย โครงสร้างสามมิติของ scFv anti-p17 ซึ่งจะมีการเปลี่ยนแปลงกรดอะมิโนสำคัญเป็นอะลานีน พบว่ากรดอะมิโนบริเวณ CDRs loops บน VH และ VL ของ mutation scFv anti-p17 ที่มี relative binding free energies ( $\Delta PB_{TOT}$ ) มากกว่า  $0.5 \text{ kcal mol}^{-1}$  เมื่อเทียบกับ wild type scFv anti-p17 จะเป็นกรดอะมิโนสำคัญที่มีอิทธิพลต่อการยึดจับกับเปปไทด์ ได้แก่ ASP31(+), ASN35(-), TRP50(-), MET100(+), LYS101(-), SER103(+), TYR104(+), SER162(+), ASN169(-), LEU185(-), LYS188(-), ASP190(-), TRP224(+), GLN225(-), และ LEU229(-) ดังรูป 1 ซึ่งกรดอะมิโนสำคัญนี้สามารถแบ่งเป็น 2 กลุ่ม (favorable (-) และ unfavorable (+)) ประกอบด้วย กลุ่มที่มี  $\Delta PB_{TOT}$  ติดลบ

มากกว่า 0.5 kcal/mol เช่น ASN35, TRP50, LYS101, ASN169, LEU185, LYS188, ASP190, GLN225, and LEU229 ซึ่งจะมี strong interaction กับเปปไทด์ อีกกลุ่มคือ กลุ่มที่มี  $\Delta PB_{TOT}$  เป็นบวกมากกว่า 0.5 kcal/mol เช่น ASP31, MET100, SER103, TYR104, SER162, and TRP224 ซึ่งจะมี poorly interaction กับเปปไทด์

ตารางที่ 3.6 Values of the free energy components for alanine scanning analysis

scFv anti-p17	$\Delta E_{LE}$	$\Delta V_{DW}$	$\Delta G_{AS}$	$\Delta PB_{SUR}$	$\Delta PB_{CAL}$	$\Delta PB_{SOL}$	$\Delta PB_{ELE}$	$\Delta PB_{TOT}$
<b>Wild Type</b>	-209.74	-71.57	-281.31	-10.17	260.42	250.25	50.68	-31.06
<b><i>Mutants in VH</i></b>								
D31A	-236.64	-71.48	-308.12	-10.16	286.34	276.18	49.71	-31.94
Y32A	-210.50	-69.13	-279.63	-10.04	258.96	248.92	48.46	-30.70
M34A	-210.12	-71.54	-281.67	-10.17	260.80	250.63	50.68	-31.04
N35A	-206.56	-70.67	-277.23	-10.22	259.05	248.84	52.49	-28.40
W50A	-208.10	-66.49	-274.59	-9.91	256.37	246.46	48.27	-28.13
I51A	-210.06	-71.55	-281.60	-10.17	260.51	250.33	50.45	-31.27
N52A	-208.90	-71.42	-280.32	-10.14	259.19	249.05	50.29	-31.27
T53A	-209.75	-71.54	-281.29	-10.17	260.16	249.99	50.41	-31.30
Y54A	-210.24	-71.54	-281.78	-10.17	260.62	250.45	50.38	-31.34
T55A	-209.94	-71.55	-281.50	-10.17	260.36	250.19	50.42	-31.31
E57A	-223.66	-71.55	-295.21	-10.17	273.97	263.80	50.31	-31.41
T59A	-209.64	-71.48	-281.12	-10.17	260.06	249.89	50.42	-31.23
Y60A	-209.71	-71.56	-281.26	-10.17	260.14	249.97	50.43	-31.30
D62A	-223.87	-71.57	-295.44	-10.17	274.39	264.22	50.52	-31.22
E63A	-224.46	-71.57	-296.03	-10.17	275.13	264.96	50.67	-31.07
F64A	-209.80	-71.56	-281.36	-10.17	260.21	250.03	50.40	-31.33
K65A	-196.56	-71.57	-268.13	-10.17	247.00	236.83	50.43	-31.30
S99A	-209.22	-71.32	-280.54	-10.17	259.77	249.59	50.54	-30.95
M100A	-207.22	-68.58	-275.80	-9.88	253.97	244.09	46.75	-31.71
K101A	-170.84	-68.77	-239.60	-10.32	221.11	210.80	50.28	-28.81
S103A	-207.61	-71.45	-279.06	-10.17	256.63	246.46	49.02	-32.60
Y104A	-212.25	-70.47	-282.73	-10.08	259.49	249.41	47.24	-33.32
D31A	-236.64	-71.48	-308.12	-10.16	286.34	276.18	49.71	-31.94
Y32A	-210.50	-69.13	-279.63	-10.04	258.96	248.92	48.46	-30.70
M34A	-210.12	-71.54	-281.67	-10.17	260.80	250.63	50.68	-31.04
N35A	-206.56	-70.67	-277.23	-10.22	259.05	248.84	52.49	-28.40
W50A	-208.10	-66.49	-274.59	-9.91	256.37	246.46	48.27	-28.13
I51A	-210.06	-71.55	-281.60	-10.17	260.51	250.33	50.45	-31.27
N52A	-208.90	-71.42	-280.32	-10.14	259.19	249.05	50.29	-31.27
T53A	-209.75	-71.54	-281.29	-10.17	260.16	249.99	50.41	-31.30
Y54A	-210.24	-71.54	-281.78	-10.17	260.62	250.45	50.38	-31.34
T55A	-209.94	-71.55	-281.50	-10.17	260.36	250.19	50.42	-31.31
E57A	-223.66	-71.55	-295.21	-10.17	273.97	263.80	50.31	-31.41
T59A	-209.64	-71.48	-281.12	-10.17	260.06	249.89	50.42	-31.23
Y60A	-209.71	-71.56	-281.26	-10.17	260.14	249.97	50.43	-31.30
D62A	-223.87	-71.57	-295.44	-10.17	274.39	264.22	50.52	-31.22
E63A	-224.46	-71.57	-296.03	-10.17	275.13	264.96	50.67	-31.07

F64A	-209.80	-71.56	-281.36	-10.17	260.21	250.03	50.40	-31.33
K65A	-196.56	-71.57	-268.13	-10.17	247.00	236.83	50.43	-31.30
S99A	-209.22	-71.32	-280.54	-10.17	259.77	249.59	50.54	-30.95
M100A	-207.22	-68.58	-275.80	-9.88	253.97	244.09	46.75	-31.71
K101A	-170.84	-68.77	-239.60	-10.32	221.11	210.80	50.28	-28.81
S103A	-207.61	-71.45	-279.06	-10.17	256.63	246.46	49.02	-32.60
Y104A	-212.25	-70.47	-282.73	-10.08	259.49	249.41	47.24	-33.32
<b><i>Mutants in VL</i></b>								
K138A	-197.07	-71.57	-268.64	-10.17	247.47	237.30	50.41	-31.34
S139A	-209.66	-71.57	-281.22	-10.17	260.08	249.91	50.42	-31.32
S140A	-209.84	-71.57	-281.41	-10.17	260.26	250.09	50.42	-31.32
Q141A	-209.93	-71.56	-281.49	-10.17	260.45	250.28	50.52	-31.20
S142A	-210.03	-71.57	-281.60	-10.17	260.47	250.30	50.44	-31.30
L143A	-209.78	-71.53	-281.31	-10.17	260.31	250.14	50.54	-31.17
L144A	-209.75	-71.55	-281.30	-10.17	260.15	249.98	50.40	-31.32
S146A	-209.31	-71.37	-280.68	-10.09	259.12	249.02	49.81	-31.66
D147A	-218.26	-71.37	-289.63	-10.15	268.77	258.62	50.51	-31.01
K149A	-194.49	-71.30	-265.80	-10.15	244.72	234.56	50.22	-31.23
T150A	-209.60	-71.55	-281.15	-10.17	260.06	249.89	50.46	-31.26
F151A	-209.06	-69.58	-278.64	-9.93	257.50	247.57	48.44	-31.07
L152A	-209.85	-71.55	-281.40	-10.17	260.33	250.16	50.48	-31.24
N153A	-209.82	-71.03	-280.85	-10.29	264.08	253.79	54.26	-27.05
L169A	-209.72	-69.50	-279.22	-10.16	259.66	249.51	49.94	-29.71
V170A	-208.85	-71.55	-280.41	-10.17	259.21	249.04	50.36	-31.36
S171A	-210.53	-71.57	-282.10	-10.17	260.95	250.78	50.42	-31.32
K172A	-171.54	-71.07	-242.61	-10.05	222.39	212.33	50.85	-30.28
L173A	-209.86	-71.55	-281.40	-10.17	260.28	250.10	50.42	-31.30
D174A	-249.95	-71.04	-320.99	-10.18	305.36	295.17	55.41	-25.81
W208A	-208.67	-71.20	-279.87	-10.17	258.43	248.25	49.76	-31.61
Q209A	-209.22	-71.36	-280.58	-10.17	260.61	250.44	51.39	-30.14
T211A	-210.09	-71.38	-281.47	-10.17	260.26	250.09	50.17	-31.38
H212A	-208.95	-71.32	-280.27	-10.16	259.58	249.42	50.63	-30.85
L213A	-209.64	-68.37	-278.01	-10.05	258.84	248.79	49.20	-29.21
Q215A	-209.08	-70.30	-279.38	-10.27	258.28	248.01	49.20	-31.37
T216A	-209.78	-71.56	-281.34	-10.17	260.20	250.03	50.42	-31.31

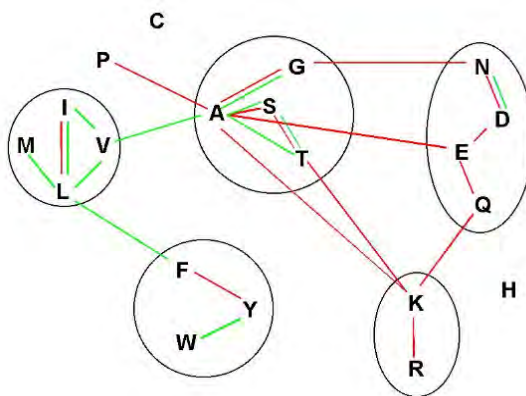


รูปที่ 3.6 Histograms reporting the binding free energies for the mutations. The total bar height reflects relative binding free energies of each amino acids in CDRs loops with wild type of scFv anti-p17 whose mutation to alanine by alanine scanning mutagenesis.

### 3.5 Amino acids mutation of scFv anti-p17 for improving the specificity of binding affinity

จาก MM-PBSA and Alanine scanning พบว่า กรดอะมิโนสำคัญที่มีอิทธิพลต่อการยึดจับกับ เปปไทด์ ได้แก่ ASP31, ASN35, TRP50, MET100, LYS101, SER103, TYR104, SER162, ASN169, LEU185, LYS188, ASP190, TRP224, GLN225, HIS228 และ LEU229 ดังนั้นขั้นตอนต่อมาจะทำการ mutation บนเรซิดิวสำคัญนี้ เพื่อปรับปรุงความสามารถในการจับ โดยการทำนายจาก solvent exposed area ร่วมกับความมีขั้วของหมู่แทนที่ของกรดอะมิโนต่างๆ ตามรูปที่ 1 พบว่า กรดอะมิโนสามารถแบ่งออกได้เป็น 5 กลุ่มด้วยกัน ซึ่งจะมีความสัมพันธ์กับสมบัติทางกายภาพของหมู่แทนที่ของกรดอะมิโน ขั้นตอนแรกได้ทำการ Mutation ที่ตำแหน่ง 31 ของ scFv anti-p17 พบว่า mutation ที่เป็นไปได้ในตำแหน่งนี้ได้แก่ D31E, D31N, D31G, D31T, D31K, D31Y, D31F, D31W, D31V และ D31M จากนั้นนำ mutated scFv anti-p17 ที่ได้ไปทำการด็อกกิงกับเปปไทด์ทั้ง 9 ชนิด ประกอบด้วย HIV-1 epitope เริ่มต้นที่ C-terminal (DTGHSSQVSNY) บน p17 (p17.1) และ mutants เปปไทด์อีก 8 ชนิดที่พบตามธรรมชาติ (p17.2-p17.9) ด้วย CDOCKER ซึ่งเป็นเทคนิค grid-based molecular docking ที่คำนวณด้วย CHARMm โดยรีเซพเตอร์จะแข็งเกร็ง ในขณะที่เปปไทด์จะสามารถยืดหยุ่นได้ จากค่า cdocker interaction energy ดังตารางที่ได้จากการด็อกกิงด้วย CDOCKER พบว่า เปปไทด์ 17.1, p17.3 และ p17.7 เมื่อทำคอมเพล็กซ์กับ mutated scFv anti-p17 จาก Aspartic acid เป็น Glycine (D31G) จะมีค่า cdocker interaction energy มากกว่าทำคอมเพล็กซ์กับ wild type scFv anti-p17 ดังนั้น D31G น่าจะเป็นตัวเลือกที่ได้จากการ mutate หมู่กรดอะมิโนในส่วน CDRs ของ scFv anti-p17 เพื่อปรับปรุงความสามารถในการจับกับเปปไทด์ และนำไปใช้เปรียบเทียบกับข้อมูลทางห้องปฏิบัติการ





รูปที่ 3.7 การเลือกเปลี่ยนหมู่อะมิโนตามขนาด และความมีขั้ว แสดง buried residues (ส่วนสีเขียว) และ exposed residues (ส่วนสีแดง)

ตารางที่ 3.7 The results of cdocker interaction energy of all mutations

scFv anti-p17	-CDOCKER_INTERACTION_ENERGY								
	p17.1	p17.2	p17.3	p17.4	p17.5	p17.6	p17.7	p17.8	p17.9
Wild Type	66.87	81.77	67.29	73.42	74.10	64.78	65.20	71.88	76.54
L213A	73.61	72.55	65.09	67.08	61.78	67.31	73.89	73.50	71.75
L213C	64.40	79.01	71.99	78.50	59.60	57.30	70.17	67.50	65.67
L213E	79.44	63.81	72.03	77.04	59.43	64.79	71.34	64.94	82.21
L213F	70.73	76.97	92.12	60.17	62.75	67.30	72.43	70.55	74.31
L213K	65.20	64.71	60.51	71.07	76.04	67.81	61.12	60.70	73.14
L213N	65.49	70.59	80.16	83.44	65.45	63.57	70.04	58.19	68.38
L213P	77.95	70.22	64.86	59.44	55.70	67.82	73.54	59.96	68.50
L213Q	69.50	60.00	59.17	73.35	57.35	78.12	62.23	65.97	71.38
M100A	69.94	76.86	68.40	65.16	78.27	81.56	62.94	61.73	66.30
M100C	59.26	81.04	67.03	65.39	70.93	66.81	66.97	70.26	69.37
M100F	57.21	71.08	63.18	61.79	84.59	68.71	83.86	80.05	78.31
M100H	72.48	73.06	98.85	71.38	75.78	73.78	65.68	73.65	71.55
M100K	70.70	63.07	68.69	78.34	74.72	73.71	67.82	63.27	72.68
M100P	78.08	78.42	64.86	69.44	85.09	68.38	67.79	64.38	78.11
M100Q	73.02	78.05	69.46	82.77	68.87	74.20	61.36	65.98	82.42
M100R	69.78	72.14	72.64	85.19	70.63	72.99	66.85	54.33	66.46
M100Y	65.07	94.29	62.43	69.28	71.87	82.33	75.93	70.42	74.65
M100E	74.17	65.20	71.08	71.15	69.93	79.48	86.44	68.93	89.21
M100W	63.93	67.04	82.82	72.32	71.37	78.86	71.30	58.07	78.85
H212D	63.83	71.12	74.83	90.60	63.85	75.31	68.23	65.00	66.93
H212E	64.48	75.85	57.87	65.90	57.01	61.15	71.22	71.42	72.27
H212K	75.68	66.31	71.54	61.03	67.91	74.51	73.90	65.68	73.32
H212N	67.38	69.12	76.10	75.97	53.50	64.45	60.22	60.14	62.80
H212Q	78.12	69.68	64.89	66.04	62.00	74.61	63.84	77.18	80.78
H212T	72.74	62.71	85.33	86.19	60.07	68.76	67.19	62.44	69.01
H212V	67.89	63.07	59.86	67.17	74.39	69.28	69.58	57.95	69.92

H212Y	61.23	78.98	62.22	86.99	64.73	58.98	58.08	64.89	65.99
K101A	71.04	70.30	63.46	76.30	62.81	62.67	68.44	65.93	73.09
K101E	78.62	66.84	64.87	66.83	57.31	64.97	63.46	77.20	73.58
K101Q	63.38	76.44	51.29	68.61	57.10	58.70	69.17	77.24	62.97
K101V	54.70	76.20	80.34	68.29	58.17	61.82	72.91	77.13	67.12
K101H	70.74	55.45	54.83	71.03	64.18	74.84	61.64	65.17	74.36
K101R	76.06	68.51	66.27	83.46	74.93	71.24	75.49	69.08	80.36
K101T	65.00	67.29	54.88	61.65	59.63	56.89	52.10	61.61	82.03
K101Y	71.61	77.35	77.25	75.61	68.68	69.66	53.47	67.60	79.59
N153G	70.92	69.05	71.96	75.77	69.82	69.33	65.75	70.97	64.53
N153K	71.73	74.51	101.76	79.20	63.95	81.82	86.80	79.43	78.94
N153Q	63.91	75.38	70.50	74.34	69.84	71.53	69.44	83.21	72.02
N153T	75.29	80.37	62.08	63.91	70.70	65.77	68.97	67.14	78.41
N153V	85.46	76.15	63.99	73.95	69.93	74.90	58.73	74.28	65.98
N153Y	74.89	76.84	57.19	69.67	54.01	64.76	71.52	84.45	70.71
D31E	66.08	74.24	76.13	62.38	71.05	73.33	65.03	73.88	75.00
D31N	65.38	61.50	61.74	69.07	70.81	82.05	69.88	56.70	61.65
D31G	67.56	77.99	76.34	67.59	67.80	68.17	66.51	57.51	66.98
D31T	71.08	74.65	64.55	75.51	64.19	71.92	60.70	67.15	69.17
D31K	57.87	76.63	81.75	68.88	55.05	70.98	60.42	57.63	65.94
D31Y	70.01	82.63	64.76	63.24	69.97	62.30	61.26	68.50	73.44
D31F	65.37	67.11	71.61	63.12	61.27	59.29	56.31	59.73	54.22
D31W	51.78	72.22	73.11	69.51	64.02	80.63	61.48	58.86	65.05
D31V	65.30	60.49	67.53	65.12	71.73	63.82	63.93	55.81	61.40
D31M	58.37	71.26	55.34	63.39	62.66	59.35	62.11	69.73	65.13
E57D	63.21	63.96	78.13	70.36	68.09	66.55	55.08	74.98	74.35
E57G	63.01	66.47	74.57	68.73	60.39	70.57	70.20	64.84	60.51
E57T	70.79	71.12	78.72	70.61	66.85	61.91	70.80	66.13	66.55
E57K	61.90	69.80	73.62	69.50	71.59	58.19	69.44	61.10	68.65
E57Y	61.48	73.73	78.19	69.73	77.26	62.39	66.66	63.97	61.64
E57V	75.69	65.77	78.91	73.74	59.83	54.41	55.04	70.34	69.60
D62E	66.54	82.97	88.43	65.00	67.92	64.81	69.04	67.11	83.13
D62N	70.75	60.42	58.97	69.06	57.76	62.49	68.32	63.56	83.39
D62G	64.34	80.47	68.50	74.28	57.38	56.89	62.34	64.62	64.93
D62T	58.04	70.11	79.40	78.15	74.85	62.42	69.03	58.94	60.44
D62K	65.67	68.56	80.72	56.44	76.13	81.39	62.06	59.85	67.93
D62Y	69.42	86.96	79.49	70.36	72.46	67.58	78.92	71.54	57.83
D62F	65.66	66.11	65.70	62.96	66.34	84.23	70.93	71.40	80.27
D62W	90.87	78.52	65.24	67.54	57.58	57.26	63.72	62.58	59.81
D62V	65.43	68.83	61.33	87.15	70.03	60.99	69.06	58.32	67.99
D62M	63.64	63.11	76.60	62.93	62.18	65.85	60.78	59.91	76.01
E63D	63.05	70.90	71.44	71.20	75.23	63.83	58.33	63.52	60.23
E63G	63.33	89.75	76.76	71.89	60.94	62.21	68.51	62.19	69.26
E63T	66.63	71.98	80.95	73.66	65.69	65.41	70.22	61.53	68.59
E63K	62.66	66.02	76.97	63.60	61.82	73.95	66.85	65.61	62.49
E63Y	73.13	85.81	80.61	68.48	60.67	58.62	72.42	69.05	74.44
E63V	66.03	85.78	72.86	69.08	62.93	63.44	58.20	65.38	64.35

K65A	66.39	56.42	70.71	79.50	65.01	59.14	63.94	58.64	81.28
K65E	72.78	73.35	65.93	64.20	63.48	63.58	61.09	59.67	68.62
K65Q	78.32	53.92	66.48	69.71	86.62	59.77	56.54	64.30	69.66
K65V	63.27	64.58	54.43	68.89	61.25	60.04	64.60	64.04	85.93
K65H	64.55	54.25	58.76	59.94	70.43	65.73	62.34	65.20	72.02
K65R	63.46	73.20	67.71	61.43	69.57	69.21	57.24	62.20	67.98
K65T	61.06	90.98	67.53	73.67	63.39	61.91	49.97	61.27	67.52
K65Y	64.22	63.01	65.20	69.36	67.55	58.47	70.47	59.70	66.56
K138A	65.92	79.97	82.99	60.05	57.92	67.08	69.19	61.09	70.85
K138E	64.94	67.29	62.97	64.16	69.34	61.08	73.80	61.30	73.28
K138Q	77.58	77.44	67.38	60.68	58.40	54.00	50.41	54.44	65.20
K138V	67.05	63.38	67.20	82.91	55.97	62.06	59.49	57.24	71.29
K138H	68.94	58.54	82.93	62.85	66.96	64.12	54.71	77.88	56.77
K138R	71.39	53.76	75.37	68.31	60.82	66.75	68.52	58.99	70.32
K138T	63.19	80.28	63.91	64.05	70.75	64.04	55.28	67.03	63.47
K138Y	63.31	70.67	68.96	71.24	65.15	76.12	65.26	69.21	64.71
D147E	59.81	56.74	56.97	61.32	78.71	60.12	64.23	64.84	65.36
D147N	64.06	68.36	70.51	79.79	59.82	69.46	57.31	64.74	69.55
D147G	72.23	64.02	74.41	65.59	61.13	60.73	74.11	63.38	70.91
D147T	54.06	55.38	44.26	62.01	55.01	59.27	61.11	69.10	64.21
D147K	60.42	72.07	73.58	61.98	54.50	80.54	58.45	66.42	81.12
D147Y	61.12	60.57	68.46	61.13	61.46	53.65	68.63	66.12	54.66
D147F	65.87	65.34	51.51	75.79	69.04	89.10	75.57	63.41	69.95
D147W	71.32	63.97	56.81	67.44	64.41	68.18	63.17	71.17	61.32

- การเปรียบเทียบระหว่าง D31G สำหรับระบบ p17.1 และ p17.9

ได้เลือกศึกษาการเปลี่ยนนิวเตชัน D31G สำหรับระบบ p17.1 และ p17.9 เพิ่มเติมเนื่องจากพบว่ามี ความแตกต่างในข้อมูลการทดลองระหว่าง Competitive ELISA and Indirect ELISA ในหัวข้อ 3.6 ซึ่งทำให้จำเป็นต้องระมัดระวังในการเลือกใช้ข้อมูล ELISA และควรพิจารณาข้อมูลทั้งสองแบบประกอบกัน

ตารางที่ 3.8 Vdw and electrostatic contribution of binding free energy of 17.1 and 17.9 (wild type, mutant)

Peptide Names	scFv-p17	Binding free energy in each contribution (kcal/mol)							
		$\Delta E_{LE}$	$\Delta V_{DW}$	$\Delta G_{AS}$	$\Delta P_{B_{SUR}}$	$\Delta P_{B_{CAL}}$	$\Delta P_{B_{SOL}}$	$\Delta P_{B_{ELE}}$	$\Delta P_{B_{TOT}}$
p17.1 ( <sup>121</sup> DTGHSSQVSQNY <sup>132</sup> )	wild type	-249.35	-44.24	-293.59	-8.22	270.15	261.93	20.80	-31.65
	D31G	-278.26	-36.55	-314.81	-7.02	299.58	292.57	21.33	-22.25
p17.9 ( <sup>121</sup> DTGNNSQVSQNY <sup>132</sup> )	wild type	-239.51	-34.19	-273.70	-6.93	254.92	247.98	15.41	-25.72
	D31G	-119.30	-23.20	-142.49	-3.76	127.21	123.46	7.92	-19.04

ตารางที่ 3.9 Binding free energy, Cdocker energy, Experiment Indirect ELISA, Competitive ELISA of 17.1, 17.9 (wild, mutant)

Peptide	scFv	Cdocker energy (kcal/mol)	Binding free energy ( $\Delta P_{TOT}$ ) (kcal/mol)	Experimental	
				Indirect ELISA (cfu/ml)	Competitive ELISA (1ug/ml)
p17.1	wild	-110.460	-31.65	1.822	0.067
	D31G	-117.893	-22.25	1.878	0.137
p17.9	wild	-98.877	-25.72	0.241	0.093
	D31G	-100.970	-19.04	1.184	0.186

จากตารางที่ 3.8, 3.9 และรูปที่ 3.8 พบว่า

-ค่า binding energy ในช่วง -19 - -32 kcal/mol ไม่แสดงความแตกต่างในการยึดจับหากใช้ข้อมูลจากการทดลอง competitive ELISA

-แต่จาก alanine scanning รูปที่ 3.6 การเปลี่ยน D31A มีผลต่อการยึดจับ แย่ลง (ค่า +) และเมื่อทำ point mutation D31G และเปรียบเทียบการยึดจับ พบว่าส่งผลให้การจับเปลี่ยนแปลงในทางที่ดีขึ้น ดูจาก รูปที่ 3.8 decomposition energy of Asp31 สอดคล้องกับ indirect ELISA ซึ่งเปลี่ยนจาก 0.241 (wild) เป็น 1.184 (D31G)

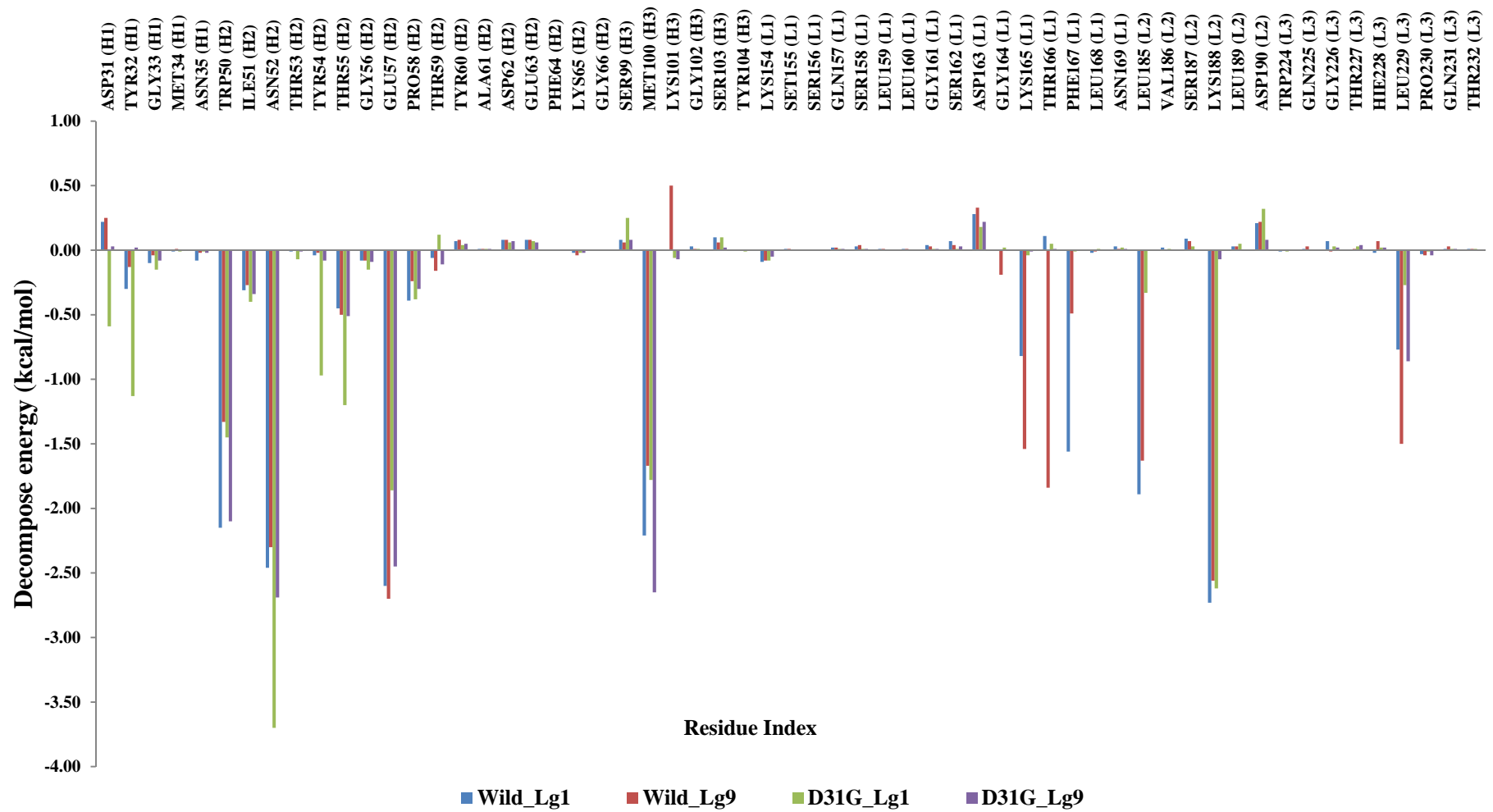
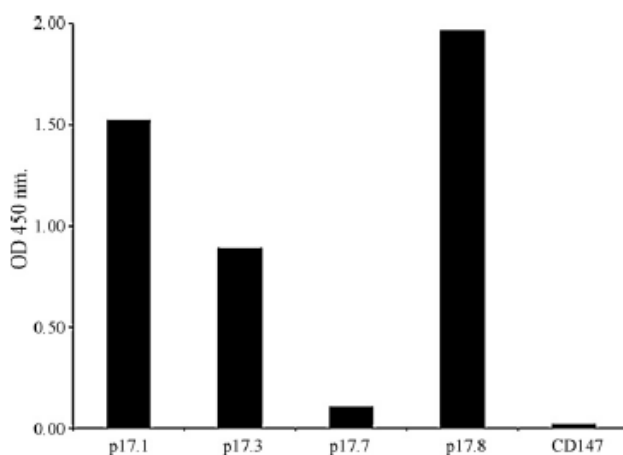


Figure 3.8 Decomposition energy of amino acids in the CDRs 17.1, 17.9 for wild and mutant (D31G) scFv anti-p17

### 3.6 Competitive ELISA and Indirect ELISA

การจำลองแบบพลศาสตร์เชิงโมเลกุลจะศึกษาค่าพลังงานการยึดจับและพฤติกรรมเคลื่อนที่แบบ Dynamic ระหว่างโครงสร้างเชิงซ้อนของ scFv anti-p17 กับ peptide p17.1, p17.3, p17.7 และ p17.8 โดยใช้โปรแกรม Amber version 9 โดยจะหาประสิทธิภาพการยึดจับจากอันตรกิริยาที่เกิดขึ้นของโมเลกุลเชิงซ้อนระหว่าง scFv anti-p17 และ peptide epitopes กับตัวทำละลายด้วยวิธี molecular dynamics (MDs) simulation ที่สัมพันธ์กับข้อมูลการตรวจด้วยวิธี ELISA พบว่า peptide epitopes ทั้งหมดสามารถจับกับโมเลกุลของ scFv anti-p17 ได้ดี ดังนั้น peptide p17.8, p17.1, p17.3 และ p17.7 เป็นต้น และในกรณีของ CD147 peptide จะมี *Positive Signal* น้อย แสดงว่า CD147 peptide ไม่สามารถยับยั้งการยึดเหนี่ยวจากอันตรกิริยาที่เกิดขึ้นระหว่างโมเลกุลเชิงซ้อนของ scFv anti-p17 และ peptide epitopes อื่นๆ



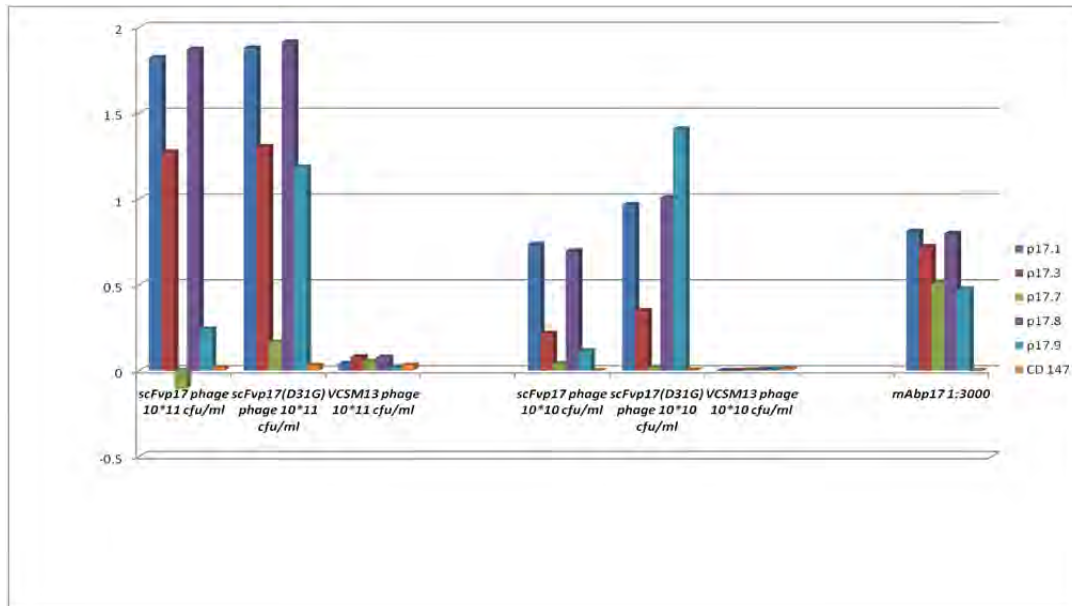
รูปที่ 3.9 ค่าแอกติวิตี้จาก competitive ELISA การทดลองครั้งที่ 1

ตารางที่ 3.7 ค่าแอกติวิตี้ของ scFv และ mAb จาก competitive ELISA การทดลองครั้งที่ 2

	Elisa PI (%) ,scFv	Elisa PI (%) ,mAb	PMF (Cache) Kcal/mol	Cdocker Kcal/mol	MMPBSA
P17.1	90.1	90.7	-902.11		-29.98
P17.2			-899.18		
P17.3	59.6	73.4	-882.65		-8.82
P17.4			-898.71		
P17.5			-843.51		
P17.6			-846.12		
P17.7	60.1	74.5	-829.94		-8.77
P17.8	87.6	92.4	-926.79		-26.38
P17.9	88.5	89	-841.02		

ตารางที่ 3.8 Competitive ELISA ทดลองซ้ำครั้งที่ 3 และผลการทำมิวเตชัน D31G

	p17.1	p17.3	p17.7	p17.8	p17.9	CD 147
scFvp17 phage $10^{*11}$ cfu/ml	1.822	1.272	-0.106	1.871	0.241	0.018
scFvp17(D31G) phage $10^{*11}$ cfu/ml	1.878	1.304	0.165	1.914	1.184	0.03
VCSM13 phage $10^{*11}$ cfu/ml	0.04	0.08	0.052	0.076	0.02	0.031
scFvp17 phage $10^{*10}$ cfu/ml	0.734	0.217	0.039	0.696	0.113	0
scFvp17(D31G) phage $10^{*10}$ cfu/ml	0.966	0.348	0.019	1.006	1.407	0.003
VCSM13 phage $10^{*10}$ cfu/ml	-0.002	-0.002	0.003	0.006	0.007	0.01
mAbp17 1:3000	0.811	0.722	0.509	0.798	0.475	-0.005



รูปที่ 4.0 Indirect ELISA

ข้อสังเกตจากการทดลอง

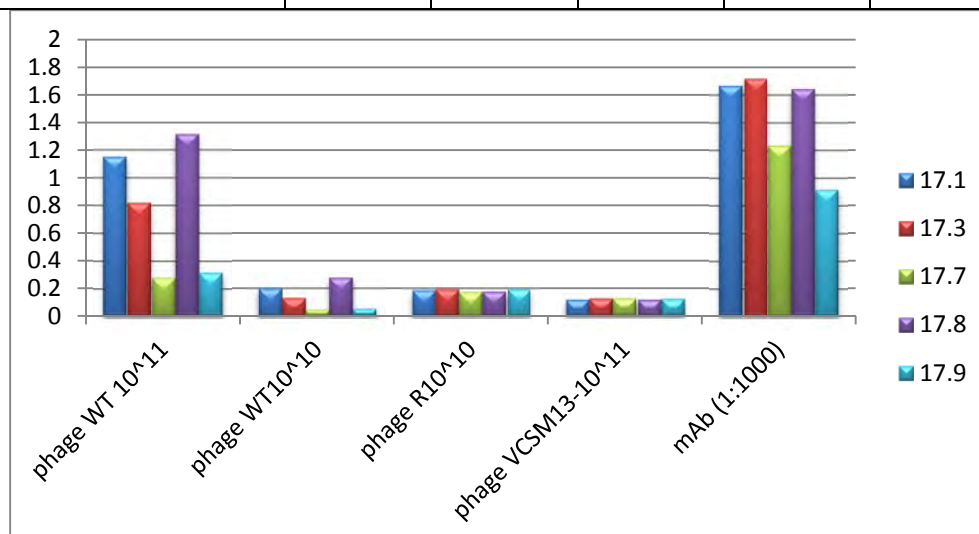
1. Affinity สำหรับ 17.9 ไม่ได้เพิ่มขึ้น
2. ควรใช้ผลการทดลองทั้ง indirect ELISA และ competitive ELISA ประกอบกัน
3. ทั้ง D31G และ wild-type ให้ผล inhibition ด้อยอยู่แล้วในการจับกับ 17.9 โดยวิธี competitive ELISA
4. D31G จับกับ 17.9 ได้ดีขึ้นใน "indirect ELISA" อาจมีสาเหตุจาก flexibility ที่เพิ่มขึ้น ทำให้ access peptide epitope ได้ดีขึ้น ทั้งนี้ต้องเปรียบเทียบข้อมูลการ docking ใหม่ อาจพบว่า aa. set บน 17.9 ที่ใช้ interact กับ D31G อาจแตกต่างกันจาก wild-type scFv17
5. จาก indirect ELISA คาดว่า epitope ของ 17.9 ที่ recognize โดย D31G อาจไม่ติดแนบอยู่บน solid phase ทำให้ D31G เข้าจับได้ง่ายขึ้นกว่า wild-type ที่มองไม่เห็น epitope ที่ติดอยู่กับ solid phase
6. จะเห็นว่าผลสอดคล้องกัน คือ 17.7 การยัดจับไม่มีการเปลี่ยนแปลงจากทั้งสองวิธี แต่ 17.9 น่าสนใจในกรณี indirect ELISA ที่ D31G เพิ่มความสามารถในการจับ

- การดัดแปลงยีนแอนติบอดีชนิด scFv เพื่อเพิ่มความสามารถในการจับกับพีลิจิตเปปไทด์ของเชื้อเอชไอวี โดยหวังผลในการรักษาในระดับยีน

การเปลี่ยนลำดับเบสของยีนที่กำหนดการสร้างซิงเกิลเซนเอพัวแอนติบอดีโดยวิธี Site-directed mutagenesis และการศึกษาแอกติวิตีเมื่อเปลี่ยนแปลงลำดับโคดอน (codon) จากเมทไทโอนีน เปลี่ยนเป็น เป็น อาร์จินีน (M100R) และ ไกลซีน (M100G) รายละเอียดทางการทดลองทำโดยภาควิชาภูมิคุ้มกันวิทยา คณะเทคนิคการแพทย์ มหาวิทยาลัยเชียงใหม่

ตารางที่ 3.9 ผลการเปลี่ยนจากกรดอะมิโนเมทไทโอนีนเป็นอาร์จินีน

Phage	p17.1	p17.3	p17.7	p17.8	p17.9
phage WT10 <sup>11</sup>	1.155	0.823	0.283	1.318	0.32
phage WT10 <sup>10</sup>	0.204	0.135	0.051	0.279	0.058
phage R10 <sup>10</sup>	0.189	0.196	0.176	0.18	0.197
phage VCSM13-10 <sup>11</sup>	0.118	0.126	0.127	0.118	0.123
mAb (1:1000)	1.671	1.722	1.239	1.647	0.92

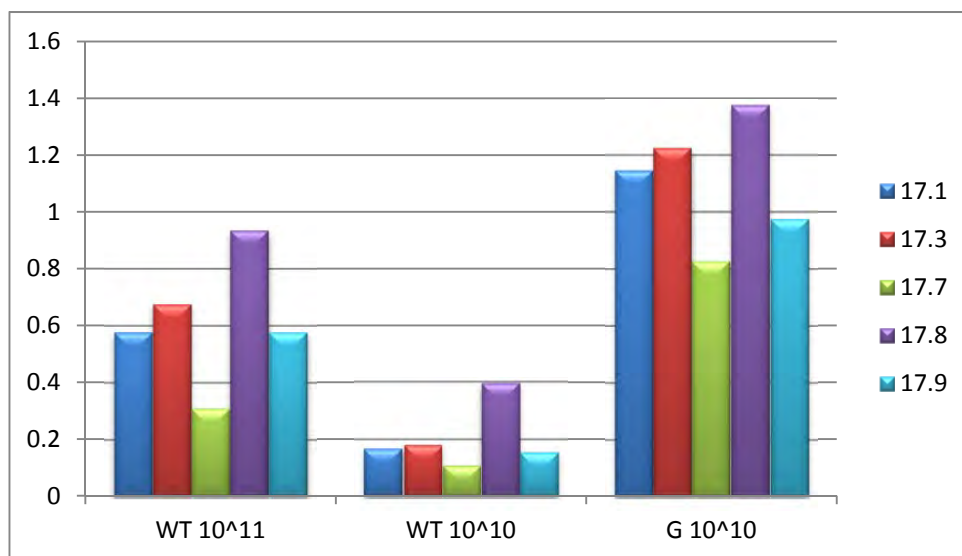




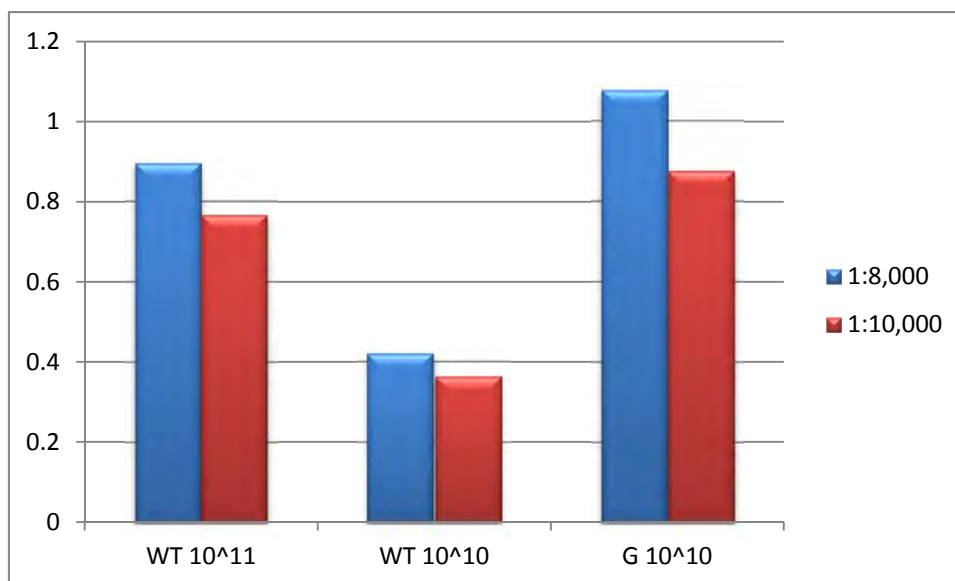
ผลจากการทดสอบการจับของรีคอมบิแนนท์เฟจจากการเปลี่ยนแปลงกรดอะมิโนเมทไทโอนีนตำแหน่งที่ 100 เป็นกรดอะมิโนอาร์จินีน ที่ความเข้มข้นเดียวกัน ( $10^{10}$  CFU/ml) พบว่า ค่าการดูดกลืนแสงมีระดับต่ำมากซึ่งแสดงให้เห็นถึงไม่มีการจับกันระหว่างรีคอมบิแนนท์แอนติบอดีและแอนติเจน แต่ที่ความเข้มข้นของเฟจดั้งเดิม (Wild type)  $10^{11}$  CFU/ml ค่าการดูดกลืนแสงมีระดับที่สูงขึ้นซึ่งระดับการจับกับแอนติเจนเรียงลำดับจากมากไปน้อยได้แก่  $17.8 > 17.1 > 17.3 > 17.9 > 17.7$

ตารางที่ 4.0 ผลการเปลี่ยนจากกรดอะมิโนเมทไทโอนีนเป็นไกลซีน

Phage	17.1	17.3	17.7	17.8	17.9
<b>WT <math>10^{11}</math></b>	0.579	0.677	0.311	0.936	0.578
<b>WT <math>10^{10}</math></b>	0.169	0.182	0.11	0.397	0.156
<b>G <math>10^{10}</math></b>	1.148	1.227	0.829	1.378	0.977



ผลจากการทดสอบการจับ (bind) ของรีคอมบิแนนท์เฟจจากการเปลี่ยนแปลงกรดอะมิโนเมทไทโอนีนตำแหน่งที่ 100 (เฟจเมทไทโอนีน) เป็นกรดอะมิโนไกลซีน (เฟจไกลซีน) ที่ความเข้มข้นเดียวกัน ( $10^{10}$  CFU/ml) พบว่า ค่าการดูดกลืนแสงของเฟจไกลซีนมีระดับสูงกว่าเฟจดั้งเดิม เมื่อเพิ่มความเข้มข้นของเฟจดั้งเดิมเป็น  $10^{10}$  CFU/ml ค่าการดูดกลืนแสงของเฟจไกลซีนยังสูงกว่าเฟจดั้งเดิม แต่ลักษณะการจับกับแอนติเจน p17 มีลักษณะการจับที่เหมือนกัน นั่นคือ จับได้ดีจากมากไปน้อยดังนี้  $17.8 > 17.1 > 17.3 > 17.9 > 17.7$  ซึ่งอาจเนื่องมาจากการแสดงออกของรีคอมบิแนนท์ของเฟจไกลซีนมากกว่าเฟจดั้งเดิม หรืออาจเกิดจากความสามารถของเฟจไกลซีนมากกว่าเฟจดั้งเดิม หรืออาจเกิดจากการนับปริมาณเฟจที่ไม่ถูกต้อง ซึ่งสิ่งที่ต้องการพิสูจน์ต่อไปคือการนับเฟจถูกต้องจริงหรือไม่โดยการทดสอบโดยวิธี ELISA ซึ่งทดสอบโดยเคลือบเฟจทั้งสองชนิดบนผิวของอิลิซาเพลท (ELISA plate) โดยตรง ที่ความเข้มข้นเท่าๆกันแล้วทำการตรวจติดตามเฟจโดยการเติมคอนจูเกตที่ความเข้มข้น 1:8,000 และ 1:10,000 แล้วทำการตรวจวัดค่าการดูดกลืนแสงที่ได้ ซึ่งให้ผลดังนี้



ที่ความเข้มข้นเดียวกัน ( $10^{10}$  CFU/ml) ที่เท่ากันของเฟจดั้งเดิมและเฟจไกลซีน ค่าการดูดกลืนแสงมีความแตกต่างกันมาก แสดงให้เห็นถึงการนับปริมาณเฟจที่มีความคลาดเคลื่อนค่อนข้างสูง ถ้าปริมาณเฟจเท่ากันจริงค่าการดูดกลืนแสงจะมีปริมาณเท่าๆ กัน ซึ่งเมื่อเพิ่มปริมาณเฟจของเฟจดั้งเดิมเป็น  $10^{11}$  CFU/ml ค่าการดูดกลืนแสงใกล้เคียงกับเฟจไกลซีนที่ความเข้มข้น  $10^{10}$  CFU/ml นั้นแสดงให้เห็นถึงปริมาณเฟจที่แท้จริงของเฟจดั้งเดิมมีค่าน้อยกว่าความเป็นจริงประมาณ 10 เท่าเมื่อเทียบกับเฟจไกลซีน

### 3.7 การปรับปรุง scFv p17 ด้วยการ grafting บน SS-bond free scFv

ปัจจุบันส่วนของแอนติบอดี ที่เรียกว่า single chain Fv (scFv) สามารถประยุกต์ใช้ใน gene therapy โดยการใช้เทคนิค intracellular antibodies (intrabodies) strategy โดยที่ scFv สามารถไปยังเป้าหมายที่จำเพาะทั้ง cellular cytoplasm, endoplasmic reticulum (ER), nucleus และ mitochondria แต่ส่วนใหญ่ประสบความสำเร็จที่บริเวณ ER จากการศึกษาพบว่า intrachain disulfide bond ใน scFv ที่มีส่วนต่อการม้วนตัว และความเสถียร โดยปกติแล้วไม่เสถียรใน reducing environment ที่พบในไซโตพลาสซึม ทำให้ไม่สามารถทำหน้าที่ได้ ดังนั้นการพัฒนา scFv ที่สามารถใช้งานได้ใน ไซโตพลาสซึมจึงเป็นโจทย์วิจัยที่สำคัญ จากการทดลองมีผู้พบ scFv ที่ไม่มีไดซัลไฟด์ และยังคงทำงานได้ในไซโตพลาสซึม และได้ศึกษาการ graft CDRs ที่จับกับแอนติเจนไปบน framework regions ของ SS-bond free scFvs แต่ส่วนใหญ่แล้วมักไม่ประสบความสำเร็จ การจำลองโดยวิธีคอมพิวเตอร์ในการ grafting และการวิเคราะห์พลศาสตร์ที่มีผลกระทบต่อพันธะไดซัลไฟด์ น่าจะมีส่วนช่วยประเมินความเสถียรของโครงสร้างโมเดลที่สร้างขึ้นได้

#### วัตถุประสงค์

เพื่อสร้างโครงสร้าง antibody จากการ graft CDRs ของ scFv anti-p17 บน framework SS-bond free scFvs และคำนวณพลังงานยึดจับ และศึกษาจำลองพลศาสตร์ของโครงสร้างในน้ำ

### ขั้นตอนการคำนวณ

1. หาโครงสร้างของแอนติบอดี framework SS-bond free scFvs ที่เหมาะสมจากฐานข้อมูล โปรตีน (Protein Data Bank, PDB)
2. Refinement โครงสร้าง scFv ด้วยกลศาสตร์เชิงโมเลกุล (Molecular Mechanics, MM) และ พลศาสตร์เชิงโมเลกุล (Molecular Dynamics, MD)
3. เลือกบริเวณ CDRs ของ scFv anti-p17 เพื่อ graft ลงบน framework SS-bond free scFvs และคำนวณพลังงานยึดจับ
4. จำลองพลศาสตร์ของโครงสร้าง scFv ในน้ำ

### ผลการทดลอง

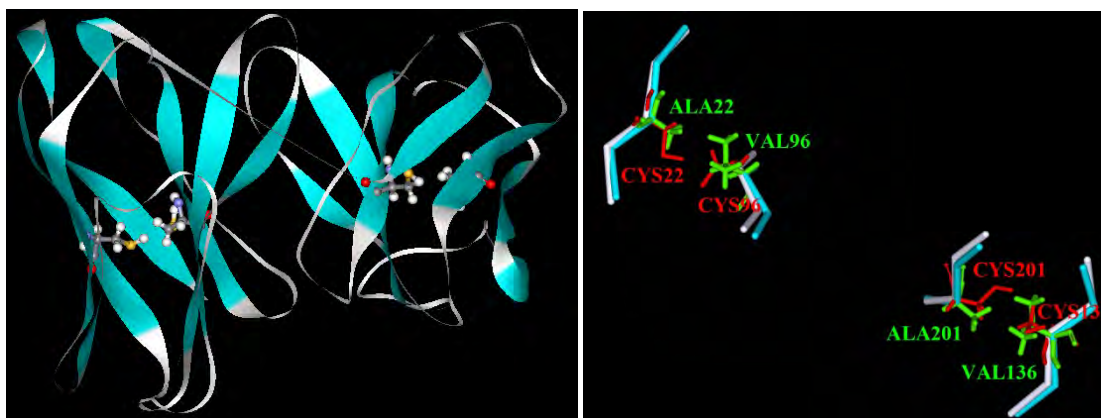
จากการศึกษาข้อมูลการทดลองพบว่า Anti-Ras scFv (2VH5.pdb) เป็นโครงสร้างที่ไม่มี intra-domain disulfide โดยที่ Cysteine เป็น ALA22-VAL96 ใน VH และ VAL136-ALA201 in VL และเมื่อโมเดลโดยการ grafting บริเวณ CDRs ของ scFv anti-p17 ลงบนบริเวณ loops ของ anti-Ras และคำนวณพลังงานยึดจับพบว่า ได้ผลดังตารางที่ 4.1 พบว่าโมเดลที่ 2 และ 3 เป็นโมเดลที่ให้พลังงานยึดจับที่ดี โดยโมเดลที่ 2 เป็นไปตามการ alignment โดยตรง ส่วนโมเดลที่ 3 เป็นโมเดลที่ขยับ CDRs ของ scFv anti-p17 ให้อยู่บริเวณปลายของ loop ของ anti-Ras ซึ่งจากโครงสร้าง 3 มิติน่าจะเป็นโมเดลที่เหมาะสมมากกว่า

ตารางที่ 4.1 Models for selected CDRs anti-p17 loop sequences grafting on selected position of anti-Ras scFv

Model No.	Anti-Ras scFv replacement position	Anti-p17 scFv loop selection for grafting	Interaction Energy (kcal/mol)	Note
1	VH;31-35, 50-64, 98-103 VL;137-152,161-166,202-209	VH;31-36, 50-66, 99-104 VL;138-153,169-174,208-216	-45.51	CDRs anti-p17 graft on anti-Ras Graft position based on alignment
2	<b>VH;30-34, 50-64, 98-103</b> <b>VL;137-152, 161-166, 202-209</b>	<b>VH;31-36, 50-66, 99-104</b> <b>VL;138-153,169-174,208-216</b>	<b>-60.77</b>	<b>CDRs anti-p17 graft on anti-Ras Graft position based on alignment</b>
3	<b>VH;28-33, 46-62, 98-103</b> <b>VL;135-150, 160-165, 202-210</b>	<b>VH;31-36, 50-66, 99-104</b> <b>VL;138-153,169-174,208-216</b>	<b>-54.50</b>	<b>CDRs anti-p17 graft on extended CDRs of anti-Ras scFv</b>
4	VH;24-37, 43-66, 94-106 VL;131-153, 155-173, 200-212	VH;26-39, 45-68, 94-106 VL;135-157,163-181,206-218	-38.65	Extended CDRs anti-p17 graft on extended CDRs of anti-Ras scFv
5	VH; <b>26-33</b> , 51-58, 97-104 VL;140-145,164-166,202-210	VH;31-36, 50-66, 99-104 VL;138-153,169-174,208-216	-41.51	CDRs anti-p17 graft on CDR of anti-Ras
6	VH; <b>29-35</b> , 51-58, 97-104 VL;140-145,164-166,202-210	VH;31-36, 50-66, 99-104 VL;138-153,169-174,208-216	-49.70	CDRs anti-p17 graft on CDR of anti-Ras
7	VH;26-33, 51-58, 97-104 VL;140-145,164-166,202-210	-	-42.26	Anti-Ras binding with p17 wild type, no grafting
8	-	VH;31-36, 50-66, 99-104 VL;138-153,169-174,208-216	-66.87	Anti-p17 scFv interaction with p17 wild type

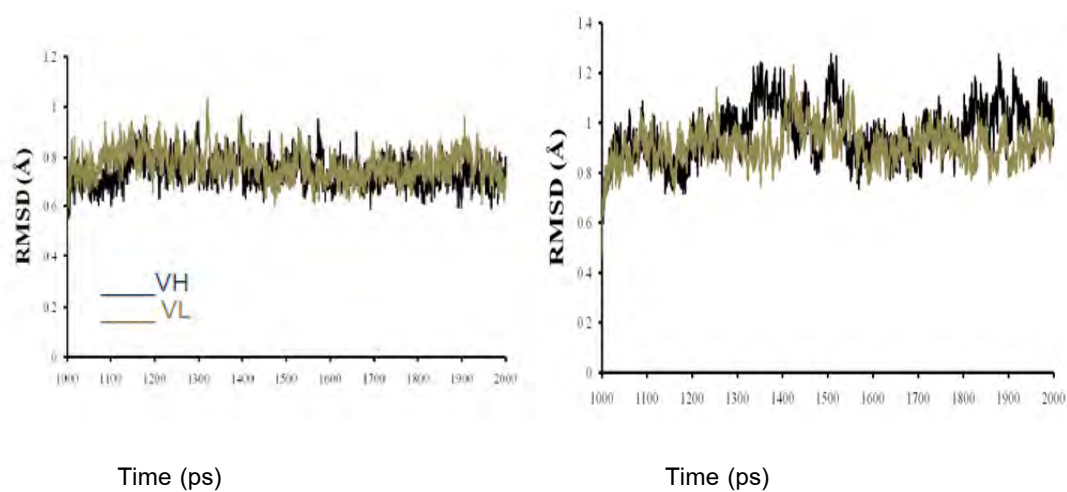
และได้ศึกษาการจำลองพลศาสตร์ของโครงสร้าง scFv ในน้ำของระบบ มี/ไม่มี intra-domain disulfide ของระบบ ดังต่อไปนี้

- 1) **Anti-Ras scFv** (2UZI.pdb) with intra-domain disulfide, CYS22-CYS96 in VH and CYS136-CYS201 in VL, were modeled to be in sulfhydryl form
- 2) **Anti-Ras scFv** (2VH5.pdb) without intra-domain disulfide, ALA22-VAL96 in VH and VAL136-ALA201 in VL
- 3) **Anti-p17 scFv** with intra-domain disulfide CYS22-CYS96 in VH and CYS137-CYS207 in VL were modeled to be in sulfhydryl form.
- 4) **Anti-p17 scFv** without intra-domain disulfide, ALA22-VAL96 in VH and VAL137-ALA207 in VL

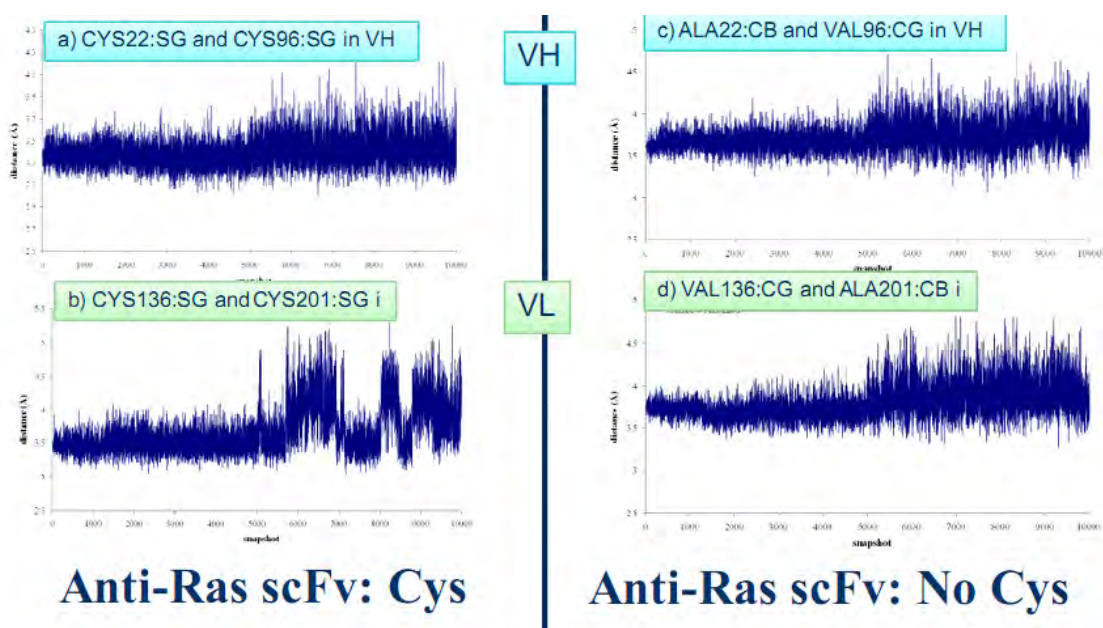


รูปที่ 4.1 โครงสร้าง Anti-Ras scFv ที่มี/ไม่มี intra-domain disulfide

เมื่อจำลองพลศาสตร์ในน้ำและศึกษาพลศาสตร์การเคลื่อนของโครงสร้างโดยการวิเคราะห์ root mean square displacement (RMSD) ใน VH และ VL สำหรับระบบ anti-Ras scFv ดังแสดงในรูปที่ 4.2 พบว่าระบบที่ไม่มีไดซัลไฟด์มีการเคลื่อนที่มากกว่า และเมื่อวิเคราะห์ระยะทางเฉพาะระหว่างบริเวณ intra-domain รูปที่ 4.3 พบว่าระยะ intra-domain ของ VH ค่อนข้างคงที่เมื่อเปรียบเทียบกับ VL แต่สำหรับระบบ anti-p17 scFv รูปที่ 4.4 แสดงการเปลี่ยนแปลงโครงสร้างค่อนข้างมากเมื่อเทียบกับ anti-Ras scFv ยืนยันได้จากการเปรียบเทียบ RMSD ของโครงสร้างก่อน และหลังการจำลองพลศาสตร์ในรูปที่ 4.5

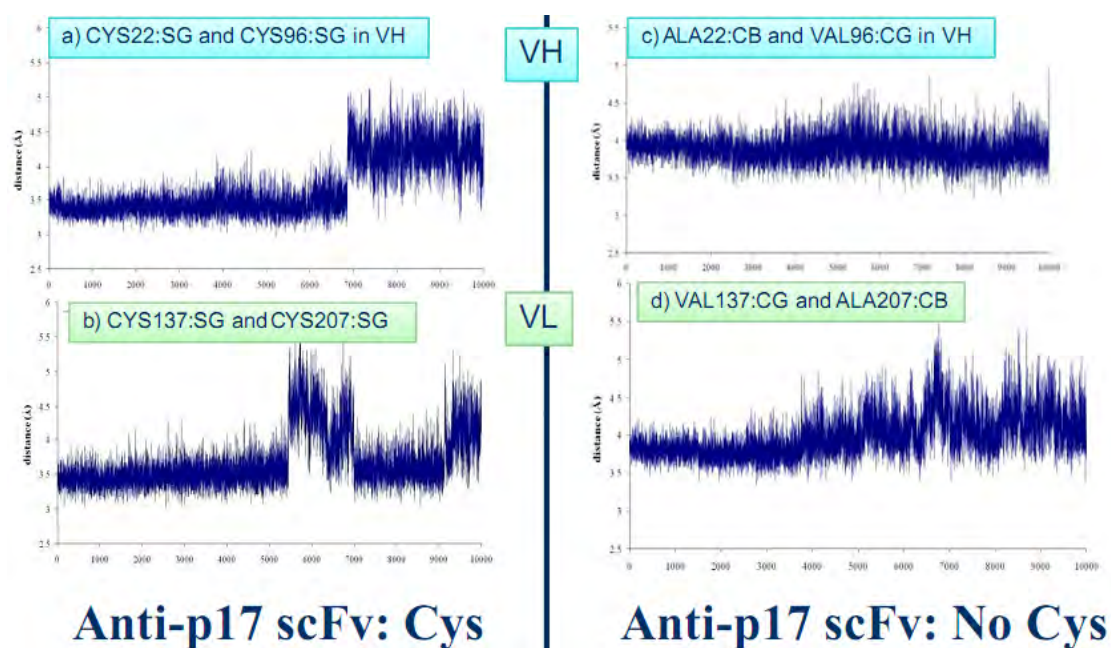


รูปที่ 4.2 RMSD ใน VH และ VL ของ Anti-Ras scFv with (left) / without (right) intra-domain disulfide

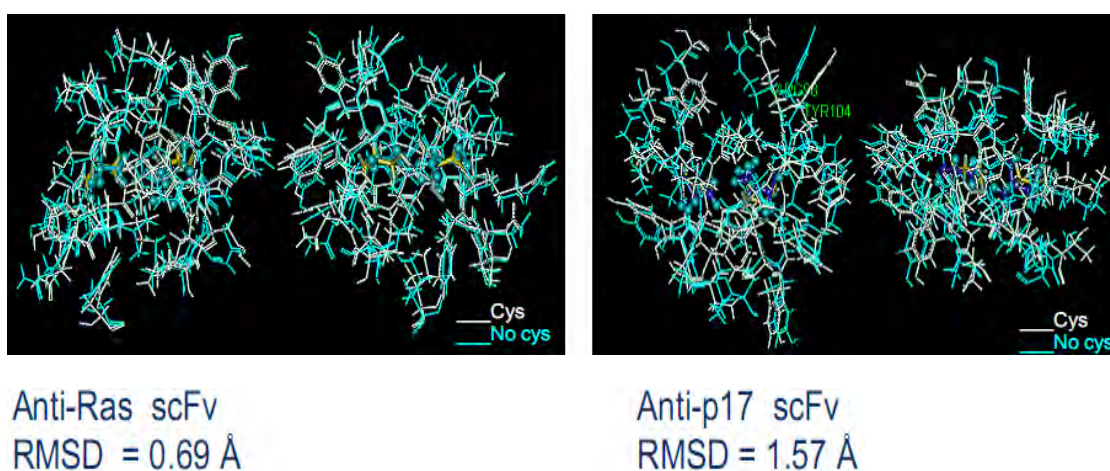


รูปที่ 4.3 ระยะทางระหว่างบริเวณ intra-domain ในระบบ anti-Ras scFv





รูปที่ 4.4 ระยะทางระหว่างบริเวณ intra-domain ในระบบ anti-p17 scFv



รูปที่ 4.5 RMSD ก่อนและหลังการจำลองพลศาสตร์ของ anti-Ras scFv และ anti-p17 scFv

เมื่อคำนวณพลังงานยึดจับในน้ำของทั้งสี่ระบบ ดังตารางที่ 4.2 พบว่า ระบบ anti-Ras scFv ทั้งมีและไม่มีไดซัลไฟด์ มีความเสถียรมากกว่า และค่าพลังงานไม่ต่างกันมากนัก ต่างจากระบบ anti-p17 scFv

ตารางที่ 4.2 พลังงานยึดจับสำหรับระบบ anti-Ras scFv และ anti-p17 scFv

	Anti-Ras scFv				Anti-p17 scFv			
	Cys		No Cys		Cys		No Cys	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
ELE	-20.82	10.36	-20.21	9.52	-227.32	14.07	-239.12	17.98
VDW	-100.55	5.02	-100.24	4.89	-72.28	4.32	-66.47	4.26
GAS	-121.37	11.54	-120.45	10.71	-299.6	14.09	-305.59	18.29
PBSUR	-11.98	0.34	-11.92	0.3	-11.02	0.29	-10.17	0.37
PBCAL	57.67	9.45	61.65	8.16	280.63	15.14	261.26	16.58
PBSOL	45.69	9.42	49.74	8.02	269.61	15.13	251.08	16.42
PBELE	36.86	4.44	41.44	4.31	53.31	10.59	22.14	7.18
PBTOT	<b>-75.68</b>	5.1	<b>-70.71</b>	4.99	<b>-29.99</b>	8.68	<b>-54.50</b>	6.79
GBSUR	-11.98	0.34	-11.92	0.3	-11.02	0.29	-10.17	0.37
GB	63.71	9.62	65.1	8.62	273.07	13.31	262.06	16.25
GBSOL	51.73	9.61	53.19	8.5	262.04	13.28	251.88	16.11
GBELE	42.89	3.78	44.9	3.54	45.75	6.37	22.94	5.44
GBTOT	-69.64	4.31	-67.26	4.64	-37.55	5.03	-53.7	5.47

นอกจากนี้ ได้วิเคราะห์พันธะไฮโดรเจนบริเวณรอบ intra-domain disulfide และทำให้ทราบหุมุม อะมีโนที่ตรง  
โครงสร้าง intra-domain ให้อยู่ด้วยกันเปรียบเทียบระหว่างระบบที่มีและไม่มีไดซัลไฟด์ บริเวณที่แตกต่างกันได้  
แสดงกรอบสีแดงไว้ ข้อมูลนี้จะเป็นประโยชน์ในการปรับปรุงโครงสร้าง scFv ต่อไป

ตารางที่ 4.3 Hydrogen bond analysis

Hydrogen Bond Analysis in Anti-Ras scFv					
Disulfide System			No Disulfide System		
	DONOR	ACCEPTOR		DONOR	ACCEPTOR
	(res@atom)	(res@atom)		(res@atom)	(res@atom)
			% Occupied		
				%occupied	distance
CYS22	7@O	21@H	21@N	99.92	2.916
	22@O	79@H	79@N	99.2	2.943
	79@O	22@H	22@N	91.52	2.84
	5@O	23@H	23@N	90.82	3.099
	21@O	7@H	7@N	87.8	2.828
	23@O	5@H	5@N	83.54	3.133
	21@O	7@HG	7@OG	51.48	2.873
CYS96	95@O	37@H	37@N	100	2.86
	97@O	35@H	35@N	99.5	2.963
	35@O	97@H	97@N	98.34	2.968
	37@O	95@H	95@N	90.86	2.875
	96@O	106@H	106@N	85	3.125
CYS136	118@O	137@H	137@N	98.92	3.005
	184@O	136@H	136@N	96.62	2.931
	135@O	120@H	120@N	95.14	2.984
	137@O	118@H	118@N	94.78	3.05
	136@O	184@H	184@N	92.52	3.119
	120@O	135@H	135@N	89.86	2.965
	183@OD1	137@HE	137@NE	89.6	2.933
	183@OD1	137@HH2	137@NH2	70.16	3.002
	183@OD2	137@HE	137@NE	42.16	3.12
	120@O	135@HG1	135@OG1	37.94	2.83
	184@O	136@HG	136@SG	22.3	3.296
	183@OD2	137@HH2	137@NH2	18.62	3.069
	118@O	136@HG	136@SG	18.06	3.236
	119@O	120@HG	120@OG	13.26	2.758
CYS201	200@O	149@H	149@N	99.8	2.93
	202@OE1	149@HH	149@OH	99.54	2.763
	201@O	212@H	212@N	98.88	2.946
	202@O	147@H	147@N	97.1	2.981
	149@O	200@H	200@N	96.42	3.09
	147@O	202@H	202@N	71.92	3.04
	202@OE1	147@HD22	147@ND2	24.64	3.039
	213@OE1	214@H	214@N	9.24	3.291
ALA22	7@O	21@H	21@N	99.8	2.915
	22@O	79@H	79@N	99.78	2.946
	79@O	22@H	22@N	99.44	2.874
	21@O	7@H	7@N	97.02	2.855
	5@O	23@H	23@N	95.68	3.029
	23@O	5@H	5@N	93.86	3.078
	21@O	7@HG	7@OG	54.54	2.722
	21@O	7@HG	7@OG	12.56	2.927
VAL96	95@O	37@H	37@N	100	2.891
	35@O	97@H	97@N	99.24	2.977
	97@O	35@H	35@N	98.4	3.065
	37@O	95@H	95@N	93.32	2.848
	96@O	106@H	106@N	90.22	3.136
	97@O	35@HD22	35@ND2	7.6	3.07
VAL136	135@O	120@H	120@N	99.82	2.951
	120@O	135@H	135@N	99.56	2.921
	184@O	136@H	136@N	99.5	2.896
	136@O	184@H	184@N	99.18	3.07
	118@O	137@H	137@N	98.44	3.038
	137@O	118@H	118@N	94.88	2.995
	183@OD1	137@HE	137@NE	84.68	2.93
	120@O	135@HG1	135@OG1	69.88	2.812
	183@OD1	137@HH2	137@NH2	60.76	3.042
	183@OD2	137@HE	137@NE	50.98	3.122
	183@OD2	137@HH2	137@NH2	25.16	3.056
	119@O	120@HG	120@OG	14.58	2.757
	138@O	137@HH1	137@NH1	11.82	2.931
ALA201	202@OE1	149@HH	149@OH	99.9	2.745
	200@O	149@H	149@N	99.9	2.926
	201@O	212@H	212@N	99.52	2.925
	149@O	200@H	200@N	97.3	3.073
	202@O	147@H	147@N	94.12	3.102
	147@O	202@H	202@N	66.36	3.04

3.9 การปรับปรุง scFv anti-p17 ด้วยการ grafting CDRs ของ scFv anti-p17 บน scFv ที่ละลายได้ดี (scFv13-R4)

ScFv13-R4

MAEVQLVESG GSLVKPGGSL RLSCAASGFT FSNYSMNWVR QAPGKGLEWI 50  
 SSISGSSRYI YYADFKGRF TISRDNATNS LYLQMNLSRA EDTAVYYCVR 100  
 SSITIFGGGM DVWGRGTLVT VSSGGGGSGG GSGGGGSGQS VLTQPASVSG 150  
 SPGQSITISC AGTSSDVGGY NYVSWYQQHP GKAPKLMIE DSKRPSGVSN 200  
 RFGSKSGNT ASLTISGLQA EDEADYYCSS YTTRSTRVFG GGTKLAVLG 249

ScFvp17

QIQLVQSGPE LKKPGETVEI SCKASGYTFT DYGMNWMKQA PGKSLKWMGW 50  
 INTYTGEPTY ADEFKGRFAF SLETSASTAY LQINNLSKED MATYFCRSM 100  
 KGSYWGQGTI VTVSAGGGGS GGGGSGGGGS DVVMTQTPLT LSVTIGQPAS 150  
 ISCKSSQSL GSDGKTFLNW LLQRPQGSPK RLIYLVSCLD SGVPDRFTGS 200  
 GSGTDFTLKI SRVEADLGV YYCWQGTHLP QTFGGGTTKLE IK 242

หมายเหตุ: ส่วนสีเหลืองคือ linker

ScFvp17 grafting เป็นลำดับอะมิโนของแอนติบอดีที่ต้องการสร้าง โดยมี frame work เป็นส่วนที่ละลายได้ดีมาจาก ScFv13-R4 และส่วนยึดจับ CDRs ของ ScFvp17

MAEVQLVESG GSLVKPGGSL RLSCAASGFT FSDYGMNWR QAPGKGLEWI 50  
 SWINTYTGEPT YADEFKGRF TISRDNATNS LYLQMNLSRA EDTAVYYCVR 100  
 SMKGSYWGRT TLVTVSSGGG GSGGGGSGGG GSQSVLTQPA SVSGSPGQSI 150  
 TISCKSSQSL LGSDGKTFLN WYQHPGKAP KLMIYLVSCL DGVSNRFGS 200  
 KSGNTASLTI SGLQAEDAD YYCWQGTHLP QTVFGGGTTKL AVLG 244



เมื่อเทียบลำดับอะมิโนในฐานข้อมูลพบว่าควรใช้แม่แบบ

ส่วน VH: จาก pdb|1T2J|A Chain A, Crystal Structure Of A Human Vh Domain

Length=116 Score = 186 bits (471), Expect = 2e-48, Method: Compositional matrix adjust.

Identities = 95/121 (78%), Positives = 101/121 (83%), Gaps = 5/121 (4%)

ส่วน VL: จาก pdb|2JB5| Chain L, Fab Fragment In Complex With Small Molecule Hapten,

Crystal Form-1 Score = 184 bits (467), Expect = 7e-48, Method: Compositional matrix adjust.

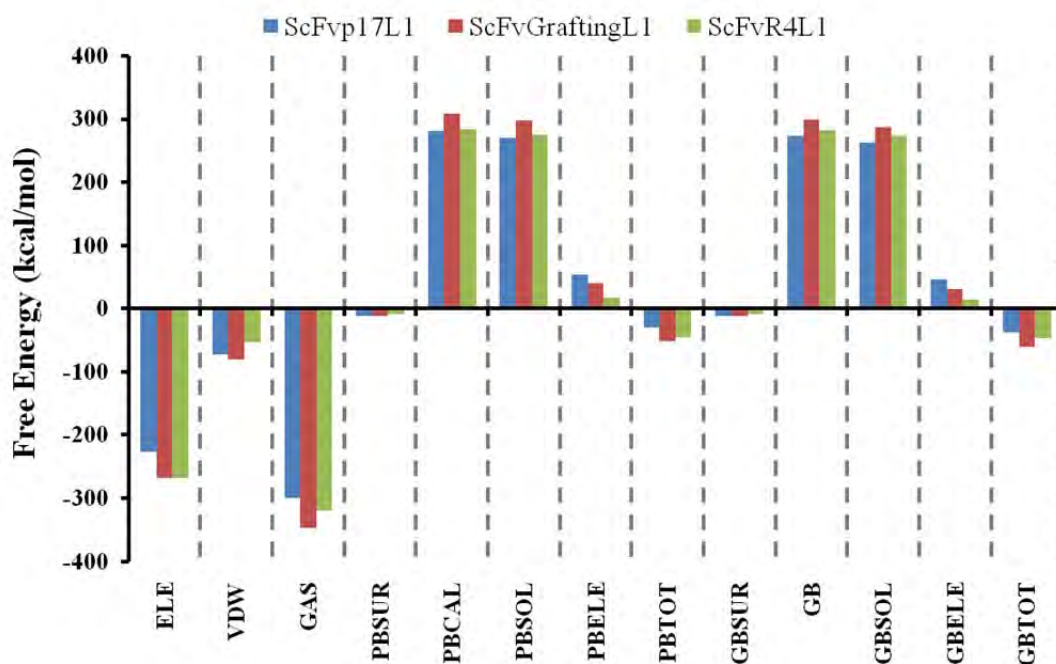
Identities = 94/109 (86%), Positives = 96/109 (88%), Gaps = 1/109 (0%)

ส่วนโครงสร้าง VH+VL โดยรวม: จาก pdb|2GHW|B Chain B, Crystal Structure Of Sars Spike Protein Receptor Binding Domain In Complex With A Neutralizing Antibody

Length=247 Score = 249 bits (637), Expect = 4e-67, Method: Compositional matrix adjust.

Identities = 137/248 (55%), Positives = 169/248 (68%), Gaps = 23/248 (9%)

เมื่อศึกษาการยึดจับของโครงสร้างเชิงซ้อน **ScFv p17 (wt)**, **ScFv p17 Grafting** และ **ScFv13-R4** กับ ligand L1 (wt) ได้ผลดังรูป และตารางด้านล่าง



รูปที่ 4.6 กราฟแสดงค่า Binding energy ของ ScFvp17 (a), ScFvGrafting (b) and ScFvR4 (c) complexes with L1 ligand

ตารางที่ 4.4 ค่า Binding energy เปรียบเทียบระหว่าง 3 โมเดล

Method	Contribution	Model Names		
		ScFvp17 (wt)	ScFvGrafting	ScFvR4
MM	$\Delta E_{LE}$	-227.32	-268.11	-267.27
	$\Delta V_{DW}$	-72.28	-79.65	-52.65
	$\Delta G_{AS}$	-299.6	-347.76	-319.92
PBSA	$\Delta P_{B_{SUL}}$	-11.02	-11.71	-8.93
	$\Delta P_{B_{CAL}}$	280.63	308.8	284.17
	$\Delta P_{B_{SOL}}$	269.61	297.09	275.23
	$\Delta P_{B_{ELE}}$	53.31	40.69	16.9
	$\Delta P_{B_{TOT}}$	-29.99	-50.66	-44.68
GBSA	$\Delta G_{B_{SUL}}$	-11.02	-11.71	-8.93
	$\Delta G_{B_{CAL}}$	273.07	298.68	282.14
	$\Delta G_{B_{SOL}}$	262.04	286.97	273.2
	$\Delta G_{B_{ELE}}$	45.75	30.57	14.87
	$\Delta G_{B_{TOT}}$	-37.55	-60.79	-46.71

### 3.9 การสร้างโครงสร้าง scFv G46 และศึกษาการยึดจับกับ gag-pol sequence

ได้สร้างโครงสร้างแอนติบอดีสายเดี่ยวชนิดใหม่ สำหรับ HIV epitope ที่บริเวณ C-Terminal ของ P17 (scFv G46) ซึ่งเป็นแอนติบอดีสายเดี่ยวที่ยึดกับ epitope บน p17 ที่ไม่จำเป็นต้องถูก cleave ออกจาก p24 และแอนติบอดีนี้ไม่จับกับ p24 ที่ cleave ออกจาก p17 และทำนายบริเวณที่ใช้ยึดจับกับ ศึกษาการยึดจับกับของแอนติบอดีดังกล่าวกับ gag-pol sequence โดยใช้ molecular docking ระหว่าง scFv G46 กับ peptide สายสั้น, 10 amino acids residues จาก gag-pol sequence โดยผลการคำนวณแรงแคติวิตี้ของแอนติบอดี สามารถนำไปใช้เปรียบเทียบกับข้อมูลการทดลอง

**Receptor:** G46m3.pdb , scFv binds with p17p24, bind กับ epitope บน p17 ที่ไม่จำเป็นต้องถูก cleave ออกจาก p24. It's not bind with p24 alone,

**Ligand:** P17\_24 sequence, scanning window 10 amino acids sequence, 1L6N.pdb

M<sup>1</sup>GARASVLSGGELDKWEKIRLRPGGKKQYKLKHIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQT  
GSEELRSLYNTIAVLYCVHQRIDVKDTKEALDKIEEQNKSKKKAQQAADTGNNQVSQ<sup>80</sup>NYPIVQNLQGQ  
MVHQAISPRTLNAWVKVVEEKAFSPVPMFSALSEGATPQDLNTMLNTVGGHQAAMQMLKETINEEAAEW  
DRLHPVHAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTS<sup>230</sup>

**Method:** Molecular Docking with Cdocker, DS 2.5

ID	Peptide Sequence	Cdocker Interaction Energy (kcal/mol)	Deviation from straight chain RMSD heavy atoms (Å)
a1	MGARASVLSG	-32.26	11.61
a2	SVLSGGELDK	-37.03	7.60
a3	GELDKWEKIR	-46.29	8.50
a4	WEKIRLRPGG, linear	-55.56	11.61
a4a	LRPGGKKQYK	fail	-
a5	KKQYKLKHIV, linear	-64.06	7.97
a6	LKHIVWASRE, helix	-60.32	10.45
a7	WASRELERFA, helix	-60.17	10.58
a8	LERFAVNPGL	-39.16	9.48
a9	VNPGLLETSE	fail	-
a10	LETSEGCRQI	fail	-
a11	GCRQILGQLQ	-48.76	9.88
a12	LGQLQPSLQT	-40.76	8.73
a13	PSLQTGSEEL	-41.39	6.88
a14	GSEELRSLYN	-44.32	9.52
a15	RSLYNTIAVL	-39.83	9.02
a16	TIAVLYCVHQ	-37.22	8.22
a17	YCVHQRIDVK	-43.84	8.11
a18	RIDVKDTKEA	-29.45	6.13
a19	DTKEALDKIE	-37.54	8.39
a20	LDKIEEQNK, helix	-53.39	6.91
a21	EEQNKSKKKA	-50.12	13.90
a22	SKKKAQQAQA	-44.46	13.05
a23	QQAADTGNN	-47.41	5.80
a24	D <sup>+</sup> TGNSQVSQ <sup>+</sup> , linear	-60.85	10.24
a25	SQVSQNYPIV, linear	-51.99	8.05
a26	NYPIVQNLQG	-34.69	6.65
a27	QNLQGQMVHQ	-33.93	9.98
a28	QMVHQAI SPR	-50.10	8.02
a29	AISPRTLNAW	-46.56	13.34
a30	TLNAWVKVVE	-41.97	7.30
a31	VKVVEEKAFS	-47.43	7.40
a32	EKA FSPEVIP, loop, ~linear	-55.02	7.66
a33	PEVIPMFSAL	-42.17	6.82
a34	MFSALSEGAT	-37.69	7.68
a35	SEGATPQDLN	-47.78	6.56
a36	PQDLN TMLNT	-44.39	7.32
a37	TMLNTVGGHQ	-39.89	7.72
a38	VGGHQAA MQM	-43.82	8.10
a39	AAMQMLKETI	-47.50	8.02
a40	LKETINEEAA	-41.18	8.53
a41	NEEAAEWDR L	-43.30	8.30
a42	EWDR LHPVHA	-37.21	9.03
a43	HPVHAGPIAP	-40.11	7.97
a44	GPIAPGQMRE	-44.31	7.47
a45	GQMREPRGSD	-34.64	6.37
a46	PRGSDIAGTT	-48.29	11.49
a47	IAGTTSTLQE	-32.33	7.50
a48	STLQE QIGWM, helix	-54.24	9.03
a49	QIGWMTHNPP	-40.57	5.05
a49	THNPPIPVGE	-32.78	7.55
a50	IPVGEIYKRW	-47.15	10.35
a51	IYKRWILGL	-45.38	15.44
a52	IILGLNKIVR, helix	-53.19	7.89
a53	NKIVRMYSPT	-42.55	8.98

**Results:**

**Suggested epitope:**

Linear part; a5, a24 sequences

Helix part; a6, a7, a20 sequences

### 3.10 การศึกษาอันตรกิริยาระหว่าง antibody-CD147

ได้สร้างโครงสร้างสามมิติระหว่างสารเชิงซ้อน scFv-M6-1B9 and CD147 โดยใช้โปรแกรม Global Range Molecular Matching (GRAMM) และศึกษาอันตรกิริยาที่เกิดขึ้น โดยทำนายอิพิโทป ในสื่บริเวณ โดย domain 1 ของ CD147 อาจเป็นบริเวณที่สำคัญต่อการยึดจับกับ M6-1B9 หากแต่การระบบนี้ยังต้องศึกษาต่อทางการทดลองจากห้องปฏิบัติการมาสนับสนุนผลการคำนวณที่ได้

#### 4. บทสรุป

การสร้างโครงสร้างแอนติบอดีสามารถทำได้ทั้งจากโปรแกรมสำเร็จรูป และจากการให้ข้อมูลลำดับอะมิโน โหนดเว็บไซต์เพื่อสร้างเป็นโครงสร้างสามมิติ บริเวณโครงสร้างที่พบว่าโมเดลได้ยากได้แก่ บริเวณ flexible loops โดยเฉพาะอย่างยิ่ง ที่บริเวณยึดจับหรือบริเวณ complementary determining regions (CDRs) จำเป็นต้องพิจารณา และประเมินโครงสร้างที่ได้อย่างระมัดระวัง งานวิจัยนี้ได้สร้างโครงสร้างแอนติบอดีหลายชนิดได้แก่ scFv anti-p17, scFv G46 ที่ยึดกับ HIV epitope บน p17 ที่ไม่จำเป็นต้องถูก cleave ออกจาก p24, antibody-CD147 จากการทำ homology modeling เทียบกับแม่แบบที่มีลักษณะคล้ายคลึงกัน นอกจากนี้ยังศึกษาพลังงานการยึดจับในน้ำโดยการจำลองพลศาสตร์ คำนวณพลังงานอิสระจาก MM-PBSA และ MM-GBSA โดยพบหมู่อะมิโน CDRs of scFv anti-p17 ที่สำคัญและมีค่า absolute relative decomposed energy มากกว่า 2.00 kcal/mol เมื่อเปรียบเทียบกับ wild type ได้แก่ MET100, LYS101, ASN169, HIS228, and LEU229 อีกทั้งผลการทำ alanine scanning ยังสามารถระบุอิทธิพลของ side chain ในบริเวณ CDRs ได้อีกด้วย ข้อมูลดังกล่าวเป็นประโยชน์ในการนำไปใช้ปรับปรุงโครงสร้างแอนติบอดีให้มีประสิทธิภาพยึดจับกับแอนติเจนได้ดีขึ้น โดยผลการเปลี่ยน M100R และ M100G ทำให้ scFv anti-p17 มีประสิทธิภาพการยึดจับได้ดีขึ้นกับ HIV-1 แอนติเจนที่พบในธรรมชาติ สอดคล้องกับผลการทดลอง indirect ELISA ข้อควรระวังในการเปรียบเทียบผลการทดลองพบว่า competitive ELISA และ indirect ELISA ให้ผลการทดลองที่แตกต่างกัน จำเป็นต้องใช้ทั้งสองข้อมูลในการประเมินผลที่ได้

นอกจากนี้ยังศึกษาการปรับปรุงความเสถียรของ scFv ใน cytoplasm เนื่องจากปัจจุบันส่วนของแอนติบอดี ที่เรียกว่า single chain Fv (scFv) สามารถประยุกต์ใช้ใน gene therapy โดยใช้เทคนิค intracellular antibodies (intrabodies) strategy โดยที่ scFv สามารถไปยังเป้าหมายที่จำเพาะทั้ง cellular cytoplasm, endoplasmic reticulum (ER), nucleus และ mitochondria แต่ส่วนใหญ่ประสบความสำเร็จที่บริเวณ ER จากการศึกษาพบว่า intrachain disulfide bond ใน scFv ที่มีส่วนต่อการม้วนตัว และความเสถียร โดยปกติแล้วไม่เสถียรใน reducing environment ที่พบในไซโตพลาสซึม ทำให้ไม่สามารถทำหน้าที่ได้ ดังนั้นการพัฒนา scFv ที่สามารถใช้งานได้ใน ไซโตพลาสซึมจึงเป็นโจทย์วิจัยที่สำคัญ งานวิจัยนี้ได้ทำการ grafting ใช้ CDRs ของ scFv anti-p17 แทนลงไปบน CDRs บน framework ของ scFv ที่เสถียรได้ใน cytoplasm โดยไม่ต้องมีพันธะไดซัลไฟด์ (scFv anti-Ras) และการใช้ CDRs ของ scFv anti-p17 แทนลงไปบน CDRs บน framework ของ scFv ที่ละลายได้ดีใน cytoplasm (scFv13-R4) ซึ่งกำลังทำการทดลองเปรียบเทียบผลจากทางห้องปฏิบัติการ จากข้อมูลการทดลองก่อนหน้านี้พบว่า แม้ว่าเมื่อเปลี่ยน cysteine เป็น val และ ala พบว่า scFv จะค่อนข้างเสถียรขึ้นเนื่องจาก reduced dithiol form ของ scFv fragment แต่ก็ไม่จำเป็นว่า การเปลี่ยน disulfides in scFv fragments เป็นคู่ Val-Ala pairs จะทำให้ปรับปรุง effect of cytoplasmically expressed intrabodies เสมอไปในทุกกรณี หากการวิเคราะห์และจำลองพลศาสตร์ของโครงสร้างแอนติบอดี น่าจะมีส่วนช่วยประเมินความเสถียรของโครงสร้างโมเดลที่สร้างขึ้นได้

## 5. Output จากโครงการวิจัยที่ได้รับทุนจาก สกว.

### 1. ผลงานตีพิมพ์ในวารสารวิชาการนานาชาติ (ระบุชื่อผู้แต่ง ชื่อเรื่อง ชื่อวารสาร ปี เล่มที่ เลขที่ และ หน้า) หรือผลงานตามที่คาดไว้ในสัญญาโครงการ

- Lee V.S.; Tue-ngeun, P.; Nangola, S.; Kitidee, K.; Jitonnorn, J.; Nimmanpipug, P.; Jiranusornkul, S.; Tayapiwatana, C. Pairwise decomposition of residue interaction energies of single chain Fv with HIV-1 p17 epitope variants. *Mol. Imm.*, 2010, 47, 982-900. (impact factor 2008 = 3.742)
- Lee V.S.; Tue-ngeun, P.; Nangola, S.; Nimmanpipug, P.; Jiranusornkul, S.; Tayapiwatana, C. "Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning" (to be submitted)
- Lee V.S.; Tue-ngeun, P.; Nimmanpipug, P.; Jiranusornkul, S.; Tayapiwatana, C. "Stabilization of Anti-Ras scFv fragments with/without disulfide bond through theoretical modeling and molecular dynamics simulations investigation" (to be submitted)
- Lee, V. S.; Jitonnorn, J.; Kasinrer, W.; Nimmanpipug, P., Tayapiwatana, C. Modeling of CD147 binding sites against ScFv: a guide for epitope mapping. (to be submitted)

### 2. การนำผลงานวิจัยไปใช้ประโยชน์

#### 2.1 เชิงสาธารณะ (มีเครือข่ายความร่วมมือ/สร้างกระแสความสนใจในวงกว้าง)

- ก่อให้เกิดความร่วมมือ MOU ระดับมหาวิทยาลัย ระหว่าง Univesiti Sains Malaysia และ มหาวิทยาลัยเชียงใหม่ ในงานด้าน molecular immulogy ทั้งทางด้านทฤษฎีและการทดลอง กับ Institute for Research in Molecular Medicine, INFORMM, Malaysia
- โครงการไทยเข้มแข็ง "โรคเอดส์" ในการออกแบบโมเดลแอนติบอดีบริเวณ cleavage site
- โครงการนักวิจัยแกนนำ ในการออกแบบโมเดลแอนติบอดี

#### 2.2 เชิงวิชาการ (มีการพัฒนาการเรียนการสอน/สร้างนักวิจัยใหม่)

มีการสนับสนุนนักวิจัยใหม่

- ผศ. ดร. ปิยรัตน์ นิมนานพิภักดิ์ ในงานวิจัยด้าน antibody-designed ankyrin
- ดร.สุพัฒน์ จิรนุสรณ์กุล ในงานวิจัยด้าน DNA-Zn finger
- ดร.สุภาภรณ์ ดอกไม้ศรีจันทร์ ในงานวิจัยด้าน molelcular modeling
- นางสาวพันธ์ทิพย์ ถือเงินกำลังศึกษาในระดับปริญญาเอก ด้านการออกแบบแอนติบอดี ณ มหาวิทยาลัยเชียงใหม่
- นายเฉลิมพล คำปา ศึกษาต่อระดับปริญญาเอก ในสาขาเกี่ยวข้องกับ biomolecular design ณ ประเทศญี่ปุ่น

#### 2.3 ผลงานตีพิมพ์ในวารสารวิชาการในประเทศ การเสนอผลงานในที่ประชุมวิชาการ

- Wahab, H. A.; Abballah, H. H.; Mavri, J.; Lee, V. S. "Implicit solvent models for studying the chemical reaction between isoflavones and guanine, 15th International Annual Symposium on Computational Science and Engineering (ANSCSE15), Bangkok University, Bangkok, Thailand, March 30-May 1, 2011 (Oral).
- Tue-nguen, P.; Nimmanpipug, P.; Nangola, S.; Tayapiwatana, C.; Lee, V. S. "Identification of the potential mutation residues of the scFv anti-P17 by alanine scanning, 15th International Annual

Symposium on Computational Science and Engineering (ANSCSE15), Bangkok University, Bangkok, Thailand, March 30-May 1, 2011 (Oral).

- Tue-nguen, P.; Nimmanpipug, P.; Nangola, S.; Tayapiwatana, C.; Lee, V. S. "Computational alanine scanning and binding free energy for predicting mutations of scFv anti-p17", The International Congress for Innovation in Chemistry (PERCH-CIC Congress VII), at Jomtien Palm Beach Hotel, Pattaya, Chonburi, May 4-7, 2011. (Poster)
- V. S. Lee, P. Nimmanpipug, K. Kodchakorn, and C. Tayapiwatana "Molecular Dynamics Simulations of Antibody ScFv Fragments without Disulfide Bond" 14th International Annual Symposium on Computational Science and Engineering (ANSCSE14), Mae Fah Luang University, Chiangrai, Thailand, March 23-26, 2010.
- Panthip Tue-ngeun, Vannajan Sanghiran Lee, Kanchanok Kodchakon, Piyarat Nimmanpipug and Chatchai Tayapiwatana, "Computational Alanine Scanning of ScFv Anti-P17 and HIV-1 P17 Epitope" the International Conference of Computational Methods for Science and Technology (ICCST-I 2010), August 4-6, 2010, Duang Tawan Centara, Chiang Mai, Thailand. (oral presentation)
- Lee V.S.; Tue-ngeun, P.; Nangola, S.; Kitidee, K.; Jitonnorn, J.; Nimmanpipug, P.; Jiranusornkul, S.; Tayapiwatana, C., "Molecular dynamics simulations of molecular recognition in single chain Fv with HIV-1 p17 epitope variants" the Pure and Applied Chemistry International Conference 2010 (PACCON 2010), January 21-23, 2010, Sunee Grand Hotel and Convention Center, Ubonratchathani, Thailand.
- Lee, V. S.; Kodchakorn, K.; Nimmanpipug, P.; Tayapiwatana, C.; "Molecular Dynamics Simulations of Antibody ScFv Fragments without Disulfide Bond" Proceeding of the 10<sup>th</sup> Annual National Symposium on Computational Science Engineering (ANSCSE14), Mae Fah Luang University, Chiang Rai, Thailand, March 23-26, 2010.
- Lee, V.S.; Tue-ngeun, P.; Nangola, S.; Kitidee, K.; Jitonnorn, J., Nimmanpipug, P. Supat Jiranusornkul, S.; Tayapiwatana, C. "Structure and dynamics of anti-P17 single chain Fv" The 4<sup>th</sup> Annual Symposium of the Protein Society of Thailand, August 26-28, 2009, Chulabhorn Research Institute Conference center, Thailand (Oral presentation)
- Tue-ngeun, P.; Jiranusornkul, S.; Jitonnorn, J.; Jaiinphon, K.; Nimmanpipug, P.; Kasinrer, W.; Tayapiwatana, C.; Lee, V. S. Molecular Dynamics of single-chain antibody to HIV epitope at c-terminal of P17, PURE AND APPLIED CHEMISTRY INTERNATIONAL CONFERENCE 2009 (PACCON 2009), January 14-16, 2009, Naresuan University, Phitsanulok, Thailand. (Oral)
- Jiranusornkul, S.; Lee, V.S; Tue-nguen, P.; Jitonnorn, J.; Jaiinphon, K.; Nimmanpipug, P.; Kasinrer, W.; Tayapiwatana, C. Computational Modelling of Single-Chain Antibody to HIV Epitope at C-Terminal of P17, NanoThailand Symposium 2008, 6th - 8th November 2008, Bangkok, Thailand. (poster)

## 2.4 กิจกรรมอื่นๆ ที่เกี่ยวข้อง

### Invited speaker

- วิทยาการอบรมเชิงปฏิบัติการเคมีเชิงคำนวณครั้งที่ 6 ณ คณะวิทยาศาสตร์ มหาวิทยาลัยมหาสารคาม ซึ่งจัดขึ้นระหว่างวันที่ 19 – 22 ตุลาคม 2552
- Lee V.S.; Tue-ngeun, P.; Nangola, S.; Kitidee, K.; Jitonnorn, J.; Nimmanpipug, P.; Jiranusornkul, S.; Tayapiwatana, C., “Molecular dynamics simulations of molecular recognition in single chain Fv with HIV-1 p17 epitope variants” the Pure and Applied Chemistry International Conference 2010 (PACCON 2010), January 21-23, 2010, Sunee Grand Hotel and Convention Center, Ubonratchathani, Thailand.
- Lee, V. S.; Kodchakorn, K.; Nimmanpipug, P.; Tayapiwatana, C.; “Molecular Dynamics Simulations of Antibody ScFv Fragments without Disulfide Bond” Proceeding of the 10<sup>th</sup> Annual National Symposium on Computational Science Engineering (ANSCSE14), Mae Fah Luang University, Chiang Rai, Thailand, March 23-26, 2010.
- Vannajan S. Lee, “Prospective of Molecular Dynamics Simulations in Drug Design” National Seminar and Workshop on Computer Aided Drug Design: Virtual Screening, Malaysian Institute of Pharmaceutical and Nutraceutical Science (IPHARM) and University of Sains Malaysia, Penang, Malaysia, December 2-8, 2010.
- Vannajan S. Lee, “Virtual screening using self organizing map”, The International Seminar on Herbal Medicine, Universitas Padjadjaran, Indonesia, November 6, 2010.
- Vannajan S. Lee, “Molecular Dynamics Simulations in Drug Design” Bandung Institute of Technology, Indonesia, November 4, 2010.
- Vannajan S. Lee, “Biomolecular and material modeling at Chiang Mai University” the International Conference of Computational Methods for Science and Technology (ICCST-I 2010), Duang Tawan Centara, Chiang Mai, Thailand, August 4-6, 2010.
- Lee, V.S.; Nimmanpipug, P.; Yu, L. “Interactions of low energy particles from plasma and ion-beam with A-DNA and silk: looking through molecular simulations”, Department of Physics, Mahidol University, Thailand. June 8, 2010.

### ผลงานตีพิมพ์ในวารสารระดับนานาชาติอื่นๆ ที่อ้างอิงสกว.

- Jitonnorn, J.; Mulholland, A. J.; Nimmanpipug, P.; Lee, V. S. “QM/MM Modeling of Substrate-Assisted Catalysis in Family 18 Chitinases: Conformational Changes and the Role of Asp142 in Catalysis in ChiB”, *Biochemistry*, 2011, 50 (21), 4697–4711 (Impact factor 2010 = 3.2).
- Jitonnorn, J.; Mulholland, A. J.; Nimmanpipug, P.; Lee, V. S. “Hybrid QM/MM study on the deglycosylation step of chitin hydrolysis catalysed by chitinase B from *Serratia marcescens*”, *Maejo Int. J. Sci. Technol*, 2011, 5(01), 47-57. (ISSN 1905-7873). (impact factor 2009 = 0.22)
- Ngaojampa, C.; Nimmanpipug, P.; Yu, L.; Anuntalabhochai, S.; Lee, V.S. “Combined quantum-mechanics/molecular-mechanics (QM/MM) dynamics simulation of A-DNA double strands irradiated by ultra-low-energy carbon ions”, *NIMB*, 2011, 269, 504–510.



- Dokmaisrijan, S.; Nimmanpipug P.; Lee, V. S., The Gas Phase Conformations and Vibrational Spectra of Valine, Leucine and Isoleucine: An Ab Initio Study, *Journal of Molecular Structure (THEOCHEM)*, 2010, 963, 130-136. (impact factor = 1.216 source : Journal Citation Reports, 2009)
- Monajjemi, M.; Lee, V. S.; Honarparvar, B.; Khaleghian, M.; Mollamin F. "Theoretical description of electromagnetic non-bonded interactions of radical, cationic, and anionic  $\text{NH}_2\text{BHNBNH}_2$  inside the B18N18 nano-ring", *Journal of Physical Chemistry C*, 2010, 2010, 114 (36), 15315–15330. (impact factor 2009 = 4.224)
- Yana, J.; Nimmanpipug, P.; Chirachanchai, S.; Dokmaisrijan, S.; Vannarat, S.; Vilaithong, T.; and Lee V. S. "Molecular dynamics simulations of Krytox-Silica-Nafion composite for high temperature fuel cell electrolyte membranes" *Polymer* 2010, 51, 4632-4638. (impact factor 2009 = 3.573)
- Ngaojampa, C.; Nimmanpipug, P.; Yu, L.; Anantalabhochai, S.; Lee, V. S. Molecular Simulations of Interactions between Low-energy Nitrogen Ions and A-DNA in Vacuum, *J. of Mol. Graph. Model.* 2010, 28, 533-539. (impact factor 2008 = 2.347)
- Lee, V.S.; Kodchakorn, K.; Jitonnorn, J.; Nimmanpipug, P.; Kongtawelert, P.; Premanode, B. Influence of Metal Cofactors and Water on the Catalytic Mechanism of Creatininase- Creatinine in Aqueous Solution from Molecular Dynamics Simulation and Quantum Study, *J. Comput Aid Mol Des*, 2010, 24, 879–886. (impact factor 2009 = 3.835)
- Wanichacheva, N.; Kamkaew, A.; Watpathomsub, S.; Lee, V.S.; Grudpan, K. 2-[3-(2-aminoethylsulfanyl)propylsulfanyl]ethanamine bearing dansyl subunits: An efficient, simple, and rapid fluorometric sensor for the detection of mercury(II) ions, *Chemistry Letters*, 2010, 39, 1099-1101.
- Wanichacheva, N.; Watpathomsub, S.; Lee, V. S.; Grudpan, K. Synthesis of a Novel Fluorescent Sensor Bearing Dansyl Fluorophores for the Highly Selective Detection of Mercury (II) Ions, *Molecules*, 2010, 15, 1798-1810. (impact factor 2009 = 1.738)
- Sangprasert, W., Lee, V. S.; Boonyawan, D.; Tashiro, K.; Nimmanpipug, P. Sulfur Hexafluoride Plasma Surface Modification of Gly-Ala and Ala-Gly as *Bombyx Mori* Silk Model Compounds: Mechanism Investigations, *J. of Mol. Struct*, 2010, 963, 130-136. (impact factor 2008 = 1.594 )
- Nokthai, P.; Shank, L.; Lee, V. S. Molecular Modeling of Peroxidase and Polyphenol Oxidase: Substrate Specificity and Active Site comparison *Int. J. Mol. Sci.* 2010, 11, 3266-3276. (impact factor = 1.387 source : Journal Citation Reports, 2009)
- Khomhoi, P.; Sangprasert, W.; Lee, V. S.; Nimmanpug, P. Theoretical Study of the Bombyx mori Silk Surface Functionalization: Quantum calculation of Glycine-Alanine unit reacting with fluorine and molecular dynamic simulation of wettability, *Chiang Mai Journal of Science*, 2010, 37(1), 106-115.
- Kareuhanon, W.; Lee, V. S.; Nimmanpipug, P.; Tayapiwatana, C.; Pattarawarapan, M. Synthesis of Molecularly Imprinted Polymers for Nevirapine using Dummy Template Imprinting Approach, *Chromatographia*, 2009, 70, 1531-1537 (impact factor 2007= 1.145).
- Tue-ngeun, P.; Lee V.S.; Nimmanpipug, P.; Chaijaruwanich, J.; Prasitwattanaseree, S.; Traisathit, P. FTIR and chemometric tools for the classification of Thai wines, *Maejo International Journal of Science and Technology*, 2009, 3(03), 446-458.

- Nimmanpipug, P., Lee, V. S.; Wolschann, Peter,, Hannongbua, S. Litchi Chinensis-derived terpenoid as anti-HIV-1 protease agent: Structural design from molecular dynamics simulations, *Molecular Simulation*, 2009, 35, 673-680. (impact factor = 1.133 source : Journal Citation Reports, 2007)
- Lee, V. S.; Nimmanpipug, P.; Mollaamin, F.; Kungwan, N.; Thanasanvorakun, S.; Monajjemi , M. Investigation of SWCNT's electrical properties and normal mode analysis: Dielectric effects, *Russ. J. Phys. Chem A*. 2009, 83(1), 2288-2296. (impact factor = 0.477 source : Journal Citation Reports, 2007)

### ภาคผนวก

- ก. Reprint Molecular Immunology 2009
- ข. Manuscript “Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning”
- ค. Manuscript “Stabilization of Anti-Ras scFv fragments with/without disulfide bond through theoretical modeling and molecular dynamics simulations investigation”
- ง. Manuscript “Modeling of CD147 binding sites against ScFv: a guide for epitope mapping”



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: [www.elsevier.com/locate/molimm](http://www.elsevier.com/locate/molimm)

## Pairwise decomposition of residue interaction energies of single chain Fv with HIV-1 p17 epitope variants

Vannajan Sanghiran Lee<sup>a,\*</sup>, Panthip Tue-ngeun<sup>a</sup>, Sawitree Nangola<sup>b</sup>, Kuntida Kitidee<sup>b</sup>, Jitrayut Jittonom<sup>a</sup>, Piyyarat Nimmanpipug<sup>a</sup>, Supat Jiranusornkul<sup>c</sup>, Chatchai Tayapiwatana<sup>b,d,\*\*</sup>

<sup>a</sup> Computational Simulation and Modeling Laboratory (CSML), Department of Chemistry and Center for Innovation in Chemistry, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>b</sup> Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>c</sup> Department of Pharmaceutical Sciences, Faculty of Pharmacy, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>d</sup> Biomedical Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

### ARTICLE INFO

#### Article history:

Received 27 October 2009

Received in revised form 7 November 2009

Accepted 13 November 2009

Available online 21 December 2009

#### Keywords:

HIV-1

p17

Epitope

Single chain Fv

Pairwise decomposition

### ABSTRACT

Computational assisted modeling was carried out to investigate the importance of specific residues in the binding site of scFv. In this study, scFv against HIV-1 epitope at the C-terminal on p17 (scFv anti-p17) was used as a candidate molecule for evaluating the method. The wild-type p17 and its nine natural mutants were docked with scFv anti-p17. Potential mean force (PMF) scores predicted the most favorable binding interaction, and the correlation agreed well with the corresponding activity data from the peptide based ELISA. In the interaction with solvent molecules, the 3D structures of scFv anti-p17 and selected peptide epitopes were further investigated by molecular dynamics (MDs) simulation with the AMBER 9 program. Post-processing of the snapshot at equilibrium was performed to evaluate the binding free energy and pairwise decomposition or residue-based energy calculation of complexes in solution using the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) protocol. Our results demonstrated that the specific residues located in the complementary determining regions (CDRs) of scFv anti-p17, MET100, LYS101, ASN169, HIS228, and LEU229, play a crucial role in the effective binding interaction with the absolute relative decomposed energy more than 2.00 kcal/mol in comparison to the original substrate.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

The Gag p55 polyprotein of human immunodeficiency virus type 1 (HIV-1) plays a critical role in HIV-1 assembly and maturation. In the assembly step, the Gag polyprotein is directly targeted to the cell membrane to produce Gag dimerization or multimerization. Afterward, the viral-encoded protease cleaves the Gag precursor into the functional proteins: the matrix (MA or p17), capsid (CA or p24), nucleocapsid (NC) and p6 domains to form the infectious virus. The matrix protein (p17) is involved in many steps of

the HIV life cycle, especially the assembly and maturation steps (Ganser-Pornillos et al., 2008; Bukrinskaya, 2004). Tewari and co-workers investigated an intrabody strategy using the single chain antibody fragment (scFv) derived from hybridoma-secreting anti-p17 antibody inside HIV-1 infected cells, which interfered with the viral replication process. This antibody specifically binds to the C-terminal epitope (DTGHSSQVSQNY) of the p17 domain (Tewari et al., 1998). The scFv platform can be expressed in both prokaryotic and eukaryotic systems and is able to be engineered in order to improve its functional affinity and stability (Quintero-Hernandez et al., 2007; Pavoni et al., 2006; Park et al., 2006). The potential applications of scFv have been explored in many areas of research, including diagnostics and gene therapy (Inui et al., 2009; Shen et al., 2008; Wang et al., 2008; Depetris et al., 2008; Stocks, 2005).

Computational approaches and protein structural analysis can provide relevant information about the functional roles of the scFv residues (Arcangeli et al., 2008). Fast protocols using force field based scoring functions and knowledge-based approaches for predicting binding affinities of protein–ligand complexes have been established. Some popular scoring functions for estimating binding affinities of protein–ligand complexes are DOCK (Ewing et al., 2001), AutoDock (Morris et al., 1998), PMF (Muegge and Martin,

\* Corresponding author. Current address: Computational Simulation and Modelling Laboratory (CSML), Department of Chemistry and Center for Innovation in Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand. Tel.: +66891100216.

\*\* Corresponding author. Current address: Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand. Tel.: +66818845141.

E-mail addresses: [vannajan@gmail.com](mailto:vannajan@gmail.com) (V.S. Lee), [ptuengeun@gmail.com](mailto:ptuengeun@gmail.com) (P. Tue-ngeun), [sawitree2727@hotmail.com](mailto:sawitree2727@hotmail.com) (S. Nangola), [kitidee.010@hotmail.com](mailto:kitidee.010@hotmail.com) (K. Kitidee), [jitrayut.018@gmail.com](mailto:jitrayut.018@gmail.com) (J. Jittonom), [piyyaratn@gmail.com](mailto:piyyaratn@gmail.com) (P. Nimmanpipug), [supatjira@hotmail.com](mailto:supatjira@hotmail.com) (S. Jiranusornkul), [asimi002@hotmail.com](mailto:asimi002@hotmail.com) (C. Tayapiwatana).

1999), GOLD (Jones et al., 1997), LUDI (Bohm, 1998), FlexX (Rarey et al., 1996), and Ligscore (Krammer et al., 2005). Since most of the reactions of biological interest occur in water, the evaluation of the solvent effect represents an important aspect of the analytical and numerical molecular modeling approaches. Such computational approaches that consider the solvent as a part of a simulation system can be divided into several major groups: (i) continuum electrostatic methods, (ii) explicit solvent models with microscopic detail, and (iii) hybrids of the first two methods (Kollman, 1993; Smith and Pettitt, 1994; van Gunsteren et al., 1994; Tomasi and Persico, 1994; Leach, 1996; Gao, 1996; Levy and Gallicchio, 1998). An additional term is sometimes added to the force field to account for specific hydrogen bonding interactions. The more sophisticated methods have three-body nonadditive terms added. These methods include molecular dynamics (MD) (Pearlman et al., 1995; Brooks et al., 1983), Monte Carlo (MC) methods (Jorgensen, 1996), or a combination of these for sampling, and represent an advantage in terms of precision. These approaches can yield a reliable approximation method that allows one to (i) use a single trajectory of a complex between a receptor and a ligand; (ii) extract energies for the complex and all binding components by mapping their coordinates from the single trajectory; (iii) scan the sites of interest for the 'hot spots' in the receptor–ligand interface; and (iv) evaluate what the changes to the binding free energy would be upon modifications/mutations of the residues at the binding interface.

With the major advances in computer processing and clustering techniques, we are now able to perform molecular simulations of a large biomolecular system on a reasonable time scale. In this study, scFv anti-p17 was simulated based on molecular modeling of its homologue structure. The antibody–antigen complex models were generated using the flexible docking program incorporating binding activity data obtained from the peptide ELISA. The scFv anti-p17 structure with its epitopes at the C-terminal on the p17 fragment of HIV-1 and its natural mutant epitopes were analyzed in detail. In addition, the dynamic simulation method was applied to analyze the antibody–antigen interacting surface and to quantify the energetic nature of the complexes resulting from protein–ligand binding in water by applying the Molecular Mechanics–Poisson–Boltzmann Surface Area (MM-PBSA) proto-

col. The latter analysis can provide interesting information, such as electrostatic and van der Waals energies, solvation energies and entropic contributions at the binding interface. Our goal is to improve our understanding of the molecular basis of antigen recognition by scFv anti-p17.

## 2. Experimental

### 2.1. Computer assisted modeling

#### 2.1.1. Homology modeling

The primary sequence of the scFv anti-p17 protein has previously been obtained by Tewari et al. (1998). The sequence of the light (VL) and heavy (VH) chain variable domains of scFv anti-p17 were compared with the primary sequences of all immunoglobulins deposited in the Protein Data Bank using the BLAST program (Altschul et al., 1997). The Complementary Determinant Region (CDR) definition of scFv anti-p17 variable domains was investigated using the Kabat method (Kabat et al., 1983).

The best match for the VH of scFv anti-p17 was the VH of idiotype–anti-idiotype Fab complex (pdb id:1iai), sharing 82% of sequence identity with the template, whereas the most homologous VL of scFv anti-p17 was the VL of the Fab fragment of a neutralizing antibody directed against an epitope of gp41 from HIV-1 (pdb id:1nld), sharing 96% of sequence identity with the template. These structures were used as templates for homology modeling of the 3D structure of scFv anti-p17 using the MODELLER program. The orientations of the VH and VL chains were generated by superposition using a crystal structure of the anti-DNA binding antibody (pdb id:2gki) as a scaffold template.

To minimize the steric clashes, the structure of scFv anti-p17 was subjected to energy minimization with 500 steps of steepest descent followed by 500 steps of conjugate gradient until the convergence criterion of 0.05 kcal/mol/Å was obtained, using the AMBER03 force field (Case et al., 2006). Structural validation of the scFv anti-p17 was checked using PROCHECK; more than 92% of the residues were in the most favored regions of the Ramachandran plot, and overall G-factors were inside the expected regions for structures with 2.0 Å resolutions. It is generally acknowledged that antigen binding occurs in a variety of ways, and thus leads to

**Table 1**

PMF scores and residues on CDR loops of scFv at 4.5 Å from each of four peptide epitopes. The common amino acids are in boldface type.

Peptide names	Peptide sequences	PMF score (kcal/mol)	Amino Acid in 4.5 Å from peptide epitope	
			Nonpolar hydrophobic	Polar hydrophilic
p17.1	<sup>121</sup> DTGHSSQVSQNY <sup>132</sup>	−902.11	GLY33 (H1), <b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , <b>GLY226 (L3)</b> , <b>LEU229 (L3)</b>	<b>SER99 (H3)</b> , <b>SER103 (H3)</b> , <b>ASP163 (L1)</b> , <b>ASP190 (L2)</b> , <b>THR227 (L3)</b>
p17.2	<sup>121</sup> DTGHSNQVSQNY <sup>132</sup>	−899.18	<b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , GLY161 (L1), LEU185 (L2), <b>GLY226 (L3)</b> , <b>LEU229 (L3)</b>	<b>THR59 (H2)</b> , <b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , <b>SER103 (H3)</b> , SER162 (L1), <b>ASP163 (L1)</b> , <b>THR227 (L3)</b> , HIS228 (L3)
p17.3	<sup>121</sup> DTGHSSQISQNY <sup>132</sup>	−882.65	<b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , PHE167 (L1), <b>GLY226 (L3)</b> , <b>LEU229 (L3)</b>	<b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , <b>ASP163 (L1)</b> , TYR184, <b>ASP190 (L2)</b> , SER191, <b>THR227 (L3)</b> , GLN231 (L3)
p17.4	<sup>121</sup> DTGHNSQVSQNY <sup>132</sup>	−898.71	GLY33 (H1), <b>MET100 (H3)</b> , <b>GLY226 (L3)</b> , <b>LEU229 (L3)</b>	ASP31 (H1), TYR32 (H1), ASN52 (H2), <b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , HIS228 (L3), GLN231 (L3)
p17.5	<sup>121</sup> NTGHSSQVSQNY <sup>132</sup>	−843.51	<b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , <b>PHE167 (L1)</b> , LEU185 (L2), <b>LEU229 (L3)</b>	ASN35 (H1), <b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , ASN169 (L1), TYR184, <b>ASP190 (L2)</b> , SER191, <b>THR227 (L3)</b> , <b>HIS228 (L3)</b>
p17.6	<sup>121</sup> DTGNSSQVSQNY <sup>132</sup>	−846.12	GLY33 (H1), <b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , <b>LEU229 (L3)</b>	<b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , LYS165 (L1), TYR184, SER191
p17.7	<sup>121</sup> DTGHSSQASQNY <sup>132</sup>	−829.94	<b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , <b>PHE167 (L1)</b> , <b>GLY226 (L3)</b>	<b>THR59 (H2)</b> , <b>LYS101 (H3)</b> , LYS165 (L1), LYS188 (L2), <b>THR227 (L3)</b> , HIS228 (L3)
p17.8	<sup>121</sup> DTGHSKQVSQNY <sup>132</sup>	−926.79	<b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , <b>PHE167 (L1)</b> , <b>GLY226 (L3)</b> , <b>LEU229 (L3)</b>	<b>THR59 (H2)</b> , <b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , <b>SER103 (H3)</b> , SER162 (L1), <b>ASP163 (L1)</b> , <b>THR227 (L3)</b>
p17.9	<sup>121</sup> DTGNNSQVSQNY <sup>132</sup>	−841.02	<b>TRP50 (H2)</b> , <b>MET100 (H3)</b> , <b>PHE167 (L1)</b> , LEU185, <b>LEU229 (L3)</b>	<b>THR59 (H2)</b> , <b>SER99 (H3)</b> , <b>LYS101 (H3)</b> , LYS165, ASN169 (L1), TYR184, <b>ASP190 (L2)</b> , SER191, <b>HIS228 (L3)</b>

The underlined letters are the mutated residue in each sequence compared to the wild type.



a differing arrangement of antibody conformations. In scFv anti-p17, there are actually six unique hypervariable units. Each of the chains contains three of the six loops that form the binding groove. The hypervariable portions of the loops on the heavy chain are designated H1(31–35), H2(50–66) and H3(99–104) while those on the light chain are L1(154–169), L2(185–190), and L3(224–232) (Tewari et al., 1998). These regions are also known as complementary determining regions (CDRs), which have a higher binding affinity to the antigen.

### 2.1.2. Molecular docking

Nine peptides, an original HIV-1 epitope at the C-terminal (<sup>121</sup>DTGHSSQVSQNY<sup>132</sup>) on p17 and eight natural mutants (Table 1), were built partly based on a crystal structure from the Protein Databank (pdb id: 1kj4). The initial structures of the nine modeled peptides were energy minimized (1000 steps of Adopted Basis Newton Raphson (ABNR)) using CHARMM force field with a RMS gradient of 0.01 kcal/(Å mol) in Discovery Studio 1.7. Structures of scFv anti-p17 (a homology model) complexed with the peptides were constructed using the docking procedure in the BioMedCaChe 2.0 (Fujitsu, Inc.) program, in which the CDR loops (L1–L3 and H1–H3) were defined as the potential binding sites. Both the peptides and the binding sites were set to be flexible during the docking simulation. Each of the docking complexes were energetically evaluated based on the potential of mean force (PMF) that describes the potential energies involving bond stretching, angle bending, torsional, and non-bonded interactions such as Amber van der Waals and hydrogen bond interactions of molecules (Muegge and Martin, 1999). The PMF scores of each peptide were evaluated by a genetic algorithm with a population size of 50, crossover rate of 0.80, elitism of 5, mutation rate of 0.2, and the maximum cycle generation was set to be 40,000. The size of the grid box was 30 × 30 × 30 Å. Finally, the complex structures were analyzed and the interaction energy between the peptides and antibody was calculated.

### 2.1.3. Molecular dynamics (MD) simulations and binding free energy calculation

MD simulations were carried out at the molecular mechanics level using the AMBER03 force field as implemented in the AMBER9 suite of programs (Case et al., 2005). Structures of antibody-peptide were solvated in a cubic box of TIP3P water extending at least 10 Å in each direction from the solute, and the cut-off distance was kept to 12 Å to compute the nonbonded interactions. All simulations were performed under periodic boundary conditions (Weber et al., 2000), and long-range electrostatics were treated by using the particle-mesh-Ewald method (Darden et al., 1993; Essmann et al., 1995). The time step was set to 1 fs and the trajectory was recorded every 0.1 ps.

Prior to MD simulations, the systems were relaxed by a series of steepest descent (SD) and conjugated gradient (CG) minimizations. The 2-ns MD simulations were performed based on each of the minimized systems by gradually heating over 60 ps from 0 to 310 K with the protein atoms fixed using a force constant of 5 kcal/mol/Å<sup>2</sup>. Then, a 200 ps pressure-constant period (NPT) was applied to obtain an equilibrated density of the constrained protein atoms. The following step was a 40 ps-volume-constant period (NVT) at a force constant of 2.5 kcal/mol/Å<sup>2</sup> followed by 100 ps dynamics at a force constant of 1.25 kcal/mol/Å<sup>2</sup>. Finally, a 1.6 ns unrestrained MD simulation (no force applied on any protein atoms) was performed for each fully flexible system in the NVT ensemble at a constant temperature of 310 K. A total of 500 snapshots were collected at 1 ps-intervals from the last 500 ps of MD for binding free energy analysis.

Based on the selected MD snapshots, the binding free energy for each antibody-peptide system could be estimated using MM-PBSA (Molecular Mechanics Poisson–Boltzmann Surface Area) (Kollman et al., 2000) and MM-GBSA (Molecular Mechanics Generalized

Born Solvent Area) (Chong et al., 1999). The binding free energies ( $\Delta G_{\text{binding}}$ ) were determined from the free energies of the complex, protein and peptide according to the equation:

$$\Delta G_{\text{binding}} = \Delta G_{\text{water}}(\text{complex}) - [\Delta G_{\text{water}}(\text{protein}) + \Delta G_{\text{water}}(\text{peptide})]$$

The binding free energies for each species in turn were estimated from the absolute molecular mechanical energies ( $E_{\text{MM}}$ ), the solvation free energies ( $G_{\text{PB/GB}} + G_{\text{nonpolar}}$ ) and the vibration, rotation and translation entropies. Each of these terms was calculated as follows:

$$\Delta G_{\text{water}} = E_{\text{MM}} + \Delta G_{\text{solvation}} - T\Delta S;$$

$$G_{\text{solvation}} = G_{\text{solvation-electrostatic}} + G_{\text{nonpolar}};$$

$$E_{\text{MM}} = E_{\text{internal}} + E_{\text{electrostatic}} + E_{\text{vdW}};$$

$$E_{\text{internal}} = E_{\text{bond}} + E_{\text{angle}} - E_{\text{torsion}}$$

where the internal energy  $E_{\text{internal}}$  has three contributions:  $E_{\text{bond}}$ ,  $E_{\text{angle}}$ , and  $E_{\text{torsion}}$ , which represent the strain energy in bonds, angles and torsion angles caused by their deviation from the equilibrium values;  $E_{\text{electrostatic}}$  and  $E_{\text{vdW}}$  are the electrostatic and van der Waals interaction energies, respectively;  $-T\Delta S$  is the change of conformational entropy upon peptide binding, which is not considered here because of its high computational demand and relatively low accuracy of prediction (Hou et al., 2002). All energies are averaged along the MD trajectories.

$E_{\text{MM}}$  was determined with the *sander* module of the AMBER suite with an infinite cut-off for all interactions. For the MM-PBSA calculations,  $\Delta G_{\text{PB}}$  was calculated with a built-in module, the *pbsa* program in AMBER9 which solves the Poisson–Boltzmann equation. The grid size for the PB calculations was 0.5 Å. In the MM-GBSA calculations,  $\Delta G_{\text{GB}}$  was calculated using the GB model with the parameters developed by Tsui and Case (2000). The values of the interior and exterior dielectric constants were set to 1 and 80, respectively.  $\Delta G_{\text{np}}$  was estimated based on the solvent accessible surface area (SASA) as  $\Delta G_{\text{np}} = 0.0072 \times \text{SASA}$ , using the *molsurf* program (Kabat et al., 1983) and (Case et al., 2006). The scFv anti-p17/peptide interaction energy profiles were generated by decomposing the total binding free energies into residue-residue interaction pairs by the MM-GBSA decomposition process in the *mm.pbsa* program of AMBER9 (Gohlke et al., 2003; Hou et al., 2008).

## 2.2. Affinity determination

### 2.2.1. Vector construction

A vector for expressing scFv that specifically binds to the C-terminal epitope (<sup>121</sup>DTGHSSQVSQNY<sup>132</sup>) of the p17 fragment of HIV (scFv anti-p17) was constructed. Briefly, to generate the gene encoding scFv anti-p17 as described previously (Tewari et al., 1998), total RNA was extracted from hybridoma cells, MH-VM33C9/ATCC HB8975, using an RNeasy Mini kit (Qiagen, Hilden, Germany) and the first-strand cDNA was synthesized using a Transcriptor High Fidelity cDNA synthesis kit (Roch, Mannheim, Germany). The resulting cDNA was further amplified for the VH and VL fragments using specific primers, and the two fragments were linked together using Fw VHP17 and Rev VLP17 primers by overlapping PCR, which resulted in the completed fragment encoding scFv anti-p17. The fragment was treated with *Sfi*I restriction enzyme and cloned into *Sfi*I-treated pComb3X phagemid vector (a gift from Dr. C.F. Barbas, Scripps Research Institute, USA) resulting in the pComb3X-scFvp17 vector. The sequence of scFv anti-p17 was analyzed by standard sequencing methods (1st Base, Singapore).

The pComb3X-scFvp17 vector was subsequently transformed into the non-suppressor *Escherichia coli* strain HB2151.

The primers used for the amplification reactions were as follows: (1) Fw VHP17 (5'-ATATGCTAGCGGCCAGGCGGCCAGATC-CAGTTGGTGCAGT-3'), (2) Rev VHP17 (5'-CGACCCTCCACCGCG-GACCCGCCACCTCCAGACCTCCGCCACCTGCA GAGACAGTGACC-AGAGTCCC-3') for V<sub>H</sub> fragment generation, (3) Fw VLP17 (5'-GGGTCCGGCGGTGGAGGTCGGATGTTGTATGACCCGACTCCA-3') and (4) Rev VLP17 (5'-TATAAGCTTTCATTAAGCGTAGTCCGGAACGT-CGTACGGGTACTGGCCGCCTGCCTTTGATTCCAGCTTGGTACCTCC-3') for V<sub>L</sub> fragment generation.

## 2.2.2. Preparation of soluble scFv anti-p17

The soluble scFv anti-p17 was produced by expressing pComb3X-scFvp17 vector in the non-suppressor *E. coli* strain HB2151. The bacterial cells harboring the vector were grown in 10 ml of Terrific broth (1.2% (w/v) tryptone, 2.4% (w/v) yeast extract, 0.4% [w/v] glycerol, 17 mM KH<sub>2</sub>PO<sub>4</sub>, 72 mM K<sub>2</sub>HPO<sub>4</sub>) containing ampicillin (100 µg/ml) at 37 °C for 18 h with shaking. One hundred microliters of precultured bacteria were inoculated in 100 ml of the same medium containing 1% (w/v) glucose and ampicillin (100 µg/ml), with shaking continued at 37 °C until an optical density (OD) of 1.5 at 600 nm was reached. To induce the protein expression, IPTG was added to the culture at a final concentration of 1 mM. After induction, the bacteria were grown at 25 °C for 20 h. The culture supernatant containing extracellular soluble scFv anti-p17 was collected by centrifugation at 5000 × g for 30 min at 4 °C. Protein was precipitated with saturated (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> in an ice bath and concentrated with Amicon Ultra centrifugal filter units (Millipore, Cork, Ireland). Finally, the concentrated protein was reconstituted with 1.5 ml of 0.15 M PBS, pH 7.2. The concentrated protein was separated in 12% SDS-PAGE under reducing conditions, and transferred to a nitrocellulose membrane (GE Healthcare, Buckinghamshire, UK). The blotted membrane was blocked with 5% skimmed-milk in PBS for 1 h at room temperature (RT) with shaking and then treated with anti-HA antibody (Sigma-Aldrich, St. Louis, MO). After incubation, the membrane was washed 5 times with washing buffer (0.05% Tween-20 in PBS) and peroxidase-conjugated goat anti-mouse immunoglobulin antibodies were added to the membranes. The peroxidase reaction was visualized using a SuperSignal West Pico Substrate (Pierce, Rockford, USA).

## 2.2.3. Evaluation of the binding activity of scFv anti-p17 by ELISA

To evaluate the binding activity of scFv anti-p17 with mutant peptides, four peptides were synthesized (GenScript, Piscataway, New Jersey, USA) and tested with the scFv protein using a standard ELISA procedure. Peptide p17.1 represented the wild-type epitope while another peptide represented the mutant peptides. The amino acid sequences of all synthetic peptides are shown in Table 1. Briefly, 100 µl of 50 µg/ml of each peptide in coating buffer (0.1 M NaHCO<sub>3</sub>, pH 8.6) were added to microtiter plates (NUNC, Roskilde, Denmark) and incubated overnight at 4 °C. The coated wells were then blocked with 200 µl of blocking buffer (2% BSA in TBS) for 1 h at RT. The wells were washed five times with washing buffer (0.05% Tween-20 in TBS). 100 µl of 200 µg/ml of scFv anti-p17 protein in blocking buffer were added to each well and incubated for 1 h at RT. After incubation, the excess antibody was eliminated by washing. Subsequently, the wells were incubated for 1 h at RT with 100 µl of HRP-conjugated goat anti-mouse IgS antibody (KPL, Maryland, USA) diluted 1/3000 in blocking solution. Wells were then washed again prior to adding 100 µl of 3,3',5,5'-tetramethyl-benzidine (TMB) substrate. The optical densities (OD) at 450 nm were measured by an ELISA plate reader (TECAN, Austria) after adding 100 µl of 1 N HCl.

To assess the binding affinity between scFv anti-p17 and mutant peptides, a peptide competitive ELISA was performed. The same

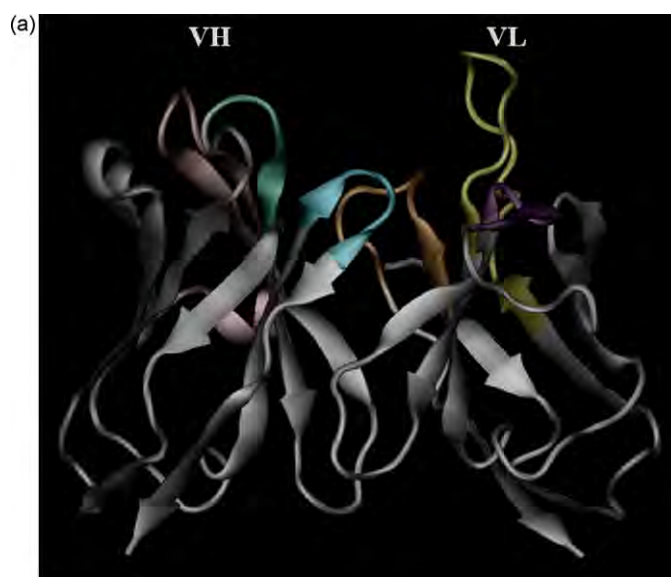
procedure was followed as described for the peptide ELISA; the scFv protein was mixed with 100 ng/ml of each mutant peptide and incubated for 1 h at RT. 100 µl of the mixture were applied into individual peptide p17.1 pre-coated wells and incubated for 1 h at RT. After washing the wells, the bound scFv anti-p17 was monitored by adding 100 µl of HRP-conjugated goat anti-mouse IgS antibody. Wells were then washed again and 100 µl of TMB substrate were subsequently added for 45 min. The enzymatic reaction was stopped by adding 100 µl of 1 N HCl. The OD at 450 nm was measured by an ELISA plate reader. The OD values were converted to percentage inhibition values (PI) by using the following formula:

$$PI = 100 - \left( \left( \frac{B}{B_0} \right) \times 100 \right) B$$
 and  $B_0$  are the OD values of scFv with peptide inhibitor and without peptide inhibitor respectively.

## 3. Discussion

### 3.1. Homology modeling of scFv and complex model

We selected VH and VL fragments of anti-p17 antibody from hybridoma cells MH-VM33C9/ATCC HB8975 for analysis. We had also prepared recombinant scFv anti-p17 by cloning the gene encoding VH and VL into a prokaryotic expression vector. For the homology modeling, the scFv antibody with the peptide epi-



(b) VH sequence	
QIQLVQSGPELKPKGETVEISCKASGYTFTDYGMNWMKQAPGKSLKWMGW	50
CDR H2	
INTYTGEPTVADEFKGRFAFSLETSASTAYLQINNLSKSEDMATYFCSRSM	100
CDR H3	
KGSYWGQGLTVTUSA	115
VL sequence	
DVVMQTPTLTSLVTIGQPASISCKSSQSLLSGDKFTFLNWLQRPGQSPK	50
CDR L2	
RLIYLVSKLDSGVPDRFTGSGSGTDFTLKISRVEADLGVVYCWQGTXLPL	100
CDR L3	
QTFGGGTGLEIK	112

**Fig. 1.** Molecular models of scFv anti-p17. (a) 3D structural model for scFv anti-p17. VH and VL domains are colored white. The hypervariable binding loops are colored green (H1), pink (H2), cyan (H3), yellow (L1), purple (L2), and orange (L3). (b) Amino acid sequences of the hypervariable binding VH (1–114) and VL (131–241) domains of scFv anti-p17 and an artificial linker shown in bold (GGGS)<sub>3</sub>; the loops in the heavy chain (CDRH1: 31–35, CDRH2: 50–66, and CDRH3: 99–104), and the loops in the light chain (CDRL1: 154–169, CDRL2: 185–190, and CDRL3: 224–232) of the scFv antibody. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).



**Table 2**

Relationship between the binding activities from competitive ELISA and those from MM-PBSA methodology at 310 K.

Method	Contribution	Peptide Names				CD147
		p17.1	p17.3	p17.7	p17.8	
MM	$\Delta(L)$	−227.32	−174.34	−169.29	−91.34	
	$\Delta(VDW)$	−72.28	−55.46	−48.01	−68.95	
	$\Delta(GAS)$	−299.60	−229.81	−217.30	−160.29	
PBSA	$\Delta(PB_{SOL})$	−11.02	−9.73	−8.54	−10.56	
	$\Delta(PB_{CAL})$	280.63	230.71	217.07	144.47	
	$\Delta(PB_{SOL})$	269.61	220.99	208.53	133.91	
	$\Delta(PB_{ELE})$	53.31	56.37	47.78	53.13	
	$\Delta(PB_{TOT})$	−29.98	−8.82	−8.77	−26.38	
GBSA	$\Delta(GB_{SOL})$	−11.02	−9.73	−8.54	−10.56	
	$\Delta(GB_{CAL})$	273.07	220.02	207.41	136.89	
	$\Delta(GB_{SOL})$	262.05	210.29	198.87	126.33	
	$\Delta(GB_{ELE})$	45.75	45.68	38.12	45.55	
	$\Delta(GB_{TOT})$	−37.55	−19.52	−18.43	−33.97	
	Experimental value <sup>a</sup>	75.94	55.25	44.60	79.46	7.94

ELE, electrostatic interactions; VDW, van der Waals interactions between the fragments; GAS, addition ELE + VDW + INT being the binding enthalpic contributions in vacuo;  $PB_{SOL}$ , nonpolar contribution to solvation;  $PB_{CAL}$ , polar contribution of solvation;  $PB_{SOL}$ , the  $PB_{SOL}$  +  $PB_{CAL}$ ;  $PB_{ELE}$ ,  $PB_{CAL}$  + ELE addition;  $PB_{TOT}$ , total binding free energy calculated by the MM-PBSA method.

<sup>a</sup> Competitive ELISA: PI (%)

tope was assembled and modeled based on a homology modeling approach. The three dimensional structure and the amino acid sequence of scFv anti-p17 are shown in Fig. 1. The models of scFv, nine peptide epitopes, and the scFv-peptide complexes were generated separately. The sequences of mutated peptides that were obtained from the GenBank database comprised the following positions: the single mutation of S126N, V128I, S125N D121N, H124N, V128A, S126K, and the double mutation of H124N and S125N. The sequences of all peptide epitopes that were initially positioned outside the binding region were docked against scFv in the same manner. The PMF scores (Table 1) of the complex structures were calculated with the BioMedCaChe 2.0 (Fujitsu, Inc.) program, where flexible peptide epitopes were docked into flexible side chain proteins. The peptide epitopes and the side chains of the amino acids of the CDR domains were kept flexible during the docking simulation. The PMF scores were in the range of −829.939 to −926.793 kcal/mol. The peptides bound in two orientations, where the N-terminal (p17.1, p17.2, p17.4–p17.6, and p17.8) and the C-terminal (p17.3, p17.7, and p17.9) of peptide sequences were directed toward the binding pocket. All interactions of the optimum docking structures reflected the negative binding energies in all models, indicating favorable binding in all complexes. Less favorable binding of peptide 17.3 and p17.9 to scFv was due to their inverted binding of the N-terminal to the outside of the binding pocket. The C-terminal sequences of p17.3 and p17.9 had better fits in the binding pocket. A similar binding pattern as for the substrate with the N-terminal sequence towards the inside of the pocket was also observed, but had higher binding energy scores. Interestingly, we found that mutation of a particular epitope at S125K of peptide p17.8 caused the maximum enhancement of the binding energy. The interacting amino acids of scFv within 4.5 Å from an individual docked peptide are listed in Table 1. The common binding residues of scFv are composed of the combination of hydrophilic and hydrophobic amino acids: Trp50 (H2), Thr59 (H2), Ser99 (H3), Met100 (H3), Lys101 (H3), Ser103 (H3), Asp163 (L1), Phe167 (L1) Asp190 (L2), Gly226 (L3), Thr227 (L3), His228 (L3), and Leu229 (L3), as indicated in boldface type.

### 3.2. Comparison of calculated binding free energy with experimental data

The DNA encoding scFv fragment of anti-p17 was successfully generated and cloned into a pComb3X phagemid vector,

resulting in the pComb3X-scFvp17 vector. This vector was subsequently transformed in non-suppressor *E. coli* (HB2151) for soluble expression of scFv anti-p17. The bacterial cells harboring pComb3X-scFvp17 vector were cultured and induced with 1 mM IPTG. The soluble protein in the culture supernatant was precipitated by ammonium precipitation and concentrated with an Amicon Ultra centrifuge filter. The scFv anti-p17 was expressed in its soluble form by induction, secreted into culture supernatant and revealed by western immunoblotting. A band with a molecular weight of approximately 30 kDa, corresponding to the molecular size of scFv, was detected in the concentrated protein (data not shown).

In order to investigate protein binding efficiency, peptide ELISA was employed to demonstrate the binding activity of scFv anti-p17 to its target peptide (p17.1), and the chosen mutant peptides (p17.3, p17.7 and p17.8) (Fig. 2). Positive signals were observed in all peptide coated wells, indicating that this recombinant scFv could bind to all mutant peptides. Peptide p17.8 gave the highest signal followed by p17.1, p17.3 and p17.7, respectively. All soluble mutant peptides were able to inhibit the binding between the scFv and immobilized p17.1 peptide (Table 2), as revealed by the percentage inhibition value (PI), but the CD147 peptide had no significant inhibitory effect. Peptides p17.1 and p17.8 exhibited the highest

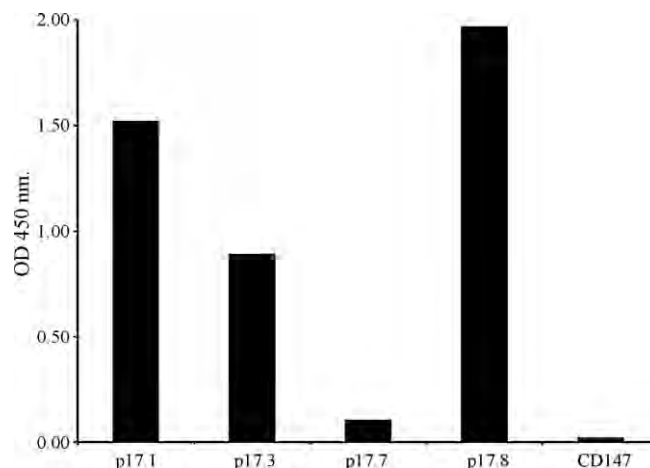


Fig. 2. The binding activity of soluble scFv anti-p17 from peptide ELISA.

inhibitory effects compared to the other two peptides, p17.3 and p17.7, at the same concentration.

Comparison of experimental activities with the results derived from MM-PBSA and MM-GBSA calculations suggested that the experimental value for MM-PBSA had a nearby correlation ( $r^2 = 0.88$ ) with the calculated binding free energy of MM-GBSA ( $r^2 = 0.90$ ), as shown in Table 2. Initially, favorable configurations for forming a scFv anti-p17–antigen complex system were built by molecular docking. The binding structures of the ligand to its receptor were analyzed based on the energy of the ligand or receptor. From the PMF scores, we selected four peptide epitopes consisting of one wild-type peptide (p17.1) and three mutated peptides (p17.3, p17.7, and p17.8) for further investigation by MDs and peptide ELISA. Peptide p17.7 had the lowest score and p17.8 had the highest score, whereas p17.3 had very similar score to that of the wild-type peptide. To understand the binding interaction in water, the binding free energies of those complexes were simulated by molecular dynamics simulations (MDs), and the Molecular Mechanics Poisson–Boltzmann Surface Area methodology was applied to calculate the binding free energy of all residues of the complexes. Table 2 lists the terms that contributed the calculation of binding free energy for the selected complex. The value of  $PB_{TOT}$  was used to compare the simulation with the experimental results. The more negative the value, the more favorable the binding. The binding energies identified by the MM-PBSA protocol were ranked as follows: peptide p17.1 < p17.8 < p17.3 < p17.7 with the values of  $-29.98$ ,  $-26.38$ ,  $-8.82$ ,  $-8.77$ , and kcal/mol, respectively. The results were consistent with PMF scores from molecular docking data, which divided the mutants into two groups of high and low activities. The binding free energy indicated highly favorable binding of scFv with peptides p17.1 and p17.8, about 18 kcal/mol more negative than p17.3. Consequently, we identified the p17.3, p17.4, p17.5, p17.6, p17.7 and p17.9 as the low affinity binding peptides, whereas the p17.1, p17.2, and p17.8 were identified as the high affinity binding peptides with our scFv.

The major contributions to the binding free energy arise from the electrostatic energy, as calculated by the molecular mechanic (MM) force field (ELE); from the electrostatic contribution to the solvation free energy, as calculated by PB ( $PB_{CAL}$ ); and van der Waals contribution from MM (VDW). For the four binding peptides, both van der Waals and electrostatic energies were quite varied among the low and high activities groups, indicating that both terms are factors determining the binding activity. Peptide p17.8 had a somewhat lower electrostatic contribution ( $-91.34$  kcal/mol). Among the other sequences, however, the combination with VDW and the sum of nonpolar and polar contributions to solvation ( $PB_{SOL}$ ) resulted in a total negative binding free energy. This supported a favorable scFv–peptide complex in pure water. The result does not equal the real binding free energy since we did not estimate the entropy contribution to binding in this study.

### 3.3. Decomposition of energy on the amino acid residues in CRD loops and specific contact upon binding

The interpretation of macroscopic data in terms of microscopic interactions of scFv with a peptide binding sequence can be done by decomposing the calculated binding free energies as a sum of components that correspond to the contributions of different energy terms or different parts of the system. Therefore, important residues of anti-HIV p17 scFv will show strong interactions in association with its antigen. To gain further insight into the key residue interactions, the overall agreement between the calculated and experimental values for the ligands obtained by the MM-PBSA approach allows us to be optimistic of the results when estimating protein–ligand interactions. Table 3 and Fig. 3 illustrate the results of this analysis, plotting the relative decomposed energies versus common amino acids on the CDR loops of scFv anti-p17 and each of four peptide epitopes. All amino acids in the scFv sequences were found to exhibit positive or negative influences on binding to the substrate molecules. Several residues of the wild-type

**Table 3**  
Common interacting residues found among the natural peptide sequences in the docking study.

Residue	Loop	Decomposed energy (kcal/mol)			
		p17.1	p17.3	p17.7	p17.8
ASP31	H1	0.26	0.35 (0.09)	0.28 (0.02)	0.24 (−0.02)
TYR32	H1	0.04	−0.05 (−0.09)	0.03 (−0.01)	0.01 (−0.03)
GLY33	H1	0.01	−0.08 (−0.09)	−0.02 (−0.03)	−0.11 (−0.12)
ASN35	H1	0.14	−0.14 (−0.28)	−0.09 (−0.23)	−0.34 (−0.48)
TRP50	H2	−1.73	<b>−3.76 (−2.03)</b>	−2.05 (−0.32)	<b>−4.90 (−3.17)</b>
ASN52	H2	0.03	−0.25 (−0.28)	−0.05 (−0.08)	−0.84 (−0.87)
THR59	H2	0.04	−0.56 (−0.60)	−0.57 (−0.61)	−0.61 (−0.65)
SER99	H3	−0.18	−0.17 (0.01)	0.09 (0.27)	−0.62 (−0.44)
MET100	H3	−3.80	<b>−1.02 (2.78)</b>	−3.27 (0.53)	−3.59 (0.21)
LYS101	H3	−2.91	<b>−0.56 (2.35)</b>	−1.10 (1.81)	<b>4.24 (7.15)</b>
SER103	H3	−0.11	0.06 (0.17)	0.03 (0.14)	−1.31 (−1.2)
GLY161	L1	0.01	−0.01 (−0.02)	−0.07 (−0.08)	−0.02 (−0.03)
SER162	L1	0.11	0.03 (−0.08)	−0.07 (−0.18)	0.20 (0.09)
ASP163	L1	0.40	0.28 (−0.12)	0.39 (−0.01)	−0.27 (−0.67)
LYS165	L1	−0.50	−0.16 (0.34)	−0.26 (0.24)	0.46 (0.96)
PHE167	L1	−2.26	−3.14 (−0.88)	−1.59 (0.67)	<b>−4.87 (−2.61)</b>
ASN169	L1	−2.66	<b>0.00 (2.66)</b>	<b>0.09 (2.75)</b>	<b>−0.22 (2.44)</b>
TYR184	–	−2.41	−2.06 (0.35)	−1.44 (0.97)	−1.79 (0.62)
LEU185	L2	−0.07	−1.24 (−1.17)	−1.13 (−1.06)	−2.06 (−1.99)
LYS188	L2	−0.14	−0.07 (0.07)	−0.30 (−0.16)	0.02 (0.16)
ASP190	L2	0.26	−0.37 (−0.63)	0.20 (−0.06)	−0.10 (−0.36)
SER191	–	−0.03	0.00 (0.03)	0.00 (0.03)	0.01 (0.04)
GLY226	L3	−0.18	−0.57 (−0.39)	−0.07 (0.11)	−1.29 (−1.11)
THR227	L3	−0.30	−0.32 (−0.02)	−1.85 (−1.55)	<b>−2.90 (−2.6)</b>
HIS228	L3	−3.93	<b>0.14 (4.07)</b>	<b>−0.42 (3.51)</b>	<b>−1.12 (2.81)</b>
LEU229	L3	−4.53	<b>−0.70 (3.83)</b>	<b>−1.15 (3.38)</b>	<b>−2.07 (2.46)</b>
GLN231	L3	−0.52	<b>−2.75 (−2.23)</b>	−1.20 (−0.68)	0.14 (0.66)

Relative energy to p17.1 is in parenthesis. The amino acids which have significant contribution with the absolute relative energy larger than 2 Kcal/mol are indicated in bold letters.

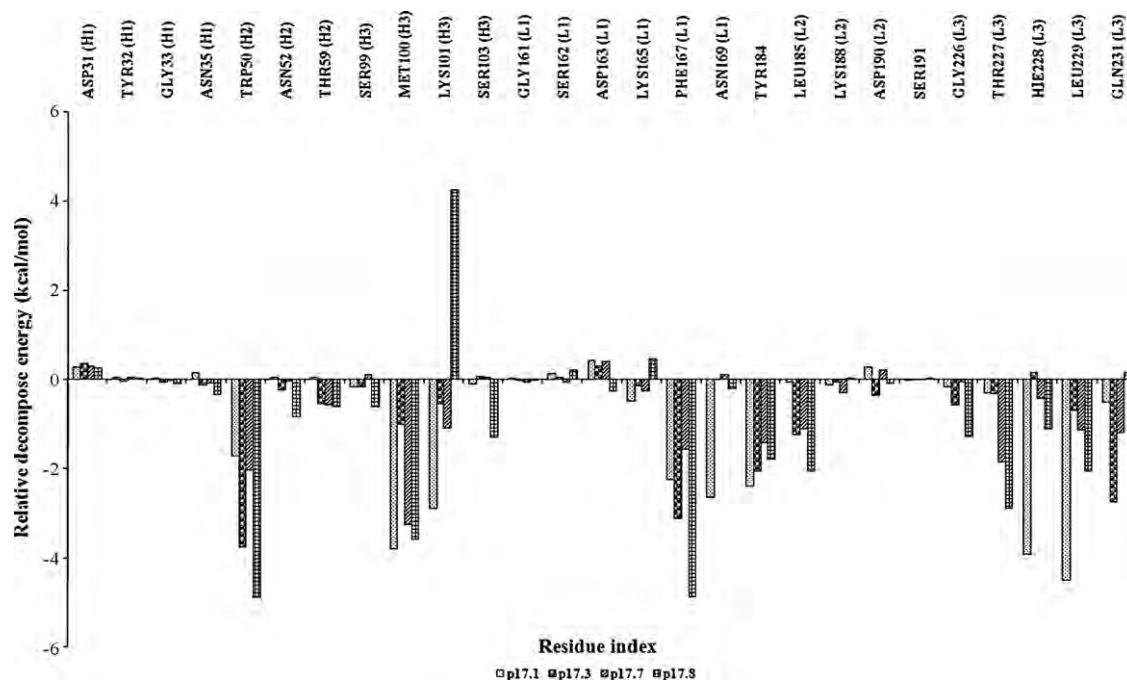


Fig. 3. List of the decomposed energies of the common amino acids in CDR loops of scFv.

p17.1 (Trp50, Met100, Lys101, Phe167, Asn169, Thr227, His228, Leu229 and Gln231) were verified to have significant effective contributions with the absolute relative energy larger than 2 kcal/mol for the stabilization energy as highlighted in Table 3. There was strong interdependence of the effects of the individual residues in the epitope sequences. Among three mutated peptides, poor binding and/or weak interaction, with relative energy above 2 kcal/mol in comparison with wild-type p17.1, was obtained from Met100, Lys101, Asn169, His228, and Leu229 of scFv. Better binding was indicated by the lower decomposed energy with the absolute relative energy below 2 kcal/mol. The difference between high and low affinity binding depends on the interaction of each sequence with the amino acids in the CDR region of scFv. The Trp50 of scFv exhibited more binding interaction with both p17.3 (V128I) and

p17.8 (S125K) than the wild-type peptide. Sequence p17.1 shows the strongest binding among all sequences, with Met100, Lys101, Asn169, His228, and Leu229, whereas sequence p17.8 exhibited the strongest binding with different amino acids such as Trp50, Phe167 and Thr227. This is compared to only one strong binding interaction among other sequences with Gln231 as observed in sequence 17.3 and no strong binding interaction with 17.7 resulted in a low binding affinity. These interactions make a significant contribution to the overall binding. As mentioned in the previous section, the decomposed energies of some residues in scFv have positive values with peptide epitopes. A particularly strong case, with decomposed energy of 4.24 kcal/mol, was observed for p17.8 binding with Lys101. However, the sum of all those energy terms resulted in favorable binding. Overall we found that Met100(H3), Lys101(H3),

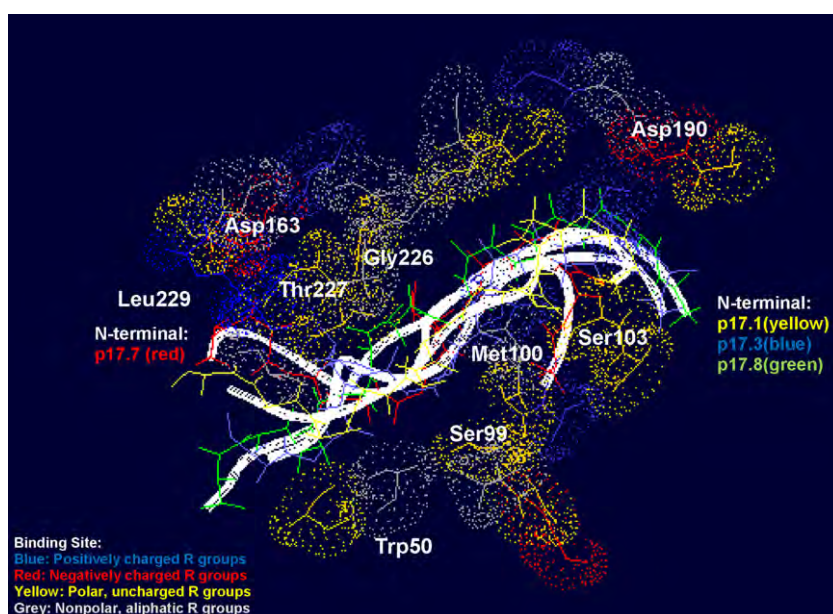


Fig. 4. The residues interaction of the final MD complex structures between scFv anti-p17 and four peptide epitopes.



Asn169(L1), His228(L3), and Leu229(L3) exhibited strong interactions with peptide p.17.1. From the MD structure in the last 500 ps, we also observed strong hydrogen bonding interactions with more than 95% occupancy for R181 in L2 with D121 and Q127 of the substrate sequence.

Further analysis of the final MD complex structures is shown in Fig. 4. As expected, p17.3 bound in different patterns among other sequences since its C-terminal has binding sites oriented toward the inside of the pocket instead of the N-terminal. Our study using pairwise decomposition of residue interaction energies has allowed us to gain insight into the interactions between anti-p17 single chain Fv with its peptide epitopes for HIV-1.

#### 4. Conclusions

Computer models were combined with laboratory experiments for the efficient determination and the identification of the most important residues for scFv in binding with natural peptide substrates. ScFv anti-p17 was built from its X-ray structure homologue, and the complexes of scFv anti-p17 with its natural epitope were generated using a flexible docking method. The efficacy of combining the scFv antibody with peptide epitopes according to the potential mean force scoring correlated well with peptide ELISA results. Molecular dynamics simulations were performed on selected peptides to evaluate their interaction in water. Overall structural changes of binding peptides in response to enzyme binding in water were investigated by the root mean square displacement. Poorly binding peptides exhibit a larger root mean square displacement than do tightly binding peptides. MM-PBSA and pairwise decomposition energies were calculated from post-analysis of molecular dynamics structures. The calculated binding free energies concurred well with experimentally determined high binding affinity (decomposed energy >2 kcal/mol) and low binding affinity of investigated peptides to anti-p17 scFv. Several amino acids, MET100, LYS101, ASN169, HIS228, and LEU229, in the complementary determining regions (CDRs) were defined for their major contribution to the binding efficiency of natural HIV epitope at the C-terminal on p17. This technique could be applied to elucidate the most important amino acids involved in the binding of scFv with its target molecules. Moreover, the identified amino acids can be modified for improving the binding activity.

#### Acknowledgements

The authors would like to express grateful acknowledgement the financial support from the Thailand Research Fund (TRF), the Research Chair Grant of National Sciences and Technology Development Agency (Thailand), the Center for Innovation in Chemistry (PERCH-CIC), Commission on Higher Education, Ministry of Education of Thailand, and the Computational Nanoscience Consortium (CNC), the National Nanotechnology Center (NAN-OTEC) and National Science and Technology Development Agency (NSTDA), Thailand for the access to Discovery Studio Version 1.7 program package. The authors would also like to thank Dr. Dale Taneyhill for proofreading the manuscript.

**Contributors:** SN, KK, CT contributed with experimental design, performance of all experiments, and writing of the manuscript. VSL and PT participated in all calculations and JJ assisted with the antibody modeling building. VSL, PN, SJ, and CT contributed with supervising and writing of the manuscript. All authors read and approved the final manuscript.

#### References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.

Arcangeli, C., Cantale, C., Galeffi, P., Rosato, V., 2008. Structure and dynamics of the anti-AMCV scFv(F8): effects of selected mutations on the antigen combining site. *J. Struct. Biol.* 164, 119–133.

Bohm, H.J., 1998. Prediction of binding constants of protein-ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput. Aided Mol. Des.* 12, 309–323.

Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M., 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4, 187–217.

Bukrinskaya, A.G., 2004. HIV-1 assembly and maturation. *Arch. Virol.* 149, 1067–1082.

Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz, J.K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J., 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26, 1668–1688.

Case, D.A., Darden, T.A., Cheatham III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., Crowley, M., 2006. AMBER 9. University of California, San Francisco, CA.

Chong, L.T., Duan, Y., Wang, L., Massova, I., Kollman, P.A., 1999. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14330–14335.

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092.

Depetris, M., Casalis, P., Kratje, R., Etcheverrigaray, M., Oggero, M., 2008. A scFv antibody fragment as a therapeutic candidate to neutralize a broad diversity of human IFN- $\alpha$  subtypes. *J. Immunol. Methods* 334, 104–113.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L., 1995. A smooth particle mesh ewald potential. *J. Chem. Phys.* 103, 8577–8592.

Ewing, T.J.A., Makino, S., Skillman, A.G., Kuntz, I.D., 2001. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule database. *J. Comput. Aided Mol. Des.* 15, 411–428.

Ganser-Pornillos, B.K., Yeager, M., Sundquist, W.I., 2008. The structural biology of HIV assembly. *Curr. Opin. Struct. Biol.* 18, 203–217.

Gao, J., 1996. Hybrid quantum mechanical/molecular mechanical simulations: an alternative avenue to solvent effects in organic chemistry. *Acc. Chem. Res.* 29, 298–305.

Gohlke, H., Kiel, C., Case, D., 2003. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* 330, 891–913.

Hou, T., Guo, S., Xu, X., 2002. Predictions of binding of a diverse set of ligands to gelatinase-A by a combination of molecular dynamics and continuum solvent models. *J. Phys. Chem. B* 106, 5527–5535.

Hou, T., Zhang, W., Case, D.A., Wang, W., 2008. Characterization of domain-peptide interaction interface: a case study on the amphiphysin-1 SH3 domain. *J. Mol. Biol.* 376, 1201–1214.

Inui, H., Takehara, A., Doi, F., Nishi, K., Takai, M., Miyake, S., Ohkawa, H., 2009. A scFv antibody-based immunoaffinity chromatography column for clean-up of bisphenol A-contaminated water samples. *J. Agric. Food. Chem.* 57, 353–358.

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748.

Jorgensen, W.L., 1996. BOSS 3. 6. Yale University, New Haven, CT.

Kabat, E.A., Wu, T.T., Bilofsky, H., Reid-Milner, M., Perry, H., 1983. Sequences of Proteins of Immunological Interest. NIH Publications No. 369–847, Bethesda, MD.

Kollman, P.A., 1993. Free energy calculations: applications to chemical and biochemical phenomena. *Chem. Rev.* 93, 2395–2417.

Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., Cheatham 3rd, T.E., 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897.

Krammer, A., Kirchhoff, P.D., Jiang, X., Venkatachalam, C.M., Waldman, M., 2005. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graph. Model.* 23, 395–407.

Leach, A.R., 1996. Molecular Modelling: Principles and Applications. Addison Wesley Longman Ltd., Singapore.

Levy, R.M., Gallicchio, E., 1998. Computer simulations with explicit solvent: recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects. *Annu. Rev. Phys. Chem.* 49, 531–567.

Morris, G.M., Goodsell, D.S., Halliday, R., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 19, 1639–1662.

Muegge, I., Martin, Y.C., 1999. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* 42, 791–804.

Park, S.G., Jung, Y.J., Lee, Y.Y., Yang, C.M., Kim, I.J., Chung, J.H., Kim, I.S., Lee, Y.J., Park, S.J., Lee, J.N., Seo, S.K., Park, Y.H., Choi, I.H., 2006. Improvement of neutralizing activity of human scFv antibodies against hepatitis B virus binding using CDR3 V(H) mutant library. *Viral. Immunol.* 19, 115–123.

Pavoni, E., Flego, M., Dupuis, M.L., Barca, S., Petronzelli, F., Anastasi, A.M., D'Alessio, V., Pelliccia, A., Vaccaro, P., Monteriu, G., Ascione, A., De Santis, R., Felici, F., Cianfriglia, M., Minenkova, O., 2006. Selection, affinity maturation, and characterization of a human scFv antibody against CEA protein. *BMC Cancer* 6, 41.

Pearlman, D.A., Case, D.A., Caldwell, J.W., Ross, W.R., Cheatham, T.E., Ferguson, D.M., Seibel, G.L., Singh, U.C., Weiner, P., Kollman, P.A., 1995. AMBER4.1 (UCSF). University of California, San Francisco, CA.

Quintero-Hernandez, V., Juarez-Gonzalez, V.R., Ortiz-Leon, M., Sanchez, R., Possani, L.D., Becerril, B., 2007. The change of the scFv into the Fab format improves the

- stability and in vivo toxin neutralization capacity of recombinant antibodies. *Mol. Immunol.* 44, 1307–1315.
- Rarey, M., Kramer, B., Lengauer, T., Klebe, G., 1996. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 261, 470–489.
- Shen, Z., Yan, H., Zhang, Y., Mernaugh, R.L., Zeng, X., 2008. Engineering peptide linkers for scFv immunosensors. *Anal. Chem.* 80, 1910–1917.
- Smith, P.E., Pettitt, B.M., 1994. Modeling solvent in biomolecular systems. *J. Phys. Chem.* 98, 9700–9711.
- Stocks, M., 2005. Intrabodies as drug discovery tools and therapeutics. *Curr. Opin. Chem. Biol.* 9, 359–365.
- Tewari, D., Goldstein, S.L., Notkins, A.L., Zhou, P., 1998. cDNA encoding a single-chain antibody to HIV p17 with cytoplasmic or nuclear retention signals inhibits HIV-1 replication. *J. Immunol.* 161, 2642–2647.
- Tomasi, J., Persico, M., 1994. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* 94, 2027–2094.
- Tsui, V., Case, D.A., 2000. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* 122, 2489–2498.
- van Gunsteren, W.F., Luque, F.J., Timms, D., Torda, A.E., 1994. Molecular mechanics in biology: from structure to function, taking account of solvation. *Annu. Rev. Biophys. Biomol. Struct.* 23, 847.
- Wang, G.P., Qi, Z.H., Chen, F.P., 2008. Treatment of acute myeloid leukemia by directly targeting both leukemia stem cells and oncogenic molecule with specific scFv-immunolipoplexes as a deliverer. *Med. Hypotheses* 70, 122–127.
- Weber, W., Hünenberger, P., McCammon, J., 2000. Molecular dynamics simulations of a polyaniline octapeptide under Ewald boundary conditions: influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B* 104, 3668–4575.

## **Structural analysis of scFv anti-p17 point mutations guided from computational alanine scanning**

Panthip Tue-ngeun<sup>1</sup>, Piyarat Nimmanpipug<sup>1,2</sup>, Narin Lawan<sup>1</sup>, Sawitree Nangola<sup>3</sup>, Chatchai Tayapiwatana<sup>3,4\*</sup>, and Vannajan Sanghiran Lee<sup>1,2\*</sup>

<sup>1</sup>Computational Simulation Modelling Laboratory (CSML), Department of Chemistry and Center of Excellence for Innovation in Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand, 50200.

<sup>2</sup>Thailand Center of Excellence in Physics, Commission on Higher Education, 328 Si Ayutthaya Road, Bangkok, 10400, Thailand

<sup>3</sup>Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand 50200

<sup>4</sup>Biomedical Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

E-mail address of the corresponding author:

Vannajan Sanghiran Lee: vannajan@gmail.com

Chatchai Tayapiwatana: asimi002@chiangmai.ac.th

### **Abstract**

A computational alanine scanning mutagenesis was submitted to quest for the residues in the energetic contribution of the mutation sites in scFv anti-p17 binding free energy with p17 c-terminal peptide alternatives. In this study, the MMPB-SA (molecular mechanics/Poisson-Boltzmann surface area) approach was calculated for the free energy variations on alanine mutation in the CDR loops of scFv anti-p17. The interactions and conservations of the mutational sites were also validated. The result demonstrated several influenced amino acid residues in the scFv binding pocket i.e. ASP31, ASN35, TRP50, MET100, LYS101, SER103, TYR104, SER162, ASN169, LEU185, LYS188, ASP190, TRP224, GLN225, and LEU229. A candidate point mutation was selected at the position 31 from aspartic acid to glycine (D31G) since this residue is a key residue in CDR and strongly effect by sidechain from alanine scanning result. The analyzed docking interaction energy of D31G with a peptide sequence (DTGNNSQVSQNY) demonstrated the favorably interaction energy comparing with the original scFv. Whereas the binding activity of original and D31G mutant against DTGNNSQVSQNY remained the same by competitive ELISA, D31G mutant remarkably improved the binding activity by the indirect ELISA. These findings raised some concerns on the correlation between the theoretical and experimental results.

**Keywords:** p17, Epitope, Single chain Fv, Alanine Scanning, HIV-1

## 1. Introduction

The Gag p55 polyprotein of human immunodeficiency virus type 1 (HIV-1) plays a critical role in HIV-1 assembly and maturation. In the assembly step, the Gag polyprotein is directly targeted to the cell membrane to produce Gag dimerization or multimerization. Afterward, the viral-encoded protease cleaves the Gag precursor into the functional proteins: the matrix (MA or p17), capsid (CA or p24), nucleocapsid (NC) and p6 domains to form the infectious virus. The HIV matrix protein, p17, is of particular interest because of its multiple roles in HIV-1 replication cycle. It has been reported that antibodies to p17 epitopes are able to inhibit HIV replication and spreading [1]. However, the low affinity with some strains has been observed and lead to our study for the new potential antibody. In this study, theoretical modeling and molecular dynamics simulations investigation of scFv against HIV-1 epitope at C-terminal on p17 (scFv anti-p17) has been performed to specify the key residues in the binding by computational alanine scanning to investigate the effect of side chain atoms of the residues in CDR loops of scFv anti-p17.

## 2. Experimental

### 2.1 Computational details

#### 2.1.1 Molecular docking

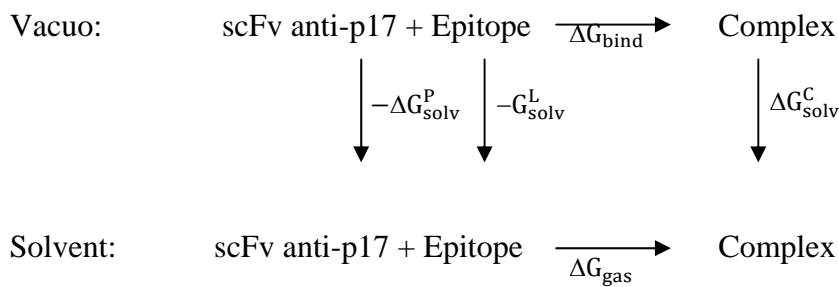
The primary sequence of the scFv anti-p17 protein has previously been obtained by [1] and the homology modeling has been built (refer to ref [2]). The general docking protocol and potential functions employed in CDOCKER have been described in prior articles [3,4]. In this work, docking of the peptides to scFv anti-p17 was conducted using CDOCKER. CDOCKER is a grid-based molecular docking method that employs CHARMm. The receptor is held rigid while the ligands are allowed to flex during the refinement. Ligands are assumed to have already been roughly docked into the receptor binding site. The scFv against HIV-1 epitope at the c-terminal on p17 (scFv anti-p17) and nine peptides (an original HIV-1 epitope at the C-terminal (DTGHSSQVSQNY) on p17 and eight natural mutants) have been taken and calculated the interaction energy by using molecular docking simulations. The active site pocket of the receptor was found on the CDRs of scFv anti-p17 by the Discovery Studio 2.0 (Accelrys Software Inc.). A site sphere radius of 25 Å was set to assign the binding pocket and the ligand partial charge method for assigning partial charges to the ligands during forcefield assignment was CHARMm. Other parameters were set as default.

#### 2.1.2 Molecular dynamics simulation and binding free energy calculations

Molecular Dynamics (MD) simulations were carried out at the molecular mechanics level using the AMBER03 force field as implemented in the AMBER9 suite of programs [5, 6]. Structures of antibody-peptide were solvated in a cubic box of TIP3P water extending at least 10 Å in each direction from the solute, and the cut-off distance was kept to 12 Å to compute the nonbonded interactions. All simulations were performed under periodic boundary conditions [7], and long-range electrostatics were treated by using the particle-mesh-Ewald method [8, 9]. The time step was set to 1 fs and the trajectory was recorded every 0.1 ps. Prior to MD simulations, the systems were relaxed by a series of steepest descent (SD) and conjugated gradient (CG) minimizations. The 2-ns MD simulations were performed based on each of the minimized systems by gradually heating over 60 ps from 0 to 310 K with the protein atoms fixed using a force constant of 5 kcal/mol/Å<sup>2</sup>. Then, a 200 ps

pressure-constant period (NPT) was applied to obtain an equilibrated density of the constrained protein atoms. The following step was a 40 ps-volume-constant period (NVT) at a force constant of 2.5 kcal/mol/Å<sup>2</sup> followed by 100 ps dynamics at a force constant of 1.25 kcal/mol/Å<sup>2</sup>. Finally, a 1.6 ns unrestrained MD simulation (no force applied on any protein atoms) was performed for each fully flexible system in the NVT ensemble at a constant temperature of 310 K. A total of 500 snapshots were collected at 1 ps intervals from the last 500 ps of MD for binding free energy analysis.

Based on the selected MD snapshots, the binding free energy for each antibody-peptide system could be estimated using MM-PBSA [10]. The binding free energy,  $\Delta G_{\text{bind}}$ , is written as the sum of the gas phase contribution,  $\Delta G_{\text{gas}}$ , the desolvation free energy of the system upon binding,  $\Delta G_{\text{desolv}}$ , and an entropic contribution,  $-T\Delta S$  (Figure 1):



**Figure 1** The thermodynamic cycle used for MM-PBSA calculations

$$\begin{aligned}
 \Delta G_{\text{bind}} &= \Delta G_{\text{gas}} - \Delta G_{\text{solv}}^{\text{P}} - \Delta G_{\text{solv}}^{\text{L}} + \Delta G_{\text{solv}}^{\text{C}} \\
 &= \Delta H_{\text{gas}} - T\Delta S - \Delta G_{\text{solv}}^{\text{P}} - \Delta G_{\text{solv}}^{\text{L}} + \Delta G_{\text{solv}}^{\text{C}} \\
 \Delta H_{\text{gas}} \approx \Delta E_{\text{gas}} &= \Delta E_{\text{elec}} + \Delta E_{\text{vdW}} + \Delta E_{\text{intra}}
 \end{aligned}$$

where the term  $\Delta H_{\text{gas}}$  contains the van der Waals ( $\Delta E_{\text{vdW}}$ ) and electrostatic ( $\Delta E_{\text{elec}}$ ) interaction energies between the two partners in the complex, and the internal energy variation (including bond, angle, and torsional angle energies) between the complex and the isolated molecules ( $\Delta E_{\text{intra}}$ ), respectively;  $-T\Delta S$  is the change of conformational entropy upon peptide binding, which is not considered here because of its high computational demand and relatively low accuracy of prediction [11]. All energies are averaged along the MD trajectories.  $\Delta G_{\text{desolv}}$  is the difference between the solvation free energy,  $\Delta G_{\text{solv}}$ , of the complex and that of the isolated parts.  $\Delta G_{\text{solv}}$  is divided into the electrostatic,  $\Delta G_{\text{elec,solv}}$ , and the nonpolar,  $\Delta G_{\text{np,solv}}$ , contributions:  $\Delta G_{\text{solv}} = \Delta G_{\text{elec,solv}} + \Delta G_{\text{np,solv}}$

For the MMPBSA calculations,  $\Delta G_{\text{elec,solv}}$  was calculated with a built-in module, the pbsa program in AMBER9 which solves the Poisson-Boltzmann equation. The grid size for the PB calculations was 0.5 Å. The values of the interior and exterior dielectric constants were set to 1 and 80, respectively.  $\Delta G_{\text{np,solv}}$  was estimated based on the solvent accessible surface area (SASA) as  $\Delta G_{\text{np,solv}} = 0.0072 \times \text{SASA}$ , using the molsurf program. The scFv anti-p17/peptide interaction energy profiles were generated by decomposing the total binding free energies into residue-residue interaction pairs by the MM-GBSA decomposition process in the mm\_pbsa program of AMBER9.



### **2.1.3 Alanine Scanning**

Alanine scanning [12], a computational method of systematic alanine substitution, has been particularly useful for the identification of functional epitopes. Substitution with alanine removes the side chain atoms of the residues in CDR loops of scFv anti-p17. All the alanine mutant structures were obtained by truncating the residue in the hypervariable portions of the loops on the heavy chain (H1-H3) and light chain (L1-L3) of scFv anti-p17. Then the modified parameter files were generated again by using the LEaP module [13]. This was extrapolated to the snapshots collected from the trajectories resulting from the MD simulations by using the script mm\_pbsa.pl implemented in the AMBER package. From the decomposed energy and alanine scanning result, the point mutation has been selected and investigated.

## **2.2. Affinity determination**

### **2.2.1 Competitive ELISA**

### **2.2.2 Indirect ELISA**

## **3. Discussion**

### **3.1 Alanine scanning**

In comparison with wild type, the calculated binding free energy and alanine scanning analysis demonstrated the importance of the residues of scFv anti-p17 in the binding pocket which are ASP31, ASN35, TRP50, MET100, LYS101, SER103, TYR104, SER146, ASN153, LEU169, LYS172, ASP174, TRP208, GLN209, and LEU213. In previous work, the results of calculated binding free energies found that five amino acids, MET100, LYS101, ASN169, HIS228, and LEU229, were defined for their major contribution to the binding efficiency of natural HIV epitope at the C-terminal on p17 [2]. In this calculation with alanine scanning, the result demonstrated the importance of the residues of scFv anti-p17 are ASP31, ASN35, TRP50, MET100, LYS101, SER103, TYR104, SER162, ASN169, LEU185, LYS188, ASP190, TRP224, GLN225, and LEU229. Therefore, from both results the importance of the mutation's location has 16 positions in the CDRs of scFv anti-p17. The mutation of amino acids in the CDRs of scFv anti-p17 has been initially selected at position 31. Mutation of aspartic acid residues in position 31 of scFv anti-p17 (wild type) was mutated to glutamic acid, asparagine, glycine, threonine, lysine, tyrosine, phenylalanine, tryptophan, valine, and methionine, respectively. Finally, the mutated scFv anti-p17 were docked with nine peptides by CDOCKER (data not shown). The calculated docking interaction energy between single mutation from aspartic acid 31 to glycine (D31G) and peptide sequence (DTGNNSQVSQNY) has shown the favorably interaction energy compared to wild type which has been correlate well with the indirect ELISA. Detail on analysis will be discussed in the following section.

### **3.2 Structural analysis of scFv anti-p17 point mutations**

### **3.3 Comparison of calculated binding free energy with experimental data**

We report a structural analysis of three human amyloid-beta binding alcohol dehydrogenase (ABAD) mutations, identified in patients with 2-methyl-3-hydroxybutyryl-coA dehydrogenase (MHBD) deficiency. Mapping of the mutations (R130C, L122V, and N247S) on ABAD crystal structure revealed that they occur in the interfaces of the enzyme tetramer. The wild-type and mutant enzymes were then subjected to molecular dynamics

simulations with the intention of studying the local effects of the mutations on protein structure.

121DTGHSSQVSQNY132

121DTGNNQVSQNY132

Our results provide an explanation for the strong effect of the R130C mutation on protein stability, evidenced from experimental results. Possibly, the primary effect of this mutation is to impair dimer assembly, as it changes the hot spot character of position 130 to null spot and causes the loss of important hydrogen bonds mediated by the R130 side chain, including a conserved interface hydrogen bond. The other two mutations do not significantly change the energetic contribution of residues 122 and 247 to subunit association, but they are predicted to cause structural changes that affect the enzymatic activity.

#### 4. Conclusion

The identification of the key residues of scFv in the complementarity determining regions (CDRs) from the computational alanine scanning and binding free energy calculation can be used in design the new potential scFv anti-p17. From the result, new antibody can be designed by mutating the potential amino acid residues in CDRs of scFv anti-p17 for improving the binding activities in further studies.

#### ACKNOWLEDGMENTS

The authors would like to express grateful acknowledgement to the Thailand Research Fund (TRF), Computational Nanoscience Consortium (CNC), and National Nanotechnology Center (NANOTEC Thailand) for the access to Discovery Studio Version 2.0 program package. The authors also acknowledge a financial support from the Center for Innovation in Chemistry (PERCH-CIC), and Commission on Higher Education, Ministry of Education of Thailand.

#### REFERENCES

1. Tewari, D., Goldstein, S.L., Notkins, A.L., and Zhou, P., *J. Immunol.*, 1998, **161**, 2642-2647.
2. Lee, V. S., Tue-ngeun, P., Nangola, S., Kitidee, K., Jitonnorn, J., Nimmanpipug, P., Jiranusornkul, S., and Tayapiwatana, C., *Mol. Imm.*, 2010, **47**, 982-990.
3. Vieth, M., Hirst, J.D., Kolinski, A., and Brooks III, C.L., *J. Comput. Chem.*, 1998, **19**, 1612.
4. Vieth, M., Hirst, J.D., Dominy, B.N., Daigler, H., and Brooks III, C.L., *J. Comput. Chem.*, 1998, **19**, 1623.
5. Case, D.A., Cheatham III, T.E., Darden, T., Gohlke, H., Luo, R., Merz, J.K.M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J., *J. Comput. Chem.*, 2005, **26**, 1668-1688.
6. Case, D.A., Darden, T.A., Cheatham III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Pearlman, D.A., and Crowley, M., AMBER 9, University of California, San Francisco, CA, 2006.
7. Weber, W., Hünenberger, P., and McCammon, J., *J. Phys. Chem. B*, 2000, **104**, 3668-4575.
8. Darden, T., York, D., and Pedersen, L., *J. Chem. Phys.*, 1993, **98**, 10089-10092.

9. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L., *J. Chem. Phys.*, 1995, **103**, 8577-8592.
10. Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., and Cheatham 3rd., T.E., *Acc. Chem. Res.*, 2000, **33**, 889-897.
11. Hou, T., Guo, S., and Xu, X., *J. Phys. Chem. B*, 2002, **106**, 5527-5535.
12. Kortemme, T., Kim, D.E., Baker, D., *Computational alanine scanning of protein-protein interfaces*, *Sci STKE* pl2, 2004.
13. Saíz-Urra, L., Cabrera, M. A., and Froeyen, M., *J. Mol. Graph. Model*, 2011, **29**(5), 726-739.

### Table Caption

**Table 1** Vdw and electrostatic contribution of binding free energy of 17.1 and 17.9 (wild type, mutant) (CYX)

**Table 2** Binding free energy, Cdocker energy, Experiment Indirect ELISA, Competitive ELISA of 17.1, 17.9 (wild, mutant) (CYX)

### Figure Caption

**Figure 1** Histograms reporting the binding free energies for the mutations. The total bar height reflects relative binding free energies of each amino acids in CDRs loops with wild type of scFv anti-p17 whose mutation to alanine by alanine scanning mutagenesis.

**Figure 2** Decomposition energy of amino acids in the CDRs 17.1, 17.9 for wild and mutant D31G scFv anti-p17

**Figure 3** Structural comparison of (a) wild-type scFv anti-p17 with p17.1, (b) wild-type scFv anti-p17 with p17.9, (c) mutant D31G with p17.1, and (d) mutant D31G with p17.9 D31G

**Table 1**

Peptide Names	scFv-p17	Binding free energy in each contribution (kcal/mol)							
		$\Delta E_{LE}$	$\Delta V_{DW}$	$\Delta G_{AS}$	$\Delta P_{B_{SUR}}$	$\Delta P_{B_{CAL}}$	$\Delta P_{B_{SOL}}$	$\Delta P_{B_{ELE}}$	$\Delta P_{B_{TOT}}$
p17.1 ( <sup>121</sup> DTGHSSQVSQNY <sup>132</sup> )	wild type	-249.35	-44.24	-293.59	-8.22	270.15	261.93	20.80	-31.65
	D31G	-278.26	-36.55	-314.81	-7.02	299.58	292.57	21.33	-22.25
p17.9 ( <sup>121</sup> DTGNNSQVSQNY <sup>132</sup> )	wild type	-239.51	-34.19	-273.70	-6.93	254.92	247.98	15.41	-25.72
	D31G	-119.30	-23.20	-142.49	-3.76	127.21	123.46	7.92	-19.04

**Table 2**

Peptide	scFv	Cdocker energy (kcal/mol)	Binding free energy ( $\Delta P_{B_{TOT}}$ ) (kcal/mol)	Experimental	
				Indirect ELISA (cfu/ml)	Competitive ELISA (1ug/ml)
p17.1	wild	-110.460	-31.65	1.822	0.067
	D31G	-117.893	-22.25	1.878	0.137
p17.9	wild	-98.877	-25.72	0.241	0.093
	D31G	-100.970	-19.04	1.184	0.186

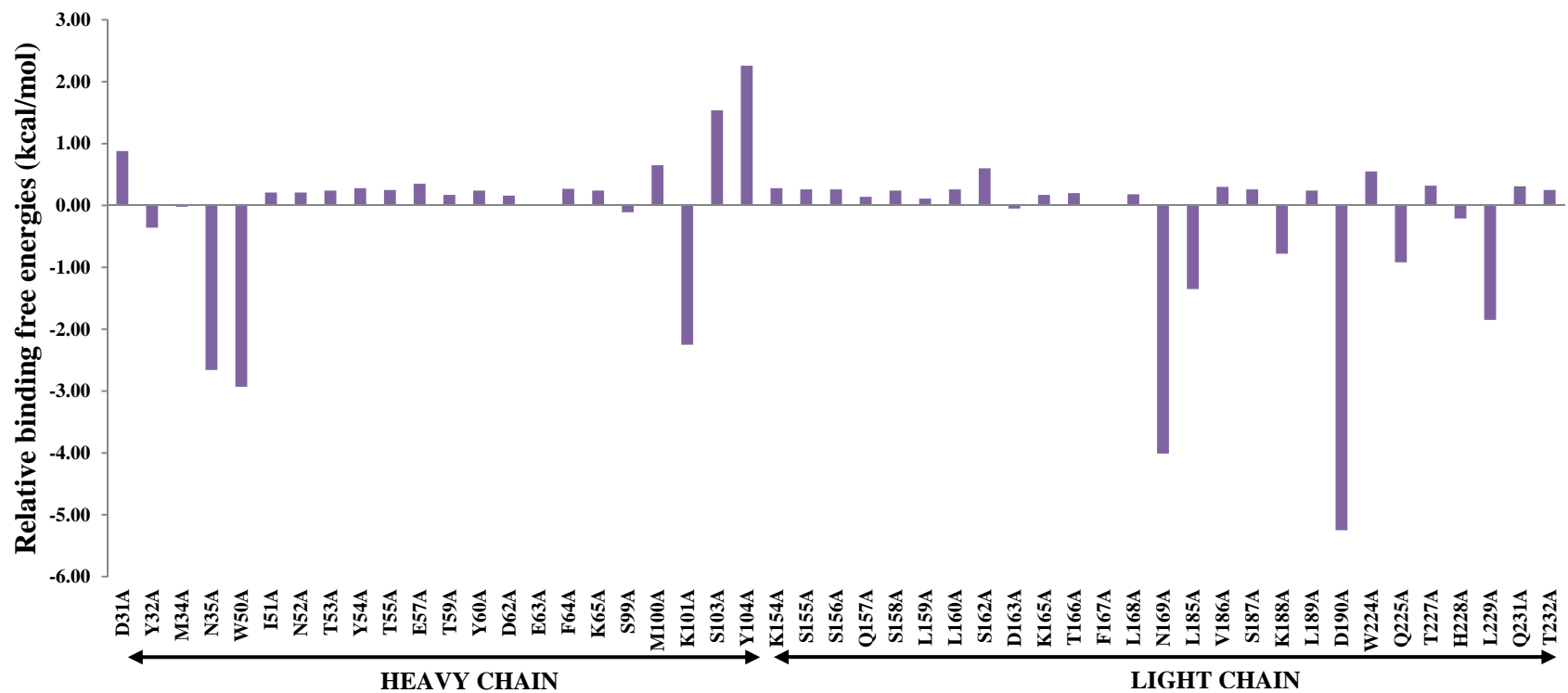
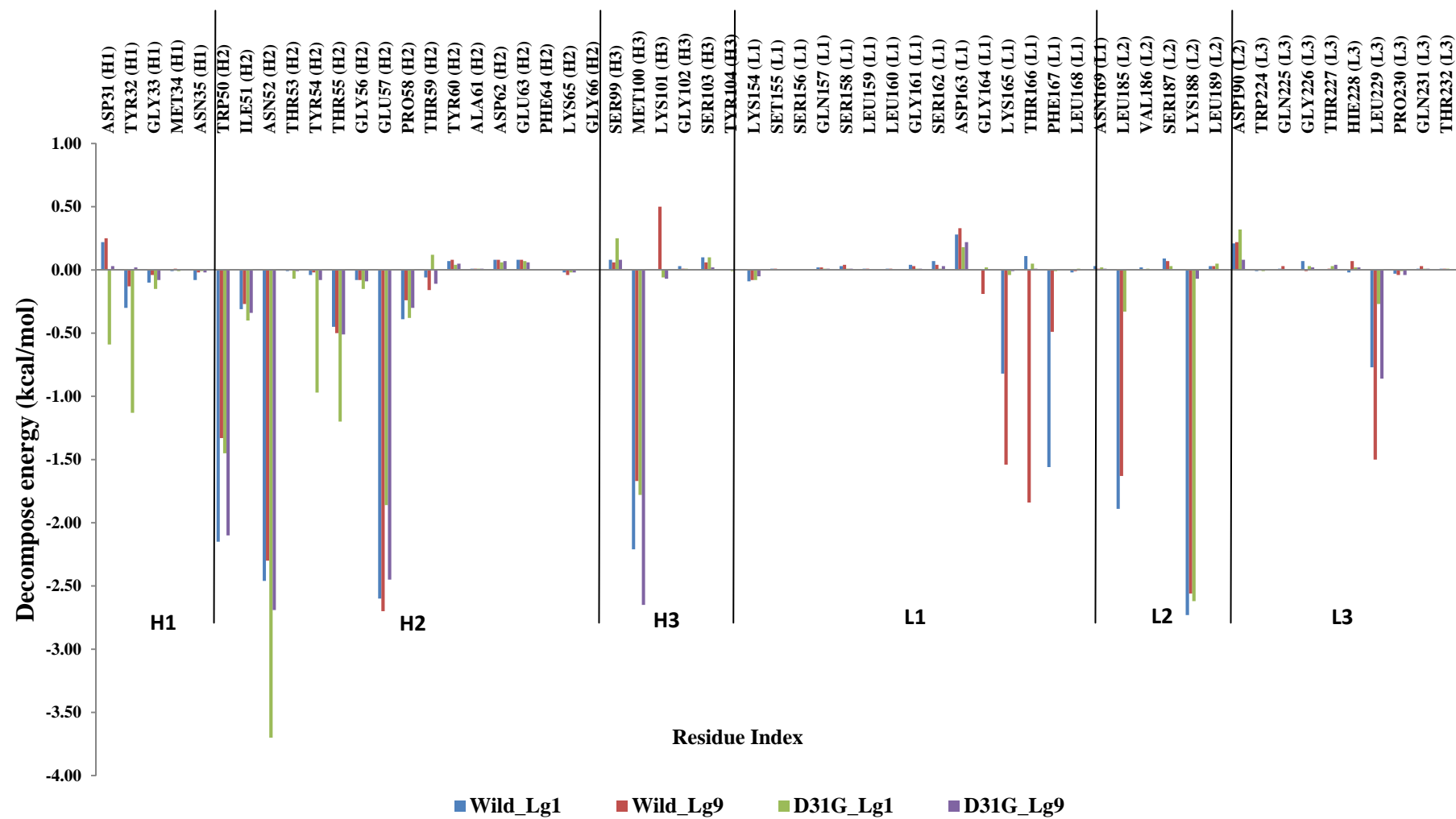


Figure 1



**Figure 2**

**Figure 3**

## **Stabilization of Anti-Ras scFv fragments with/without disulfide bond through theoretical modeling and molecular dynamics simulations investigation**

**V. S. Lee<sup>1, 2, C</sup>, K. Kodchakorn<sup>1</sup>, P. Nimmanpipug<sup>1, 2</sup>, and C. Tayapiwatana<sup>2, 3, C</sup>**

<sup>1</sup>Computational Simulation and Modeling Laboratory (CSML), Department of Chemistry and Center for Innovation in Chemistry, at the Faculty of Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>2</sup>Thailand Center of Excellence in Physics, CHE, Ministry of Education, Bangkok 10400, THAILAND

<sup>3</sup>Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences,

<sup>4</sup>Biomedical Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency at the Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200, Thailand

<sup>C</sup>E-mail: vannajan@gmail.com; Fax: 6653-892277; Tel. 6689-1100216  
asimi002@hotmail.com; Fax: 6653-946042; Tel. 6681-8845141

### **Abstract**

Intracellularly expressed antibody fragments (intrabodies) have been used as powerful tools for clinical applications and for the functional analysis of proteins inside the cell. Among several types of intrabodies, single chain fragment variables (scFv) composed of only the variable regions (VH or VL) of antibodies are the smallest and thus the easiest to design. However, normal antibody fragments do not form disulfide bonds in the cytoplasm and usually are unable to achieve a stable native fold in the absence of the disulfide bonds. Recently, the crystal structure of anti-RAS VH and VL fragments without disulfide bonds after substitution of cysteine residues and the wild type VH and VL with intact disulfide bonds showed no structural differences between the two types of the VH and VL. There is great interest in engineering antibody fragments that will fold and are stable under reducing conditions, and that could serve as framework to which other specificities could be grafted. We have undertaken a molecular dynamics simulations to investigate the two types of such systems with and without disulfide bonds. The structural analysis in term of distance geometry analysis, hydrogen bond analysis, residue interactions, the disulfide strain energy and binding affinity between the domain, were observed in order to explain the stability of the antibody scFv fragments.

**Keywords:** Disulfide bond, Antibody engineering, ScFv fragment, Molecular dynamics simulation, Molecular modeling.

## Introduction

Due to its specificity, antibody molecule is valuable for diagnosis and therapy. Currently antibody variable fragments i.e. single chain Fv (scFv) has been applied for gene therapy to manipulate microbial and cellular signaling pathways by intracellular antibodies (intrabodies) strategy. The scFv can practically be directed to the specific cellular cytoplasm, endoplasmic reticulum (ER), nucleus and mitochondria. The ER environment is the natural milieu for antibody expression. Correspondingly, a number of successes were reported when the specific targets of intrabody were in ER. However, several technical hurdles have thus far restricted their more pervasive use. The intrachain disulfide bond in scFv structure, which contributes to protein folding and stability are typically not thermodynamically favored in the reducing environment of the cytoplasm. The failure to fold limits both the production of the functional scFvs and their targeting against antigens in a redox condition. Thus, expression of functional scFv in the cytoplasm of mammalian cell is potentially of challenge for intrabody perspectives. Interestingly in 2005 Shaki-Loewenstein et al. proposed the universal strategy to express functional scFv using maltose binding protein (MBP) fusion partner in redox environment [ref]. The MBP motive acted as chaperone in promoting the correctly fold of scFv in cytoplasm. However the inverse effect of MBP has not been elucidated so far.

In addition the folding state of individual scFv is an intrinsic property demonstrated by Rajpal et al [ref]. The functional cysteine-less scFv anti-caspase3 retained in cytoplasm was obtained in this study. Many attempts have been made to graft the six CDRs with defined antigen combining sites onto the framework regions on such a SS-bond free scFvs. However, this approach often fails to produce active molecules and requires further mutagenesis to recover the specificity of the CDR-donor antibody. Although certain successful cases were demonstrated for the functional grafted scFvs, the technology is cumbersome and may not be applied in general. More recently the X-ray crystallographic data of wild-type and SS-bond free scFv anti-RAS were generated [ref]. This information will be applied in our study to determine the possible interaction compartments which support in stabilizing the conformational structure of SS-bond free scFv. It will be compared with a representative scFv anti-HIV p17 c-terminal which accumulates in a misfolded and inactive state when expresses in cytoplasm. The acquired information will be potentially useful in rational design of the stable and functional scFv in redox environment.

Because antibodies ordinarily are designed to be secreted from the cell, intrabodies require special alterations, including the use of single-chain antibodies, modification of immunoglobulin domains for hyperstability, selection of antibodies resistant to the more reducing intracellular environment, or expression as a fusion protein with maltose binding protein or other stable intracellular proteins. The efficiency of the expression of scFv in the cytoplasm is often hampered by misfolding, degradation, or aggregation of scFv due to reduced conditions in the cytoplasm (Cattaneo and Biocca, 1999). In addition, due to the lack of disulfide bonds, scFv molecules often fail to exhibit proper folding (Lobato and Rabbitts, 2003). To circumvent these hurdles, several approaches have been exploited to enhance the production of stable and functional intrabodies in the cellular environment. Recently, it was reported the single domain fragments maintain their functions as intrabodies without intradomain disulfide bonds when expressed inside the cells. In our study, we perform the molecular dynamics simulation of antibody scFv fragments with/without disulfide bond to explain the stability and to determine the key residues in the variable region of the antibodies contained the highly conserved intradomain disulfide bonds that hold the  $\beta$ -sheet structure associated with antigen binding. The dynamics



investigation of the crystal structure of anti-RAS VH and VL fragments after substitution of cysteine residues and the wild type VH and VL with intact disulfide bonds in comparison with the some selected antibody which is not stable in cytoplasm were investigated. In addition, this theoretically technique allows the insight understanding and selection of the stable antibody fragments from molecular dynamics simulations.

### **Molecular modeling and molecular dynamic simulations**

The structures of the disulfide and disulfide-free anti-Ras scFv were modeled based on the X-ray structures PDB file 2UZI.pdb and 2VH5.pdb, respectively. In comparison, another theoretical model of unstable scFv in cytoplasm, anti-p17 scFv were investigated. Under the reducing conditions of the cytoplasm, (Gilbert, 1990) the conserved intradomain disulfide bond of antibody domains (Williams & Barclay, 1988) cannot form. Indeed, cytoplasmically expressed scFv fragments in COS cells, plants and bacteria were shown to carry free sulfhydryl groups (Biocca et al., 1995; Martineau et al., 1998; Tavladoraki et al., 1999). As the highly conserved intradomain disulfides are critical for the stability of scFv fragments (Proba et al., 1997; Glockshuber et al., 1992), only intrinsically very stable scFv fragments will be able to fold correctly in sufficient amounts to be active as cytoplasmic intrabodies. This is consistent with the finding that many cytoplasmic intrabodies show low expression levels and limited half-lives (Cattaneo & Biocca, 1999), and therefore only weak inhibitory effects. Moreover, many fragments selected under oxidizing conditions (for example, by phage display) fail to be active as intrabodies at all in the reducing cytoplasm (Visintin et al., 1999), probably because they are not stable enough. Others only prove to be active in the cytoplasm when expressed at lower temperature (20°C), but not at 30°C (Visintin et al., 1999), which would prevent many in vivo applications. To investigate such a system, we have performed the molecular dynamics simulations for the following systems;

1) **Anti-Ras scFv** (2UZI.pdb) with intra-domain disulfide, CYS22-CYS96 in VH and CYS136-CYS201 in VL, were modeled to be in sulfhydryl form

2) **Anti-Ras scFv** (2VH5.pdb) without intra-domain disulfide, ALA22-VAL96 in VH and VAL136-ALA201 in VL

3) **Anti-p17 scFv** with intra-domain disulfide CYS22-CYS96 in VH and CYS137-CYS207 in VL were modeled to be in sulfhydryl form.

4) **Anti-p17 scFv** without intra-domain disulfide, ALA22-VAL96 in VH and VAL137-ALA207 in VL

MD simulations were carried out at the molecular mechanics level using the AMBER03 force .eld as implemented in the AMBER9 suite of programs (Case et al., 2005). Structures of antibody-peptide were solvated in a cubic box of TIP3P water extending at least 10 Å in each direction from the solute, and the cut-off distance was kept to 12 Å to compute the nonbonded interactions. All simulations were performed under periodic boundary conditions (Weber et al., 2000), and long-range electrostatics were treated by using the particle-mesh-Ewald method (Darden et al., 1993; Essmann et al., 1995). The time step was set to 1 fs and the trajectory was recorded every 0.1 ps. Prior to MD simulations, the systems were relaxed by a series of steepest descent (SD) and conjugated gradient (CG) minimizations. The 2-nsMDsimulations were performed based on each of the minimized systems by gradually heating over 60 ps from 0 to 310K with the protein atoms fixed using a force constant of 5 kcal/mol/Å<sup>2</sup>. Then, a 200 ps pressure-constant period (NPT) was applied to obtain an equilibrated density of the constrained protein atoms. The following step was a 40 ps-volume-constant period (NVT) at a force constant

of 2.5 kcal/mol/ Å<sup>2</sup> followed by 100 ps dynamics at a force constant of 1.25 kcal/mol/ Å<sup>2</sup>. Finally, a 1.6 ns unrestrained MD simulation (no force applied on any protein atoms) was performed for each fully flexible system in the NVT ensemble at a constant temperature of 310 K. A total of 500 snapshots were collected at 1 ps intervals from the last 500 ps of MD for binding free energy analysis. Based on the selected MD snapshots, the binding free energy for each antibody-peptide system could be estimated using MM-PBSA (Molecular Mechanics Poisson–Boltzmann Surface Area) (Kollman et al., 2000) and MM-GBSA (Molecular Mechanics Generalized Born Solvent Area) (Chong et al., 1999). The binding free energies between VH and VL ( $\Delta G_{\text{binding}}$ ) were determined from the free energies of the complex, protein and peptide according to the equation:  $\Delta G_{\text{binding}} = \Delta G_{\text{water}}(\text{VH+VL}) - \Delta G_{\text{water}}(\text{VH}) + \Delta G_{\text{water}}(\text{VL})$ . The binding free energies for each species in turn were estimated from the absolute molecular mechanical energies (EMM), the solvation free energies (GPB/GB +  $G_{\text{nonpolar}}$ ) and the vibration, rotation and translation entropies. Each of these terms was calculated as follows:

$$\begin{aligned}\Delta G_{\text{water}} &= \text{EMM} + \Delta G_{\text{solvation}} - T\Delta S; \\ G_{\text{solvation}} &= G_{\text{solvation-electrostatic}} + G_{\text{nonpolar}}; \\ \text{EMM} &= E_{\text{internal}} + E_{\text{electrostatic}} + E_{\text{vdW}}; \\ E_{\text{internal}} &= E_{\text{bond}} + E_{\text{angle}} - E_{\text{torsion}}\end{aligned}$$

where the internal energy  $E_{\text{internal}}$  has three contributions:  $E_{\text{bond}}$ ,  $E_{\text{angle}}$ , and  $E_{\text{torsion}}$ , which represent the strain energy in bonds, angles and torsion angles caused by their deviation from the equilibrium values;  $E_{\text{electrostatic}}$  and  $E_{\text{vdW}}$  are the electrostatic and van der Waals interaction energies, respectively;  $-T\Delta S$  is the change of conformational entropy upon peptide binding, which is not considered here because of its high computational demand and relatively low accuracy of prediction (Hou et al., 2002). All energies are averaged along the MD trajectories. EMM was determined with the sander module of the AMBER suite with an infinite cut-off for all interactions. For the MMPBSA calculations,  $\Delta G_{\text{PB}}$  was calculated with a built-in module, the pbsa program in AMBER9 which solves the Poisson–Boltzmann equation. The grid size for the PB calculations was 0.5 Å. In the MM-GBSA calculations,  $\Delta G_{\text{GB}}$  was calculated using the GB model with the parameters developed by Tsui and Case (2000). The values of the interior and exterior dielectric constants were set to 1 and 80, respectively.  $\Delta G_{\text{np}}$  was estimated based on the solvent accessible surface area (SASA) as  $\Delta G_{\text{np}} = 0.0072 \text{ SASA}$ , using the molsurf program (Kabat et al., 1983) and (Case et al., 2006). The scFv anti-p17/peptide interaction energy profiles were generated by decomposing the total binding free energies into residue-residue interaction pairs by the MM-GBSA decomposition process in the mm pbsa program of AMBER9 (Gohlke et al., 2003; Hou et al., 2008).

## Results and discussions

### Dynamics of system

Fully functional sulfhydryl, and cysteine-free derivative by replacing the disulfide forming cysteine residues in VH and VL with the amino acid combination valine-alanine in both domains of the intrinsically of a stable anti-RAS scFv from protein data bank were simulated in water by molecular dynamics simulation. Intra domain in anti-Ras system. Schematic representation in Figure 1 of the immunoglobulin heavy and light-chain variable domain fold displaying nine  $\beta$ -strands in each VH (VHa1-VHb9) and VL (VHa1-VHb9). The possible presence of Cys residues and disulfide bonds between the  $\beta$ -sheets is marked in ball and stick model. The root mean square displacement (RMSD) of VH in anti-Ras of VH in Figure 2(left) show significantly stable structure while the RMSD of VL in anti-Ras (Figure 2, right) and both chain of anti-p17 scFv (Figure 3) are fluctuated. In consistent with the selected intra-domain distance, S-S in sulfhydryl and C-C in cysteine-free system (Figure 4 and 5), the stable distance in VH of anti-Ras during production dynamics of 1 ns in Figure 4 (left) was observed in sulfhydryl, and cysteine-free system. Not all scFv can enhance the stability just only replacing cys with val and ala. The cysteine-free anti-GCN4(SS22) in 4D5 scFv fragment (Wörn, A., and Plückthun, A. (1998) *FEBS Lett.* **237**, 357–361) caused a smaller decrease in reporter gene activity than the anti-GCN4 wild-type. Thus, the reduced dithiol form of the wildtype scFv intrabody, as present in the reducing environment of the cytoplasm, performed better *in vivo* than the anti-GCN4(SS22). Replacing the disulfides in scFv fragments with Val-Ala pairs therefore does not necessarily cause an improved effect of cytoplasmically expressed intrabodies, even though Val-Ala pairs had been found to be slightly more stable compared with the reduced dithiol form of the 4D5 scFv fragment (Wörn, A., and Plückthun, A. (1998) *FEBS Lett.* **237**, 357–361). The overall stability of a two-domain protein such as a scFv fragment depends on the intrinsic stability of the VL and VH domains as well as on the stability of the interface. However, apparently this effect is antibody-specific or overcompensated by the aggregation tendency of the cysteinefree variant, and these differences require a more detailed investigation. Our results suggest that, in principle, a disulfide-free fully functional derivative of any scFv can be obtained and the low fluctuation of the distance in VH of sulfhydryl system could reflect the stability of the disulfide-free after valine-alanine substitution .

### **Role of intra- and inter-domain interaction and hydrogen bond analysis in stabilization of anti-scFv. How much interactions need to stabilize inter-domain of scFv?**

It is a challenge for protein engineers to understand the factors influencing scFv stability in more detail, in order to be able to produce highly stable fragments, following either rational or evolutionary approaches. The significantly improved activity of scFv fragments with increased *in vitro* stability during extracellular and intracellular applications *in vivo* has been demonstrated in a number of examples (Benhar & Pastan, 1995; Willuda et al., 1999; Wong et al., 2000) and further underlines the fact that stability engineering is worth the effort. The challenges in engineering antibody scFv fragments for higher stability are not fundamentally different from those in stability engineering of other proteins (reviewed by Faganin, 1995; Shaw & Bott, 1996; Rubingh, 1997; Forrer et al., 1999), where also rational and evolutionary approaches have been used. However, in the case of antibody fragments, and scFvs in particular, several advantages are apparent, which may even make them useful as a model system for extracting general features about protein stability. Firstly, there are probably more crystal structures solved than for any other class of related topology (not counting the point mutants of some enzymes). Secondly, the domains do interact, but can also be expressed and studied individually, allowing a quantification of the contribution of the domain interactions to stability. Thirdly, the conserved disulfide bonds can be taken out and thus the domains can be studied as "low-stability" and "high-stability" versions. Stabilization of scFv fragments for intracellular applications can either be achieved by introducing stabilizing point mutations or by CDR grafting onto a superior scFv framework. The latter strategy has also been successfully employed for stabilizing a fluorescein-binding scFv fragment (Jung, S., and Plückh, A. (1997) *Protein Eng.* **10**, 959–966). Additionally, attaching a constant domain to the scFv fragments may also improve scFv stability, because constant domains may possibly provide additional extrinsic domain stabilization to the scFv. Although not tested by quantitative *in vitro* stability measurements yet, it is possible that it was this increased stability, which caused improved effects of cytoplasmically expressed intrabodies in some cases, where scFv constant domain fusions were expressed, and improved performance as cytoplasmic intrabodies was noticed (Cohen, P., Mani, J. C., and Lane, D. P. (1998) *Oncogene* **17**, 2445–2456, Tewari, D., Goldstein, S. L., Notkins, A. L., and Zhou, P. (1998) *J. Immunol.* **161**, 2642–2647). However, it is also possible that increased folding yield is responsible for the improved effect of constant domain fusions, since the presence of the constant domain covers the hydrophobic V-C interface of the variable domain, which is solvent exposed in the scFv fragment and can contribute to aggregation (Jung, S., and Plückh, A. (1997) *Protein Eng.* **10**, 959–966).

In the role of intra- and inter-domain interaction as in Table 1, from the x-ray structure the strong interaction in VH of anti-Ras scFv (-681 kcal/mol) was observed in comparison with the disulfide free system (-644) which mainly due to the loss in the electrostatic moiety. However, in the VL of anti-Ras favorable binding in disulfide free system with lower interaction (-524 compared to -495 kcal/mol) between intra-domain were found which may be the cause of the low fluctuation of the intra-domain distance in VL from a short molecular dynamics simulation (Figure 3, bottom right). Detailed investigation of the interaction energy of  $\beta$ -sheet between intra-domain has been done in Table 2, indicated the strong interaction between  $\beta$ -sheet  $\alpha 1$ .... and  $\beta$ -sheet  $\alpha 1$ . The key amino acids that hold the  $\beta$ -sheet structure of disulfide free anti-Ras in VH are.....and in VL are while .....are the key amino acids in holding the VH and VL to stay together. Analysis of the hydrogen bond as in Table 3 also confirmed the similar hydrogen bond interaction between disulfide and disulfide-free anti-

Ras system. The highlights in bold letter are the different hydrogen bond between the two systems.

Discuss more figure 6, RMSD of the original and final MD structure of anti-p17 show higher deviation and VH and VL seem to be move apart in the no cysteine sytem.

### Geometry and the strain of the disulfide bonds

One critical hallmark influencing the intrinsic domain stability of both VH and VL is the intradomain disulfide bond, which is formed by the highly conserved Cys residues at framework positions **H22/H92** and **L23/L88**, respectively (Williams & Barclay, 1988) (numbering according to Kabat et al., 1991). The stabilizing contribution of intradomain disulfides in antibody domains has been calculated to be in the range of 4-6 kcal/mol (Goto & Hamaguchi, 1979; Goto et al., 1987; Frisch et al., 1996). The correct native structure can, in principle, also be reached in the absence of the disulfide (s), as shown by X-ray crystallography for the disulfide-free REIV-k domain (Usón et al., 1997) and by functional studies for disulfide-free variants of the levan-binding scFv A48 (Proba et al., 1998) and the p185HER2-ECD-binding 4D5 scFv fragments (Wörn & Plückthun, 1998b). However, many scFv fragments cannot fold to the native structure without disulfide bonds (e.g. Glockshuber et al., 1992) because they are not stable enough to "survive" reduction or removal of the disulfide bonds energetically. Nevertheless, some exceptions have been reported. These proteins are either intrinsically stable enough (Wörn & Plückthun, 1998b; Tavladoraki et al., 1999) or their overall stability had to be increased beforehand by rational engineering (as in case of the isolated REIV-k domain; Frisch et al., 1996) or by evolutionary approaches (Proba et al., 1998; Martineau et al., 1998). Completely cysteine-free scFv fragments have been produced (Proba et al., 1998; Wörn & Plückthun, 1998b). Val-Ala pairs were selected from 25 possible combinations as being the best substitution for the VL disulfide bond in the cysteine-free scFv fragment A48 (Proba et al., 1998). Indeed, a cysteine-free variant of the scFv 4D5, where the two intradomain disulfide bonds had been replaced with Val-Ala pairs, was found to be slightly more stable than the reduced dithiol form of the cysteine-containing 4D5 (Wörn & Plückthun, 1998b). However, in a model system established in yeast, such a Val-Ala variant of a GCN4-binding scFv fragment turned out to be less active as cytoplasmic intrabody than the corresponding cysteine containing variant (Wörn et al., 2000). Thus, replacing the intradomain disulfides with Val-Ala pairs does at least not generally increase the performance of scFv fragments expressed under reducing conditions, and the structural reasons for this difference in behavior need to be studied further. Besides the obvious stabilizing effect of the intradomain disulfide bond, many single point mutations have been reported to increase the intrinsic stability of scFv domains. Some of these residues will be described in the section on domain stability. In term of the geometry and strain of disulfide bond as reported earlier for the CD4, a member of the immunoglobuline superfamily of receptors

#### Disulfide exchange in domain 2 of CD4 is required for entry of HIV-1

Lisa J. Matthias<sup>1</sup>, Patricia T.W. Yam<sup>1</sup>, Xing-Mai Jiang<sup>1</sup>, Nick Vandegraaff<sup>1</sup>, Peng Li<sup>1</sup>, Pantelis Pountourios<sup>1</sup>, Neil Donoghue<sup>1</sup> and Philip J. Hogg<sup>1</sup>

Published online: 1 July 2002, corrected 9 July 2002 (details online); doi:10.1038/18115

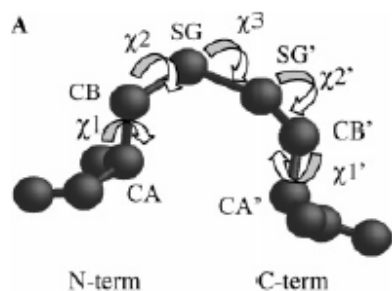
nature paper), calculation of the dihedral strain energies off the disulfide bonds reveals that the D2 bond has a high strain energy (4.74 kcal/mol) compare to the D1 (2.28 kcal/mol) and D4 (1.71 kcal/mol). A high dihedral strain energy (of D2) correlates with the ease of reduction of the disulfide bond.

Katz, B.A. & Kossiakoff, A. The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J. Biol. Chem.* **261**, 15480–15485 (1986).

D2 disulfide bond could be reduced

easily on the surface of cell. In our case, we have calculate in the similar manner from the equation,

$$E \text{ (kcal/mol)} = 2(1 + \cos(3\chi_1)) + 2(1 + \cos(3\chi_{1'})) + (1 + \cos(3\chi_2)) + (1 + \cos(3\chi_{2'})) + 3.5(1 + \cos(2\chi_3)) + 0.6(1 + \cos(3\chi_3)) \quad (1)$$



where  $E$  is the strain energy in kcal/mol and  $\chi_1, \chi_2, \chi_3, \chi_{2'}$ , and  $\chi_{1'}$  are the standard side chain torsion angles of disulfide bonds. As a result, the strain energy in VH and VL of anti-Ras and anti-p17 scFv are ....., ....., ....., ....., respectively.

**Which other residues, not in the intra-domain related to the stabilization of scFv structure?**

**I haven't analysed this part yet ka**

Interface stability and the domain-swapping phenomenon Insufficient stability of the VH-VL interface of scFv fragments has often been suggested as a main cause of irreversible scFv inactivation (Reiter et al., 1994a), since transient opening of the interface, which would be allowed by the peptide linker, exposes hydrophobic patches that favor aggregation. Fv fragments have been reported to dissociate into VH and VL domains with KD values ranging from 10<sup>-9</sup> to 10<sup>-6</sup> M (reviewed by Pluckthun, 1992), and this would allow dynamic and transient "opening" of the scFv assembly. The different strategies that have been followed to increase interface stability in scFv fragments are summarized below. A different but related aspect influencing scFv stability is the monomer-dimer equilibrium (Figure 3). Understanding this aspect is important in order to be able to obtain homogeneous scFv preparations. ScFv fragments which favor dimerization can sometimes be desired to create heterodimeric or homodimeric dimers, which have been termed diabodies. They can be used to create bispecific fragments or homodimers with increased avidity and decreased renal clearance time (reviewed by Hudson, 1999; Colcher et al., 1998). Many scFv fragments have been reported to form dimers to some extent, where the VH domain of one chain is paired with the VL domain of another chain and vice versa (Essig et al., 1993; Holliger et al., 1993; Whitlow et al., 1993, 1994; Desplancq et al., 1994) and the dynamics of this process have been investigated (Arndt et al., 1998). The fraction of dimer formed is dependent on linker length (with short linkers between Cys and ten amino acid residues forming almost quantitatively dimer), but also on external parameters such as ionic strength, pH and the presence or absence of antigen (Arndt et al., 1998). Also, different scFv fragments with intermediate linker lengths differ greatly in the amount of dimer formation; the molecular details of this phenomenon are not yet understood. The monomeric form of scFv with normal 15-20 residues linkers is believed to be thermodynamically most stable. However,

depending on expression conditions, ionic strength, and pH the scFv dimer can initially be dominating and eventually remain trapped in this state in a protein-independent manner (Arndt et al., 1998). The role of the linker is discussed again below. Intuitively, diabodies would appear at first to be "more stable" than the monomeric form of the molecule, yet experimental results point to a more complicated picture. The thermodynamic stability of the anti-digoxin scFv 26-10 has been examined with stepwise increasing linker lengths from zero to 25 residues (in units of (Gly4-Ser)) by J. Huston et al. (personal communication; M.-S. Tai, H. Oppermann, W. F. Stafford & J. S. Huston, unpublished results). Saturating antigen served as a noncovalent linker, greatly reducing the difference in stability between monomer and dimer. However, in the absence of antigen, stability differences were quite marked. According to its transition curve in GdnHCl equilibrium unfolding, the normal scFv monomer, which was predominantly formed with linker lengths of 15, 20, and 25 amino acid residues, had denaturation midpoints at nearly twice the denaturant concentration as those of the diabodies, which formed with the 0, 5, and 10-mer linkers. In order to understand the basis for this difference between diabody and scFv stability, the corresponding unlinked 26-10 Fv fragment was also analyzed (J. Huston, personal communication). The midpoint of the transition curve of the Fv fragment was essentially the same as the midpoints of the 26-10 diabodies. For the scFv, a slight increase of stability was observed with increasing linker length, but they were all decidedly more stable than the diabodies, based on the transition midpoints. It might be noted that the VL domain of the Fv fragment was expressed as a fusion to a domain of protein A, but this fusion partner did not contribute to the transition curve of the Fv, but did improve bacterial expression of the free VL domain. The differences in stability between the Fv fragment, the scFv monomer and the less stable scFv dimer emphasize that diabodies are topologically more similar to unlinked Fv fragments than to scFv fragments. Although the chains of a diabody are associated by a tandem repeat of Fv contacts, their overall stability is at best only marginally higher than that of the Fv fragment comprising them. In contrast, the scFv is topologically a single polypeptide chain, where the domains are held at a high local concentration, recently estimated to be in the low millimolar range (Jaeger & Pluckthun, 1999a). Taken together, these experiments suggest that a diabody made from a weakly associated Fv fragment may also be poorly stable, and may thus be unsuitable for many applications in vivo.

## Conclusions

### **Table Caption**

**Table 1** Interaction energy of VH and VL, intra-domain in VH, in VL for anti-Ras, anti-p17, and selected scFv structures.

**Table 2** Interaction energy of  $\beta$ -sheet between intra-domain in anti-Ras system

**Table 3** Hydrogen bond analysis between intra-domain in anti-Ras system

### **Figure Caption**

**Figure 1** Intra domain in anti-Ras system. Schematic representation of the immunoglobulin heavy and light-chain variable domain fold displaying nine  $\beta$ -strands in each VH (VHa1-VHb9) and VL (VLa1-VLb9). The possible presence of Cys residues and disulfide bonds between the  $\beta$ -sheets is marked in ball and stick model

**Figure 2** Root mean square displacement of VH and VL in anti-Ras and anti-p17 systems

**Figure 3** Distance analysis between intra-domain in VH and VL for anti-Ras and anti-p17 systems during molecular dynamics simulation. Positions of key amino acids in the model structure in disulfide (thin lines) and disulfide-free (thick lines) are indicated in the insert picture

Binding free energy between VH and VL in anti-Ras and anti-p17 systems

**Figure 4** Distance analysis between intra-domain in VH and VL for anti-Ras and anti-p17 systems during molecular dynamics simulation. Positions of key amino acids in the model structure in disulfide (thin lines) and disulfide-free (thick lines) are indicated in the insert picture

**Figure 5** Binding free energy of VH and VL in anti-Ras and anti-p17 scFv

**Figure 6** Structural comparison of initial and average structure from molecular dynamics simulation of sulfhydryl and disulfide-free system for anti-RAS and anti-p17 scFv.



Table 1

[illegible]

Table 2

	intradomain		Energy (kcal/mol)
2vH5cys	VHA (1+2, 3, 8, 7)	VHB )4, 5, 6, 9, 10)	
	VLA (11, 12, 17, 16)	VLB (13, 14, 15, 18)	
	VHA	1+2	
		3	
		8	
		7	
	VHB	4	
		5	
		6	
		9	
		10	
	VLA	13	
		14	
		15	
		18	
	VLB	11	
		12	
		17	
		16	
2vH5Nocys			

Table 3

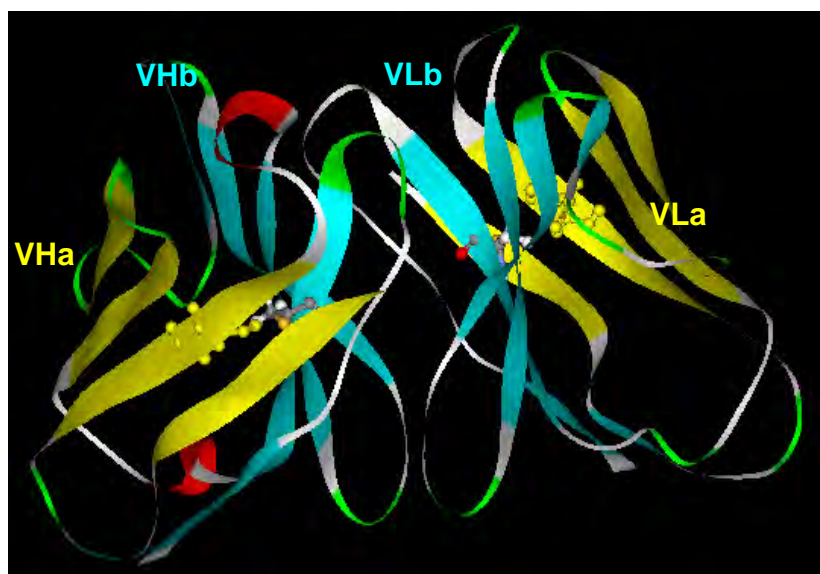
# Hydrogen Bond Analysis in Anti-Ras scFv

## Disulfide System

	DONOR	ACCEPTORH	ACCEPTOR	% Occupied	Distance (Å)
	(res@atom)	(res@atom)	(res@atom)		
CYS22	7@O	21@H	21@N	99.92	2.916
	22@O	79@H	79@N	99.2	2.943
	79@O	22@H	22@N	91.52	2.84
	5@O	23@H	23@N	90.82	3.099
	21@O	7@H	7@N	87.8	2.828
	23@O	5@H	5@N	83.54	3.133
	21@O	7@HG	7@OG	51.48	2.873
CYS96	95@O	37@H	37@N	100	2.86
	97@O	35@H	35@N	99.5	2.963
	35@O	97@H	97@N	98.34	2.968
	37@O	95@H	95@N	90.86	2.875
	96@O	106@H	106@N	85	3.125
CYS136	118@O	137@H	137@N	98.92	3.005
	184@O	136@H	136@N	96.62	2.931
	135@O	120@H	120@N	95.14	2.984
	137@O	118@H	118@N	94.78	3.05
	136@O	184@H	184@N	92.52	3.119
	120@O	135@H	135@N	89.86	2.965
	183@OD1	137@HE	137@NE	89.6	2.933
	183@OD1	137@HH121	137@NH12	70.16	3.002
	183@OD2	137@HE	137@NE	42.16	3.12
	120@O	135@HG1	135@OG1	37.94	2.83
	184@O	136@HG	136@SG	22.3	3.296
	183@OD2	137@HH121	137@NH12	18.62	3.069
	118@O	136@HG	136@SG	18.06	3.236
	119@O	120@HG	120@OG	13.26	2.758
CYS201	200@O	149@H	149@N	99.8	2.93
	202@OE1	149@HH	149@OH	99.54	2.763
	201@O	212@H	212@N	98.88	2.946
	202@O	147@H	147@N	97.1	2.981
	149@O	200@H	200@N	96.42	3.09
	147@O	202@H	202@N	71.92	3.04
	202@OE1	147@HD22	147@ND2	24.64	3.039
	213@OE1	214@H	214@N	9.24	3.291

## No Disulfide System

	DONOR	ACCEPTORH	ACCEPTOR	%occupied	distance
	(res@atom)	(res@atom)	(res@atom)		
ALA22	7@O	21@H	21@N	99.8	2.915
	22@O	79@H	79@N	99.78	2.946
	79@O	22@H	22@N	90.44	2.874
	21@O	7@H	7@N	97.02	2.855
	5@O	23@H	23@N	95.68	3.029
	23@O	5@H	5@N	93.86	3.078
	5@O	7@HG	7@OG	54.54	2.722
	21@O	7@HG	7@OG	12.56	2.927
VAL96	95@O	37@H	37@N	100	2.891
	35@O	97@H	97@N	99.24	2.977
	97@O	35@H	35@N	98.4	3.065
	37@O	95@H	95@N	93.32	2.848
	96@O	106@H	106@N	90.22	3.136
	97@O	35@HD22	35@ND2	7.6	3.07
VAL136	135@O	120@H	120@N	99.82	2.951
	120@O	135@H	135@N	99.56	2.921
	184@O	136@H	136@N	99.5	2.896
	136@O	184@H	184@N	99.18	3.07
	118@O	137@H	137@N	98.44	3.038
	137@O	118@H	118@N	94.88	2.995
	183@OD1	137@HE	137@NE	84.68	2.93
	120@O	135@HG1	135@OG1	69.88	2.812
	183@OD1	137@HH121	137@NH12	60.76	3.042
	183@OD2	137@HE	137@NE	50.98	3.122
	183@OD2	137@HH121	137@NH12	25.16	3.056
	119@O	120@HG	120@OG	14.58	2.757
	138@O	137@HH11	137@NH1	11.82	2.931
AT A 201	200@O	149@HH	149@N	99.9	2.926
	201@O	212@H	212@N	99.52	2.925
	149@O	200@H	200@N	97.3	3.073
	202@O	147@H	147@N	94.12	3.102
	147@O	202@H	202@N	66.36	3.04



Cys22 (VH A2) Cys96 (VH B8) Ds-s = 3.38 Å  
 Cys136 (VL A2) Cys201 (VL B8) Ds-s = 3.46 Å

Figure 1

## RMSD of VH and VL in Anti-Ras scFv

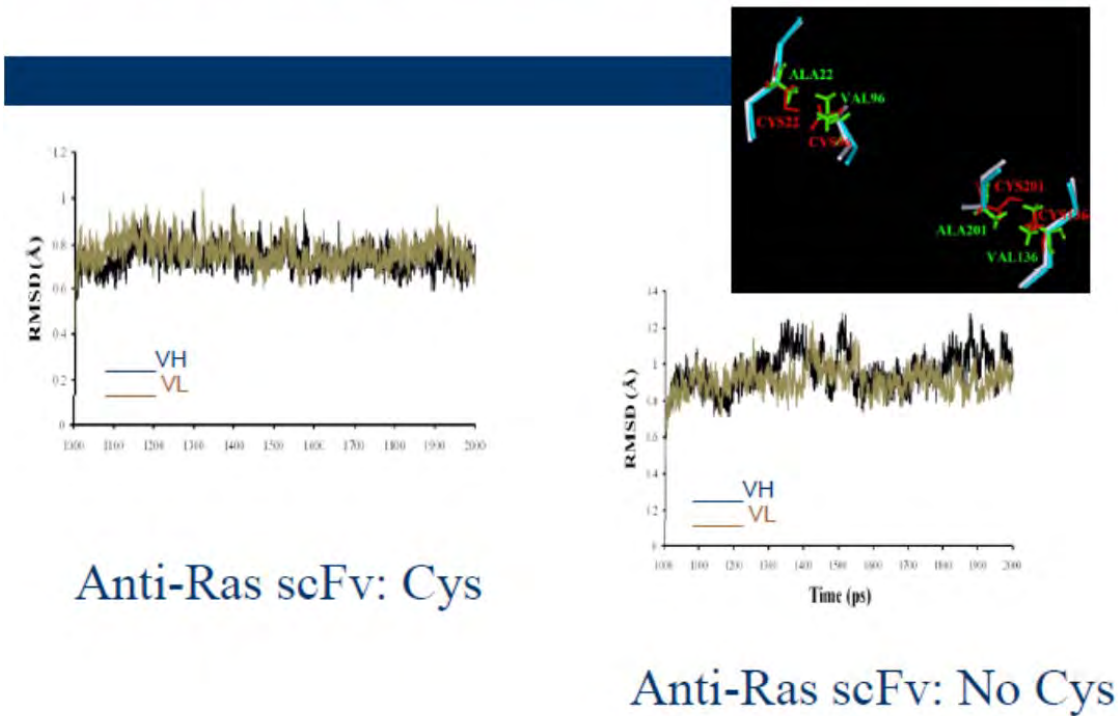


Figure 2

## Distance Analysis between intra-domain in VH and VL during MD Simulations

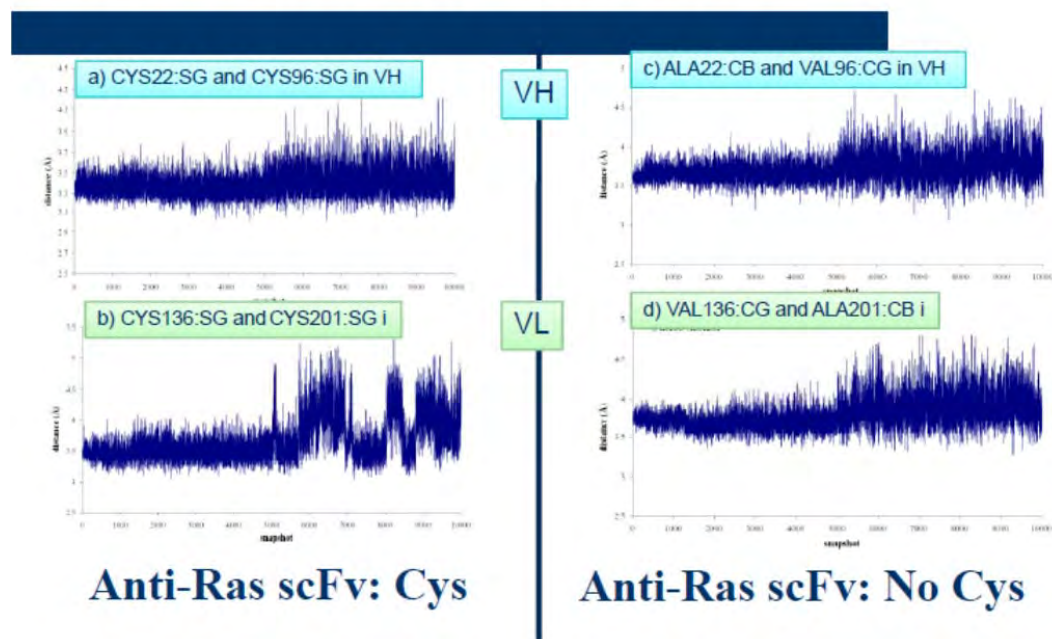


Figure 3



## Distance Analysis between intra-domain in VH and VL during MD Simulations

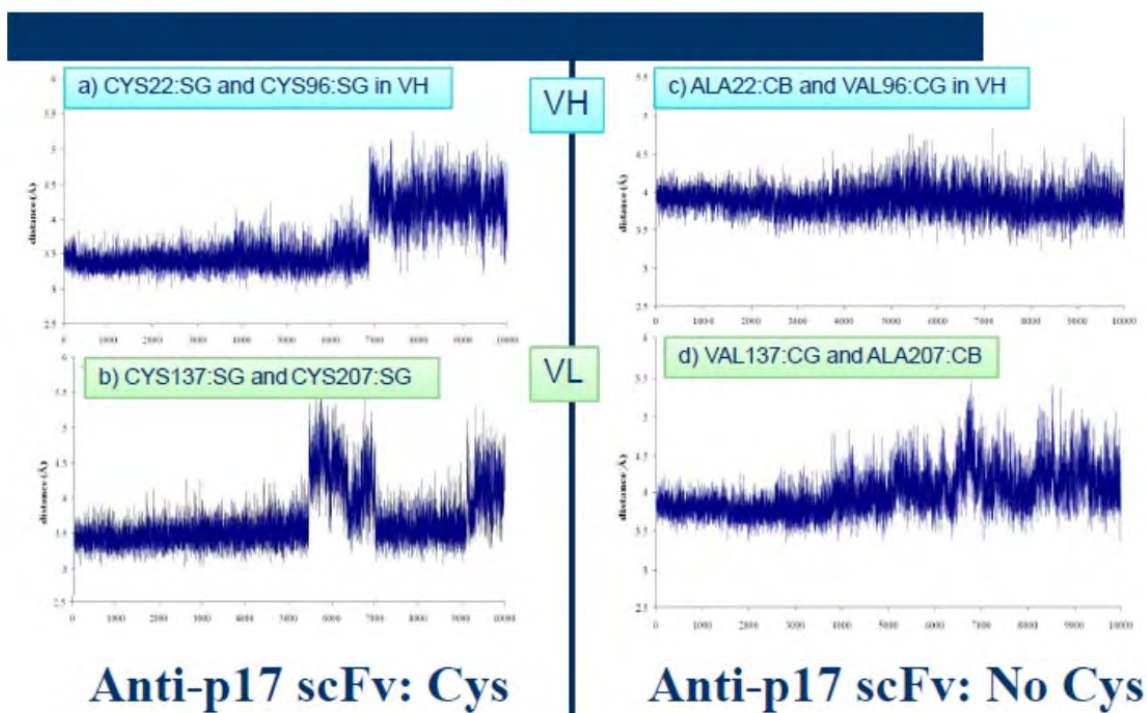
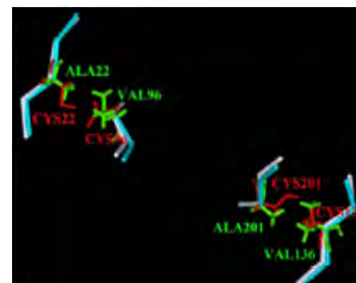
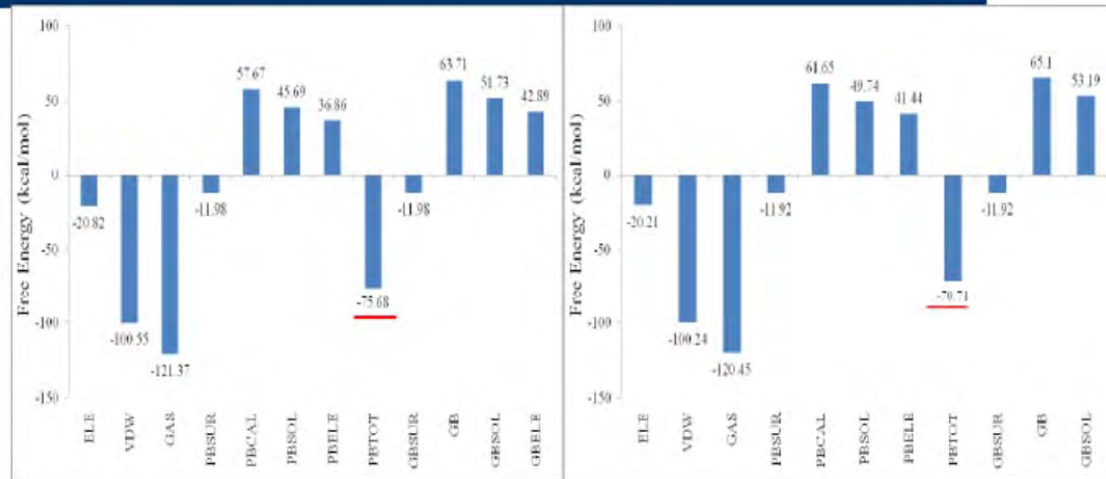


Figure 4





# Binding free energy between VH and VL in anti-ras scFv



Anti-Ras scFv: Cys

Anti-Ras scFv: No Cys

Figure 5



# Binding free energy between VH and VL in anti-p17 scFv

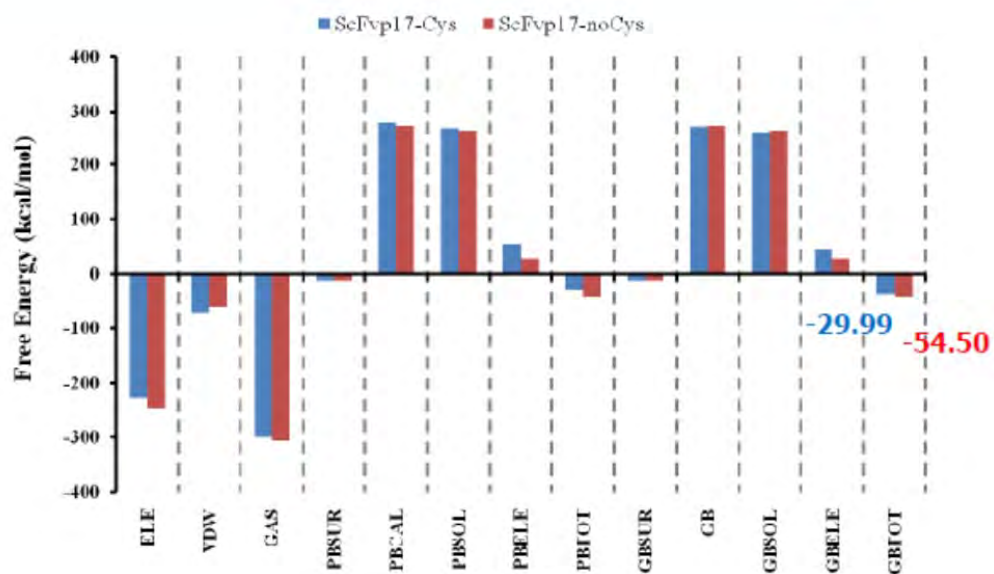


Figure 5 cont.

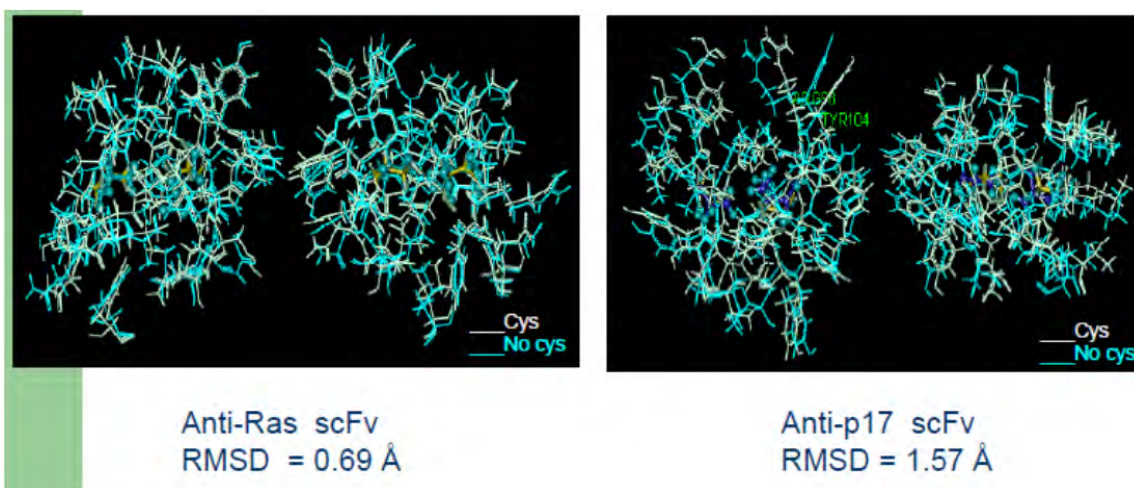


Figure 6

## Modeling of CD147 Binding Sites against ScFv: A Guide for Epitope Mapping

Vannajan Sanghiran Lee<sup>1\*</sup>, Jitrayut Jitonnorn<sup>1</sup>, Panthip Tue-ngeun<sup>1</sup>, Watchara Kasinrer<sup>2</sup>, Piyarat Nimmanpipug<sup>1</sup>, Chatchai Tayapiwatana<sup>2,3\*\*</sup>

<sup>1</sup>Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200 Thailand

<sup>2</sup>Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, 50200 Thailand

<sup>3</sup>Biomedical Engineering Center, Chiang Mai University, Chiang Mai, 50200 Thailand

Corresponding Authors:

\* +66 53943341-5 ext. 117, E-mail: vannajan@gmail.com,

\*\* +66 818845141, E-mail: asimi002@hotmail.com

**Keywords:** monoclonal antibody, scFv, CD147, homology modeling, protein-protein docking, bioinformatics

### ABSTRACT

A three dimensional model of CD147 antigen was constructed, based on homologous templates retrieved from Protein databank using the restraint based modeling program MODELLER 8v1. The CDR loops of the single-chain variable fragment (scFv) from a monoclonal antibody to CD147 clone M6-1B9 was structurally simulated *in situ* using the Web Antibody Modeling (WAM) server. An overall quality assessment of each refined model was made by various analysis programs, including PROCHECK, PROSA and VERIFY3D. Recognition binding sites on CD147 and the interactive side chains on scFv were identified using the Global Range Molecular Matching (GRAMM) program. The whole protein complexes were then freely relaxed with unrestrained minimization and the binding residues on CD147 were identified within 5Å of the referenced antibody binding site. Epitope prediction using a computer guided docking program suggested that four peptide regions of domain 1 of CD147 may play a critical role in the binding activity of M6-1B9.

## INTRODUCTION

The human cell surface molecule CD147 is a type I integral membrane protein of the immunoglobulin (Ig) superfamily [1-3]. It is widely expressed on various cell types and is a lymphocyte activation-associated molecule [3-5]. Currently several monoclonal antibodies (mAbs) have been generated to target this molecule for functional studies. By employing these established mAbs, certain domains on CD147 have been characterized. Kasinrerker *et al.* reported two consecutive CD147 domains that resulted in different activations i.e. induction of homotypic cell aggregation and inhibition of lymphocyte proliferation after anti-CD3 induction with two mAb panels [4, 6, 7]. A negative regulatory function for CD147 in T cells has been shown using either M6-1E9 or M6-1B9 anti-CD147 mAbs which recognized CD147 domain 1 [7]. More recently, Ku *et al.* demonstrated the inhibition of matrix metalloproteinase production from COS-7 cells after engagement of the well-characterized CD147 epitopes [8]. These findings have significance for the future applications in cancer and transplantation immunotherapy. Nevertheless, immunoglobulin gene and amino acid sequences involved in the CDR domains of the established hybridomas have never been investigated. Recently, our group has cloned CD147 specific single-chain Fv (scFv) from a hybridoma, M6-1B9, and expressed it in *E. coli*. The inhibiting activity of soluble scFv with the original monoclonal antibody, M6-1B9, was demonstrated. Interestingly, suppression of OKT3-induced T cell proliferation using monomeric soluble scFv was also observed. These results suggested that the generated soluble scFv carries a CD147-specific paratope recognizing a native epitope on the CD147 molecule [9]. The amino acid sequence of scFv-M6-1B9 was deduced from the acquired nucleotide sequence encoding the light (L) and heavy (H) chain variable segments. The complementary determining regions (CDRs) were determined *via* the Kabat numbering pattern [10].

Herein we describe the three-dimensional structures that have not previously been elucidated for scFv-M6-1B9 and CD147 complex. The modeled proteins were docked by the software GRAMM [11, 12] on a PC-LINUX workstation in order to study the interaction between scFv and CD147. Possible interactions of scFv with the potential epitopes on CD147 domain 1 could thus be predicted. The predicted data will be used for comparison with B-lymphocyte linear epitope prediction protocols. As this is the first report regarding the structural basis of the interaction of scFv and CD147, our studies may contribute towards the understanding of the immunological activity of scFv against CD147.

## RESULTS AND DISCUSSION

### scFv-M6-1B9 model and validation

In the present study, specific CD147 scFv was constructed and used to model using the information from the PDB database. The computer model of scFv was constructed using the Web Antibody Modeling (WAM) server (<http://antibody.bath.ac.uk/index.html>) [13]. However, the method did not provide the linker for the generated model. Therefore, the linker was built separately by using the AMBER program (Case *et al.*, 2006). This linker was included into the constructed ScFv model.

The 3D structure of the scFv-M6-1B9 model with linker is shown in Fig. 1(a). This model includes two structural domains corresponding to the VH and VL domains. The linker connected the N-terminal of the heavy chain to the C-terminal of the light chain (see Fig. 1). It is generally acknowledged that antigen binding occurs in a variety of ways, and thus leads to a differing arrangement of antibody

conformations. In scFv, there are actually six unique hypervariable units. Each of the hypervariable regions is surrounded by less variable stretches, which are structurally  $\beta$ -sheets. Each of the chains contains three of the six loops that form the binding region. The hypervariable portions of the loops on the light chain are designated L1(22-33), L2(49-55) and L3(88-96) while those on the heavy chain are H1(134-143), H2(158-167), and H3(207-212) [14-17] (Fig. 1). These regions are also known as complementary determining regions (CDRs); they have a higher binding affinity to the antigen. The positions of amino acid for the CDR loops (L1-L3 and H1-H3), which correspond to the antigen-binding loops are highlighted in Fig. 1(b).

The reliability analysis for the CD147 scFv modeled structure was checked with various validation methods. The stereochemical quality of the scFv-M6-1B9 model was evaluated by PROCHECK [18] as shown in Fig. 2(a). Dihedral angles ( $\Phi$ ,  $\Psi$ ) of all residues were located in most favored (162 residues, 81.8%), additionally allowed regions (31 residues, 15.7%), generously allowed regions (3 residues, 1.5%) and disallowed regions (2 residues, 1%) of the Ramachandran plot. There were two residues (Ala83 and Thr181) in the disallowed region; Thr181 is located in CDR H2 (175-184) and Ala83 is outside the CDR loops.

All other criteria, such as peptide bond planarity and hydrogen bond energy, were within values statistically expected for proteins with a resolution below 2 Å (data not shown). Additionally, the evaluation of the quality of the modeled structure with ProSA-web [19, 20] yielded an energy graph below zero (see Fig. 2(b)) and a z-score of -6.79 (Fig. 2(c)), which corresponds to scores expected for a native fold of a sequence length of 239 amino acids. This supports the assumption that the modeled protein represents a natively folded structure. Finally, the modeled structure was evaluated for the compatibility of each of the amino acid residues as checked by VERIFY3D [21] (Fig. 2(d)) indicating a positive score for all residues. From all validation tests, then, the structure model of CD147 scFv seems reasonable at the present level of theory.

### CD147 model and validation

An optimal sequence alignment is essential to the success of homology modeling. Three reference proteins, neural cell adhesion molecule from *Rattus norvegicus* (1EPF), junctional adhesion from *Homo sapiens* (1NBQ), and junctional adhesion from *Mus musculus* (1F97), were used to model the structure of human CD147 molecule. The sequence identities for the three reference proteins compared with the CD147-HM were 21.10%, 17.73% and 16.06%, respectively (Table 1). With a low score in sequence identity, we created a sequence identity table to evaluate the similarity among the templates. As shown in Table 1, most templates share a small identity (~16-19%) according to the difference in their species, except for JA-HS and JA-MM. However, those templates were quite different in their sequence and structure, as illustrated in Fig. 3(a) and 3(b). In this case, multiple templates should be used to improve the accuracy of our model. Therefore, we chose all of those proteins as the multiple templates for modeling CD147-HM (see Fig. 3(a)). The secondary structure was calculated using the DSSP program [22] (Fig. 3(c)). It was found that the enzyme was a beta-sandwich fold and contains 18  $\beta$ -strands and 2  $\alpha$ -helices. The disulfide bonds for the CD147-HM model were predicted using the DiANNA web-server. This analysis predicts two disulfide bonds; (1) Cys20 and Cys60, (2) Cys105 and Cys164, which form in each subunit of CD147-HM.

The quality of the homology model was further evaluated by PROCHECK [18], ProSA-web [19, 20] and VERIFY3D [21]. The Ramachandran plot (Fig. 4(a)), revealed a good quality stereochemistry. The  $\Phi$ ,  $\Psi$  torsion angles of 78.6 % of the residues had values within the most favored areas. 18.3 % of the residues had values within additionally allowed regions and 3.1 % of the residues had values within generously allowed regions of the Ramachandran plot. The PROSA energy as shown in Fig. 4(c) indicated some mistakes in the modeled structure. This is due to the bad alignment according to the low sequence identity. However, the z-score for the structure of the model (Fig. 4(b)) was in an acceptable level (-3.02). This result indicated that the modeled protein was still close to the natively folded structure. Further explanation for some errors on the modeled structure was found by the validation test of VERIFY3D. The result from VERIFY3D for the target and the templates is shown in Fig. 4(d). The overall structure of the target model looks fine, except for one region at residues 130-150. The source of negative score for the target model was probably from the template of 1EPF, which showed an energy trend similar to those from the target. However, the overall quality of the structure model of CD147 is reasonable and is within the criteria of each of the validation programs. In comparison with the recent PDB structure (3B5H.pdb), our model structure shares a common structure.

### **Molecular interactions between the scFv-M6-1B9 and CD147**

The structure of scFv-M6-1B9 and CD147 conformation was modeled as described in Materials and Methods and evaluated as mentioned in the earlier sections. In order to predict the mode of interaction between scFv and CD147, the model of scFv-M6-1B9 was docked with CD147. The molecular docking of CD147 into the binding region of scFv was done using the global range molecular matching (GRAMM) program [11, 12]. As a result, the CD147 ScFv antibody complexes clustered into four distinct groups. The four best molecular docking models in each cluster, after energy minimizations using steepest descent and conjugated gradient methods, are shown in Fig. 5. ScFv is shown in green with CDR loops colored in yellow. The different subdomains, VH and VL, of scFv are indicated. Domains I and II of the CD147 antigen are shown in red and blue, respectively. The complex conformation conformed to the rule of lock-key relationship between antigen and antibody. A list of the amino acid residues of CD147 predicted to be interacting within 5Å of the scFv CDR and hydrogen bond interaction are presented in Table 2. The maximum difference in the binding energy of 30 kcal/mol ranges from -581 to -611 kcal/mol for the four complex models. These interactions are reflected in the negative binding energy, which is highly favorable for the complex between scFv and all models. However, from the molecular structure analysis, the predicted interaction of the first domain of CD147 to scFv [15] is found in only model II and IV, which scFv binds near to the CDRs conserved residue indicated the better complex conformations. Since more interacting amino acid residues in domain 1 of CD147 and strong H-bonding interactions which hold the first domain to the scFv in comparison to model IV, model II would seem to be the better complex model. Besides, more complexes in cluster II suggests the possibility of finding such a pattern. In detailed analysis of the model II – scFv-M6-1B9 complex, a conformation is analogous to that seen in the crystal structure of the lysozyme - antibody (PDB code: 1a2y [23]) which also binds to the CDRs of scFv. The interacting residues of CD147 are Phe<sup>6</sup>, Val<sup>9</sup>, Glu<sup>10</sup>, Asp<sup>11</sup>, Leu<sup>12</sup>, Val<sup>40</sup>, Leu<sup>41</sup>, Gly<sup>62</sup>, Glu<sup>63</sup>, His<sup>81</sup>, Gly<sup>82</sup>, Pro<sup>83</sup>, Pro<sup>84</sup>, Arg<sup>85</sup>, Val<sup>86</sup>, and Arg<sup>145</sup>. Interestingly, the residues involved in interaction from the CDRs of scFv

are mainly from the L1, L3 and H2. Possible hydrogen bond interactions associated in two CDR regions, L1 (O:V40 of CD147 with H:S24 of L1 and O:E63 of CD147 with H:S26 of L1), and L3 (O:E10 of CD147 with H:Y93 of L3) were observed. The orientation of model II - scFv-M6-1B9 is illustrated in Fig. 6, showing the amino acids suggested as most important for ligand interaction in 5Å vicinity from CDRs. Hydrogen bonding and hydrophobic residues are also indicated in the figure. Although the most probable anti-CD147 scFv structure has been predicted, there will be many possible stable complexes at equilibrium. Therefore, the low energy structure should be taken into an account to find the interacting peptide sequences of CD147 that are associated with binding to scFv-M6-1B9; we will discuss in the next section.

### **Epitope prediction with computer guided docking program**

Defining epitopes theoretically or experimentally is challenging [24-28]. X-ray crystallography of antibody–antigen complexes is perhaps the best way to validate differential antigen disruption results, but it will require tremendous effort, resources and time. Therefore, we utilized the commonly used and well-accepted docking method to conjecture epitope locations on an antigen. We propose that the ability of a protein model with a particular loop conformation to dock with a complementary ligand is a useful criterion. For each docking structure of the anti-CD147 ScFv antibody from GRAMM, amino acid residues within 5Å binding region of the CDR domain (L1-L3 and H1-H3) were examined from the lowest interaction energy of the 50 complex models, and the common associated binding regions are shown in Fig. 7. Eight peptide regions (A-H) of CD147 interacted with scFv. Four potential epitopes, Group A, B, C and D, which resided in the first domain of CD147, are the preferred binding types [15]. The complex conformations of the four putative binding epitopes and scFv were sequentially analyzed. The deduced peptide regions A, B, C, and D from the minimized scFv-model II of CD147 are selected to perform flexible docking for both antibody and antigen using CAChe 6.1.12.33, as described in Materials and Methods. Predicted binding energies using PMF scores with united atoms are reported in Table 3. The corresponding complexes with the peptide ligands are also predicted. These interactions are reflected in the binding energy of interaction between scFv and peptide ligands, which is highly favorable for Group A-D. The best complex model is Group C with the PMF binding energy of -186.91 kcal/mol. As Thr181 in the H2 loop conformation in the models is located in a disallowed region of Ramachadran plot, the change in CDR H2 conformation may result in the loss of a few interactions. Interactions between the CD147 and other CDR loops may be maintained, despite the change in the shape of CDR H2. However, this effect should be overcome with the flexible docking, which allows both peptides and the CDR loops to be moved. A negative favorable intermolecular energy reflects that peptides A-D have realistic binding modes. The calculated intermolecular energies decrease (become more favorable) from peptide D > A > B, and C, respectively. The patterns of hydrogen bonding and hydrophobic interactions between the key residues within 3Å of scFv and peptide ligands in each case (A-D) are shown (Fig. 8(a) - 8(d)). Although each pattern is quite different, a general trend showing the interaction with the H1-H3 loop is apparent, as shown in Fig. 8. Several of the important hydrogen bonds of H2 and H3 to CDRs, namely, Y157, S177, Y184, R223, and Y229 and the scFv-M6-1B9 residues involved in the hydrophobic contacts, such as Y93, Y157, A158, Y182, Y184, F224, R225, Y229, as listed for each peptide in Table 4, are observed in the predicted complexes.



According to a non-glycosylated continuous epitope of CD147 recognized by M6-1B9 mAb [29], we compared our docking results with an antibody epitope prediction from the Immune Epitope Database and Analysis Resource, IEDB ([http://tools.immuneepitope.org/tools/bcell/iedb\\_input](http://tools.immuneepitope.org/tools/bcell/iedb_input)). Many epitope prediction protocols for CD147 were tested, i.e., Emini Surface Accessibility [30], Kolaskar and Tongaonkar Antigenicity [31] and Bepipred Linear Epitope Prediction [32] based on antigenicity [33], hydrophilicity [34], inverted hydrophobicity [35, 36], accessibility [30], and secondary structure [37, 38]. Among these methods, it has been shown recently that Bepipred gives more accurate prediction than other protocols [32]. Therefore, we have reported this protocol in comparison with our docking method, as listed in Table 3. Three predicted epitope sets were generated from Bepipred. Interestingly, the predicted results from other protocols are similar to those of the Bepipred protocol. Two of three predicted epitopes contain the overlapping regions i.e. <sup>47</sup>PGQKTEFKVDSDDQWG<sup>62</sup> by Emini Surface Accessibility [30] and <sup>77</sup>NIQLHGPPRVKAVKSS<sup>92</sup> by Kolaskar and Tongaonkar Antigenicity [31].

From the docking method, similar results of the two overlapping regions of the shorter peptide Group C(<sup>57</sup>SDDQWGEYSCVFL<sup>69</sup>) and Group D(<sup>81</sup>HGPPRVKA<sup>88</sup>) with the epitope set 2(<sup>45</sup>ALPGQKTEFKVDSDDQWG<sup>62</sup>) and set 3(<sup>78</sup>IQLHGPPRVKAV<sup>89</sup>), respectively, agreed with those of the other protocols. The docking method also suggested two new predicted epitopes of Group A (<sup>7</sup>TTVEDLGSK<sup>15</sup>) and B (<sup>36</sup>KGGVVLKEDA<sup>45</sup>). It is interesting to note that, our four predicted peptides contain the hydrophobic residues such as Cys, Leu and Val, as suggested by Kolaskar and Tongaonkar [31] using a semi-empirical method for prediction of antigenic determinants on protein antigens. These hydrophobic residues found on the surface are likely to be a part of antigenic sites. This result implies that our docking prediction is in good agreement with the other linear epitope prediction protocols, although both predictions were clearly different in their backgrounds. In fact, the docking method is based on the three dimensional structural information of both antigen and antibody and their binding interaction, whereas the linear epitope prediction is based on the statistical information of an antigen from its antigenicity, hydrophilicity, inverted hydrophobicity, accessibility, and secondary structure. Combining the statistical and structural information of both antigen and antibody would provide more reliable epitope prediction for further testing.

## Conclusions

Based on the availability of structural information generated by the antigen-antibody docking method, it may become possible to conjecture epitope locations on an antigen. In our study, the unknown crystal structure scFv-M6-1B9 and CD147 were first generated using a homology modeling method. Then, the overall quality of each refined model was evaluated by different analysis programs including PROCHECK, PROSA and VERIFY3D. The scFv-M6-1B9/CD147 complexes were identified using the Global Range Molecular Matching (GRAMM) program. The key interacting peptide regions were mapped from the lowest interaction energy of the 50 complex models and then the peptide ligands were docked to scFv-M6-1B9. Our calculation suggests that the four epitopes, <sup>7</sup>TTVEDLGSK<sup>15</sup>, <sup>36</sup>KGGVVLKEDA<sup>45</sup>, <sup>57</sup>SDDQWGEYSCVFL<sup>69</sup>, and <sup>81</sup>HGPPRVKA<sup>88</sup> of CD147 could bind with a high affinity to scFv-M6-1B9. In agreement with other protocols from the statistical information, the latter two potential epitopes can be selected as a basis for further testing.

## Computational Methods

### 3D model building of scFv-M6-1B9 antibody

A single-chain variable fragment homology model for anti-CD147 was built by using the Web Antibody Modeling (WAM) server (<http://antibody.bath.ac.uk/index.html>) [13]. Heavy and light chain variable fragments (VH and VL) were modeled using the most sequence-homologous known framework structures. For conformations of the CDRs, five loops (all except H3) were identified as belonging to specific canonical classes and modeled on those with maximum sequence homology within the corresponding class. Non-canonical and H3 loops were built using the CAMAL (Combined Antibody Modeling Algorithm) method, consisting of a combined database/conformational search [39]. Canonical non-conserved and non-canonical residue side chains were modeled using the CONGEN algorithm [40]. Final energy minimizations on CAMAL-generated conformations were performed by employing the accessibility profile screen. The obtained scFv-M6-1B9 model was added all hydrogen atoms and the peptide linker between heavy and light chain which contains Ser and Gly residues was built using AMBER9 (Case *et al.*, 2006). Then, the model was performed constraint energy minimization with gradually reduced force starting from 100, 50, 25, and 0 kJ/mol/Å<sup>2</sup>. 2,500 steps of steepest descent, followed by 2,500 steps of conjugate gradient algorithm were applied during energy minimization using a ff03 force field in AMBER9.

### 3D model building of CD147 antigen

At the time of the research, there is no PDB of CD147 available. Therefore, the amino acid sequence of human CD147 antigen from *Homo sapiens* (CD147-HM) was obtained from the Swiss-Prot-TREMBL database (accession number: P35613) [41]. The templates selected were neural cell adhesion molecule from *Rattus norvegicus* (NCAM-RN) (PDB entry: 1EPF), junctional adhesion from *Homo sapiens* (JA-HS) (PDB entry: 1NBQ), and junctional adhesion from *Mus musculus* (JA-MM) (PDB entry: 1F97); these were obtained from the PDB database [42]. Multiple sequence alignment between CD147-HM and the templates used for homology modeling (provided by ClustalX [43] and represented by ESPript [22]) and the structural overlap between CD147-HM and the templates were shown in Fig. 3(a) and 3(b), respectively. Model building of CD147-HM was performed based on those templates, using the restraint based modeling program MODELLER 8v1 [44]. This program uses an automated approach to comparative modeling *via* the satisfaction of spatial restraints. In the model generating step, we used the "patch" command in MODELER to constrain coordinates of cysteines 20 and 66, and 105 and 164, which form disulfide bonds in each subunit. The disulfide bond for CD147-HM model was predicted by using the DiANNA web-server (<http://clavius.bc.edu/~clotelab/DiANNA/>) [45]. Five models were generated and assessed for their quality using the Modeller Objective Function parameter and PROCHECK [18]. The best model was retained and the side chains were repositioned in optimized conformations, using SCWRL [46]. The CD147-HM model was finally subjected to energy minimization in order to remove the geometrical strain, using parm03 force field as implemented in AMBER9. The overall stereochemical quality of the final model for CD147-HM was assessed by PROCHECK [18], ProSA-web [19, 20] and VERIFY3D [21]. The secondary structure of CD147-HM predicted by DSSP [47] and is illustrated in Fig. 3(c).



### Docking the CD147 antigen onto scFv-M6-1B9 antibody

Recognition docking sites of CD147 on scFv were identified by the Global Range Molecular Matching (GRAMM) program [11, 12] using the generic mode with a grid step of 6.8 Å and torsional rotation of 20° intervals. A total of 1,000 docked structures were generated and fifty antibody-antigen models were chosen based on their energy ranking for further analysis. The CD147 binding residues pattern of 50 protein complexes were identified within 5 Å from the CDR domains. The appropriate binding peptide regions were selected from the predicted scFv-M6-1B9/CD147 models to perform the flexible docking with CAChe 6.1.12.33 (Fujitsu, Inc.) program, where flexible or rigid ligands were docked into proteins with flexible or rigid side chains. In this study, the peptide ligands and the side chains of the amino acids of CDR domains were flexible during the docking simulation. Charges were not included in this calculation. All the atoms in each structure were redefined by CAChe's atom type. The grid resolution was 61 x 61 x 61 points with a spacing of 0.500 Å. The grid box was settled by the center of the CDR domain. The search method used a Lamarckian genetic algorithm and the population size was set to 50 in each of 40,000 generations. The atom pair interaction potentials, knowledge-based PMF-scores [48], were used to analyze inter-atomic interactions within each complex.

### ACKNOWLEDGEMENTS

This work was supported by grants from the Thailand Research Fund (TRF), the National Center for Genetic Engineering and Biotechnology (BIOTEC), the Center for Innovation in Chemistry, Postgraduate Education and Research Program in Chemistry (PERCH-CIC) and also supported by the Biomedical Engineering Center, Chiang Mai University, Thailand. We thank Dr. David Perahia of the Institut de Biochimie et de Biophysique Moléculaire et Cellulaire, Université Paris-Sud, France for his suggestions and critical reading of the manuscript. We also wish to thank Dr. Dale Taneyhill for proofreading the manuscript.

### REFERENCES

- [1] T. Miyauchi, Y. Masuzawa, T. Muramatsu, *J. Biochem. (Tokyo)*, **110**, 770 (1991).
- [2] C. Biswas, Y. Zhang, R. DeCastro, H. Guo, T. Nakamura, H. Kataoka, K. Nabeshima, *Cancer Res.*, **55**, 434 (1995).
- [3] W. Kasinrerker, E. Fiebiger, I. Stefanova, T. Baumruker, W. Knapp, H. Stockinger, *J. Immunol.*, **149**, 847 (1992).
- [4] W. Kasinrerker, N. Tokrasinwit, P. Phunpae, *Immunology*, **96**, 184 (1999).
- [5] G. Staffler, H. Stockinger, *J. Biol. Regul. Homeost. Agents*, **14**, 327 (2000).
- [6] P. Khunkeawla, S. Moonsom, G. Staffler, P. Kongtawelert, W. Kasinrerker, *Immunobiology*, **203**, 659 (2001).
- [7] S. Chiampanichayakul, P. Peng-in, P. Khunkaewla, H. Stockinger, W. Kasinrerker, *Immunobiology*, **211**, 167 (2006).
- [8] X.-M. Ku, C.-G. Liao, Y. Li, Y. X.-M., B. Yang, X.-Y. Yao, L. Wang, L.-M. Kong, P. Zhao, Z.-N. Chen, *Scand. J. Immunol.*, **65**, 435 (2007).
- [9] K. Tragoolpua, N. Intasai, W. Kasinrerker, S. Mai, Y. Yuan, C. Tayapiwatana, **8**, 5 (2008).
- [10] T. T. Wu, E. A. Kabat, *J. Exp. Med.*, **132**, 211 (1970).
- [11] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A. A. Friesem, C. Aflalo, I. A. Vakser, *Proc. Natl. Acad. Sci. USA*, **89**, 2195 (1992).
- [12] I. A. Vakser, *Protein Eng.*, **8**, 371 (1995).
- [13] N. R. Whitelegg, A. R. Rees, *Protein Eng.*, **13**, 819 (2000).

- [14] C. K. Heng, R. Y. Othman, *Bioinformation*, **1**, 118 (2006).
- [15] C. Koch, G. Staffler, R. Huttinger, I. Hilgert, E. Prager, J. Cerny, P. Steinlein, O. Majdic, V. Horejsi, H. Stockinger, *Int. Immunol.*, **11**, 777 (1999).
- [16] C. Chothia, A. M. Lesk, *J. Mol. Biol.*, **196**, 901 (1987).
- [17] R. M. MacCallum, A. C. Martin, J. M. Thornton, *J. Mol. Biol.*, **262**, 732 (1996).
- [18] R. A. Laskowski, D. S. Moss, J. M. Thornton, *J. Mol. Biol.*, **231**, 1049 (1993).
- [19] M. J. Sippl, *Proteins*, **17**, 355 (1993).
- [20] M. Wiederstein, M. J. Sippl, *Nucleic Acids Res.*, **35**, W407 (2007).
- [21] D. Eisenberg, R. Luthy, J. U. Bowie, *Methods Enzymol.*, **277**, 396 (1997).
- [22] P. Gouet, E. Courcelle, D. I. Stuart, F. Metoz, *Bioinformatics*, **15**, 305 (1999).
- [23] J. Ay, T. Keitel, G. Kuttner, H. Wessner, C. Scholz, M. Hahn, W. Hohne, *J. Mol. Biol.*, **301**, 239 (2000).
- [24] M. R. Stone, M. E. Lostrom, M. R. Tam, R. C. Nowinski, *Virology*, **96**, 286 (1979).
- [25] M. H. V. Van Regenmortel, *Methods*, **9**, 465 (1996).
- [26] M. H. Van Regenmortel, *J. Immunol. Methods*, **216**, 37 (1998).
- [27] N. S. Greenspan, E. Di Cera, *Nat. Biotechnol.*, **17**, 936 (1999).
- [28] N. S. Greenspan, *Adv. Cancer. Res.*, **80**, 147 (2001).
- [29] C. Tayapiwatana, P. Arooncharus, W. Kasinrerak, *J. Immunol. Methods*, **281**, 177 (2003).
- [30] E. A. Emini, J. V. Hughes, D. S. Perlow, J. Boger, *J. Virol.*, **55**, 836 (1985).
- [31] A. S. Kolaskar, P. C. Tongaonkar, *FEBS Lett.*, **276**, 172 (1990).
- [32] J. E. Larsen, O. Lund, M. Nielsen, *Immunome Res.*, **2**, 2 (2006).
- [33] G. W. Welling, W. J. Weijer, R. van der Zee, S. Welling-Wester, *FEBS Lett.*, **188**, 215 (1985).
- [34] J. M. Parker, D. Guo, R. S. Hodges, *Biochemistry*, **25**, 5425 (1986).
- [35] J. Kyte, R. F. Doolittle, *J. Mol. Biol.*, **157**, 105 (1982).
- [36] J. L. Cornette, K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, C. DeLisi, *J. Mol. Biol.*, **195**, 659 (1987).
- [37] P. Y. Chou, G. D. Fasman, *Adv. Enzymol. Relat. Areas. Mol. Biol.*, **47**, 45 (1978).
- [38] M. Levitt, *Biochemistry*, **17**, 4277 (1978).
- [39] A. C. Martin, J. C. Cheetham, A. R. Rees, *Proc. Natl. Acad. Sci. USA*, **86**, 9268 (1989).
- [40] R. E. Bruccoleri, M. Karplus, *Biopolymers*, **26**, 137 (1987).
- [41] B. Boeckmann, A. Bairoch, R. Apweiler, M. C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilbout, M. Schneider, *Nucleic Acids Res.*, **31**, 365 (2003).
- [42] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.*, **28**, 235 (2000).
- [43] F. Jeanmougin, J. D. Thompson, M. Gouy, D. G. Higgins, T. J. Gibson, *Trends Biochem. Sci.*, **23**, 403 (1998).
- [44] A. Sali, T. L. Blundell, *J. Mol. Biol.*, **234**, 779 (1993).
- [45] F. Ferre, P. Clote, *Bioinformatics*, **21**, 2336 (2005).
- [46] A. A. Canutescu, A. A. Shelenkov, R. L. Dunbrack, Jr., *Protein Sci.*, **12**, 2001 (2003).
- [47] W. Kabsch, C. Sander, *Biopolymers*, **22**, 2577 (1983).
- [48] I. Muegge, Y. C. Martin, *J. Med. Chem.*, **42**, 791 (1999).

**Table Caption**

**Table 1** Sequence identity of the target sequence with the template sequences of the three known crystal structures.

	CD147-HM	NCAM-RN	JA-HS	JA-MM
CD147-HM	100.00	21.10	17.43	16.06
NCAM-RN		100.00	18.35	18.81
JA-HS			100.00	64.48
JA-MM				100.00

**Table 2** The complexed energy between scFv and CD147 and the main molecular interactions between the scFv binding regions and the list of residues on CD147 within 5Å from the binding regions from the best model in each cluster.

Cluster	Numbers in Cluster	Energy (kcal/mol)	Residues of ScFv	Residues of CD147 within 5Å from ScFv loops	H-bond interaction CD147:CDRs
I	13	-611	L1	130-131	M130@O: S22@HG, N131@HD:S22@OG
			L2	-	-
			L3	143,145	R145@HH: L95@O
			H1	-	-
			H2	-	-
			H3	-	-
II	20	-606	L1	40-41,62-63,84	V40@O:S24@HG, E63@OE2:S26@HG
			L2	-	-
			L3	10-12,63, 82-86, 145	E10@OE2:Y93@HN,
			H1	-	-

			H2	6, 9-10, 81	-
			H3	-	-
III	4	-602	L1	-	-
			L2	97, 111-113, 116, 146-148	W116@HE1:S55@O, S146@OG:S55@HN
			L3	145	-
			H1	127-128, 131, 141-143	-
			H2	-	-
			H3	115-116, 143-146,	N143@O:Y229@HH, R145@O:A228@HN, R145@HH21:G227@O R145@HE:G227@O
IV	13	-581	L1	11, 143-146,	Q143@O:S29@HG R145@HN:S29@O R145@HH11:S30@O
			L2	12, 22, 61-64, 83-85	-
			L3	126-128, 143-144, 170	L127@HZ:N92@OD1,

			H1	15, 168-176	K170@HZ:S160@OG G171@HN:M159@O S172@HN:Y157@O S172@HO:P154@O D173@O:A158@HN A175@HN:S156@O
			H2	117, 119-120, 122-125, 163, 167,170-172, 175-177	Y119@O:Y184@HH Y119@O:S177@HG R163@NE:Y182@HH A175@O:S178@HG
			H3	9-11, 13, 86-89, 168-170, 173-175	D11@OD2:Y229@HH K170@HN:F224@O S169@O:R225@HN

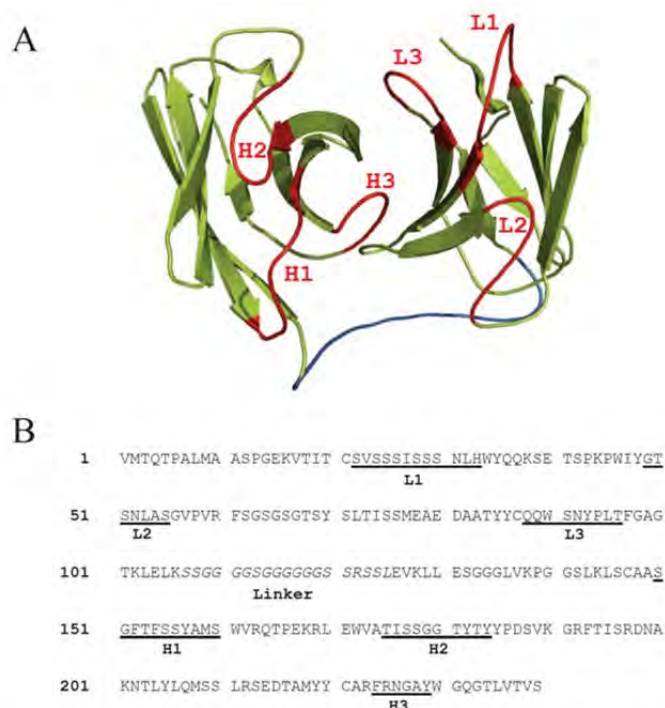
**Table 3** Interaction energy of scFv with short peptide regions from CAChe program.

Group	Residues No. of CD147	PMF Binding Energy (kcal / mol)
A	7-15 ( <sup>7</sup> TTVEDLGSK <sup>15</sup> )	-153.45
B	36-45 ( <sup>36</sup> KGGVVLKEDA <sup>45</sup> )	-171.97
C	57-69 ( <sup>57</sup> SDDQWGEYSCVFL <sup>69</sup> )	-186.91
D	81-88 ( <sup>81</sup> HGPPRVKA <sup>88</sup> )	-136.88
Set 1	23-30 ( <sup>23</sup> NDSATEVT <sup>30</sup> )	-157.90
Set 2	45-62 ( <sup>45</sup> ALPGQKTEFKVDSDDQWG <sup>62</sup> )	-184.43
Set 3	78-89 ( <sup>78</sup> IQLHGPPRVKAV <sup>89</sup> )	-194.48

**Table 4** List of the scFv- M6-1B9 key residues that interacted within a 3Å vicinity from four predicted peptide sequences and the hydrogen bond analysis of scFv-M6-1B9/CD147.

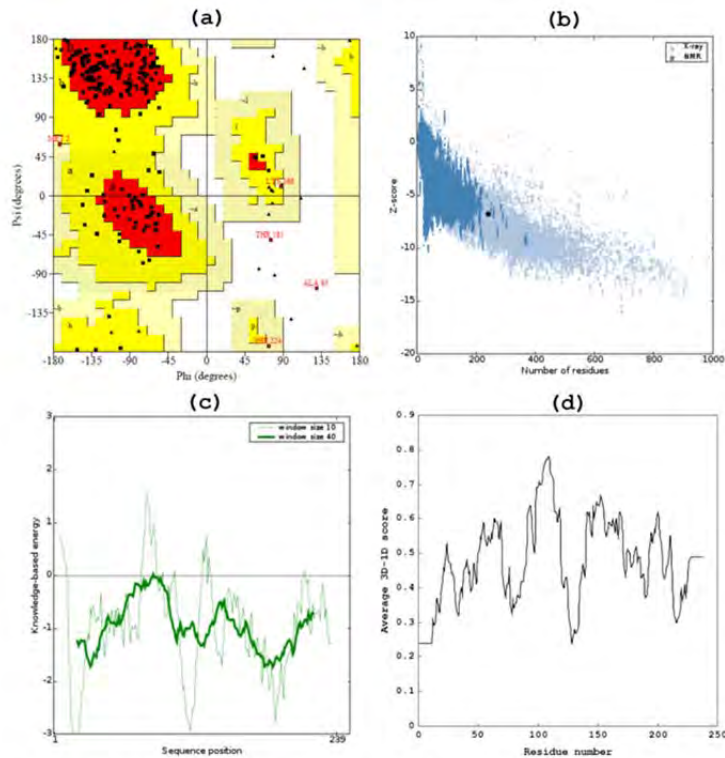
Peptide	Sequence	3Å scFv residues from peptide:CDRs	Hydrophobic contact residues of scFv-M6-1B9	H-bond (scFv:peptide)
A	<sup>7</sup> TTVEDLGSK <sup>15</sup>	Y157:H1, Y229:H3	Y93, Y157, Y229	Y157@OH:S14@NH
B	<sup>36</sup> KGGVVLKEDA <sup>45</sup>	S177, Y184:H2, Y229:H3	Y157, S177, Y182, Y184, F224, R225, Y229	S177@OH:G37@O, Y184@OH:G37@NH
C	<sup>57</sup> SDDQWGEYSCVFL <sup>69</sup>	R223, Y229:H3	Y93, Y184, R223, F224 Y229	R223NH1, Y229@OH:D58@OD1, Y229@OH:S57@O
D	<sup>81</sup> HGPPRVKA <sup>88</sup>	Y157:H1, Y184:H2, Y229:H3	Y93, Y157, A158, Y184, F224	Y184@OH:K87@O, Y229@OH:H81@NE

**Figure Caption**



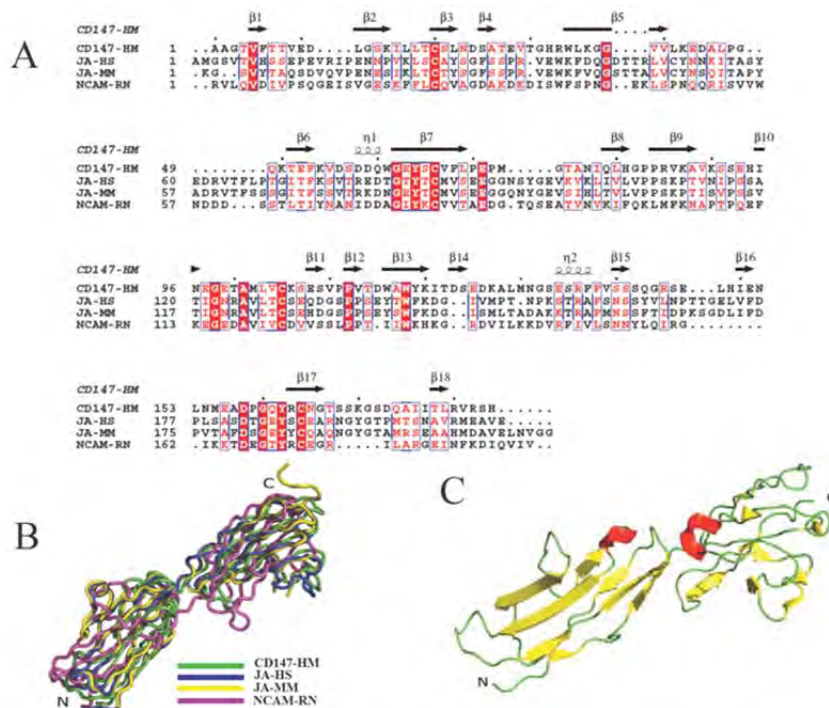
**Fig. 1 a)** The 3D structure of light and heavy chain variable regions (VL and VH) of anti-CD147 scFv. The loops are shown in red and the linker in blue. **b)** Primary structure of VL(1-106), VH(126-239), the loops in light chain (L1: 22-33, L2: 49-55 and L3: 88-96), and the loops in heavy chain (H1: 150-160, H2: 175-184, and H3: 224-229) of anti-CD147 scFv. The sequences and positions of the flexible linker, which contains Ser and Gly residues, are also shown in italics.

**Fig. 1 a)** The 3D structure of light and heavy chain variable regions (VL and VH) of anti-CD147 scFv. The loops are shown in red and the linker in blue. **b)** Primary structure of VL(1-106), VH(126-239), the loops in light chain (L1: 22-33, L2: 49-55 and L3: 88-96), and the loops in heavy chain (H1: 150-160, H2: 175-184, and H3: 224-229) of anti-CD147 scFv. The sequences and positions of the flexible linker, which contains Ser and Gly residues, are also shown in italics.



**Fig. 2** The validation of scFv model. (a) Ramachandran plot (b) PROSA z-score (c) PROSA energy (d) VERIFY3D.

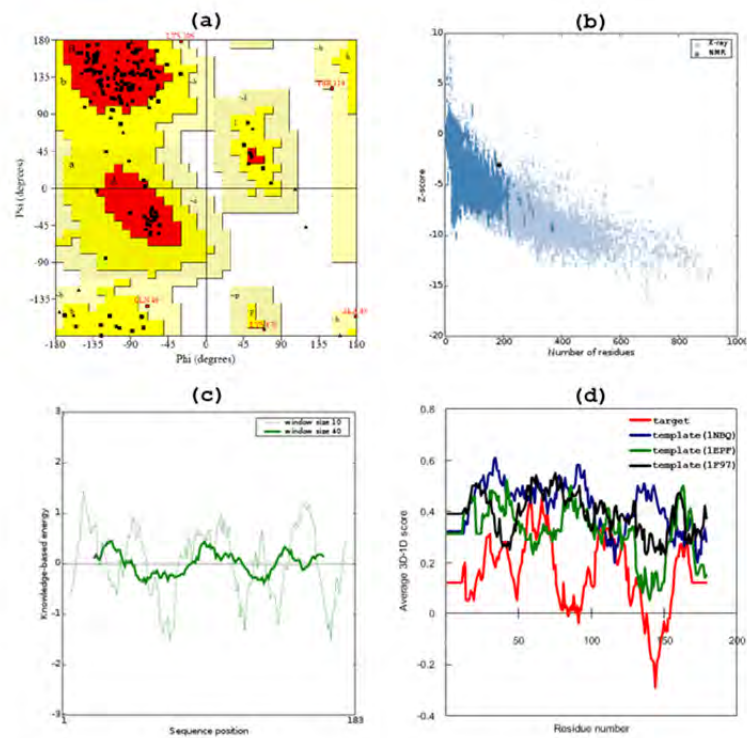
**Fig. 2** The validation of scFv model. (a) Ramachandran plot (b) PROSA z-score (c) PROSA energy (d) VERIFY3D.



**Fig. 3** Multiple sequence alignment between CD147-HM and the templates used for homology modeling. (A) the sequence alignment (B) structural overlap between CD147-HM and the templates (C) the secondary structure of CD147-HM.

**Fig. 3** Multiple sequence alignment between CD147-HM and the templates used for

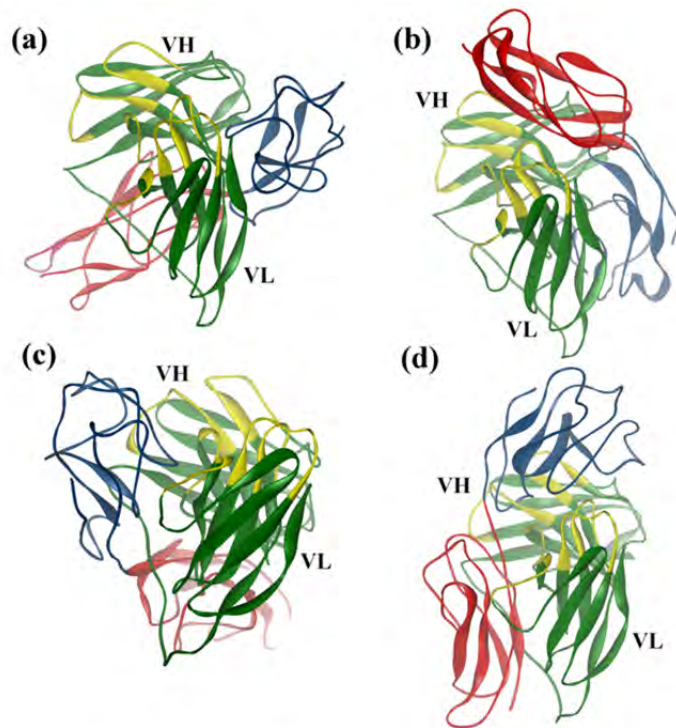
homology modeling. (A) the sequence alignment (B) structural overlap between CD147-HM and the templates (C) the secondary structure of CD147-HM.



**Fig. 4** The validation of CD147 model. (a) Ramachandran plot (b) PROSA z-score (c) PROSA energy (d) VERIFY3D.

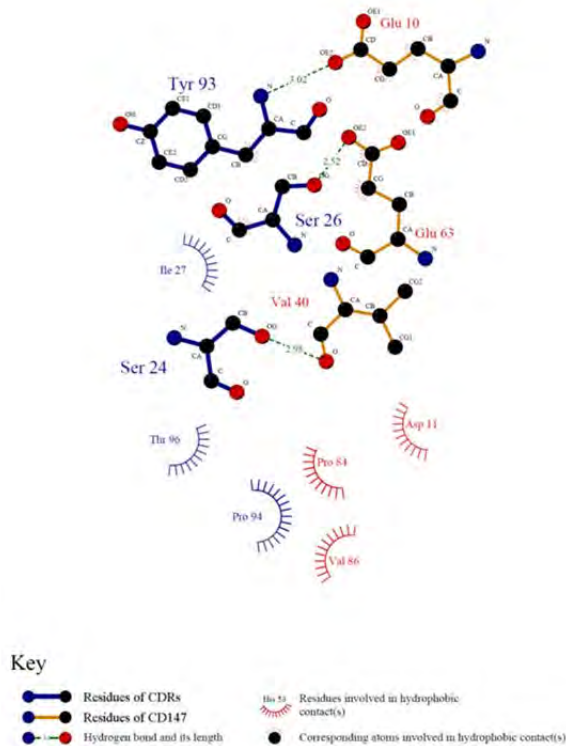
**Fig. 4** The validation of CD147 model. (a) Ramachandran plot (b) PROSA z-score (c) PROSA energy (d) VERIFY3D.





**Fig. 5** Four best molecular docking models in cluster I (a), II (b), III (c) and IV (d). ScFv are shown in green and CDRs are shown in yellow. Domain I and II of CD147 antigen are shown in red and blue, respectively.

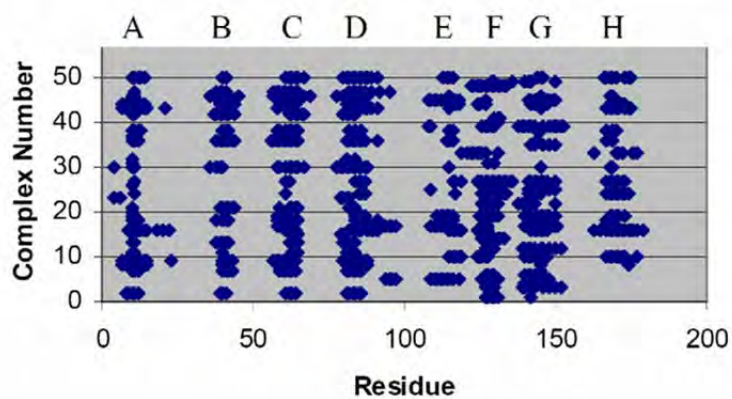
**Fig. 5** Four best molecular docking models in cluster I (a), II (b), III (c) and IV (d). ScFv are shown in green and CDRs are shown in yellow. Domain I and II of CD147 antigen are shown in red and blue, respectively.



**Fig. 6** Schematic representation of interactions between CD147 residues and CDRs residues of scFv-M6-1B9 for the docking model in cluster II. Hydrogen bond distance and hydrophobic interaction are also indicated.

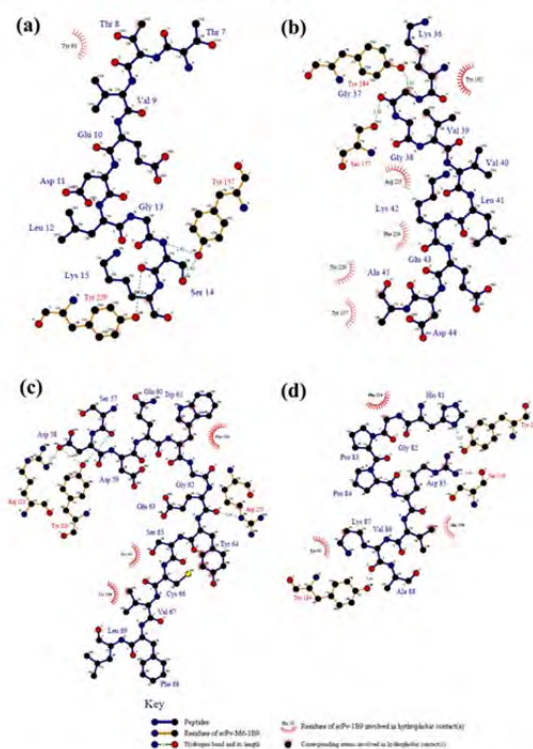


**Fig. 6** Schematic representation of interactions between CD147 residues and CDRs residues of scFv-M6-1B9 for the docking model in cluster II. Hydrogen bond distance and hydrophobic interaction are also indicated.



**Fig. 7** The important interactions of amino acid residues of CD147 from the lowest interaction energy of the 50 complex models associated in binding in 5Å region from CDRs of scFv-M6-1B9.

**Fig. 7** The important interactions of amino acid residues of CD147 from the lowest interaction energy of the 50 complex models associated in binding in 5Å region from CDRs of scFv-M6-1B9.



**Fig. 8** Schematic representation of interactions between peptide ligands Group A-D and scFv-M6-1B9. Hydrogen bond distance and hydrophobic interaction are also indicated.

**Fig. 8** Schematic representation of interactions between peptide ligands Group A-D and scFv-M6-1B9. Hydrogen bond distance and hydrophobic interaction are also indicated.