



รายงานวิจัยฉบับสมบูรณ์

การออกแบบโปรตีน zinc finger โดยวิธีทางคอมพิวเตอร์
สำหรับการจดจำลำดับเบสบนสายดีเอ็นเอ
(Computational Design of Zinc-finger Proteins
for Recognition of DNA Sequences)

อ.ดร.สุวัฒน์ จิรานุสรณ์กุล และคณะ

มีนาคม 2554

รายงานวิจัยฉบับสมบูรณ์

การออกแบบโปรตีน zinc finger โดยวิธีทางคอมพิวเตอร์
สำหรับการจดจำลำดับเบสบนสายดีเอ็นเอ
(Computational Design of Zinc-finger Proteins
for Recognition of DNA Sequences)

คณะผู้วิจัย

สังกัด

1. อ.ดร.สุวัฒน์ จิราอนุสรณ์กุล

ภาควิชาวิทยาศาสตร์เภสัชกรรม

คณะเภสัชศาสตร์ มหาวิทยาลัยเชียงใหม่

2. รศ.ดร.ชัชชัย ตะยาภิวัฒนา

ภาควิชาภูมิคุ้มกันวิทยาคลินิก

คณะเทคนิคการแพทย์ มหาวิทยาลัยเชียงใหม่

สนับสนุนโดยสำนักงานคณะกรรมการการอุดมศึกษา
และสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกอ. และ สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

กิตติกรรมประกาศ

โครงการวิจัยครั้งนี้ได้รับทุนสนับสนุนจากสำนักงานคณะกรรมการการอุดมศึกษา (สกอ.) และสำนักงานกองทุนสนับสนุนการวิจัย(สกว.) ที่เล็งเห็นความสำคัญของการศึกษาวิจัยในเชิงวิทยาศาสตร์พื้นฐาน โดยอาศัยเทคนิคทางเคมีคอมพิวเตอร์ในการจำลองแบบโครงสร้างโมเลกุลเพื่อทำนายความสามารถในการจดจำลำดับเบสบนสายดีเอ็นเอของโปรตีน zinc finger และเป็นแนวทางในการพัฒนาเป็นยารักษาโรคหรือชุดทดสอบต่อไปในอนาคต

ผู้วิจัยขอขอบคุณสำนักงานกองทุนสนับสนุนการวิจัย ที่จัดสรรงบประมาณเพื่อสนับสนุนโครงการวิจัยครั้งนี้ และขอขอบคุณสำนักงานคณะกรรมการการอุดมศึกษาและคณะเภสัชศาสตร์ มหาวิทยาลัยเชียงใหม่ที่ให้การสนับสนุนในทุกๆ ด้าน เพื่อให้ผู้วิจัยได้ทำการวิจัยครั้งนี้จนสำเร็จลุล่วง

ขอขอบคุณ รศ.ดร. ชัชชัย ตะยาภิวัดนา ภาควิชาภูมิคุ้มกันวิทยาคลินิก คณะเทคนิคการแพทย์ ที่ได้ให้ความกรุณาสละเวลารับเป็นอาจารย์ที่ปรึกษางานวิจัย และให้แรงกระตุ้นผลักดันจนงานวิจัยนี้สำเร็จลุล่วง

ขอขอบคุณ ผศ.ดร.วรรณจันทร์ แสงหิรัญ ลี ภาควิชาเคมี คณะวิทยาศาสตร์ สำหรับคำแนะนำที่เป็นประโยชน์ในการวิจัย และห้องปฏิบัติการออกแบบโมเดลและการจำลองโดยวิธีทางคอมพิวเตอร์ (CSML) ภาควิชาเคมี คณะวิทยาศาสตร์ และศูนย์นาโนเทคโนโลยีแห่งชาติ (นาโนเทค) สำหรับการอนุญาตให้ใช้โปรแกรมสำเร็จรูป Discovery Studio และเครื่องคอมพิวเตอร์คำนวณ

ขอขอบคุณนายศุภชัย ศักดิ์จักรภพ นักศึกษาปริญญาเอก ภาควิชาภูมิคุ้มกันวิทยาคลินิก คณะเทคนิคการแพทย์ ในส่วนข้อมูลที่เกี่ยวข้องในเชิงการทดลอง รวมถึงบุคคลอื่นๆ ที่เกี่ยวข้องและอาจไม่ได้กล่าวไว้ ณ ที่นี้

สุดท้ายนี้ผู้วิจัยขอขอบคุณครอบครัวและบุตรชายที่คอยเป็นกำลังใจให้ และหวังว่าการวิจัยครั้งนี้ จะเป็นประโยชน์สำหรับการวิจัยเพื่อพัฒนาประเทศไทยให้ก้าวหน้าต่อไป และหากมีข้อผิดพลาดประการใดในการทำวิจัยครั้งนี้ ผู้วิจัยขออภัย เพื่อนำมาพัฒนาต่อไป

สุพัฒน์ จิรานุสรณ์กุล

15 มีนาคม 2554

บทคัดย่อ

รหัสโครงการ : MRG5280247

ชื่อโครงการ : การออกแบบโปรตีน zinc finger โดยวิธีทางคอมพิวเตอร์ สำหรับการจดจำลำดับเบสบนสายดีเอ็นเอ

ชื่อนักวิจัย : อ.ดร.สุวัฒน์ จิรานุสรณ์กุล
ภาควิชาวิทยาศาสตร์เภสัชกรรม คณะเภสัชศาสตร์ มหาวิทยาลัยเชียงใหม่

E-mail Address : supatj@pharmacy.cmu.ac.th และ supatjira@hotmail.com

ระยะเวลาโครงการ : 24 เดือน (16 มีนาคม 2552 ถึง 15 มีนาคม 2554)

เทคนิคการจำลองพลวัตของโมเลกุลนำมาใช้เพื่อศึกษาอันตรกิริยาระหว่างโปรตีน zinc finger (ZFP) ที่ออกแบบกับสายดีเอ็นเอที่มีลำดับเบสจำเพาะสำหรับเอนไซม์อินทิเกรสของไวรัสเอชไอวี-1 และสายดีเอ็นเอที่ไม่มีลำดับเบสจำเพาะ โดยอาศัยข้อมูลโครงสร้างผลึกสามมิติจากการหักเหของรังสีเอกซ์ของสารประกอบ ZFP 6 โมดูลที่ออกแบบกับสายดีเอ็นเอ (PDB code 2I13) นำมาสร้างแบบจำลองโมเลกุลของ ZFP ที่เข้าจับกับส่วนลำดับเบส 2-long terminal repeat (2LTR) circle junction และกับลำดับเบสที่ไม่จำเพาะ จากนั้นจึงทำการจำลองพลวัตของแบบจำลอง เพื่อศึกษาข้อมูลรูปแบบการจดจำลำดับเบสของ ZFP และคำนวณหาพลังงานยึดจับสัมพัทธ์ด้วยวิธี molecular mechanics/generalized Born surface area (MM/GBSA) ซึ่งพบว่า ZFP มีพลังงานยึดจับระหว่างลำดับเบสเป้าหมายและที่ไม่ใช่เป้าหมายที่แตกต่างกันอย่างมีนัยสำคัญ หลังจากนั้นจึงทำการศึกษาพลังงานการสลายพันธะยึดจับของกรดอะมิโนแต่ละตำแหน่งบน recognition helix ของแบบจำลองโมเลกุลที่จำเพาะระหว่าง ZFP กับ 2LTR ซึ่งพบว่าหากทำการปรับเปลี่ยนกรดอะมิโน E162 ที่อยู่ในตำแหน่ง +6 ของ recognition helix ไปเป็นกรดอะมิโนที่มีประจุบวกมีแนวโน้มที่จะเพิ่มความแข็งแรงในการเข้าจับและอาจเพิ่มความจำเพาะในการเข้าจับของ ZFP ต่อเป้าหมาย 2LTR ได้อีกด้วย ดังนั้นผลการศึกษานี้จะเป็นประโยชน์ในการปรับปรุงการออกแบบโปรตีน zinc finger ให้มีความสามารถในการเข้าจับกับลำดับเบสเป้าหมายได้ดียิ่งขึ้น

คำหลัก : เอชไอวี-1; ตัวยับยั้งอินทิเกรส; โปรตีน zinc finger; 2-LTR-circle junctions; การจำลองพลวัตของโมเลกุล

Abstract

Project Code : MRG5280247

Project Title : Computational design of zinc-finger proteins for recognition of DNA sequences

Investigator : Supat Jiranusornkul, Ph.D.
Department of Pharmaceutical Sciences,
Faculty of Pharmacy, Chiang Mai University

E-mail Address : supatj@pharmacy.cmu.ac.th และ supatjira@hotmail.com

Project Period : 24 months (16 March 2009 to 15 March 2011)

The interactions of a designed zinc-finger protein (ZFP) with DNA containing the specific recognition sequence of the HIV-1 integrase and with a non-specific DNA were studied with molecular dynamics simulation techniques. Molecular models of the designed ZFP bound to the 2-long terminal repeat (2LTR) circle junction and to a non-specific sequence were constructed based on an X-ray crystallographic structure of a designed six finger motif-DNA complex (PDB code 2I13). Molecular dynamics simulations were employed to gain structural insights into DNA recognition patterns of the designed ZFP. The molecular mechanics/generalized Born surface area (MM/GBSA) approach was then used to calculate relative binding interactions. The results showed that the binding enthalpies of the specific and non-specific DNA binding to the designed ZFP differed significantly. The pairwise decomposition method was further applied to calculate the interactions between each recognition residue and the 2LTR target. Mutations of E162 at the position +6 of the recognition helix to positively charged amino acid residues are likely to enhance the binding affinity of the designed ZFP to the 2LTR site. This finding may have significant implications for zinc finger design toward improvement of its binding affinity

Keywords : HIV-1; integrase inhibitor; zinc finger protein; 2-LTR-circle junctions; molecular dynamics simulation

สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อ	ข
Abstract	ค
1. บทนำ	
1.1 โปรตีน zinc finger (ZF)	1
1.2 การจดจำลำดับเบสบนสายดีเอ็นเอของโปรตีน ZF	2
1.3 การออกแบบโปรตีน ZF ชนิด Cys ₂ His ₂	5
1.4 การสร้างแบบจำลองโมเลกุล และการจำลองพลวัต	7
1.5 Long Terminal Repeats (LTRs)	10
1.6 วัตถุประสงค์	12
2. วิธีการทดลอง	
2.1 การออกแบบโปรตีน ZF ที่จำเพาะต่อ 2LTR ของสายดีเอ็นเอไวรัส HIV-1	13
2.2 การสร้างแบบจำลองโปรตีน ZF ที่เข้าจับกับสายดีเอ็นเอ	15
2.3 การจัดเตรียมระบบและการจำลองพลวัต	15
2.4 การวิเคราะห์โครงสร้างหลังจากการจำลองพลวัต	16
2.5 การวิเคราะห์เชิงพลังงานของโมเลกุล	17
2.6 ความจำเพาะของโปรตีน ZFPv1	18
2.7 ความสามารถในการเข้าจับของโปรตีน ZFPv1	19
3. ผลการทดลอง	
3.1 ความจำเพาะเจาะจงของโปรตีน ZFPv1	21

3.2 ความสามารถในการเข้าจับของโปรตีน ZFPv1	33
4. วิจัยรณัผลการทดลอง	42
5. สรุปลผลการทดลอง	47
เอกสารอ้างอิง	48

ภาคผนวก

1. Poster: "Molecular Modeling of Zinc Finger Protein for Recognition of HIV-1 DNA Sequences"
2. Publication: "Designed Zinc Finger Protein Interacting with the HIV-1 Integrase Recognition Sequence at 2-LTR-Circle Junctions" in Protein Science, 2009.

สารบัญรูปภาพ

1.1	โครงสร้างการจัดเรียงตัวพื้นฐานของโปรตีน ZF ชนิด Cys ₂ His ₂ 1 โมดูล	1
1.2	โครงสร้างผลึกของโปรตีน Zif268 และดีเอ็นเอเป้าหมาย	2
1.3	รูปแบบการจดจำลำดับเบสระหว่างสายโซ่ข้างของกรดอะมิโนกับลำดับเบสบนสายดีเอ็นเอ	4
1.4	ภาพเปรียบเทียบโครงสร้างดีเอ็นเอ	8
1.5	ส่วน Long Terminal Repeats (LTRs) ที่ประกอบด้วยส่วน U3 R และ U5 บริเวณปลายทั้งสองของ cDNA	10
1.6	รูปแบบของ cDNA ที่พบได้หลังจากการติดเชื้อ retrovirus และการเชื่อมต่อกับยีนเจ้าบ้าน	11
2.1	ลำดับเบสในส่วน 2LTR ที่เชื่อมกันของ circular HIV-1 DNA	13
2.2	การใช้งานโปรแกรม ZF Tools	14
2.3	การเปรียบเทียบลำดับเบสและกรดอะมิโนระหว่างโครงสร้างผลึกต้นแบบ PDB id 2I13 และแบบจำลองโมเลกุลที่ศึกษา	19
3.1	โครงสร้างโมเลกุลเปรียบเทียบระหว่างชุดแบบจำลอง Zif1 และ Zif2	24
3.2	เปรียบเทียบพันธะไฮโดรเจน (< 3.5 Å) ที่อาจเป็นไปได้ในแต่ละโมดูลของ โปรตีน ZFPv1 ในแบบจำลอง Zif1 และ Zif2	27 – 28
3.3	ค่า RMSD ของโครงสร้างต่างๆ 2 ps จากช่วงเวลา 5 ns เทียบกับโครงสร้างเฉลี่ย ที่ได้จากการจำลองพลวัตแต่ละชุดแบบจำลอง ZFPv1-GAT ZFPv1-GAA ZFPv2-GAT และ ZFPv2-GAA	30
3.4	พันธะไฮโดรเจนที่พบได้ระหว่าง Arg50-DG182 Asn47-DA183 และ Gln44-DA183/DT183 จากการจำลองพลวัตของ ZFPv2-GAA และ ZFPv2-GAT	32
3.5	การเกิดพันธะไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง -1 +3 และ +6	

ของ recognition helix ทั้ง 6 โมดูลในแบบจำลอง ZFPv1-GAT	34 – 35
3.6 พลังงานการสลายพันธะยึดจับ (decompose energy) ระหว่างกรดอะมิโนในตำแหน่ง -2 ถึง +6 ของ recognition helix ทั้ง 6 โมดูลของโปรตีน ZFPv1 กับลำดับเบสเป้าหมาย 2LTR	36
3.7 ค่า RMSD ของโครงสร้างทุกๆ 2 ps จากช่วงเวลา 5 ns เทียบกับโครงสร้างเฉลี่ย ที่ได้จากการจำลองพลวัตของชุดแบบจำลอง E162D E162H E162K E162R และ E162Q	38
3.8 การเกิดพันธะไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง +6 (Lys162 และ Asp162) ของ recognition helix ในโมดูล 6 ของแบบจำลอง E162K และ E162D	39
3.9 พลังงานการสลายพันธะยึดจับของกรดอะมิโนในตำแหน่ง -2 ถึง +6 ของโมดูล 6 ในแบบจำลอง ZFPv1-GAT E162D E162H E162K E162R และ E162Q	41

สารบัญตาราง

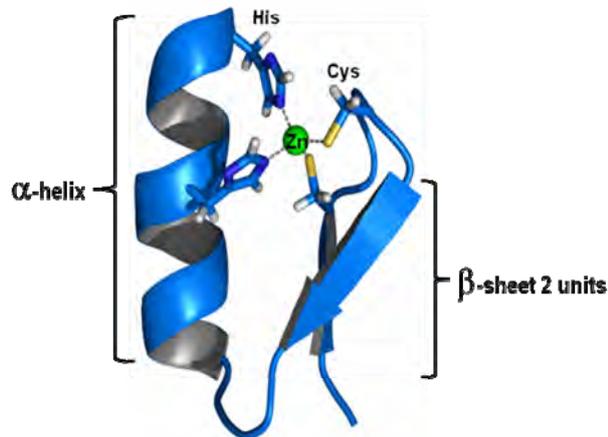
3.1	ผลวิเคราะห์ชนิดของกรดอะมิโนในตำแหน่ง -1 ถึง +6 บน recognition helix จากทั้ง 6 โมดูลที่ทำนายได้จากฐานข้อมูล ZF Tools และ ZIFDB	22
3.2	ระยะห่าง (Å) ระหว่างอะตอมของ H-bond donor และ H-bond acceptor ในโปรตีน ZFPv1 และสายดีเอ็นเอในชุดแบบจำลอง Zif1 และ Zif2	26
3.3	การเปรียบเทียบความแตกต่างของพลังงานในการยึดจับระหว่างโปรตีน ZFPv1 กับสายดีเอ็นเอในแบบจำลอง Zif1 และ Zif2	29
3.4	เปรียบเทียบค่าพลังงานการยึดจับของโมเลกุลระหว่างโปรตีนและสายดีเอ็นเอ ด้วยวิธี MM/GBSA และ MM/PBSA	31
3.5	กรดอะมิโนบน recognition helix กับเบสในตำแหน่งที่เกิดอันตรกิริยาที่ไม่เหมาะสม ในแบบจำลอง ZFPv1-GAT	37
3.6	พลังงานการยึดจับสัมพัทธ์ของแบบจำลองที่มีการปรับเปลี่ยนชนิดกรดอะมิโน เทียบกับแบบจำลองต้นแบบ ZFPv1-GAT โดยวิธี MM/GBSA	41

1. บทนำ

1.1 โปรตีน zinc finger (ZF)

โปรตีน zinc finger (ZF) จัดเป็นลักษณะพิเศษหนึ่งของการม้วนพับของโปรตีน ได้เป็นโครงสร้างตติยภูมิ (tertiary structure) ที่เกิดแรงยึดเหนี่ยวระหว่างกรดอะมิโน เช่น ซิสเทอีน (cysteine) และฮิสทีดีน (histidine) กับโลหะสังกะสีอย่างน้อย 1 อะตอม เกิดเป็นโครงสร้างที่มีลักษณะคล้ายนิ้วมือ และมีความเสถียร สามารถจัดแบ่งได้เป็น 8 ประเภทตามรูปร่างการจัดเรียงตัวของโปรตีน zinc finger^[1] ซึ่งโปรตีน ZF ชนิด Cys₂His₂ เป็นชนิดที่พบมากที่สุด พบได้มากถึง 2% ในจีโนมมนุษย์^[2] โดยเฉพาะในส่วนโปรตีน transcription factor ต่างๆ และพบว่าส่วนโปรตีน ZF นี้เป็นการจัดเรียงตัวของโปรตีนที่มีความสำคัญในการเข้าจับกับสายดีเอ็นเอ หรืออาร์เอ็นเอในกระบวนการเปลี่ยนแปลงของพันธุกรรม (gene manipulation)

โปรตีน ZF ชนิด Cys₂His₂ มีกรดอะมิโนซิสเทอีนและฮิสทีดีนอย่างละ 2 โมเลกุลในการเกิดพันธะโคออดิเนต (coordinate bond) กับโลหะสังกะสี 1 อะตอม (รูป 1.1) โดยมีลำดับกรดอะมิโนพื้นฐานความยาวประมาณ 30 กรดอะมิโนต่อโปรตีน ZF 1 โมดูล (หรือ finger) จัดเรียงในรูปแบบทั่วไปเป็น (Phe/Tyr)-X-Cys-X₂₋₅-Cys-X₃-(Phe/Tyr)-X₅-Ψ-X₂-His-X₃₋₅-His^[3] โดย X คือ กรดอะมิโนใดๆ และ Ψ คือ กรดอะมิโนที่มีสายโซ่ข้าง (sidechain) เป็นหมู่ที่มีสมบัติ hydrophobicity เช่น วาลีน (valine) ลิวซีน (leucine) ฟีนิลอะลานีน (phenylalanine) และไทโรซีน (tyrosine) เป็นต้น โครงสร้างของโปรตีน ZF นี้จะมีการม้วนพับเป็นลักษณะ ββα คือ ประกอบด้วย β-sheet 2 ชุด และ α-helix 1 ชุด รอบๆ โลหะสังกะสี ซึ่งทำให้โครงสร้างนี้เสถียรและมีความสามารถในการเข้าจับหรือจดจำลำดับเบสบนสายดีเอ็นเอได้เป็นอย่างดี

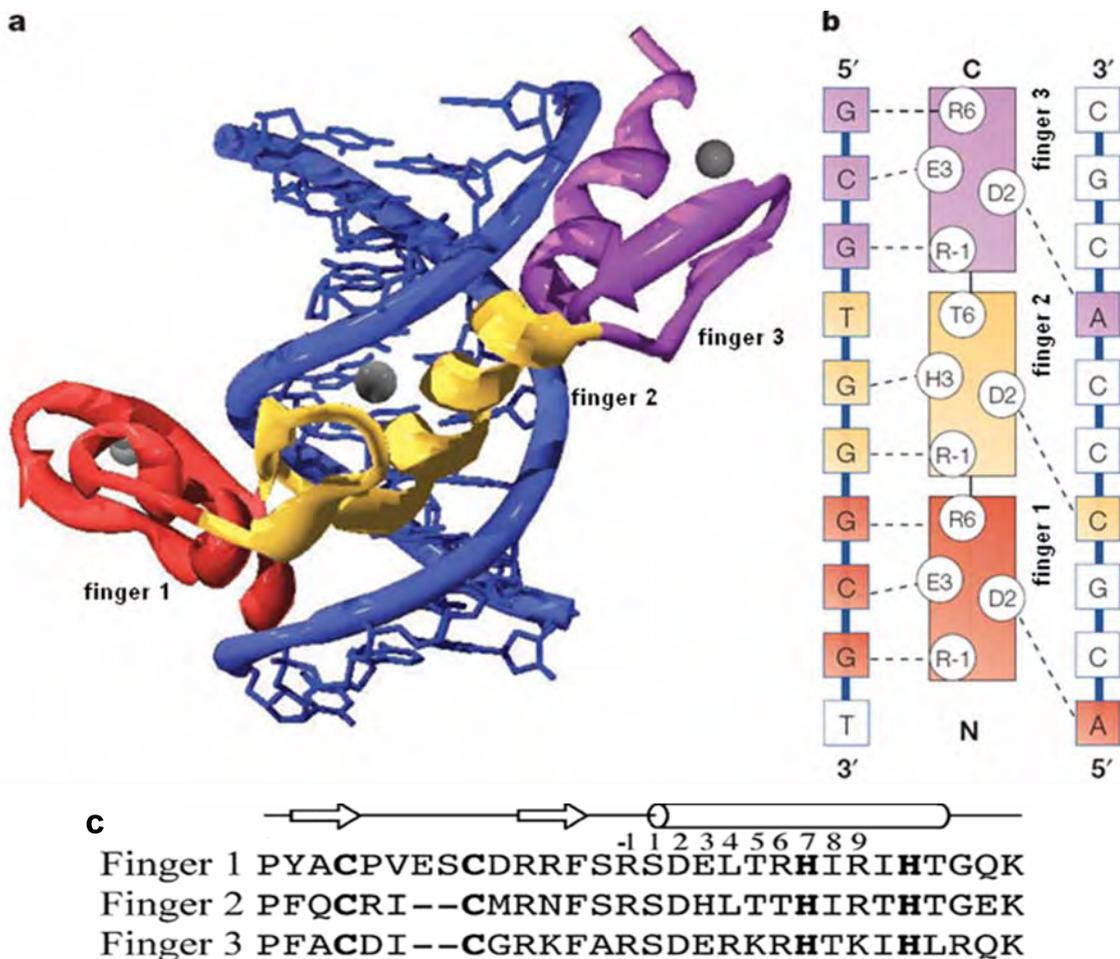


รูป 1.1 โครงสร้างการจัดเรียงตัวพื้นฐานของโปรตีน ZF ชนิด Cys₂His₂ 1 โมดูล

(ดัดแปลงจาก http://en.wikipedia.org/wiki/File:Zinc_finger_rendered.png โดย Thomas Splettstoesser)

1.2 การจดจำลำดับเบสบนสายดีเอ็นเอของโปรตีน ZF

การศึกษาความสามารถจดจำลำดับเบสบนสายดีเอ็นเอของโปรตีน ZF (DNA recognition of zinc-finger proteins) มีการพัฒนาไปอย่างมาก หลังจากการค้นพบโครงสร้างผลึกสามมิติของ transcription factor โปรตีน Zif268 ที่เข้าจับกับสายดีเอ็นเอเป้าหมายในปี 1991^[4] ซึ่งโปรตีน Zif268 จัดเป็นโปรตีน ZF ชนิด Cys₂His₂ ที่ประกอบด้วยโปรตีน ZF จำนวน 3 โมดูลที่ทำงานร่วมกันในการจดจำลำดับเบสเป้าหมาย 9 คู่เบส ได้แก่ 5'-GCG[G/T]GGCG-3'^[5, 6] ทำให้ทราบถึงการจัดเรียงตัวของโปรตีน ZF ที่มีลักษณะเป็นนิ้วมือนี้ จำเป็นต้องมีโลหะสังกะสีเป็นองค์ประกอบเสมอและโครงสร้างนี้มีขนาดและรูปร่างที่ความเหมาะสมในการเข้าจับกับบริเวณ major groove บนสายดีเอ็นเอ และที่สำคัญคือ ส่วน α -helix ในตำแหน่ง -1 ถึง +9 ในรูป 1.2(c) ของแต่ละโมดูลของโปรตีน ZF มีส่วนสำคัญที่เกี่ยวข้องกับการจดจำลำดับเบสจำนวน 3 คู่เบส (triplet) อย่างจำเพาะเจาะจง (รูป 1.2)



รูป 1.2 โครงสร้างผลึกของโปรตีน Zif268 และดีเอ็นเอเป้าหมาย (a) แบบจำลองโมเลกุลสามมิติ (b) อันตรกิริยาระหว่างกันและ (c) ลำดับกรดอะมิโนของโปรตีน ZF ทั้ง 3 โมดูลของ Zif268^[3, 7]

ข้อมูลโครงสร้างสามมิติร่วมกับการวิเคราะห์โครงสร้างการเข้าจับระหว่างโปรตีน Zif268 กับ สายดีเอ็นเอเป้าหมายในรายละเอียด ดังแสดงในรูป 1.2(a) และ (b) สามารถช่วยทำความเข้าใจและ อธิบายธรรมชาติในการทำงานของโปรตีน ZF ได้อย่างดี พบว่าโปรตีน Zif268 จำนวน 3 โมดูล ซึ่งเข้า จับกับสายดีเอ็นเอเป้าหมายที่บริเวณ major groove โดยอาศัยโครงสร้างเคมีของสายโซ่ข้างของกรดอะ มิโนในตำแหน่งที่จำเพาะของส่วน α -helix (sequence-specific) ได้แก่ ตำแหน่ง -1 +3 และ +6 ในการ เกิดอันตรกิริยากับเบสเป้าหมาย 3 คู่เบสจากตำแหน่ง 3' ไปยัง 5' ตามลำดับ อาทิ

- ในตำแหน่ง -1 และ +6 หมู่กัวนิตินที่มีประจุบวกของกรดอะมิโน Arg จะเกิดอันตรกิริยาได้ดีกับ หมู่คาร์บอนิลที่มีประจุลบของเบสกวีนีน แต่ไม่พบอันตรกิริยาระหว่างกรดอะมิโน Thr ใน ตำแหน่งที่ +6 ของโมดูลที่ 2 กับเบสไทมีน
- ในตำแหน่ง +3 กรดอะมิโน His ของโมดูลที่ 2 สามารถเกิดอันตรกิริยาได้ดีกับเบสกวีนีน หรือ หมู่กรดคาร์บอกซิลิกที่เป็นประจุลบของกรดอะมิโน Glu ของโมดูลที่ 1 และ 3 สามารถเกิด อันตรกิริยาได้ดีกับหมู่อะมิโนของเบสไซโตซีน
- ในตำแหน่ง +2 ส่วนกรดอะมิโน Asp อาจมีผลเกิดอันตรกิริยากับเบสบนสายดีเอ็นเอตรงข้ามได้

อันตรกิริยาดังกล่าวข้างต้นทำให้เกิดการเข้าจับระหว่างโปรตีน ZF กับสายดีเอ็นเอได้อย่าง จำเพาะและแข็งแรง ซึ่งลำดับกรดอะมิโนในแต่ละโมดูลนั้นมีลักษณะที่สงวนไว้อย่างดี (well-conserved) โดยมีรูปแบบทั่วไปเป็น (Phe/Tyr)-X-Cys-X₂₋₅-Cys-X₃-(Phe/Tyr)-X₅- ψ -X₂-His-X₃₋₅-His และมีความ แตกต่างกันเพียงส่วน α -helix ที่สัมพันธ์กับการเข้าจับกับลำดับเบสที่จำเพาะ ซึ่งพบว่าในกรณีของ โปรตีน ZF อื่นๆ ที่มีลำดับกรดอะมิโนที่ต่างจาก Zif268 อย่างชัดเจน (non-standard ZF) จะอาศัย กรดอะมิโนในตำแหน่ง +3 และ +6 เป็นสำคัญเช่นเดียวกับโปรตีน Zif268 แต่ในตำแหน่ง -1 และ +2 มี รูปแบบการจดจำที่ต่างออกไป^[3]

จากการวิเคราะห์ข้อมูลการเข้าจับหรือจดจำลำดับเบสของโปรตีน Zif268 สามารถนำไป เปรียบเทียบกับโปรตีน ZF ชนิดอื่นๆ ทำให้สามารถประเมินรูปแบบการจดจำที่เป็นมาตรฐานสำหรับ โปรตีน ZF ได้ ดังแสดงในรูป 1.3 ซึ่งระบุชนิดของกรดอะมิโนในตำแหน่ง -1 +2 +3 และ +6 ในส่วน α - helix ที่สามารถเกิดอันตรกิริยาได้ดีกับลำดับเบสทั้ง 4 ชนิดในตำแหน่งต่างๆ กัน บนสายดีเอ็นเอจาก 5' ไปยัง 3' และเบสบนสายดีเอ็นเอตรงข้าม โดยกรดอะมิโนที่เป็นตัวอักษรเข้ม หมายถึง ข้อมูลกรดอะมิโน ที่พบการเกิดอันตรกิริยากับเบสชนิดนั้นได้บ่อยครั้งจากการทดลอง phage display ส่วนเครื่องหมาย

ดอกจัน หมายถึง กรดอะมิโนที่พบการเกิดอันตรกิริยากับเบสชนิดนั้นที่ได้จากการทดลองทางการสร้างแบบจำลอง และเครื่องหมายคำถาม หมายถึง การเกิดอันตรกิริยาระหว่างกรดอะมิโนกับเบสนั้นยังไม่ชัดเจน

อย่างไรก็ตาม จากรูป 1.3 จะพบว่ายังมีความไม่ชัดเจนในรูปแบบการจดจำการเข้าจับระหว่างโปรตีน ZF กับสายดีเอ็นเอในหลายๆ ตำแหน่ง เช่น ตำแหน่ง +6 มีเพียงอันตรกิริยาระหว่างกรดอะมิโน Arg และเบสกวีนีนเท่านั้นที่มีความชัดเจน นอกจากนั้นแล้วยังไม่พบรูปแบบการเข้าจับระหว่างกันที่จำเพาะ และสำหรับเบสไทมีนไม่ว่าจะอยู่ตำแหน่งใดบนสายดีเอ็นเอก็ยังไม่พบกรดอะมิโนที่จำเพาะในการเข้าจับจากการศึกษาทั้งในหลอดทดลองและในเชิงทฤษฎี เป็นต้น ดังนั้น จะเห็นว่ารูปแบบการจดจำลำดับเบสของโปรตีน ZF ยังไม่สามารถสรุปได้ชัดเจน ซึ่งปัจจัยเกี่ยวข้องที่ควรพิจารณาสำหรับการศึกษาการจดจำลำดับเบสของโปรตีน ZF ได้แก่

- การเกิดอันตรกิริยาระหว่างสายโซ่ข้างของกรดอะมิโนกับเบส
- ลำดับเบสที่แตกต่างกันของสายดีเอ็นเอมีผลต่อการจัดเรียงตัวและความยืดหยุ่น (flexibility) ของสายดีเอ็นเออื่นๆ
- บทบาทของโมเลกุลน้ำในการเกิดอันตรกิริยาระหว่างกรดอะมิโนและเบส เช่น การเกิดอันตรกิริยาระหว่างสายโซ่ข้างของกรดอะมิโนกับเบสบนสายดีเอ็นเอตรงข้ามที่อาจต้องอาศัยโมเลกุลน้ำเป็นตัวเชื่อมต่อ เป็นต้น
- การเกิดอันตรกิริยาระหว่างสายโซ่ข้างที่มีประจุบวกของกรดอะมิโนกับหมู่ฟอสเฟตที่มีประจุลบบนแกนหลักของสายดีเอ็นเอ ซึ่งทำให้โปรตีน ZF สามารถเข้าจับกับสายดีเอ็นเอได้อย่างแข็งแรง และดึงให้โครงสร้างแต่ละโมดูลของโปรตีน ZF จับกับ major groove ของดีเอ็นเอได้อย่างพอเหมาะ
- ผลของโมดูลข้างเคียงที่มีผลให้การเข้าจับของแต่ละโมดูลมีความเหมาะสมต่อลำดับเบสนั้นๆ

Base Specificity

	T	C	A	G		
6	Lys?		Gln?	Arg* Lys*	5'	3'
3	Ser Ala	Asp* Thr Glu Ser*	Asn* His*	His* Lys*	5'	3'
-1	Thr Leu* His	Asp* His	Gln*	Arg*	5'	3'
2		Asp*	Asp*		3'	5'

↑

รูป 1.3 รูปแบบการจดจำลำดับเบสระหว่างสายโซ่ข้างของกรดอะมิโนกับลำดับเบสบนสายดีเอ็นเอ^[3]

1.3 การออกแบบโปรตีน ZF ชนิด Cys₂His₂

การออกแบบโปรตีน ZF (ZF protein design) เป็นเทคนิคหนึ่งที่ได้รับคามนิยมมากสำหรับการวิจัยเพื่อควบคุมการแสดงออกหรือการปรับเปลี่ยนการทำงานของยีนใดๆ ที่ต้องการ เพื่อใช้ประโยชน์ในการศึกษาทำความเข้าใจการเกิดโรคและพัฒนาเป็นยารักษาโรค โดยความพยายามใช้เทคนิคทางพันธุวิศวกรรมในการสร้างโปรตีน ZF เลียนแบบธรรมชาติ (artificial ZF protein) ที่มีความสามารถในการเข้าจับกับลำดับเบสบนสายดีเอ็นเอตามต้องการ และส่วนใหญ่จะอาศัยโครงสร้างพื้นฐานที่ได้จากโปรตีน ZF ชนิด Cys₂His₂ ซึ่งมีข้อมูลศึกษาค่อนข้างชัดเจนเกี่ยวกับโครงสร้างพื้นฐานและรูปแบบในการจดจำลำดับเบส

จากข้อมูลการศึกษาวิเคราะห์โครงสร้างของโปรตีน ZF พบว่าโครงสร้างหลักของโปรตีน ZF ประกอบด้วย 4 องค์ประกอบหลัก ได้แก่ 1) N-terminal backbone จำนวน 9 กรดอะมิโน 2) recognition helix จำนวน 9 กรดอะมิโน 3) C-terminal backbone จำนวน 5 กรดอะมิโน และ 4) linker จำนวน 5 กรดอะมิโน จากข้อมูลนี้นำไปใช้ในการออกแบบและสังเคราะห์โปรตีน ZF ในแต่ละโมดูลซึ่งมีโครงหลัก (scaffold) เช่นเดียวกัน แต่มีการเปลี่ยนแปลงลำดับกรดอะมิโนเฉพาะในส่วน recognition helix ของ α -helix ให้มีความจำเพาะกับลำดับเบสบนสายดีเอ็นเอเป้าหมายได้

หลักการทั่วไปในการออกแบบโปรตีน ZF เพื่อให้มีความจำเพาะต่อลำดับเบส สามารถแบ่งได้เป็น 2 หลักการ ได้แก่ 1) Modular Assembly และ 2) Oligomerized Pool Engineering (OPEN) วิธี Modular Assembly มีข้อดีคือ อาศัยข้อมูลที่ได้จากการทดลอง phage display และ site-directed mutation ในการเข้าจับของโปรตีน ZF จำนวน 3 โมดูลกับดีเอ็นเอเป้าหมายโดยพิจารณาความสามารถในการเข้าจับของโมดูลที่ 2 กับลำดับเบสเป้าหมายจนสามารถสรุปได้ถึงลำดับกรดอะมิโนที่เหมาะสมในส่วน recognition helix กับลำดับเบส 5'-GNN-3' และส่วนใหญ่ของ 5'-ANN-3' 5'-CNN-3' และส่วนน้อยของ 5'-TNN-3' เมื่อ N คือนิวคลีโอไทด์ใดๆ^[8-12] ร่วมกับสมมติฐานที่ว่าโปรตีน ZF 1 โมดูล มีความสามารถในการจดจำลำดับเบสได้ 3 คู่เบส โดยไม่ได้คำนึงถึงตำแหน่งของคู่เบสนั้นบนสายดีเอ็นเอหรือตำแหน่งของโมดูลในโปรตีน ZF (context-independent) ดังนั้นวิธี Modular Assembly จึงเป็นวิธีที่ง่าย สามารถออกแบบได้โดยค้นหาลำดับเบสที่เหมาะสมสำหรับใช้เป็นเป้าหมายเสียก่อน และให้โปรแกรมทำนายลำดับกรดอะมิโนที่เหมาะสมสำหรับแต่ละ recognition helix ในการเข้าจับกับแต่ละ 3 คู่เบส แล้วนำมาเชื่อมต่อกันด้วยโครงร่างหลักในส่วน N-terminal ส่วน C-terminal และส่วน linker

สำหรับแต่ละโมดูล แต่อย่างไรก็ตามวิธี Modular Assembly ยังมีข้อจำกัดถึงการทำนาย recognition helix สำหรับลำดับเบสเป้าหมายที่เป็น 5'-ANN-3' 5'-CNN-3' โดยเฉพาะ 5'-TNN-3' ซึ่งยังไม่มีอยู่ในฐานข้อมูล และนอกจากนั้นยังพบรายงานความล้มเหลวของโปรตีน ZF ที่ออกแบบด้วยวิธี Modular Assembly นี้ในการเข้าจับกับสายดีเอ็นเอเป้าหมายจากการทดลองในเซลล์^[13] ส่วนวิธี OPEN จะนิยมใช้ในการออกแบบโปรตีน ZF จำนวน 3 โมดูลและเชื่อมต่อกับเอนไซม์ nuclease เพื่อทำการตัดสายดีเอ็นเอ โดยวิธี OPEN จะคำนึงถึงตำแหน่งของโปรตีน ZF แต่ละโมดูลว่าอยู่ในลำดับที่เท่าใด (context-dependent) วิธี OPEN มีหลักการเพื่อเลียนแบบโครงสร้างของ transcription factor ส่วนใหญ่ที่มีโปรตีนที่ทำหน้าที่ตัดสายดีเอ็นเอร่วมอยู่ด้วย และนิยมใช้โปรแกรม ZIFit (Zinc Finger Targeter)^[14] ในการออกแบบ

การออกแบบและสังเคราะห์โปรตีน ZF เลียนแบบธรรมชาตินี้ มีการศึกษาวิจัยหลากหลาย ตัวอย่างเช่น การออกแบบโปรตีน ZF เพื่อให้จับกับลำดับเบส 18 คู่เบสของ checkpoint kinase 2 (CHK2) gene ซึ่งเป็นยีนที่มีส่วนเกี่ยวข้องกับการควบคุมวงจรชีวิตของเซลล์ การซ่อมแซมดีเอ็นเอ และการตายของเซลล์ พบว่า การใช้โปรตีน ZF นี้ไปจับกับยีนดังกล่าวสามารถลดการสร้าง CHK2 mRNA และโปรตีน CHK2 ลงได้มากกว่า 10 เท่าของการที่ไม่มีโปรตีน ZF^[15] นี้ ซึ่งเป็นทางเลือกหนึ่งในการพัฒนาเพื่อรักษาโรคมะเร็ง หรือการประยุกต์โปรตีน ZF ร่วมกับเอนไซม์ nuclease เพื่อตัดสายดีเอ็นเอให้เกิดเป็น double strand break ที่ยีนเป้าหมาย รวมถึงมีการพัฒนาการทำนายความสามารถในการจดจำลำดับเบสของโปรตีน ZF โดยอาศัยวิธีทางชีวสารสนเทศ^[16]

ข้อควรคำนึงในการออกแบบและประยุกต์โปรตีน ZF ในทางปฏิบัติมี 2 ประเด็นที่ต้องคำนึง คือ ต้องการให้โปรตีนที่ออกแบบนั้นมีความจำเพาะ (specificity) ต่อดีเอ็นเอเป้าหมายเท่านั้นไม่ไปเกิดการจับกับดีเอ็นเอที่ไม่ใช่เป้าหมายในร่างกาย และมีความสามารถในการเข้าจับ (binding affinity) กับดีเอ็นเอเป้าหมายได้อย่างแข็งแรงหรือสามารถเอาชนะโปรตีนหรือเอนไซม์อื่นที่สามารถเข้าจับกับดีเอ็นเอเป้าหมายเดียวกัน ดังนั้นการเลือกลำดับเบสเป้าหมายและจำนวนโมดูลของโปรตีนที่ต้องการสร้างขึ้นมาจึงมีความสำคัญ พบว่าคู่เบสความยาว 9 คู่เบส สามารถพบซ้ำได้ประมาณ 13000 ครั้งในจีโนมมนุษย์ แต่ถ้าหากความยาว 18 คู่เบสจะสามารถพบซ้ำได้เมื่อมีขนาดจีโนมที่ใหญ่กว่าจีโนมมนุษย์ถึง 20 เท่า^[17] ดังนั้นในการเลือกลำดับเบสเป้าหมายจึงนิยมให้มีความยาว 18 คู่เบสเพื่อป้องกันการเกิดการเข้าจับของโปรตีน ZF ในลำดับเบสอื่นที่ไม่ต้องการ นอกจากนั้นแล้วการที่เลือกใช้เบสเป้าหมายความยาว 18 คู่เบสทำให้ต้องมีโปรตีน ZF จำนวน 6 โมดูล ซึ่งพบว่ามีความแข็งแรงในการเข้าจับกับเป้าหมายได้ดีกว่าการใช้โปรตีน ZF จำนวน 3 โมดูลอีกด้วย^[3]

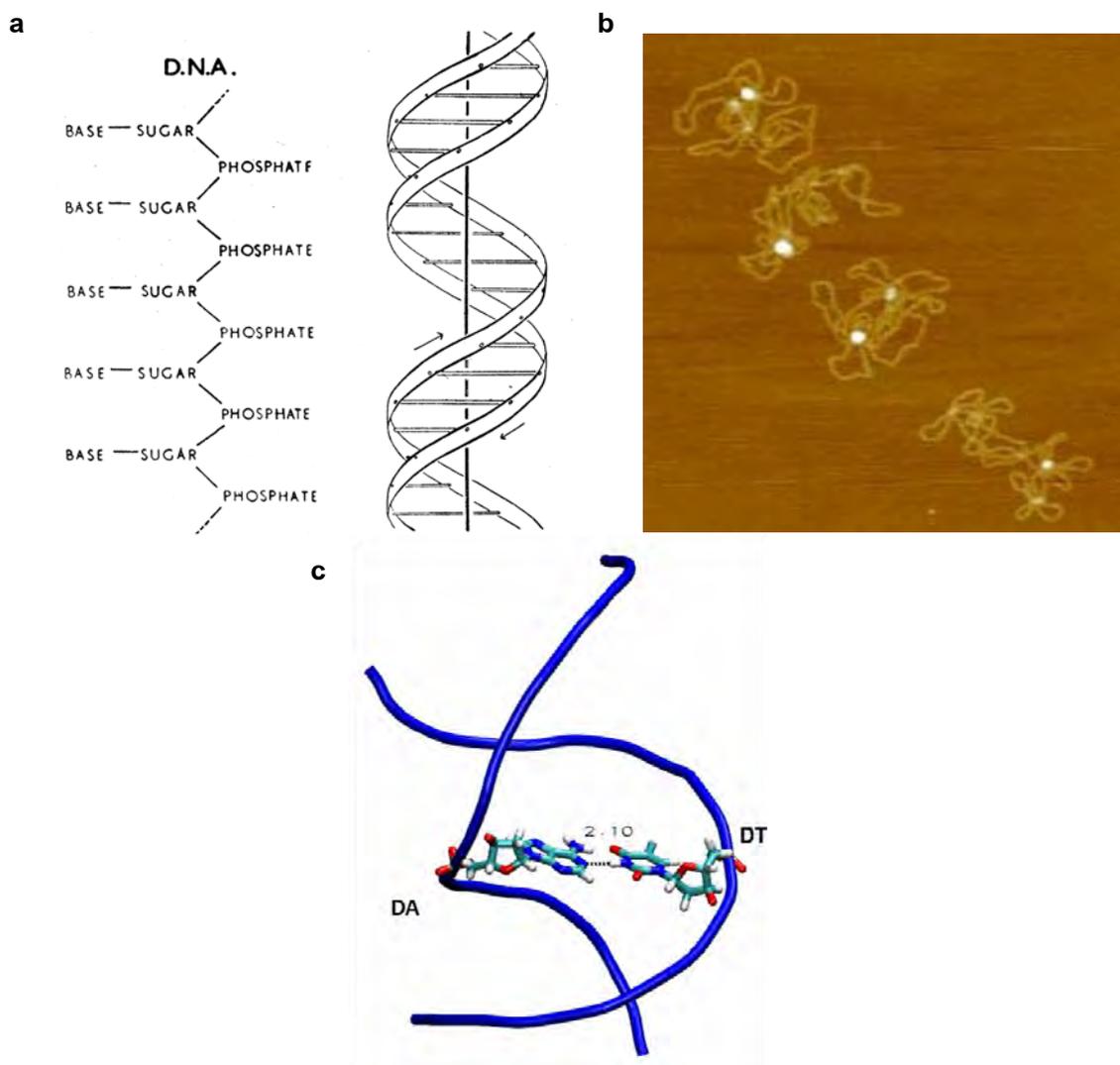
การทำนายลำดับกรดอะมิโนที่เหมาะสมของโปรตีน ZF ที่จำเพาะต่อลำดับเบสของสายดีเอ็นเอหนึ่งๆ สามารถจะกระทำได้โดยง่ายผ่านโปรแกรมสำเร็จรูปต่างๆ แต่อย่างไรก็ตามผลที่ได้เป็นเพียงโครงสร้างปฐมภูมิของโปรตีนเท่านั้น ยังไม่ทราบถึงการจัดเรียงตัวเป็นโครงสร้างตติยภูมิหรือโครงสร้างสามมิติของโปรตีนและดีเอ็นเอในขณะที่เกิดขึ้นสารประกอบและเกิดอันตรกิริยาซึ่งกันและกัน ดังนั้นการศึกษาโดยการจำลองโครงสร้างสามมิติของโปรตีนและดีเอ็นเอทางคอมพิวเตอร์ จึงมีความสำคัญในการตรวจสอบความถูกต้องของโครงสร้างปฐมภูมิที่ได้จากการทำนาย ก่อนลงมือทำการทดลองจริง ดังนั้นการใช้เทคนิคการสร้างแบบจำลองโมเลกุลและการศึกษาพฤติกรรมทางพลศาสตร์ (dynamic behaviour) ของทั้งโปรตีนและดีเอ็นเอในขณะที่ยังจับกันเกิดเป็นสารเชิงซ้อนนั้น เป็นวิธีการหนึ่งที่ทำให้สามารถเข้าใจธรรมชาติของโปรตีน ZF และสายดีเอ็นเอ เพื่อเป็นประโยชน์ในการทำนายและปรับปรุงโปรตีน ZF ที่ออกแบบได้ให้มีประสิทธิภาพสูงขึ้น โดยที่สามารถประหยัดทั้งเวลาและค่าใช้จ่ายมากกว่าการทดลองจริง

1.4 การสร้างแบบจำลองโมเลกุล และการจำลองพลวัต

การสร้างแบบจำลองโมเลกุล (molecular modelling) เป็นวิธีการหนึ่ง ในการใช้เทคนิคทางคอมพิวเตอร์มาประกอบในการอธิบายให้เห็นภาพ อธิบายหรือการประเมินถึงสมบัติต่างๆ ของโครงสร้างโมเลกุลได้ในทางทฤษฎี^[18] ซึ่งเป็นเทคนิคที่สามารถประยุกต์กับสารเคมีที่มีโมเลกุลขนาดเล็ก สารโพลิเมอร์ หรือแม้กระทั่งสารชีวโมเลกุล หลักการพื้นฐานในการสร้างแบบจำลองโมเลกุล คือ การระบุตำแหน่งของอะตอมแต่ละอะตอมในโมเลกุลในลักษณะ 3 มิติ ในแนวแกน X Y และ Z ทำให้สามารถใช้โปรแกรมสำเร็จรูปบางชนิด เช่น ชุดโปรแกรม Discovery Studio จากบริษัท Accelrys และโปรแกรม Visual Molecular Dynamics (VMD)^[19] เป็นต้น ในการทำให้เห็นเป็นภาพสามมิติได้ และยังสามารถจัดการกับโครงสร้างโมเลกุลที่สร้างขึ้นได้อย่างมีประสิทธิภาพ โดยที่ทำได้ยากหรือเป็นไปไม่ได้ในการทำการทดลองจริง ตัวอย่างเช่น การสร้างแบบจำลองโมเลกุลดีเอ็นเอที่มีลักษณะเป็นเกลียวคู่คล้ายบันไดวนที่ถูกนำเสนอโดย James Watson และ Francis Crick ในปี 1953 ในภาพ 2 มิติ หรือการสร้างเป็นแบบจำลองพลาสติก ถูกแปลงให้เป็นภาพ 3 มิติด้วยโปรแกรมคอมพิวเตอร์ ทำให้สามารถระบุถึงสมบัติต่างๆ เช่น ความยาวของพันธะไฮโดรเจนระหว่างเบสอะดีนีน (adenine) กับเบสไทมีน (thymine) ในสายดีเอ็นเอได้ ซึ่งเครื่องมือวิทยาศาสตร์ในปัจจุบันทำให้เห็นสายดีเอ็นเอได้เพียงลักษณะเป็นเหมือนเส้นด้ายด้วยเครื่อง Atomic Force Microscope (AFM) ดังแสดงในรูป 1.4 เป็นต้น นอกจากการทำให้เห็นภาพแล้วการสร้างแบบจำลองยังสามารถนำความรู้ทางเคมีของอะตอม มวลอะตอม ขนาดอะตอม การเกิดแรงหรือพันธะระหว่างอะตอมแวดล้อมซึ่งเราสามารถระบุตำแหน่งของแต่ละอะตอมได้

แล้วนั้น นำมาประเมินสมบัติต่างๆ ของโมเลกุลได้ เช่น พลังงานศักย์ในโมเลกุล เอนทาลปี (enthalpy) และสภาพขั้ว เป็นต้น

เทคนิคการคำนวณในทางการสร้างแบบจำลองโมเลกุล สามารถจำแนกได้เป็น 2 เทคนิค ได้แก่ กลศาสตร์ควอนตัม (quantum mechanics) และกลศาสตร์โมเลกุล (molecular mechanics) ในงานวิจัยนี้จะทำการคำนวณด้วยเทคนิคกลศาสตร์โมเลกุลซึ่งจะไม่คำนึงถึงการเคลื่อนที่ของอิเล็กตรอนรอบอะตอมเช่นในเทคนิคกลศาสตร์ควอนตัม กลศาสตร์โมเลกุลอาศัยหลักการพื้นฐานให้อะตอมแต่ละอะตอมเป็นลูกบอลทรงกลมที่มีมวล (mass) และประจุ (charge) ที่แน่นอน และอะตอมแต่ละอะตอมเกิดแรงกระทำต่อกันหรือเกิดพันธะกันในลักษณะคล้ายสปริงที่เชื่อมต่อกันระหว่างแต่ละอะตอม ซึ่งจะอาศัยเวลาในการคำนวณที่น้อยกว่าและเหมาะสมในการศึกษากับระบบที่เป็นสารชีวโมเลกุลขนาดใหญ่ เช่น โปรตีนหรือดีเอ็นเอ มากกว่าเทคนิคกลศาสตร์ควอนตัม



รูป 1.4 ภาพเปรียบเทียบโครงสร้างดีเอ็นเอ (a) ภาพวาดโครงสร้าง 2 มิติของดีเอ็นเอที่นำเสนอโดย

Watson และ Crick (b) ภาพโมเลกุลสายดีเอ็นเอจากเครื่อง AFM และ (c) ภาพจำลองโครงสร้าง 3 มิติของสายดีเอ็นเอแสดงอะตอมและความยาวพันธะไฮโดรเจนของเบสอะดีนีน (DA) กับเบสไทมีน (DT)

การคำนวณทางกลศาสตร์โมเลกุลมีวิธีการคำนวณได้หลายวิธีขึ้นกับแหล่งที่มาหรือหลักฐานอ้างอิงและวิธีการคำนวณที่เลือกใช้หรือที่เรียกว่า สนามพลังงาน (force field) การศึกษานี้จะอาศัยฐานข้อมูลจาก AMBER force field^[20] ซึ่งเป็นสนามพลังงานหนึ่งที่ยอมรับใช้ในการสร้างแบบจำลองโมเลกุลสำหรับสารชีวโมเลกุล มีหลักการประเมินค่าพลังงานศักย์ทางกลศาสตร์โมเลกุล (E_{MM}) ให้เป็นผลรวมของพลังงานที่เกิดขึ้นจากตัวแปรต่างๆ ได้แก่ ความยาวพันธะ มุมที่เกิดขึ้นระหว่างพันธะ 2 พันธะ มุม dihedral และ แรงแวนเดอร์วาลและศักย์ไฟฟ้าสถิตของแต่ละอะตอม ที่เบี่ยงเบนไปจากค่ามาตรฐานทางทฤษฎีเมื่ออะตอมอยู่ในภาวะสมดุล (ดูสมการ 1) เช่น ระยะห่างที่เหมาะสมระหว่างอะตอมออกซิเจนและไฮโดรเจน ค่ามุมที่เหมาะสมระหว่างอะตอมออกซิเจนและไฮโดรเจนในโมเลกุลน้ำ เป็นต้น ซึ่งข้อมูลค่ามาตรฐานเหล่านี้ถูกรวบรวมและจัดเก็บไว้ในฐานข้อมูลของสนามพลังงานที่เลือกใช้ เมื่อสามารถคำนวณค่า E_{MM} ของแต่ละโมเลกุลได้แล้ว จะทำให้ประเมินได้ถึงความเสถียรของโมเลกุลนั้นๆ เป็นอันดับแรก หรือเมื่อต้องการหาการจัดเรียงตัวของโมเลกุลในสามมิติที่มีความเสถียรในธรรมชาติ ก็สามารถทำได้โดยการลดระดับพลังงานศักย์ของโมเลกุลให้ต่ำลงมา ซึ่งเชื่อว่าจะเป็นโครงสร้างสามมิติจริงของโมเลกุลนั้นในธรรมชาติ

$$E_{MM} = \sum_{bonds} \left[K_r (r - r_{eq})^2 + \right] \sum_{angles} \left[K_\theta (\theta - \theta_{eq})^2 + \right] \sum_{dihedrals} \left[K_\phi (\phi - \phi_{eq})^2 + \right]$$

อย่างไรก็ตาม จากการคำนวณพลังงานของโมเลกุลนั้น ทำให้ทราบว่าอะตอมแต่ละอะตอมมีพลังงานอยู่ภายใน รวมถึงการเกิดอันตรกิริยากับอะตอมแวดล้อม หรือการได้รับพลังงานจากอุณหภูมิภายนอก ทำให้ทราบว่าในสภาพความเป็นจริงแล้วอะตอมแต่ละอะตอมในโมเลกุลมีแนวโน้มที่จะเคลื่อนที่อยู่ตลอดเวลา ดังนั้นถ้าต้องการศึกษาถึงพฤติกรรมทางพลศาสตร์ของโมเลกุลนั้นๆ ในช่วงระยะเวลาที่กำหนด จำเป็นต้องอาศัยเทคนิคที่เรียกว่า “การจำลองพลวัต (molecular dynamics simulation)” เพื่อให้ทราบว่าในระยะเวลาที่กำหนดนั้น อะตอมแต่ละอะตอมซึ่งมีพลังงานอยู่ภายในนั้น มีการเคลื่อนที่จากจุดเริ่มต้นอย่างไรในสภาวะที่กำหนดขึ้นในแบบจำลองโมเลกุล โดยมีหลักการพื้นฐานเป็นไปตามกฎข้อที่สองของนิวตัน คือ ความเร่งของวัตถุ (a) จะแปรผันตามแรงที่กระทำต่อวัตถุ (F) แต่จะแปรผกผันกับมวลของวัตถุ (m) ซึ่งแรงที่กระทำต่อแต่ละอะตอมสามารถคำนวณได้จากการเปลี่ยนแปลงของค่าพลังงาน E_{MM} และตำแหน่งในสามมิติ (r) ของอะตอมนั้น^[21] ดังสมการที่ 2 และผลรวมของแรงที่เกิดขึ้นต่ออะตอมแต่ละอะตอมในโมเลกุล ณ เวลา t ใดๆ จากสมการ 2 นั้น จะสามารถ

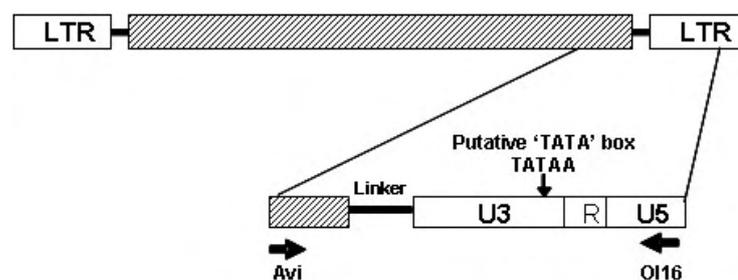
นำมาคำนวณหาค่าความเร่งที่เกิดขึ้น และเมื่อเวลาเปลี่ยนแปลงไปในแต่ละขั้นเวลา (time step, δt) จะสามารถคำนวณหาตำแหน่งในสามมิติใหม่ของแต่ละอะตอมได้ โดยอาศัย Verlet algorithm^[22] ในสมการที่ 3 ค่า δt ที่นิยมใช้ในการจำลองพลวัต คือ 1 หรือ 2 fs จากหลักการนี้ทำให้สามารถเก็บข้อมูลการเคลื่อนที่ของอะตอมแต่ละอะตอมในระบบแบบจำลองโมเลกุลที่สร้างขึ้นเทียบกับเวลาต่างๆ ได้ ซึ่งข้อมูลที่ได้นี้จะเรียกว่า trajectory เพื่อจะนำไปใช้ในการวิเคราะห์ผลพฤติกรรมทางพลศาสตร์ต่อไป

$$F = ma = -\frac{dE_{MM}}{dr} \quad (2)$$

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \quad (3)$$

1.5 Long Terminal Repeats (LTRs)

Long Terminal Repeats (LTRs) เป็นลำดับเบสซ้ำกันที่อยู่ด้านปลายทั้งสองของดีเอ็นเอไวรัส (cDNA) ที่ถูกสร้างขึ้นมาจากกระบวนการ reverse transcription จากอาร์เอ็นเอของไวรัส HIV-1 ภายในเซลล์ที่มีการติดเชื้อ ซึ่งลำดับเบสบนสาย LTR ประกอบด้วยส่วนสำคัญ 3 ส่วน คือ U3 (unique 3' RNA region) R (repeat RNA) และ U5 (unique 5' RNA region) ดังรูป 1.5 ซึ่งมีผลในการควบคุมการแสดงออกของยีนไวรัส เนื่องจากประกอบด้วยส่วน promoter ที่มี 'TATA' box เป็นองค์ประกอบ ส่วน enhancer และ ส่วน polyadenylation signals

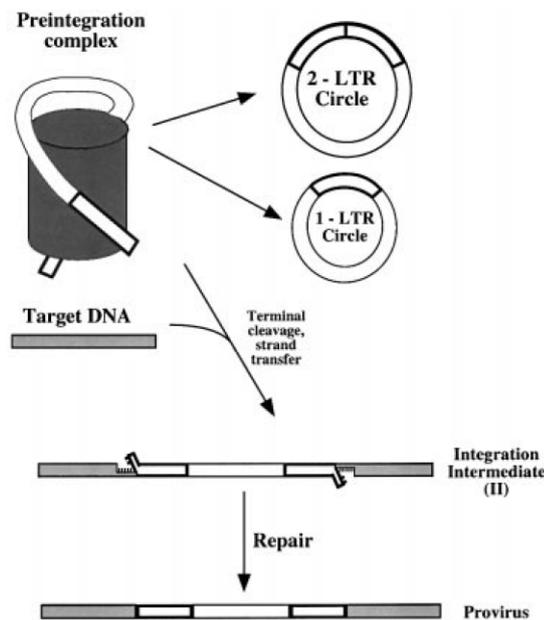


รูป 1.5 ส่วน Long Terminal Repeats (LTRs) ที่ประกอบด้วยส่วน U3 R และ U5 บริเวณปลายทั้งสองของ cDNA^[23]

ส่วน cDNA นี้ พบว่ามี 2 รูปแบบหลังจากการติดเชื้อ retrovirus (รูป 1.6) ได้แก่ 1) cDNA ที่เป็นสายตรง (linear form) ซึ่งอยู่ในรูป preintegration complexes (PICs) จะถูกเชื่อมต่อ (integrate) กับสายดีเอ็นเอของเจ้าบ้านโดยอาศัยปลาย LTR เป็นส่วนเชื่อมต่อและเอนไซม์อินทิเกรสของไวรัส HIV-1 ด้วย 2 ขั้นตอน เริ่มจากการตัดเบส จำนวน 2 คู่เบสที่ปลาย 3' จาก cDNA ออก แล้วเกิด strand

transfer เชื่อมเข้ากับยีนของเจ้าบ้านได้เป็น integration intermediate (II) แล้วจึงเกิดการซ่อมแซมของ ดีเอ็นเอได้เป็น proviral DNA และ 2) ส่วน cDNA สายตรงที่ไม่ถูกเชื่อมต่อกับยีนของเซลล์เจ้าบ้าน จะถูกซ่อมแซมด้วยกระบวนการ non-homologous DNA end-joining (NHEJ) ของเซลล์เจ้าบ้าน^[24] เชื่อมให้เป็นวง cDNA (circular form) ที่มีปลายเป็นลำดับเบส LTR ทั้งสองปลาย (2LTR) หรือหนึ่ง ปลาย (1LTR) เชื่อมต่อกัน และการที่ภายในเซลล์ที่ติดเชื้อ HIV-1 มี cDNA ที่มีปลาย LTR อิสระนั้นจะ เกิด double-strand breaks ซึ่งสามารถเหนี่ยวนำให้เซลล์เกิด apoptosis ได้ ดังนั้นการเชื่อม cDNA แบบสายตรงให้เป็นวงนั้น เป็นกระบวนการหนึ่งในการป้องกันการเกิด apoptosis ของเซลล์อีกด้วย แต่ อย่างไรก็ตามยังมีประเด็นถกเถียงถึงความคงตัวและปริมาณของ cDNA ที่เป็นวงภายในเซลล์ เนื่องจากการที่ปลาย LTR ถูกเชื่อมเข้าด้วยกันทำให้ไม่สามารถเกิดการเพิ่มจำนวนได้เมื่อเซลล์มีการ แบ่งตัว หรืออาจถูกย่อยสลายได้ด้วยเอนไซม์นิวคลีเอส (nuclease) ภายในเซลล์

จากการศึกษาหาปริมาณของ cDNA ที่ได้จากการทำ PCR โดยวิธีวิเคราะห์การเรืองแสงนั้น พบว่าเมื่อเซลล์ที่มีการติดเชื้อ HIV-1 ในระยะเริ่มต้นจะพบปริมาณของ cDNA แบบสายตรงสะสมมาก ขึ้นจนถึงปริมาณสะสมมากที่สุดภายใน 12 ถึง 24 ชั่วโมง ส่วนชนิดที่เป็นวงนั้นจะมีปริมาณสะสมจนถึง ระดับสูงที่สุดภายหลังจาก 24 ชั่วโมง^[25] และ cDNA ที่เป็นวงชนิด 2LTR นั้น พบว่ามีความคงตัวได้ดี ภายในเซลล์^[26, 27]



รูป 1.6 รูปแบบของ cDNA ที่พบได้หลังจากการติดเชื้อ retrovirus และการเชื่อมต่อกับยีนเจ้าบ้าน^[24]

1.6 วัตถุประสงค์

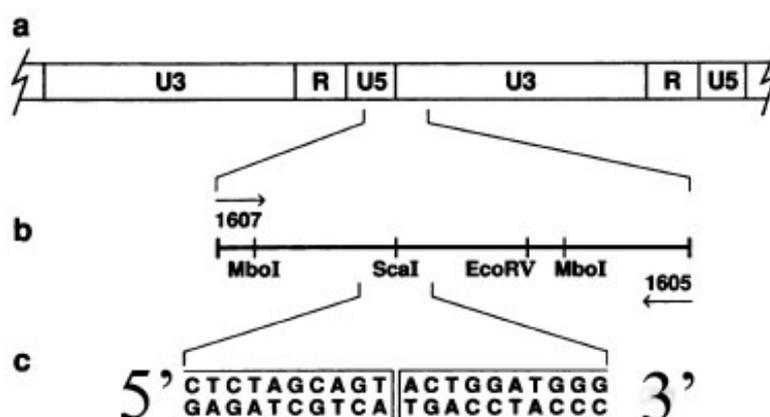
โครงการ การออกแบบโปรตีน ZF โดยวิธีทางคอมพิวเตอร์ สำหรับการจดจำลำดับเบสบนสาย ดีเอ็นเอ (Computational design of zinc-finger proteins for recognition of DNA sequences) เลือกใช้ลำดับเบสเป้าหมายในส่วน 2LTR ซึ่งเป็นส่วนที่มีความสำคัญสำหรับเอนไซม์อินทิเกรสในการ เชื่อมต่อปลายสายของ cDNA เข้ากับยีนของเซลล์เจ้าบ้าน ทำให้ดีเอ็นเอของไวรัสสามารถเพิ่มปริมาณ ได้อย่างรวดเร็ว ดังนั้นการออกแบบโปรตีน ZF เพื่อให้จับกับลำดับเบส 2LTR ในการศึกษา นี้ เป็น กระบวนการยับยั้งการเพิ่มจำนวนของดีเอ็นเอของไวรัส ที่เป็นสารตั้งต้นในธรรมชาติของเอนไซม์อินทิเกรส และมีวัตถุประสงค์ในการออกแบบโปรตีน ZF ที่มีความจำเพาะสูงในการเข้าจับลำดับเบส เป้าหมาย โดยวิธีทางคอมพิวเตอร์ ทำการศึกษาพฤติกรรมทางพลศาสตร์ (dynamic behavior) ในขณะ เกิดสารเชิงซ้อนระหว่างโปรตีน ZF และสายดีเอ็นเอที่มีลำดับเบสที่แตกต่างกัน ทำความเข้าใจถึง ความจำเพาะ (specificity) และความสามารถในการเข้าจับ (binding affinity) ของโปรตีนที่ออกแบบ ขึ้นมา รวมถึงเพื่อสร้างฐานข้อมูลเบื้องต้นสำหรับการออกแบบโปรตีน ZF ที่มีความจำเพาะและสามารถ จับกับลำดับเบสเป้าหมายได้แข็งแรงต่อลำดับเบส ซึ่งอาจนำมาพัฒนาเป็นทางเลือกหนึ่งในการรักษา โรคเอดส์จากเชื้อ HIV-1 โดยการยับยั้งการทำงานของเอนไซม์อินทิเกรส ซึ่งเป็นกระบวนการที่เกิดขึ้น เฉพาะในเซลล์ที่ติดเชื้อ HIV-1 และไม่พบกระบวนการนี้ในเซลล์ของมนุษย์ทั่วไป

2. วิธีการทดลอง

2.1 การออกแบบโปรตีน ZF ที่จำเพาะต่อ 2LTR ของสายดีเอ็นเอไวรัส HIV-1

เป้าหมายที่ใช้ในการศึกษานี้ คือ ลำดับเบสจำเพาะที่พบบริเวณที่เชื่อมต่อกันระหว่างปลาย 5' และปลาย 3' ของดีเอ็นเอไวรัส HIV-1 ที่อยู่ในรูป cDNA ที่เป็นวง ซึ่งจะมีลำดับเบสที่แน่นอนของปลาย 5' และปลาย 3' นี้ เรียกว่า บริเวณ long terminal repeat (LTR) ดังแสดงในรูป 2.1 ลำดับเบสเป้าหมายที่เหมาะสมในการใช้ในการออกแบบโปรตีน ZF ที่ได้จากการทำนายโดยเว็บไซต์ ZF Tools เป็นเบสจำนวน 18 คู่เบส คือ 5'-CTA GCA GTA CTG GAT GGG-3'^[28] ซึ่งเป็นบริเวณที่เอนไซม์อินทิเกรสของเชื้อไวรัส HIV-1 สามารถเข้าจับและทำการเชื่อมต่อเอาดีเอ็นเอของไวรัสกับของเซลล์เจ้าบ้านเข้าด้วยกัน และเกิดการเพิ่มจำนวนของดีเอ็นเอไวรัส HIV-1 ต่อไป จากลำดับเบสเป้าหมายนี้ได้นำมาใช้เป็นต้นแบบ เพื่อทำนายลำดับกรดอะมิโนของโปรตีน ZF ที่จำเพาะกับลำดับเบสเป้าหมาย (target site) โดยอาศัยเว็บไซต์ฐานข้อมูล Zinc Finger Tools (<http://www.scripps.edu/mb/barbas/zfdesign/zfdesignhome.php>) ซึ่งเป็นการทำนายลำดับกรดอะมิโนด้วยหลักการของ Modular Assembly ตัวอย่างการใช้งาน ZF Tools แสดงในรูป 2.2 ได้ผลลำดับกรดอะมิโนที่ประกอบด้วยโปรตีน ZF จำนวน 6 โมดูล (ZFPv1) ความยาว 176 กรดอะมิโน ได้แก่

LEPGEKPYKC PECGKSFSRS DKLVRHQ RTH TGEKPYKCPE CGKSFSTSGN LVRHQ RTH TG
EKPYKCPECG KSFSRNDALT EHQRTH TGEK PYKCPECGKS FSQSSSLVRH QRTH TGEKPY
KCPECGKSFS QSGDLRRHQ R THTGEKPYKC PECGKSFSQN STLTEHQ RTH TGKKTS



รูป 2.1 ลำดับเบสในส่วน 2LTR ที่เชื่อมกันของ circular HIV-1 DNA

a

Zinc Finger Tools

See us also at zincfingertools.org.

- ZF Tools home
- News
- Intro to ZFPs
- Tools overview
- DNA search tips
- ZFP design tips
- Available triplets
- Synthesize a ZFP
- Effector domains
- Nuclease design
- Citations
- About
- Send feedback
- Welcome Supat
- Sign out

Search DNA Sequence for Contiguous or Separated Target Sites

Enter a DNA sequence to be scanned for target sites (10 kb max).

Triplets to search

GNG

GN(ACT)

ANN

CNN

TNN

ZF set: Total

Search both strands

Contiguous targets

Separated targets

Minimum target size (bp): 18

Half-site size (bp): 9

Core (IUPAC bases or number of bp): NNNNNN or Enzyme site

Juxtapose: C-terms or N-terms

Design a Zinc Finger Protein

Enter the DNA target site. The ZF amino acid sequence predicted to recognize this site will be determined. This tool expects only valid triplets.

ระบุ target site

Target site 5' CTAGCAGTACTGGATGGG 3'

N-term backbone: YKCECGKSFS

C-term backbone: HQRTH

Zinc finger linker: TGEKP

N-term fixed: LEPGEKP

C-term fixed: TGKKTS

Include fixed sequences?

b

Zinc Finger Tools

See us also at zincfingertools.org.

- ZF Tools home
- News
- Intro to ZFPs
- Tools overview
- DNA search tips
- ZFP design tips
- Available triplets
- Synthesize a ZFP
- Effector domains
- Nuclease design
- Citations
- About
- Send feedback
- Welcome Supat
- Sign out

Design a ZF

N-term backbone: YKCECGKSFS

C-term backbone: HQRTH

ZF linker: TGEKP

N-term fixed: LEPGEKP

C-term fixed: TGKKTS

Using 49 triplets.

This target site does **not** have potential target site overlap issues (also consider TSO for 3' sequence to this site).

Requested target site is:
5' CTA GCA GTA CTG GAT GGG 3'

Perform a short-nucleotide NCBI BLAST with this target sequence with the genome.

Pop-up ELISA graphs on triplet rollover: Yes No

If ELISA graphs display slowly, you can cache all images (~1.5 MB total) by clicking [here](#).

Finger	Triplet	Helix
1	GGG	RSDRLVR
2	GAT	TSGNLVR
3	CTG	RNDALTE
4	GTA	QSSSLVR
5	GCA	QSGDLRR
6	CTA	QNSTLTE

ผลโปรตีน zinc finger ในส่วน recognition helix

The amino acid sequence of the ZF protein predicted to bind to the target sequence is:
LEPGEKPYKCECGKSFRRSDKLVRRHQRTHTGKEPKYKCECGKSFSTSGNLVRRHQRTHTGEPKYKCECGKSFRRNDALTEHQRTHTGERPKYKCECGKSFSSQSSSLVRRHQRTHTGERPKYKCECGKSFSSQSGDLRRHQRTHTGEPKYKCECGKSFQNSTLTEHQRTHTGKRTS

รูป 2.2 การใช้งานโปรแกรม ZF Tools

(a) การระบุเป้าหมาย (target site) ของลำดับเบสจาก 5' ไปยัง 3'

(b) ผลลัพท์การดอะมิโนของโปรตีน ZF ที่เหมาะสมกับลำดับเบสเป้าหมาย

2.2 การสร้างแบบจำลองโปรตีน ZF ที่เข้าจับกับสายดีเอ็นเอ

จากข้อมูลลำดับกรดอะมิโนของโปรตีน ZF ที่ทำนายได้จากเว็บไซต์ ZF Tools นำมาสร้างแบบจำลองโมเลกุลโดยอาศัยข้อมูลของ Segal และคณะ^[29] ที่ได้รายงานโครงสร้างสามมิติที่มีความละเอียด (resolution) สูงถึง 1.96 Å จากการทดลองตกผลึกของสารประกอบระหว่างโปรตีน ZF ความยาว 6 โมดูลที่ออกแบบโดยเว็บไซต์ ZF Tools ที่มีชื่อว่า “Aart” และสายดีเอ็นเอ 5'-ATG TAG AGA AAA ACC AGG-3' ซึ่งถูกจัดเก็บไว้ในฐานข้อมูล Protein Data Bank (PDB) ในรหัส 2I13 เป็นต้นแบบในการสร้างแบบจำลองแทนการทำด็อกกิ้ง เนื่องจากการทดลองด็อกกิ้งโดยใช้ชุดโปรแกรม Discovery Studio พบว่าโครงสร้างที่ได้มีความแตกต่างจากโครงสร้างผลึกที่มีการรายงานอย่างมาก และโดยหลักการทั่วไปแล้วโครงสร้างผลึกจะมีความน่าเชื่อถือมากกว่า เนื่องจากเป็นข้อมูลที่ได้จากการทดลองตกผลึกระหว่างสารละลายโปรตีน ZF และสารละลายของดีเอ็นเอร่วมกันซึ่งคาดว่าเป็นเหตุการณ์ที่เกิดขึ้นจริงในเซลล์เมื่อมีสารทั้งสองชนิดนี้อยู่ด้วยกัน ดังนั้นการเลือกใช้ต้นแบบจาก PDB id 2I13 จึงมีความเหมาะสมมากกว่า และทำการสร้างแบบจำลองโมเลกุลเพิ่มเติมเพื่อศึกษาความจำเพาะ (specificity) และความสามารถในการเข้าจับ (affinity) ของโปรตีน zinc finger ในหัวข้อ 2.6 และ 2.7 ต่อไป

2.3 การจัดเตรียมระบบและการจำลองพลวัต

ในการศึกษานี้แบบจำลองโครงสร้างโมเลกุลที่ได้ปรับเปลี่ยนตามต้องการแล้ว จะอาศัยชุดโปรแกรม Assisted Model Building with Energy Refinement (AMBER) version 10 และ AmberTools version 1.2 และใช้ AMBER ff03 เป็นฐานข้อมูลในการเตรียมระบบ (system preparation) ของแบบจำลอง ส่วนโลหะสังกะสีนั้นใช้ค่าตัวแปรที่รายงานโดย Stote และ Karplus^[30] ($\sigma = 1.70 \text{ \AA}$, $\epsilon = 0.67 \text{ kcal/mol}$) แทนตัวแปรมาตรฐาน ($\sigma = 1.10 \text{ \AA}$, $\epsilon = 0.0125 \text{ kcal/mol}$) จาก AMBER ff03 เพื่อให้โมเลกุลสังกะสีสามารถเกิดอันตรกิริยากับซิสเทอีนและฮิสทีดีนได้แข็งแรงมากขึ้น ทำให้โปรตีน ZF คงสภาพตลอดระยะเวลาการจำลองพลวัต (Molecular Dynamics Simulations)

การจัดเตรียมระบบอาศัยโปรแกรม leap จากชุดโปรแกรม AmberTools เริ่มจากการจัดวางโมเลกุลที่จะศึกษาลงในกล่องทรงแปดเหลี่ยมปลายตัด (truncated octahedral box) ที่แต่ละด้านของกล่องจะต้องห่างจากอะตอมของโมเลกุลที่จะศึกษาอย่างน้อย 10 Å และภายในกล่องนี้บรรจุด้วยตัวทำละลาย (solvation) คือ แบบจำลองโมเลกุลน้ำชนิด TIP3P^[31] แล้วจึงทำการสะเทินประจุรวมของทั้งระบบโดยเฉพาะส่วนหมู่ฟอสเฟตของสายดีเอ็นเอที่มีประจุลบให้เป็นกลาง (neutralization) ด้วย

โซเดียมไอออน นอกจากนั้นระบบทั้งหมดนี้หรือกล่าวคือกล่องทรงแปดเหลี่ยมปลายตัดที่ภายในบรรจุด้วยโมเลกุลของโปรตีน ZF สายดีเอ็นเอ น้ำและโซเดียมไอออน จะถูกทำให้อยู่ในสภาพ periodic boundary condition (PBC) ที่ทุกๆ ด้านของกล่องทรงแปดเหลี่ยมปลายตัดนี้ จะมีกล่องในลักษณะเดียวกันนี้ต่อไปในสามมิติอย่างไม่มีที่สิ้นสุด เพื่อให้ในระหว่างการจำลองพลวัต ถ้ามีอะตอมใดอะตอมหนึ่งเคลื่อนที่ออกจากกล่องนี้ จะมีอะตอมที่เหมือนกันนี้จากกล่องใกล้เคียงมาแทนที่เสมอเพื่อให้โมเลกุลในกล่องที่เราศึกษาที่อยู่ในสภาพสมดุลเสมอ รวมทั้งให้การจำลองพลวัตมีความคล้ายคลึงกับสภาพความเป็นจริงมากที่สุด คือ โมเลกุลที่เราศึกษามักไม่ได้อยู่เป็นโมเลกุลเดี่ยวๆ จะต้องอยู่ในลักษณะที่เป็นสารละลายที่สมดุลมีหลายๆโมเลกุลอยู่ด้วยกันเสมอ

หลังจากการเตรียมระบบเบื้องต้นแล้ว โมเลกุลที่ศึกษาจะเข้าสู่กระบวนการลดพลังงานศักย์ (energy minimisation) ของระบบให้ต่ำลง โดยอาศัยโปรแกรม sander ในชุดโปรแกรม AMBER10 ทำเป็น 2 ขั้นตอน คือ เริ่มทำการลดพลังงานของโมเลกุลตัวทำละลายและไอออนก่อน แล้วจึงทำการลดพลังงานศักย์ของโมเลกุลโปรตีน ดีเอ็นเอ ตัวละลายและไอออนอีกครั้งหนึ่ง แล้วจึงเข้าสู่การทำให้ระบบสมดุล (equilibration) โดยอาศัยหลักการเพิ่มอุณหภูมิของระบบอย่างช้าๆ จนกระทั่งถึง 300 K พร้อมทั้งลดพลังงานที่ใช้จริงโมเลกุลที่ต้องการศึกษาอย่างช้าๆ จนกระทั่งโมเลกุลนั้น สามารถเคลื่อนที่ได้อย่างอิสระ ซึ่งใช้เวลาในการจำลองอยู่ที่ 190 ps หลังจากนั้นจึงทำการจำลองพลวัตเพื่อเริ่มเก็บข้อมูลพฤติกรรมทางพลศาสตร์ของระบบแบบจำลองเป็นเวลา 5 ns ที่อุณหภูมิและความดันคงที่ (300 K และ 1 atm) โดยการเก็บข้อมูลการเคลื่อนที่ของแต่ละอะตอมในระบบทุกๆ 1 ps หรือที่เรียกว่า "trajectory" เพื่อทำการวิเคราะห์ผลการศึกษาต่อไป

2.4 การวิเคราะห์โครงสร้างหลังการจำลองพลวัต

ข้อมูล trajectory ที่บันทึกได้จะตัดเอาส่วนข้อมูลการเคลื่อนที่ของโมเลกุลน้ำ TIP3P และโซเดียมไอออนออก เพื่อความสะดวกและรวดเร็วในการวิเคราะห์ข้อมูลพฤติกรรมทางพลศาสตร์ของโปรตีน ZF และสายดีเอ็นเอ โดยอาศัยโปรแกรม ptraj จากชุดโปรแกรม AmberTools ข้อมูลพื้นฐานที่ทำการวิเคราะห์ ได้แก่ ระยะห่างระหว่าง H-bond donor และ H-bond acceptor เพื่อดูความสามารถในการเกิดพันธะไฮโดรเจนของอะตอม 2 อะตอม และค่า root mean square deviation (RMSD) เพื่อประเมินถึงเสถียรภาพของการจัดเรียงตัว (conformational stability) ของโมเลกุลที่ศึกษาในช่วงเวลาที่ทำการจำลองพลวัต ซึ่งค่า RMSD สามารถคำนวณได้จาก

เมื่อ N เป็นจำนวนของตำแหน่งในสามมิติตามแกน X Y และ Z ของอะตอมทั้งหมดในโครงสร้างโมเลกุล และ D_i เป็นระยะห่างระหว่างตำแหน่งบนแกนแต่ละแกนของอะตอม i เมื่อโมเลกุลทั้งสองโมเลกุลมาซ้อนทับกัน ตัวอย่างในการคำนวณ เช่น การหาค่า $RMSD$ ระหว่างโมเลกุลที่ศึกษา ณ เวลา 0 ps เทียบกับ ณ เวลา 2 ps มีความแตกต่างกันเท่าใด หากค่า $RMSD$ มีค่าคงที่ แสดงว่าโครงสร้างโมเลกุลที่ศึกษานั้นมีเสถียรภาพดีในการจัดเรียงตัว ไม่มีการแตกของพันธะหรือไม่มีการเคลื่อนที่ของอะตอมบางอะตอมอย่างรวดเร็วจากพลังงานศักย์ภายในอะตอมที่สูงเกินไป

2.5 การวิเคราะห์เชิงพลังงานของโมเลกุล

การคำนวณหาพลังงานในการยึดจับระหว่างโปรตีน ZF ทั้ง 6 โมดูลกับสายดีเอ็นเอเป้าหมายสามารถคำนวณได้จากวิธี Molecular Mechanics/Generalised Born Surface Area (MM/GBSA) ซึ่งเป็นแนวคิดหนึ่งในผสมผสานแนวคิดทางกลศาสตร์โมเลกุลร่วมกับการใช้แบบจำลองของตัวทำละลายชนิด generalized Born (GB) และพื้นที่ผิวของโมเลกุลที่ตัวทำละลายสามารถสัมผัสได้ (solvent accessible surface area) เพื่อการคำนวณหาพลังงานทางกลศาสตร์โมเลกุล (molecular mechanics energy) ของโมเลกุลหนึ่งๆ (G_x) จากสมการ 1 ซึ่งเป็นผลรวมของแรงแวนเดอร์วาลส์ (van der Waals interaction, E_{vdw}) แรงไฟฟ้าสถิต (electrostatic interaction, E_{es}) พลังงานภายในโมเลกุล (internal energy, E_{in}) พลังงานอิสระของระบบที่ถูกเติมด้วยตัวทำละลาย GB (G_{pol}) และพลังงานอิสระที่เป็นผลจากความไม่ชอบน้ำของโมเลกุล (G_{nonpol}) ซึ่งคำนวณได้จากพื้นที่ผิวของโมเลกุลที่ตัวทำละลายสามารถสัมผัสได้ในสมการ 2

จากหลักการคำนวณในสมการ 1 ทำให้สามารถหาค่าพลังงานทางกลศาสตร์โมเลกุลของโมเลกุลหนึ่งๆ ได้ ดังนั้นหากแยกคำนวณค่าพลังงานเหล่านี้เฉพาะของโปรตีน ZF และกำหนดให้เป็น ligand (G_{ligand}) พลังงานเฉพาะของสายดีเอ็นเอและกำหนดให้เป็น receptor ($G_{receptor}$) และพลังงานของสารประกอบของโปรตีน ZF และสายดีเอ็นเอ กำหนดให้เป็น complex ($G_{complex}$) ดังนั้นความแตกต่างระหว่างพลังงานของ complex หักลบด้วย receptor และ ligand จึงสามารถระบุได้ว่าพลังงานการยึดจับระหว่างโปรตีน ZF กับสายดีเอ็นเอ ที่เกิดขึ้น ($\Delta G_{binding}$) คำนวณได้จากสมการ 3 ยิ่งค่านี้มีค่าเป็นลบมากเท่าใด แสดงว่าการยึดจับระหว่าง ligand และ receptor มีความแข็งแรงสูง ซึ่งในการศึกษานี้จะทำ

การคำนวณหาค่า G_x และค่า $\Delta G_{binding}$ ทุกๆ 2 ps และคำนวณหาค่า G_{pol} อาศัย GB solvent model ที่พัฒนาโดย Onufriev และคณะ (igb =2)^[32] ซึ่งเหมาะสมสำหรับการคำนวณหาค่า G_{pol} ของสารชีวโมเลกุล

$$\Delta G_{binding} = G_{complex} - G_{receptor} - G_{ligand} \quad (3)$$

2.6 ความจำเพาะของโปรตีน ZFPv1

เพื่อให้ได้แบบจำลองโครงสร้างโมเลกุลของโปรตีน ZF ZFPv1 และสายดีเอ็นเอเป้าหมาย 5'-CTA GCA GTA CTG GAT GGG-3' ให้มีการจัดเรียงตัวในสามมิติใกล้เคียงกับที่มีการรายงานจากโครงสร้างผลึกของโปรตีน Aart ที่เข้าจับกับสายดีเอ็นเอ จึงทำการปรับเปลี่ยนทั้งในส่วนลำดับกรดอะมิโนจาก recognition helix และลำดับเบสจากโครงสร้างต้นแบบ PDB id 2I13 ให้เป็นไปตามลำดับกรดอะมิโนและเบสตามต้องการ (รูป 2.3) โดยทำการปรับเปลี่ยนเฉพาะส่วนโครงสร้างของหมู่สายโซ่ข้าง และคางตำแหน่งแกนหลัก (backbone) ของทั้งโปรตีนและดีเอ็นเอไว้เช่นเดียวกับโครงสร้างต้นแบบ PDB id 2I13 โดยอาศัยชุดโปรแกรม Discovery Studio และให้ชุดข้อมูลแบบจำลองของโปรตีน ZFPv1 ที่เข้าจับกับลำดับเบส 2LTR ซึ่งเป็นเป้าหมายที่จำเพาะนี้ให้เป็นชุดแบบจำลอง "Zif1" และยังสามารถทดลองปรับเปลี่ยนลำดับเบสให้เป็นลำดับที่ไม่จำเพาะต่อ ZFPv1 คือ 5'-TGC TAG CGT ATC ATC TAG-3' ให้เป็นชุดแบบจำลอง "Zif2" ซึ่งมีความแตกต่างของดีเอ็นเอเป้าหมายอย่างชัดเจนกับโปรตีน ZF เป้าหมายโมเลกุลเดียวกัน

นอกจากนั้นแล้ว พบว่าลำดับเบสที่เป็นไปได้ของบริเวณปลายสายดีเอ็นเอของเชื้อไวรัส HIV-1 ที่พบในธรรมชาติมี 2 แบบ ได้แก่ 5'-CTA GCA GTA CTG GAT GGG-3' (เรียกย่อว่าเป็น "GAT") และ 5'-CTA GCA GTA CTG GAA GGG-3' (GAA) ซึ่งมีความแตกต่างกันเพียงลำดับเบสหนึ่งตำแหน่งสำหรับการจับกับโปรตีน ZF โมดูลที่ 2 ดังนั้นจึงได้ทำการออกแบบลำดับกรดอะมิโนของโปรตีน ZF ที่จำเพาะต่อลำดับเบสของเชื้อไวรัส HIV-1 ทั้งสองแบบโดยวิธีการเดียวกับข้างต้น ได้แก่ ZFPv1 และ ZFPv2 ซึ่ง ZFPv1 เป็นโปรตีนที่ออกแบบสำหรับการเข้าจับกับลำดับเบส GAT ส่วน ZFPv2 เป็นโปรตีนที่ออกแบบมาสำหรับลำดับเบส GAA ซึ่งมีลำดับกรดอะมิโนในส่วน recognition helix ของโมดูลที่ 2 เป็น FSQSSNLVR แทน FSTSGNLVR ใน ZFPv1 จากนั้นผู้วิจัยจึงได้ทำการจำลองโครงสร้างโมเลกุลสามมิติของสารประกอบทั้งหมด 4 ชุด ได้แก่ ZFPv1-GAA, ZFPv1-GAT, ZFPv2-GAA และ ZFPv2-GAT เพื่อศึกษาความจำเพาะเจาะจงของโปรตีนที่ออกแบบได้ในการจำแนกความแตกต่างของลำดับเบสเป้าหมายเพียงหนึ่งตำแหน่ง และแบบจำลองแต่ละชุดจะทำการจัดเตรียม

ระบบ ทำการจำลองพลวัตเป็นเวลา 5 ns และทำการวิเคราะห์เชิงโครงสร้างและเชิงพลังงานหลังจาก การจำลองพลวัตต่อไปตามหัวข้อ 2.3 ถึง 2.5 ข้างต้นเช่นเดียวกัน

a 5'-ATG TAG AGA AAA ACC AGG-3' ---- (PDB id 2I13)
 5'-CTA GCA GTA CTG GAT GGG-3' ---- (2LTR)

b

			-1	3	6		
Aart	GEKP	YACPECGKS	FSRSDHLAE	HQRTH	TGEKP	32	
ZFPv1	GEKP	<u>YKCPECGKS</u>	<u>FSRSDKLVR</u>	HQRTH	TGEKP		
ZFPv2	GEKP	<u>YKCPECGKS</u>	<u>FSRSDKLVR</u>	HQRTH	TGEKP		
		YKCPECGKS	FSDKKDLTR	HQRTH	TGEKP	60	
		YKCPECGKS	FSTSGNLVR	HQRTH	TGEKP		
		YKCPECGKS	FS <u>QSS</u> NLVR	HQRTH	TGEKP		
		YKCPECGKS	FSQRANLRA	HQRTH	TGEKP	88	
		YKCPECGKS	<u>FSRNDALTE</u>	HQRTH	TGEKP		
		YKCPECGKS	<u>FSRNDALTE</u>	HQRTH	TGEKP		
		YKCPECGKS	FSQLAHLRA	HQRTH	TGEKP	116	
		YKCPECGKS	FSQSSSLVR	HQRTH	TGEKP		
		YKCPECGKS	FSQSSSLVR	HQRTH	TGEKP		
		YKCPECGKS	FSREDNLHT	HQRTH	TGEKP	144	
		YKCPECGKS	<u>FSQSGDLRR</u>	HQRTH	TGEKP		
		YKCPECGKS	<u>FSQSGDLRR</u>	HQRTH	TGEKP		
		YKCPECGKS	FSRRDALNV	HQRTH		167	
		YKCPECGKS	<u>FSQNSTLTE</u>	HQRTH			
		YKCPECGKS	<u>FSQNSTLTE</u>	HQRTH			

N-term backbone Recognition helix C-term backbone ZF linker

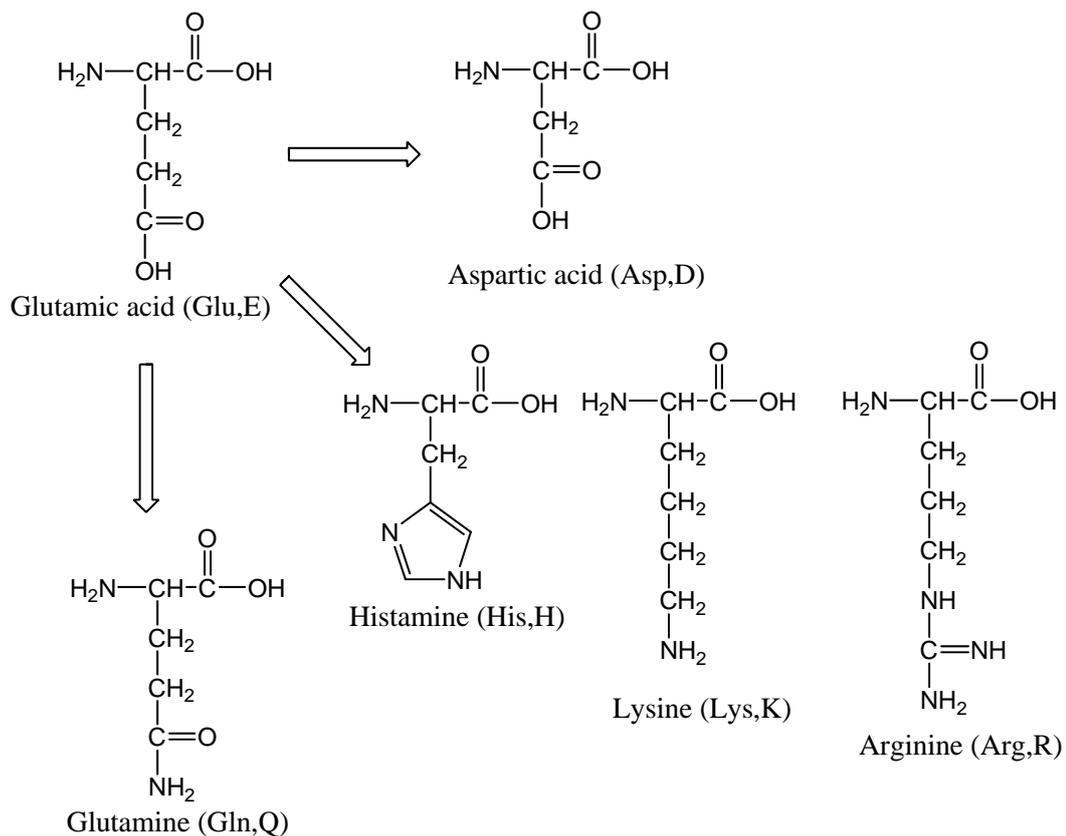
รูป 2.3 การเปรียบเทียบลำดับเบสและกรดอะมิโนระหว่างโครงสร้างผลึกต้นแบบ PDB id 2I13 และแบบจำลองโมเลกุลที่ศึกษา (a) ลำดับเบสในโครงร่างผลึก PDB id 2I13 และเบสเป้าหมาย 2LTR (b) ลำดับกรดอะมิโน Aart ZFPv1 และ ZFPv2 โดยอักษรที่ขีดเส้นใต้ คือ กรดอะมิโนที่ปรับเปลี่ยนจากโปรตีน Aart

2.7 ความสามารถในการเข้าจับของโปรตีน ZFPv1

จากข้อมูลการจำลองพลวัตของแบบจำลอง ZFPv1-GAT เป็นแบบจำลองซึ่งเป็นตัวแทนของโมเลกุลที่จำเพาะระหว่างโปรตีน ZF ที่ออกแบบกับลำดับเบสเป้าหมาย 2LTR ที่ได้มีการพิสูจน์ในห้องทดลองแล้วว่าสามารถเกิดการเข้าจับกันได้จริงจากงานวิจัยที่ตีพิมพ์ก่อนหน้านี้^[28] จึงอาศัยเทคนิคใน

การสร้างแบบจำลองเพิ่มเติมในการปรับเปลี่ยนชนิดของกรดอะมิโนบางตำแหน่งที่ให้ค่าพลังงานการยึดจับที่ไม่เหมาะสมไปเป็นกรดอะมิโนชนิดอื่น เพื่อเป็นการเพิ่มความแข็งแรงในการเข้าจับของโปรตีน ZFPv1 โดยมีสมมติฐานในการปรับเปลี่ยนชนิดของกรดอะมิโนเริ่มจากผลข้อมูลการวิเคราะห์เชิงพลังงานการยึดจับของกรดอะมิโนแต่ละหน่วยของ recognition helix ทั้ง 6 โมดูลของโปรตีน ZFPv1 (decompose energy residue by residue) กับลำดับเบสเป้าหมาย 2LTR พบว่ากรดอะมิโน Glu ที่ตำแหน่ง 162 ของโมดูลที่ 6 สามารถใช้เป็นต้นแบบในการพิสูจน์สมมติฐานดังกล่าว โดยอาศัยชุดโปรแกรม Discovery Studio ในการปรับเปลี่ยนกรดอะมิโน Glu ในตำแหน่ง 162 ไปเป็น

1. กรดอะมิโนที่มีประจุลบ แต่มีสายโซ่ข้างสั้นลง ได้แก่ Asp (E162D)
2. กรดอะมิโนที่มีประจุบวก ได้แก่ His (E162H) Lys (E162K) และ Arg (E162R)
3. กรดอะมิโนที่เป็นกลาง ได้แก่ Gln (E162Q)



แบบจำลองโมเลกุลที่มีการปรับเปลี่ยนชนิดของกรดอะมิโนในตำแหน่งที่ 162 ให้แตกต่างกันไปจากต้นแบบ ZFPv1 ทั้งหมด 5 ชุดแบบจำลองนี้ จะทำการจัดเตรียมระบบ ทำการจำลองพลวัตเป็นเวลา 5 ns หลังจากนั้นจึงทำการวิเคราะห์เชิงโครงสร้างในส่วนโมดูลที่ 6 และวิเคราะห์เชิงพลังงานหลังจากการ

จําลองผลวัติต่อไปตามหัวข้อ 2.3 ถึง 2.5

3. ผลการทดลอง

3.1 ความจำเพาะเจาะจงของโปรตีน ZFPv1

จากข้อมูลลำดับกรดอะมิโนของโปรตีน ZF ที่ทำนายได้จากเว็บไซต์ ZF Tools โดยเฉพาะในส่วนลำดับกรดอะมิโนบน recognition helix ตั้งแต่ตำแหน่ง -1 ถึง +6 ทดลองนำมาเปรียบเทียบกับผลที่ได้จากการทำนายโดย ZF Database (ZIFDB)^[33] ซึ่งเป็นฐานข้อมูลรวบรวมโปรตีน ZF ที่ได้จากธรรมชาติและที่ได้จากการทดลองสร้างโปรตีน ZF ขึ้นมาจากเทคนิคทางพันธุวิศวกรรม จากหลายกลุ่มการทดลอง อาทิ กลุ่ม Sangamo ที่ให้ผลการทำนายลำดับกรดอะมิโนบนโปรตีน ZF โดยพิจารณาถึงแต่ละโมดูลภายในโปรตีน ZF เดียวกัน จะมีผลต่อการจับจำและเข้าจับกับสายดีเอ็นเอ หรือจาก ToolGen ซึ่งรวบรวมข้อมูลจากโปรตีน ZF ที่สามารถจับกับลำดับเบสที่จำเพาะที่พบในเซลล์มนุษย์

ผลการเปรียบเทียบความถี่ของการพบชนิดของกรดอะมิโนในแต่ละตำแหน่งที่ได้จาก ZF Tools และ ZIFDB แสดงไว้ในตาราง 3.1 เพื่อตรวจสอบโดยรวมถึงกรดอะมิโนที่ทำนายได้จาก ZF Tools นั้นมีความแตกต่างเพียงใดเปรียบเทียบกับผลที่รายงานได้จาก ZIFDB ในแต่ละตำแหน่งของกรดอะมิโนบน recognition helix ของแต่ละโมดูลทั้ง 6 โมดูล ซึ่งลำดับกรดอะมิโนที่ทำนายได้จาก ZF Tools และ 3 ลำดับเบสเป้าหมาย แสดงไว้ในคอลัมน์ที่ 1 ซ้ายมือ และกรดอะมิโนที่พบในแต่ละตำแหน่งตั้งแต่ -1 ถึง +6 บน recognition helix แสดงไว้ในคอลัมน์ถัดไปตามลำดับ รวมถึงโอกาสการพบได้ของกรดอะมิโนแต่ละชนิดนั้น และกรดอะมิโนที่เป็นอักษรเข้ม คือ กรดอะมิโนชนิดที่ตรงกับที่ทำนายได้จาก ZF Tools ส่วนกรดอะมิโนบางชนิดที่มีโอกาสพบได้เท่าๆ กันในฐานข้อมูล ZIFDB ได้รวมไว้ในระดับเดียวกัน เช่น กรดอะมิโนตำแหน่ง +6 ของโมดูล 1 จาก ZF Tools ทำนายว่าควรจะต้องเป็นกรดอะมิโนอาร์จินีน (Arg หรือ R) ในการจดจำลำดับเบสตัวนั้น และจากการวิเคราะห์ด้วย ZIFDB พบว่าในตำแหน่งนี้มีรายงานความเป็นไปได้ของการพบชนิดของกรดอะมิโนเป็นอาร์จินีน (R) 64% เป็นแอสพาราจีน (N) 10% โพรลีน (P) 5% อะลานีน (A) หรือกรดแอสพาทิก (D) หรือกลัยซีน (G) หรือฮิสทีดีน (H) หรือไลซีน (K) หรือเซอรีน (S) หรือทรีโอนีน (T) หรือวาลีน (V) ชนิดละ 3% เป็นต้น ซึ่งจากข้อมูลเหล่านี้พบว่าชนิดของกรดอะมิโนที่ทำนายโดย ZF Tools ส่วนใหญ่จะเป็นชนิดที่มีโอกาสพบได้มากที่สุดอยู่แล้วจาก ZIFDB ยกเว้นโมดูล 1 ตำแหน่ง +3 พบว่าฮิสทีดีนปรากฏในฐานข้อมูล 62% แต่ ZF Tools แนะนำให้เป็นไลซีนที่พบได้เป็นลำดับสอง คือ 23% ซึ่งทั้งกรดอะมิโนฮิสทีดีนและไลซีนต่างเป็นกรดอะมิโนในกลุ่มที่เป็นต่างและมีประจุบวกเช่นกันและมักใช้แทนที่กันได้ หรือกรดอะมิโนในตำแหน่ง +2 ของโมดูล 2 เป็นกลัยซีนซึ่งพบเป็นลำดับที่ 2 แทนฮิสทีดีนที่พบมากที่สุด และมีสายโซ่ข้างที่มีขนาด

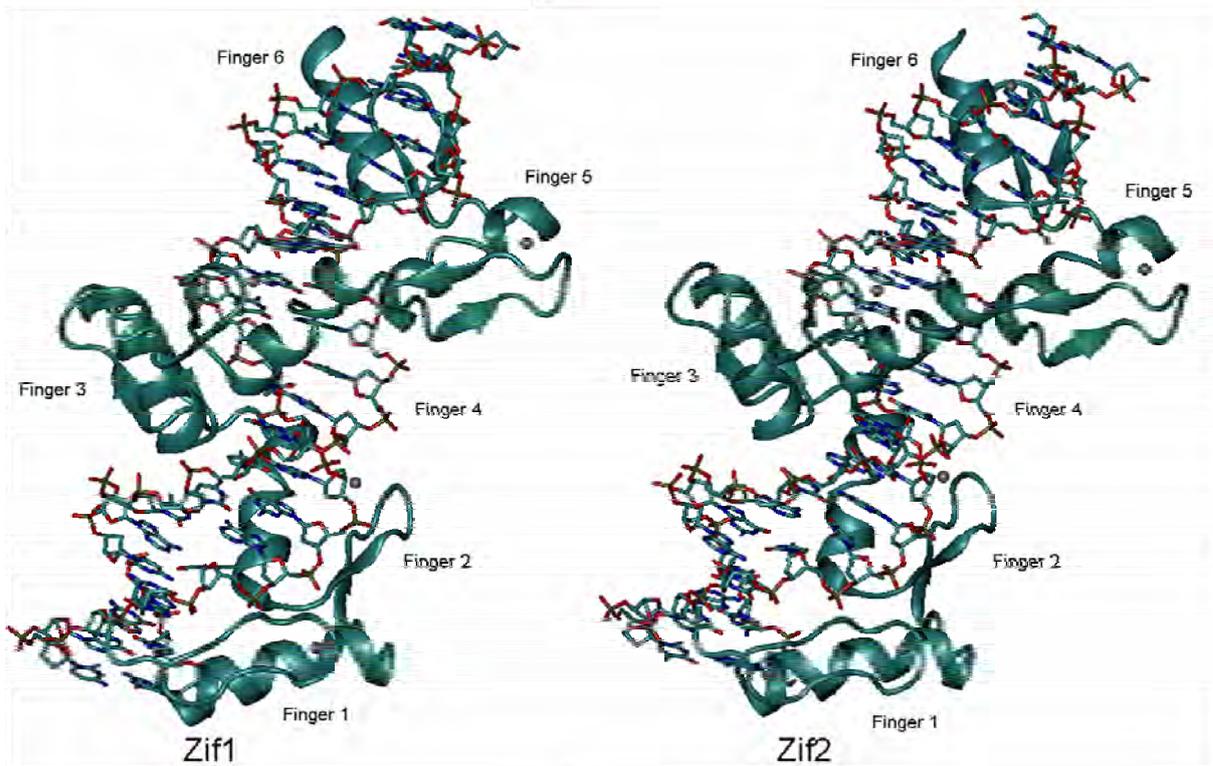
ใหญ่กว่าและมีประจุมากกว่ากลัยซีน ซึ่งอาจมีผลต่อการเข้าจับสายดีเอ็นเอตรงข้ามได้เพิ่มขึ้นหากเปลี่ยนเป็นฮิสทีดิน หรือในตำแหน่ง +5 ของโมดูล 4 และ 5

ตาราง 3.1 ผลวิเคราะห์ชนิดของกรดอะมิโนในตำแหน่ง -1 ถึง +6 บน recognition helix จากทั้ง 6 โมดูล ที่ทำนายได้จากฐานข้อมูล ZF Tools และ ZIFDB

	Position -1	Position +1	Position +2	Position +3	Position +4	Position +5	Position +6
Finger 1 5'-GGG-3' RSDKLVR	R (87%) K (13%)	S (28%) R (18%) G (13%) K/N/Q (8%) I/V (5%) A/D/M (3%)	D (54%) E (26%) S (5%) A/H/L/P/T/W (3%)	H (62%) K (23%) G (8%) R (5%) N (3%)	L (100%)	V (23%) T (21%) G (18%) A (8%) H/N/Q/R/S (5%) M/P (3%)	R (64%) N (10%) P (5%) A/D/G/H/K/S/ T/V (3%)
Finger 2 5'-GAT-3' TSGNLVR	T (33%) V (26%) L (13%) I (10%) Q (8%) S (5%) E/M (3%)	S (28%) K (18%) R (15%) G (13%) N (8%) A/Q (5%) L/M/P (3%)	H (28%) G/Q (23%) S (13%) E (8%) A/N (3%)	N (77%) R (10%) A/T (5%) H (3%)	L (97%) R (3%)	A/T/V (18%) K (13%) D/S (8%) G/Q (5%) H/M/R (3%)	R (77%) V (18%) K/T (3%)
Finger 3 5'-CTG-3' RNDALTE	R (100%)	N (50%) S (50%)	D (100%)	A (100%)	L (100%)	T (50%) R (50%)	E (100%)
Finger 4 5'-GTA-3' QSSSLVR	Q (100%)	S (61%) G (11%) A/D/P/Q/R (6%)	S (61%) G (17%) T (11%) A/Q (6%)	S (61%) A (28%) T (11%)	L (100%)	T (33%) V (22%) A (17%) I/L/Q/R/S (11%)	R (100%)
Finger 5 5'-GCA-3' QSGDLRR	Q (70%) R (13%) N (6%) L (4%) A (2%) H (2%) S (2%) T (2%)	S (26%) G (21%) R (11%) K (9%) A (8%) N/P/T (6%) Q (4%) D/M (2%)	G (32%) E/N (13%) Q/T (11%) A (9%) V (4%) C/H/S (2%)	T (43%) D (23%) E (13%) Q/S (6%) V (4%) G/I/M (2%)	L (100%)	T (25%) K (21%) R (11%) N/V (9%) G/Q (6%) D (4%) A/H/L/M/S (2%)	R (100%)
Finger 6 5'-CTA-3' QNSTLTE	Q (100%)	N (100%)	S (100%)	T (100%)	L (100%)	T (100%)	E (100%)

จากข้อมูลการเปรียบเทียบผลการทำนายชนิดของกรดอะมิโนจากฐานข้อมูลข้างต้น พบว่า โปรแกรม ZF Tools ทำนายให้ชนิดของกรดอะมิโนส่วนใหญ่เป็นชนิดที่พบมากที่สุดในระดับที่ 1 และ 2 จากฐานข้อมูลโดยเฉพาะในตำแหน่ง -1 และ +6 ที่มีความสำคัญในการเข้าจับกับเบสเป้าหมายในตำแหน่ง 3' และ 5' ตามลำดับ ส่วนกรดอะมิโนตำแหน่ง +3 ซึ่งมีผลต่อการเข้าจับกับเบสคู่กลางจาก 3 คู่เบสเป้าหมายในโมดูลที่ 1 และ 5 และกรดอะมิโนตำแหน่ง +2 ซึ่งมีผลต่อการเข้าจับกับเบสในสายดีเอ็นเอตรงข้าม โปรแกรม ZF Tools ทำนายให้ผลชนิดของกรดอะมิโนที่มีความถี่เป็นอันดับที่ 2 จากฐานข้อมูล ส่วนกรดอะมิโนตำแหน่ง +1 และ +5 เป็นตำแหน่งที่มีความแปรปรวนของชนิดกรดอะมิโนสูงที่สุด แต่ตำแหน่งเหล่านี้มีผลต่อการเข้าจับกับสายดีเอ็นเอโดยตรงน้อย แต่อาจจะมีผลต่อการจัดเรียงตัวของ recognition helix ก็เป็นไปได้ ซึ่งข้อมูลเหล่านี้อาจมีประโยชน์ในการใช้ประกอบการทดลองปรับเปลี่ยนชนิดของกรดอะมิโนบน recognition helix เพื่อเป็นการเพิ่มความจำเพาะหรือ เพิ่มความแรงในการยึดจับของโปรตีน ZF ได้

หลังจากการตรวจสอบลำดับและชนิดของกรดอะมิโนของโปรตีน ZF ที่ทำนายได้จากโปรแกรม ZF Tools จึงสร้างแบบจำลองโมเลกุลระหว่างโปรตีน ZF กับลำดับเบสเป้าหมายชนิดที่จำเพาะและไม่จำเพาะได้แก่ แบบจำลอง Zif1 และ Zif2 ตามลำดับ ซึ่งแบบจำลองทั้งสองเป็นการปรับเปลี่ยนโครงสร้างจากโครงร่างหลักต้นแบบของโปรตีน Aart ที่จับกับสายดีเอ็นเอที่มีอยู่ในฐานข้อมูล PDB โดยทั้ง Zif1 และ Zif2 มีโปรตีน ZF จำนวน 6 โมดูล คือ ZFPv1 เช่นกัน แต่มีดีเอ็นเอเป้าหมายแตกต่างกันอย่างชัดเจน คือ Zif1 มีสายดีเอ็นเอเป้าหมายเป็น 5'-CTA GCA GTA CTG GAT GGG-3' ซึ่งเป็นลำดับเบสในส่วน 2LTR ที่สามารถเกิดสารประกอบที่จำเพาะ (specific complex) กับ ZFPv1 ส่วน Zif2 มีลำดับเบสเป็น 5'-TGC TAG CGT ATC ATC TAG-3' ซึ่งเป็นลำดับเบสที่ไม่จำเพาะสำหรับ ZFPv1 และอาจเกิดสารประกอบที่ไม่จำเพาะ (non-specific complex) ขึ้นได้ หลังจากการปรับเปลี่ยนโครงสร้างได้ตามเป้าหมายแล้วจึงจัดเตรียมระบบและทำการลดพลังงานโมเลกุลลงสำหรับแบบจำลอง Zif1 และ Zif2 และสามารถคำนวณค่า root mean square deviation (RMSD) ระหว่างแบบจำลองทั้งสองที่ทำการลดพลังงานภายในโมเลกุลแล้ว ได้เท่ากับ 0.51 และ 0.94 Å ของโปรตีนและดีเอ็นเอในทั้งสองระบบ ตามลำดับ แสดงว่าการจัดเรียงตัวในแบบจำลองทั้งสองมีความคล้ายคลึงกันมากทั้งในส่วน of โปรตีน ZFPv1 และสายดีเอ็นเอ (รูป 3.1)



รูป 3.1 โครงสร้างโมเลกุลเปรียบเทียบระหว่างชุดแบบจำลอง Zif1 และ Zif2

ข้อมูลโครงสร้างความสัมพันธ์ของตำแหน่งของกรดอะมิโนกับการจดจำลำดับเบสบนสายดีเอ็นเอของโปรตีน ZF ต้นแบบ Zif268 พบว่าตำแหน่งกรดอะมิโนที่ -1 +3 และ +6 มีความสำคัญในการเข้าจับกับดีเอ็นเอเป้าหมาย ดังนั้นการพิจารณาความแตกต่างระหว่างแบบจำลอง 2 ชุดนี้เพื่อทำความเข้าใจความจำเพาะของโปรตีน ZFPv1 จึงให้ความสำคัญในการพิจารณาในส่วน recognition helix เป็นอันดับแรก โดยการวิเคราะห์โครงสร้างแบบจำลองโมเลกุลทั้งสองที่ได้หลังจากการทำการลดพลังงานแล้วเปรียบเทียบกัน ซึ่งค่าระยะห่างระหว่าง hydrogen bond (H-bond) donor และ H-bond acceptor ของอะตอมที่เกี่ยวข้อง (ระบุในวงเล็บท้ายลำดับกรดอะมิโนหรือเบส) และตำแหน่งของกรดอะมิโนบน recognition helix (ระบุในวงเล็บกำกับ) แสดงไว้ในตาราง 3.2 และโครงสร้างโมเลกุลในแต่ละโมดูลแสดงการเกิดพันธะไฮโดรเจนทั้งหมดที่ระบุไว้ในตาราง 3.2 แสดงไว้ในรูป 3.2 ประกอบ

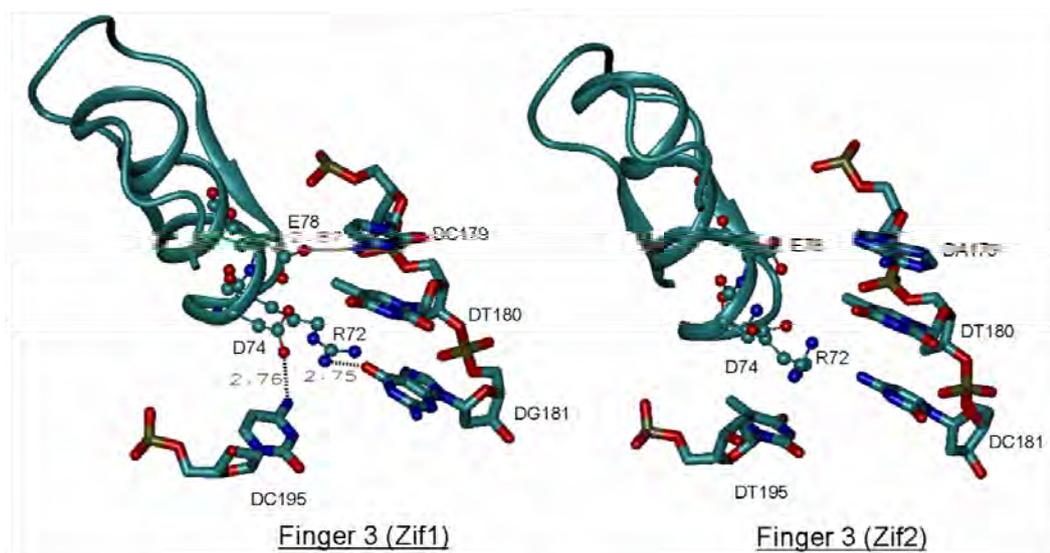
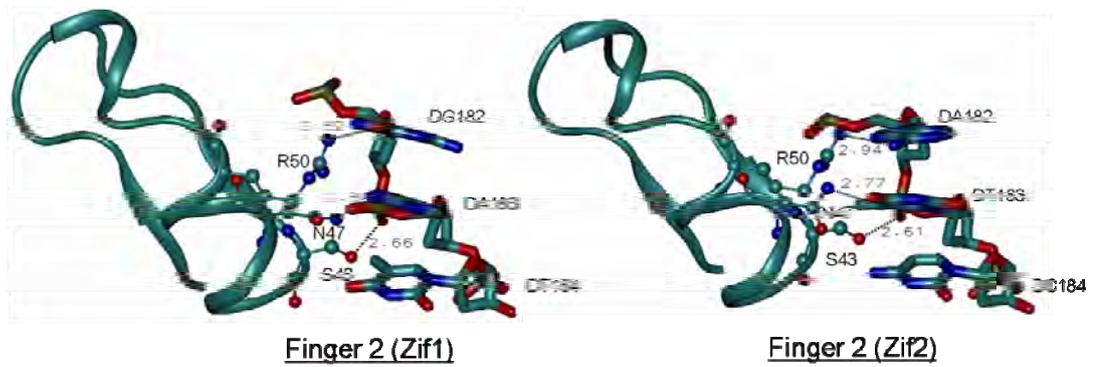
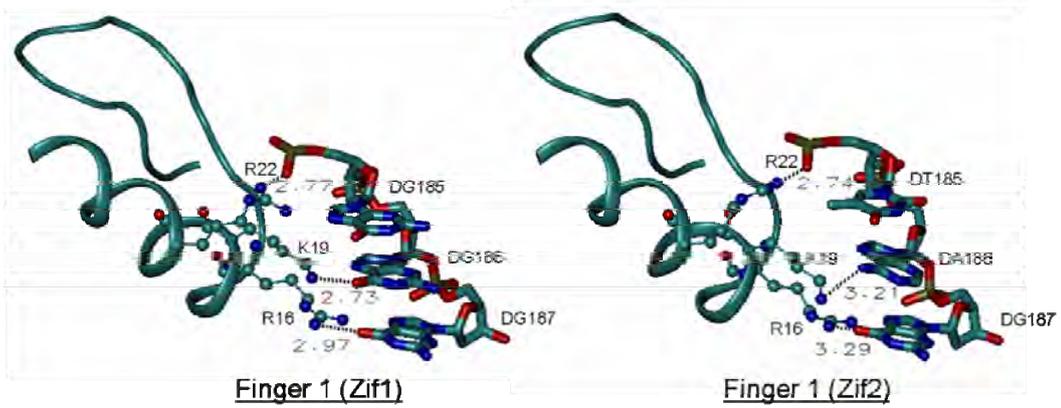
จากการวิเคราะห์โครงสร้างโมเลกุลการเกิดแรงยึดเหนี่ยวระหว่างโปรตีน ZFPv1 และสายดีเอ็นเอเป้าหมายทั้งสองในชุดแบบจำลอง Zif1 และ Zif2 พบว่าส่วนสำคัญในการทำให้เกิดความจำเพาะในการเข้าจับส่วนหนึ่งเกิดจากแรงประเภทไฟฟ้าสถิตเป็นหลัก คือ พันธะไฮโดรเจนระหว่างกรดอะมิโนกับสาย

ดีเอ็นเอโดยไม่มีโมเลกุลอื่น เช่น น้ำ มาเกี่ยวข้อง หากวิเคราะห์การเกิดพันธะในแต่ละโมดูล พบว่า อะตอมที่เกี่ยวข้องในการเกิดพันธะไฮโดรเจนนั้น จะมาจากกรดอะมิโนในตำแหน่ง -1 +3 และ +6 ของ recognition helix เป็นส่วนใหญ่ ซึ่งเป็นข้อมูลที่ทราบโดยทั่วไปอยู่แล้ว นอกจากนั้นแล้ว จะเห็นว่ามีการจัดอะมิโนในตำแหน่ง +2 ในโมดูลที่ 3 ที่พบว่าเกิดพันธะกับเบสในสายดีเอ็นเอตรงข้ามได้ตามที่เคยมีการรายงานไว้แล้ว แต่ที่สำคัญประการหนึ่งที่พบจากข้อมูล คือ กรดอะมิโนเซอรีนที่อยู่ในตำแหน่ง -2 ของ recognition helix ในโมดูลที่ 2 4 5 และ 6 มีส่วนสำคัญในการเพิ่มความแข็งแรงในการเข้าจับสายดีเอ็นเอ แต่อย่างไรก็ตามเป็นการเกิดอันตรกิริยาระหว่างหมู่ไฮดรอกซีของกรดอะมิโนเซอรีนกับหมู่ฟอสเฟต (อะตอม O2P) บนโครงสร้างแกนของสายดีเอ็นเอ (DNA backbone) ทั้งหมด ซึ่งอาจไม่มีความสัมพันธ์กับความสามารถในการจำแนกหรือความจำเพาะในการจดจำลำดับเบสที่จำเพาะ แต่ให้ผลในการเพิ่มความแข็งแรงในการเข้าจับระหว่างโปรตีน ZF กับลำดับเบสเป้าหมาย

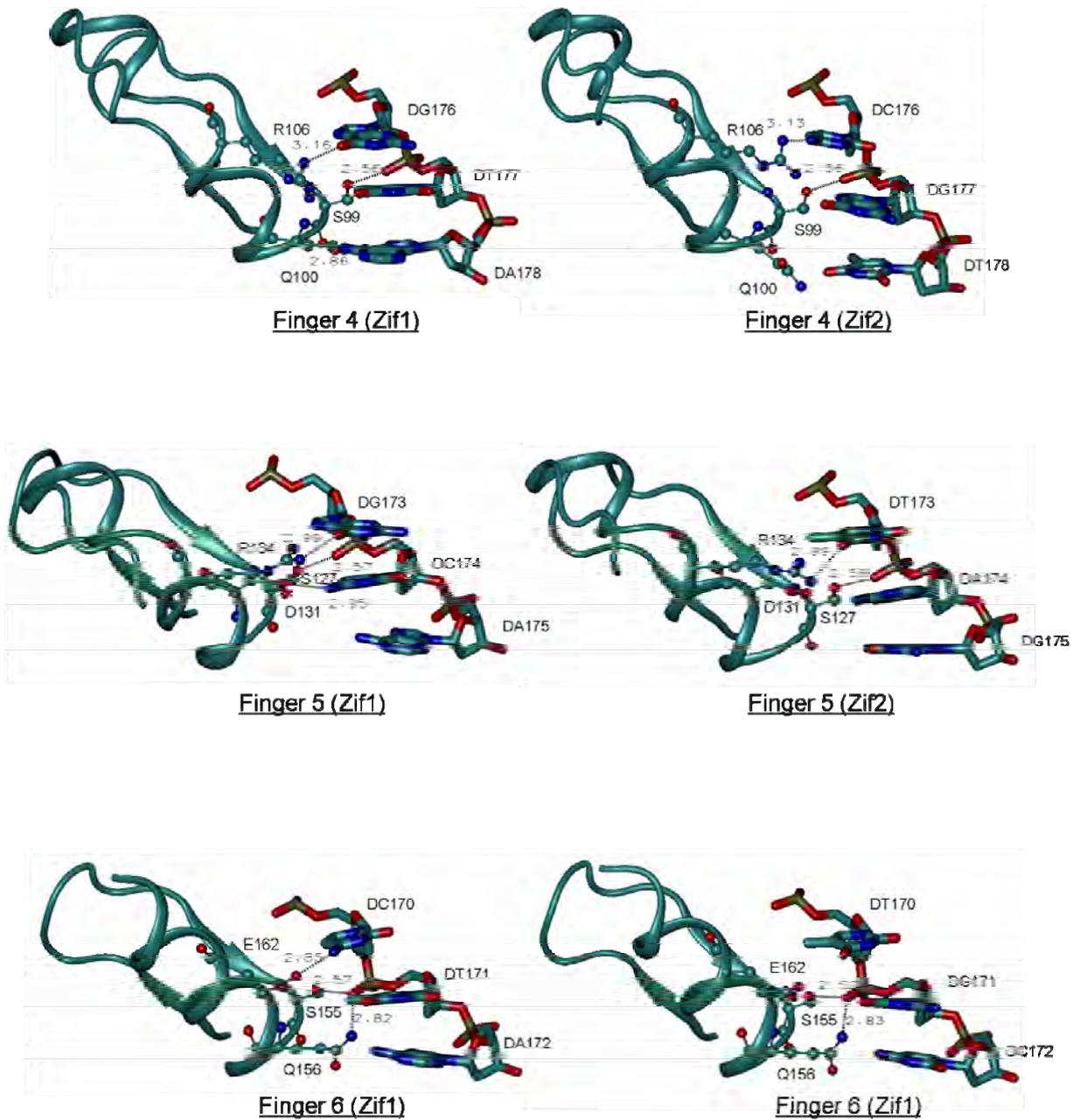
จากรูป 3.2 จะเห็นว่ารูปแบบการเข้าจับระหว่างโปรตีน ZFPv1 กับสายดีเอ็นเอทั้งสอง ไม่ได้มีความแตกต่างกันอย่างชัดเจน มีเพียงโมดูลที่ 3 ที่ใน Zif2 ไม่สามารถตรวจพบพันธะไฮโดรเจนใดๆ ในส่วน recognition helix ตำแหน่ง Arg72 (ตำแหน่ง -1) Asp74 (ตำแหน่ง +2) Glu78 (ตำแหน่ง +6) กับในส่วนของเบสเป้าหมาย ATC เลย อาจเนื่องจากการเกิดแรงผลักระหว่างประจุบวกของหมู่ควินิดีนของอาร์จินีน (Arg72) กับหมู่เอมีนของไซโตซีน (DC181) การที่หมู่เมทิลของไทมีนในตำแหน่ง DT195 อาจมีผลรบกวนหรือบดบัง (steric effect) การเกิดพันธะไฮโดรเจนได้ รวมทั้งหมู่เอมีนของอะดีนีน DA179 อยู่ในตำแหน่งที่ไม่เหมาะสมในการเกิดพันธะไฮโดรเจนกับกรดกลูตามิก Glu78 ซึ่งความแตกต่างในโมดูลที่ 3 นี้ อาจมีส่วนสำคัญในการจำแนกความแตกต่างระหว่างดีเอ็นเอที่เป็นเป้าหมายและที่ไม่ใช่เป้าหมายได้

ตาราง 3.2 ระยะห่าง (Å) ระหว่างอะตอมของ H-bond donor และ H-bond acceptor ในโปรตีน ZFPv1 และสายดีเอ็นเอในชุดแบบจำลอง Zif1 และ Zif2

	ZFPv1 residues	DNA residues	Distance (Å)		ZFPv1 residues	DNA residues	Distance (Å)
Finger 1	Arg16(NH)	Zif1:DG187(O6)	2.97	Finger 4	Ser99(OH)	Zif1:DT177(O2P)	2.56
	[ตำแหน่ง -1]	Zif2:DG187(O6)	3.29		[ตำแหน่ง -2]	Zif2:DT177(O2P)	2.56
	Lys19(NZ)	Zif1:DG186(O6)	2.73		Gln100(O)	Zif1:DA178(N6)	2.86
	[ตำแหน่ง +3]	Zif2:DA186(N6)	3.21		[ตำแหน่ง -1]	Zif2:DT178	-
	Arg22(NH)	Zif1:DG185(O2P)	2.77		Arg106(NH)	Zif1:DG176(O6)	3.16
[ตำแหน่ง +6]	Zif2:DT185(O2P)	2.74	[ตำแหน่ง +6]	Zif2:DC176(N4)	3.13		
Finger 2	Ser43(OH)	Zif1:DA183(O1P)	2.66	Finger 5	Ser127(OH)	Zif1:DC174(O2P)	2.57
	[ตำแหน่ง -2]	Zif2:DT183(O1P)	2.61		[ตำแหน่ง -2]	Zif2:DA174(O2P)	2.58
	Asn47(ND)	Zif1:DA183(O2P)	2.80		Asp131(OH)	Zif1:DC174(N4)	2.95
	[ตำแหน่ง +3]	Zif2:DT183(O2P)	2.77		[ตำแหน่ง +3]	Zif2:DA174	-
	Arg50(NH)	Zif1:DG182(O6)	2.82		Arg134(NH)	Zif1:DG173(O6)	2.99
[ตำแหน่ง +6]	Zif2:DA182(N6)	2.94	[ตำแหน่ง +6]	Zif2:DT173(O4)	2.99		
Finger 3	Arg72(NH)	Zif1:DG181(O6)	2.75	Finger 6	Ser155(OH)	Zif1:DT171(O2P)	2.57
	[ตำแหน่ง -1]	Zif2:DC181	-		[ตำแหน่ง -2]	Zif2:DG171(O2P)	2.52
	Asp74(OH)	Zif1:DC195(N4)	2.76		Gln156(NH)	Zif1:DT171(O2P)	2.82
	[ตำแหน่ง +2]	Zif2:DT195	-		[ตำแหน่ง -1]	Zif2:DG171(O2P)	2.83
	Glu78(OH)	Zif1:DC179(N4)	2.87		Glu162(OH)	Zif1:DC170(N4)	2.85
[ตำแหน่ง +6]	Zif2:DA179	-	[ตำแหน่ง +6]	Zif2:DT170	-		



รูป 3.2 เปรียบเทียบพันธะไฮโดรเจน (< 3.5 Å) ที่อาจเป็นไปได้ในแต่ละโมดูลของโปรตีน ZFPv1 ในแบบจำลอง Zif1 และ Zif2



รูป 3.2 (ต่อ) เปรียบเทียบพันธะไฮโดรเจน (< 3.5 Å) ที่อาจเป็นไปได้ในแต่ละโมดูลของโปรตีน ZFPv1 ในแบบจำลอง Zif1 และ Zif2

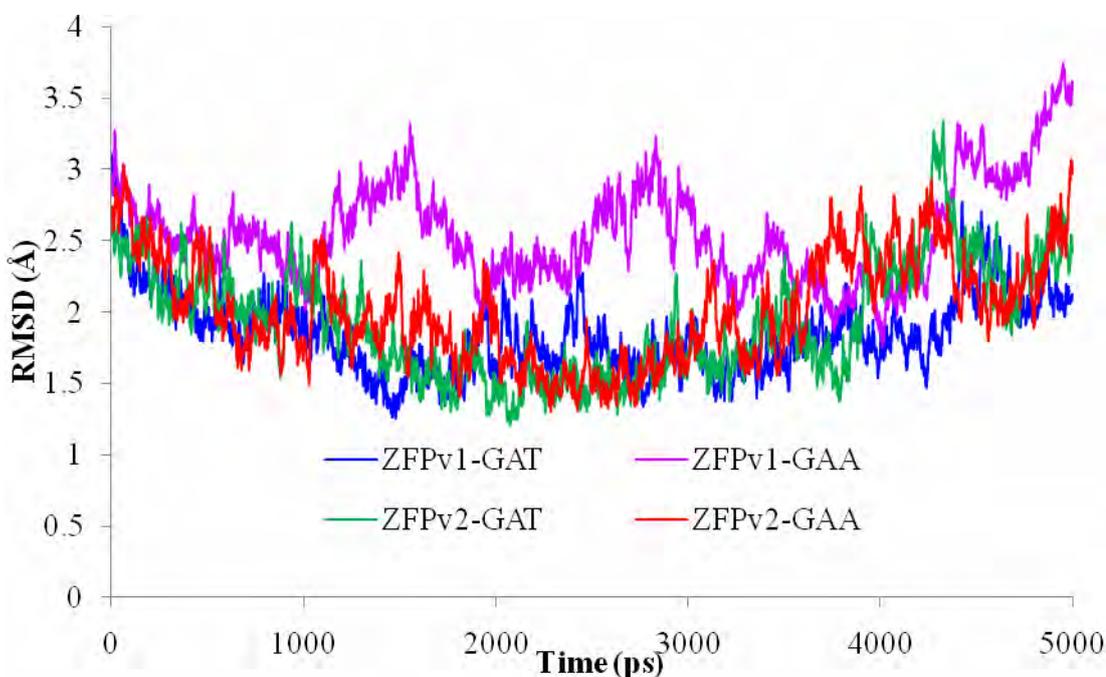
หากพิจารณารูปแบบการเกิดพันธะไฮโดรเจนเฉพาะในส่วนที่เกิดอันตรกิริยาระหว่างกรดอะมิโนกับเบสเป้าหมาย 2LTR เท่านั้น โดยไม่สนใจพันธะที่เกิดกับหมู่ฟอสเฟตของสายดีเอ็นเอ พบว่าในแต่ละโมดูลของแบบจำลอง Zif1 จะมีความยาวพันธะไฮโดรเจนที่สั้นกว่าในแบบจำลอง Zif2 ซึ่งแสดงให้เห็นถึงความแข็งแรงของพันธะที่เกิดขึ้นในสารประกอบที่จำเพาะระหว่างโปรตีน ZFPv1 กับ 2LTR ได้ดีกว่าลำดับเบสที่ไม่จำเพาะใน Zif2 ตัวอย่างเช่น ในโมดูลที่ 1 กรดอะมิโน Arg16 และ Lys19 เกิดพันธะกับ DG187 และ DG186 ใน Zif1 ได้ดีกว่า DG187 และ DA186 ใน Zif2 ส่วนโมดูลที่ 2 Arg50 เกิดพันธะไฮโดรเจนกับ DG182 (Zif1) ได้ดีกว่า DA182 (Zif2) ในโมดูลที่ 3 มีความแตกต่างกันอย่างชัดเจนดังที่กล่าวแล้วข้างต้น โมดูลที่ 4 5 และ 6 พบว่ามีลักษณะคล้ายคลึงกัน คือ Gln100 ในโมดูลที่ 4 สามารถเกิดพันธะกับ DA178 (Zif1) ได้ แต่ไม่พบพันธะใน Zif2 ส่วน Asp131 ในโมดูลที่ 5 และ Glu162 ในโมดูลที่ 6 สามารถเกิดพันธะกับเบสเป้าหมายใน Zif1 ได้ แต่ไม่เกิดพันธะกับเบสที่ไม่ใช่เป้าหมายใน Zif2

หลังจากการวิเคราะห์โครงสร้างข้างต้น จึงได้ทำการคำนวณพลังงานการยึดจับระหว่างโปรตีน ZFPv1 และสายดีเอ็นเอในแบบจำลอง Zif1 และ Zif2 ด้วยวิธี MM/GBSA จากการทดลองการจำลองพลวัตและเปรียบเทียบค่าความแข็งแรงของพลังงานในการยึดจับ ($\Delta G_{binding}$) ระหว่างโปรตีน ZFPv1 กับสายดีเอ็นเอในแบบจำลอง Zif1 และ Zif2 พบว่าความสามารถในการเข้าจับของโปรตีน ZFPv1 กับดีเอ็นเอเป้าหมาย 2LTR ใน Zif1 มีความชอบจับกันมากกว่าการจับกับดีเอ็นเอที่ไม่ใช่เป้าหมายใน Zif2 สูงมากถึง 35.0 kcal/mol (ค่า $\Delta\Delta G_{binding}$) มีปัจจัยสำคัญที่เกี่ยวข้องที่ค่า ΔE_{es} ดังนั้นจากข้อมูลพื้นฐานนี้อาจเสนอได้ว่าโปรตีน ZF ที่ออกแบบได้นี้มีความสามารถในการจำแนกลำดับเบสเป้าหมายและที่ไม่ใช่เป้าหมายได้ด้วยความชอบจับที่แตกต่างกันได้ด้วยพลังงานการยึดจับที่แตกต่างกันอย่างมาก เนื่องจากอันตรกิริยาชนิดไฟฟ้าสถิตย์ (electrostatic interactions, E_{es}) เช่น แรงระหว่างประจุบวกและลบ แรงระหว่างประจุกับสภาพขั้ว และพันธะไฮโดรเจน เป็นสำคัญ

ตาราง 3.3 การเปรียบเทียบความแตกต่างของพลังงานในการยึดจับระหว่างโปรตีน ZFPv1 กับสายดีเอ็นเอในแบบจำลอง Zif1 และ Zif2

sequence	$\Delta\Delta E_{vdw}$ (kcal mol ⁻¹)	$\Delta\Delta E_{es}$ (kcal mol ⁻¹)	$\Delta\Delta G_{pol}$ (kcal mol ⁻¹)	$\Delta\Delta G_{nonpol}$ (kcal mol ⁻¹)	$\Delta\Delta G_{binding}$ (kcal mol ⁻¹)
Zif1-Zif2	-3.1 ±1.7	-255.9 ±9.0	226.3 ±8.2	-2.3 ±0.0	-35.0 ±0.5

ในส่วนการศึกษาต่อมา ได้สร้างแบบจำลองโครงสร้างโมเลกุลสามมิติอีกจำนวน 4 ชุด ได้แก่ ZFPv1-GAT ZFPv1-GAA ZFPv2-GAT และ ZFPv2-GAA ซึ่งลำดับเบสเป้าหมาย GAT (5'-CTA GCA GTA CTG GAT GGG-3') จัดเป็นเป้าหมายสำหรับโปรตีน ZFPv1 และลำดับเบสเป้าหมาย GAA (5'-CTA GCA GTA CTG GAA GGG-3') จัดเป็นเป้าหมายสำหรับโปรตีน ZFPv2 ดังนั้นจะเห็นว่าชุดแบบจำลองที่เป็นสารประกอบจำเพาะ คือ ZFPv1-GAT และ ZFPv2-GAA ควรจะมีความแข็งแรงในการยึดจับสูงกว่าส่วนแบบจำลองที่เป็นสารประกอบที่ไม่จำเพาะ คือ ZFPv1-GAA และ ZFPv2-GAT โดยที่ความไม่จำเพาะนี้คาดว่าจะเกิดจากความแตกต่างของลำดับเบสเพียง 1 ตำแหน่ง หลังจากการปรับเปลี่ยนโครงสร้าง การจัดเตรียมระบบ การลดพลังงานศักย์ของโมเลกุล การทำให้ระบบสมดุล และการจำลองพลวัตเป็นเวลา 5 ns แล้วจึงทำการคำนวณหาค่า RMSD ของโครงสร้างทุกๆ 2 ps จากช่วงเวลา 5 ns เทียบกับโครงสร้างเฉลี่ยที่ได้จากการจำลองพลวัตแต่ละชุดแบบจำลอง (รูป 3.3) พบว่าแบบจำลอง ZFPv1-GAT ZFPv2-GAT และ ZFPv2-GAA มีค่า RMSD อยู่ในช่วงประมาณ 1.5 - 3.0 Å ส่วน ZFPv1-GAA พบว่ามีค่า RMSD สูงที่สุดอยู่ในช่วง 2.0 - 3.5 Å แต่อย่างไรก็ตาม สามารถยอมรับได้ว่าแบบจำลองทั้ง 4 ชุดอยู่ในสถานะที่เสถียร ไม่มีการแตกของพันธะในโครงสร้าง สามารถใช้เป็นตัวอย่างในการวิเคราะห์พลังงานในขั้นตอนต่อไปได้



รูป 3.3 ค่า RMSD ของโครงสร้างทุกๆ 2 ps จากช่วงเวลา 5 ns เทียบกับโครงสร้างเฉลี่ยที่ได้จากการจำลองพลวัตแต่ละชุดแบบจำลอง ZFPv1-GAT ZFPv1-GAA ZFPv2-GAT และ ZFPv2-GAA

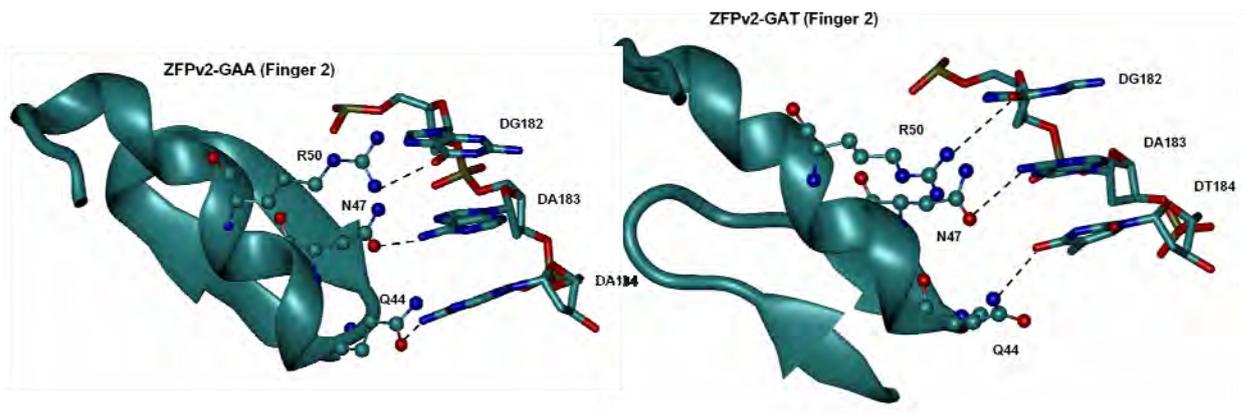
จากข้อมูลการจำลองพลวัตของแต่ละชุดแบบจำลอง และทำการวิเคราะห์เชิงพลังงานการยึดจับของโมเลกุลระหว่างโปรตีนและสายดีเอ็นเอด้วยวิธี MM/GBSA และ MM/PBSA ระหว่างแบบจำลองทั้ง 4 ชุด พบว่า สำหรับแบบจำลองโปรตีน ZFPv1 กับลำดับเบส GAT ได้ผลการทดลองเป็นไปตามคาดหมาย คือ ZFPv1 มีความชอบเข้าจับ GAT ได้มากกว่า GAA ด้วยพลังงานที่แข็งแกร่งกว่าประมาณ 8.5 kcal/mol ด้วยวิธี MM/GBSA และประมาณ 13.4 kcal/mol ด้วยวิธี MM/PBSA ซึ่งเป็นผลเนื่องมาจากความแตกต่างกันเพียง 1 คู่เบสจาก DT184 ใน GAT ไปเป็น DA184 ใน GAA กรดอะมิโน Thr44 ในตำแหน่ง -1 เป็นตำแหน่งที่รับผิดชอบในการจำแนกความแตกต่างระหว่างเบส DT184 และ DA184 แต่อย่างไรก็ตามจากการวิเคราะห์โครงสร้างของ ZFPv1-GAT (รูป 3.2 Finger 2 ของ Zif1) และ ZFPv1-GAA จากการจำลองพลวัต 5 ns ไม่พบการเกิดอันตรกิริยาใดๆ ต่อกัน และจากผลการคำนวณ RMSD ของระบบ ZFPv1-GAA มีค่าสูงกว่า ZFPv1-GAT สามารถบอกได้ว่าแต่ละอะตอมภายในโมเลกุลของโปรตีน ZFPv1 และสายดีเอ็นเอ GAA มีอิสระในการเคลื่อนที่ภายในช่วงเวลาที่ศึกษามากกว่า กล่าวคือมีการเกิดอันตรกิริยาระหว่างโมเลกุลโปรตีน ZFPv1 และสายดีเอ็นเอ GAA น้อยกว่า ZFPv1-GAT ซึ่งสามารถยืนยันได้จากผลการคำนวณค่าพลังงานการยึดจับของโมเลกุลระหว่าง ZFPv1-GAA มีความแข็งแรงน้อยกว่า ZFPv1-GAT เช่นกัน

ตาราง 3.4 เปรียบเทียบค่าพลังงานการยึดจับของโมเลกุลระหว่างโปรตีนและสายดีเอ็นเอด้วยวิธี MM/GBSA และ MM/PBSA

	$\Delta G_{binding}$ [MM/GBSA] (kcal mol ⁻¹)	$\Delta\Delta G_{binding}$ [MM/GBSA] (kcal mol ⁻¹)	$\Delta G_{binding}$ [MM/PBSA] (kcal mol ⁻¹)	$\Delta\Delta G_{binding}$ [MM/PBSA] (kcal mol ⁻¹)
ZFPv1-GAA	-318.4 ±30.0	8.5 ±35.6	-315.9 ±26.3	13.4 ±36.1
ZFPv1-GAT	-326.9 ±19.1		-329.3 ±24.8	
ZFPv2-GAA	-340.4 ±20.9	14.2 ±26.4	-347.8 ±30.4	19.4 ±38.5
ZFPv2-GAT	-354.6 ±16.1		-367.2 ±23.6	

กรณีแบบจำลองโปรตีน ZFPv2 ที่ออกแบบมาให้มีความจำเพาะต่อสายดีเอ็นเอ GAA มีการปรับเปลี่ยนชนิดของกรดอะมิโน 2 ตำแหน่ง (ดูรูป 2.3) คือ Thr44 ไปเป็น Gln44 และ Gly46 ไปเป็น Ser46 พบว่า ส่วนพันธะไฮโดรเจนระหว่าง Arg50 กับ DG182 และ Asn47 กับ DA183 มีลักษณะ

เช่นเดียวกับใน ZFPv1-GAA และ ZFPv1-GAT แต่สิ่งที่ต่างกันชัดเจน ได้แก่ Gln44 มีความชอบเข้าจับกับลำดับเบส DA184 หรือ DT184 ได้ดีกว่า Thr44 ใน ZFPv1 โดยอาศัยหมู่คาร์บอนิลของ Gln44 ในการเกิดพันธะไฮโดรเจนกับหมู่เอมีนของ DA184 (รูป 3.4 บน) และอาศัยหมู่เอมีนของ Gln44 ในการเกิดพันธะไฮโดรเจนกับหมู่คาร์บอนิลของ DT184 (รูป 3.4 ล่าง) ซึ่งเป็นสาเหตุหนึ่งที่ทำให้ค่าพลังงานการยึดจับระหว่างโปรตีน ZFPv2 กับสายดีเอ็นเอมีความแข็งแรงมากกว่า ZFPv1 ที่ตรวจไม่พบการเกิดพันธะไฮโดรเจนระหว่าง Thr44 กับลำดับเบส DA184 หรือ DT184 และการเปรียบเทียบพลังงานในการยึดจับของ ZFPv2 กับสายดีเอ็นเอ GAT มีความแข็งแรงมากกว่า GAA ประมาณ 14.2 kcal/mol ด้วยวิธี MM/GBSA ซึ่งให้ผลตรงข้ามกับที่คาดการณ์ไว้ รวมถึงการทดลองเปลี่ยนวิธีการคำนวณเป็น MM/PBSA ต่างก็ให้ข้อมูลที่คล้ายคลึงกัน คือ ประมาณ 19.4 kcal/mol ดังแสดงในตาราง 3.4 ซึ่งยังไม่สามารถอธิบายได้จากผลการคำนวณที่พบ ดังนั้นในการสรุปถึงความสามารถในจำแนกความแตกต่างหรือความจำเพาะของโปรตีน ZF ที่ออกแบบและสังเคราะห์ขึ้นนั้น ในทางทฤษฎียังไม่สามารถยืนยันความสามารถในการจำแนกได้ในกรณีที่ลำดับเบสมีความแตกต่างกันเพียงเล็กน้อยระหว่างลำดับเบสเป้าหมายและที่ไม่ใช่เป้าหมาย เช่น ในสายดีเอ็นเอ GAT และ GAA แต่ในกรณีที่ลำดับเบสแตกต่างกันอย่างชัดเจน เช่น ในแบบจำลอง Zif1 และ Zif2 พบว่าสามารถทำนายความแตกต่างได้อย่างชัดเจน โดยอาศัยวิธีคำนวณค่าพลังงานการยึดจับระหว่างโปรตีน ZF กับสายดีเอ็นเอ และพบว่าการคำนวณด้วยหลักการ MM/GBSA หรือ MM/PBSA ต่างให้ผลการคำนวณไปในแนวทางเดียวกัน แต่วิธี MM/GBSA ประหยัดเวลาในการคำนวณมากกว่า ดังนั้นในการทดลองต่อมาจึงคำนวณเฉพาะวิธี MM/GBSA



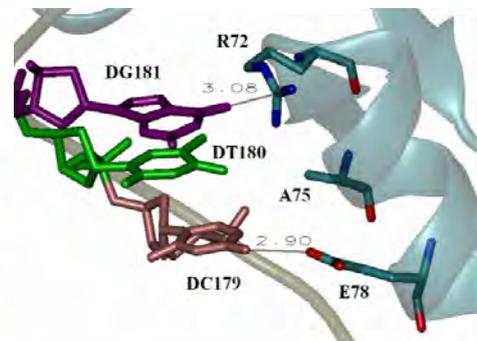
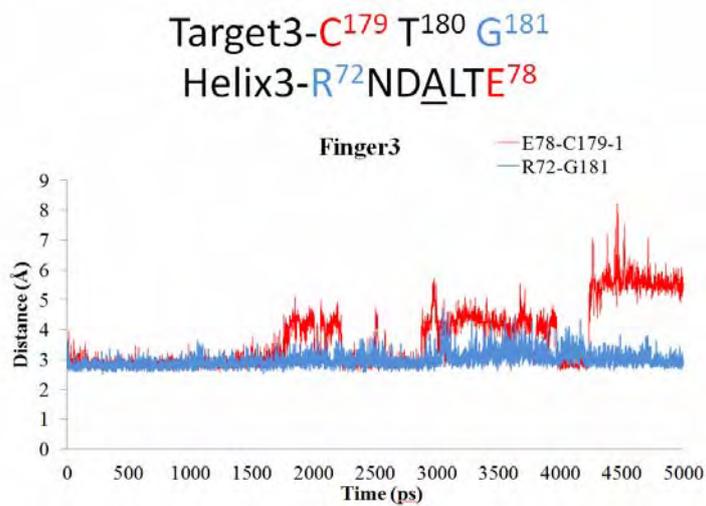
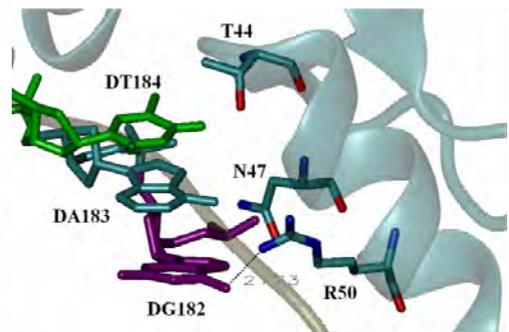
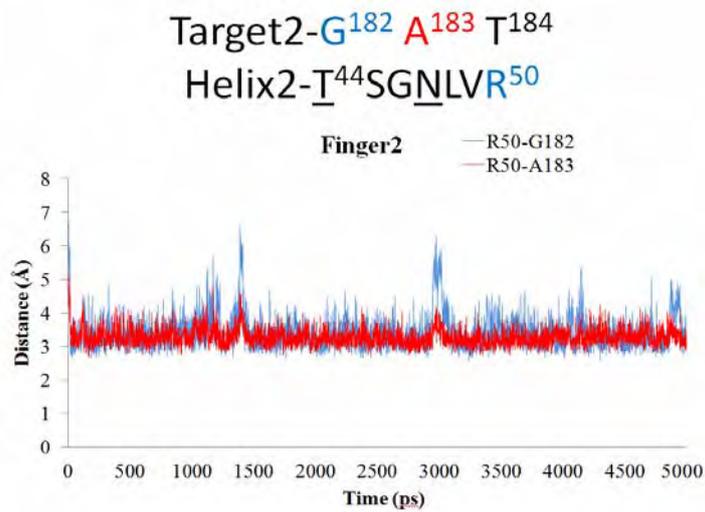
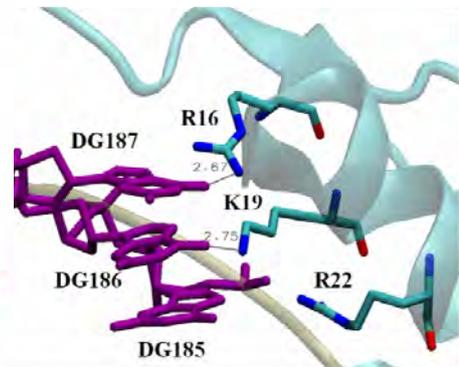
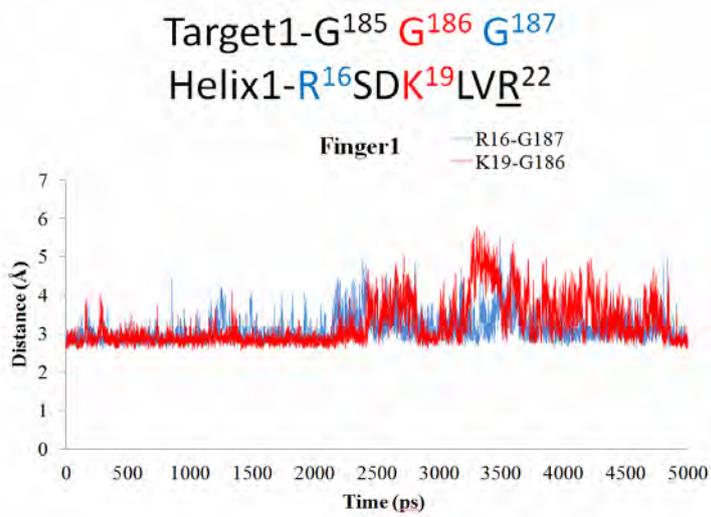
รูป 3.4 พันธะไฮโดรเจนที่พบได้ระหว่าง Arg50-DG182 Asn47-DA183 และ Gln44-DA183/DT183 จากการจำลองพลวัตของ ZFPv2-GAA และ ZFPv2-GAT

3.2 ความสามารถในการเข้าจับของโปรตีน ZFPv1

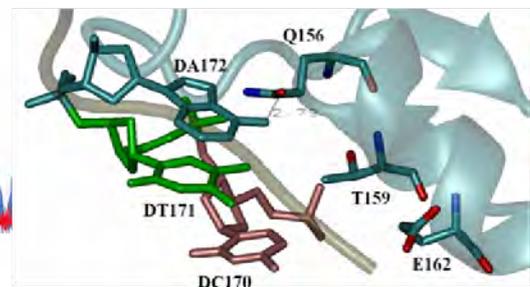
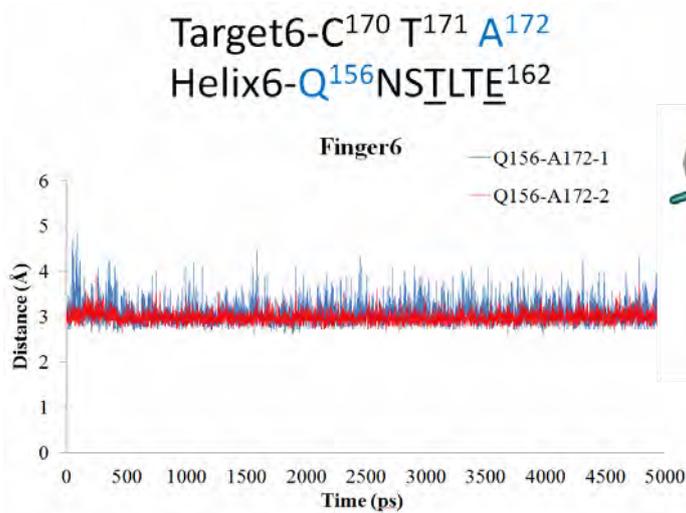
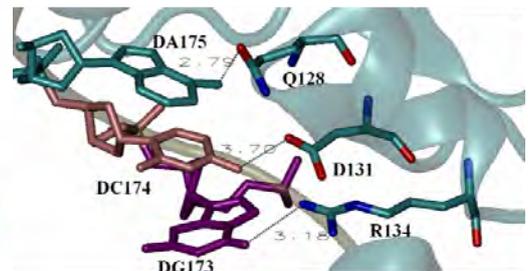
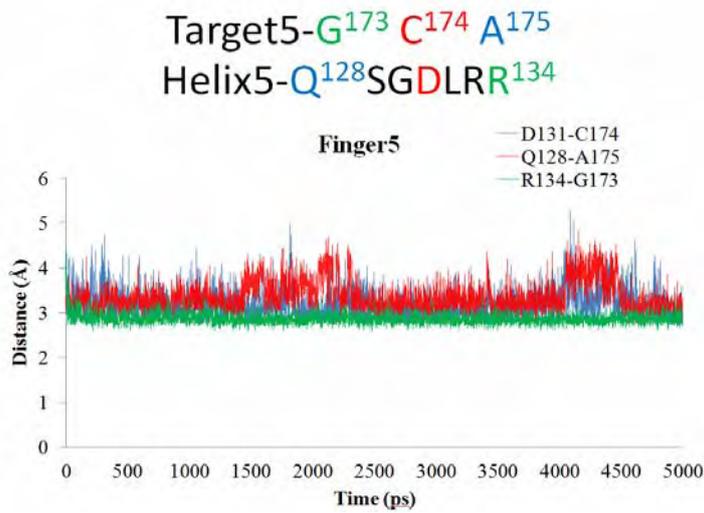
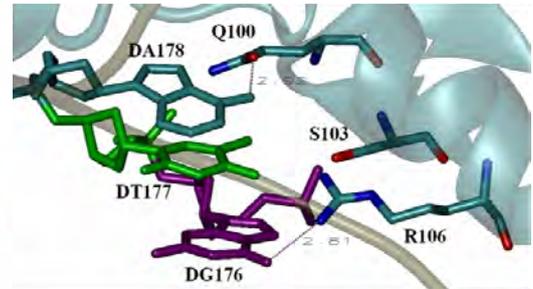
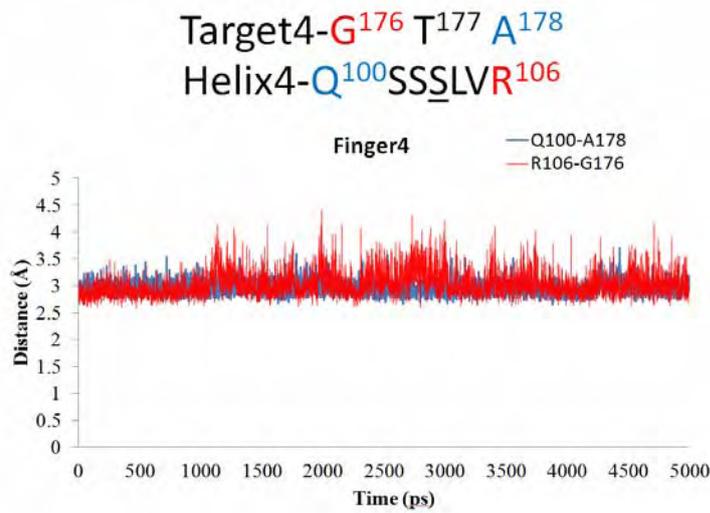
จากการจำลองพลวัตของแบบจำลอง ZFPv1-GAT ในข้อ 3.1 ซึ่งเป็นแบบจำลองที่จำเพาะระหว่างโปรตีน ZF ที่ออกแบบกับลำดับเบสเป้าหมาย 2LTR พบว่ามีความแข็งแรงในการเกิดเป็นสารประกอบระหว่างโปรตีนกับลำดับเบสเป้าหมายได้ดีกว่าลำดับเบสที่ไม่ใช่เป้าหมาย (ZFPv1-GAA) ด้วยพลังงานการยึดจับที่ต่างกันประมาณ 8.5 kcal/mol ซึ่งจัดว่าไม่มีความแตกต่างกันอย่างชัดเจนมากนัก จึงเริ่มทำการวิเคราะห์ความสามารถในการเกิดพันธะไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง -1 +3 และ +6 ของ recognition helix ทั้ง 6 โมดูลของโปรตีน ZFPv1 กับลำดับเบสเป้าหมาย 2LTR ในแบบจำลอง ZFPv1-GAT โดยคำนวณจากระยะห่างที่เป็นไปได้ระหว่างอะตอมที่มีสมบัติ hydrogen bond donor และ acceptor ในแต่ละโมดูล ดังแสดงในรูป 3.5

- โมดูล 1 พบการเกิดพันธะไฮโดรเจนระหว่างหมู่กวินิดีนของ Arg16 (ตำแหน่ง -1) กับหมู่คาร์บอนิลของ DG187 และหมู่เอมีนของ Lys19 (ตำแหน่ง +3) กับหมู่คาร์บอนิลของ DG186 ไม่พบการเกิดพันธะไฮโดรเจนในตำแหน่ง +6 ของ Arg22
- โมดูล 2 พบการเกิดพันธะไฮโดรเจนระหว่างหมู่กวินิดีนของ Arg50 (ตำแหน่ง +6) กับหมู่คาร์บอนิลของ DG182 และอาจเกิดพันธะไฮโดรเจนกับหมู่เอมีนของ DA183 ได้ด้วย ไม่พบการเกิดพันธะไฮโดรเจนในตำแหน่ง -1 ของ Thr44 และตำแหน่ง +3 ของ Asn47
- โมดูล 3 พบการเกิดพันธะไฮโดรเจนระหว่างหมู่กวินิดีนของ Arg72 (ตำแหน่ง -1) กับหมู่คาร์บอนิลของ DG181 ส่วนหมู่คาร์บอนิลของ Glu78 (ตำแหน่ง +6) อาจเกิดอันตรกิริยากับหมู่เอมีนของ DC179 ได้บ้าง แต่อาจเกิดแรงผลักรันได้ด้วย และไม่พบการเกิดพันธะไฮโดรเจนในตำแหน่ง +3 ของ Ala75
- โมดูล 4 พบการเกิดพันธะไฮโดรเจนระหว่างหมู่คาร์บอนิลของ Gln100 (ตำแหน่ง -1) กับหมู่เอมีนของ DA178 และหมู่กวินิดีนของ Arg106 (ตำแหน่ง +6) กับหมู่คาร์บอนิลของ DG176 ไม่พบการเกิดพันธะไฮโดรเจนในตำแหน่ง +3 ของ Ser103
- โมดูล 5 พบการเกิดพันธะไฮโดรเจนระหว่างหมู่คาร์บอนิลของ Gln128 (ตำแหน่ง -1) กับหมู่เอมีนของ DA175 หมู่คาร์บอนิลของ Asp131 (ตำแหน่ง +2) กับหมู่เอมีนของ DC174 และหมู่กวินิดีนของ Arg134 (ตำแหน่ง +6) กับหมู่คาร์บอนิลของ DG173

- โมดูล 6 พบการเกิดพันธะไฮโดรเจนระหว่างหมู่คาร์บอนิลของ Gln156 (ตำแหน่ง -1) กับหมู่เอมีนของ DA172 และไม่พบการเกิดพันธะไฮโดรเจนในตำแหน่ง +3 ของ Thr159 และตำแหน่ง +6 ของ Glu162

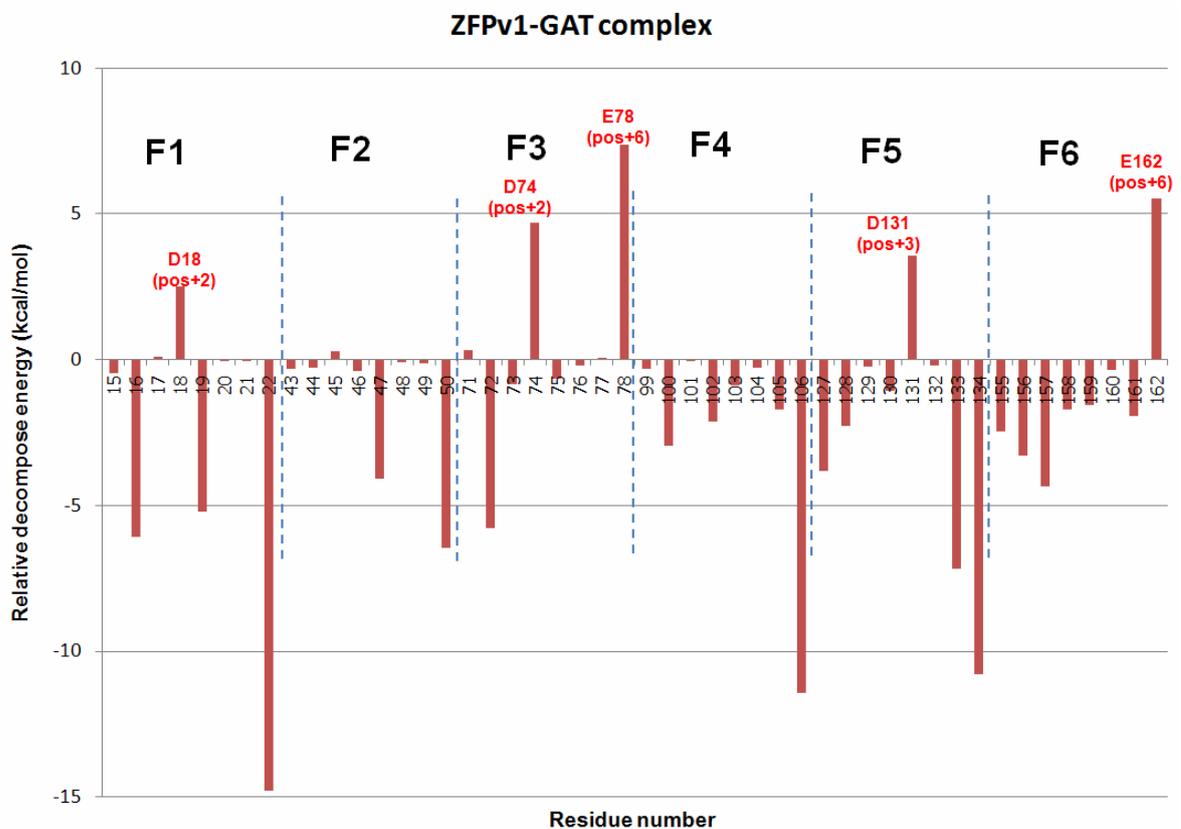


รูป 3.5 การเกิดพันธะไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง -1 +3 และ +6 ของ recognition helix ทั้ง 6 โมดูลในแบบจำลอง ZFPv1-GAT



รูป 3.5(ต่อ) การเกิดพันธะไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง -1 +3 และ +6
ของ recognition helix ทั้ง 6 โมดูลในแบบจำลอง ZFPv1-GAT

นอกจากการวิเคราะห์ระยะห่างระหว่างอะตอมที่สามารถเกิดพันธะไฮโดรเจนจากโครงสร้างโมเลกุลที่ได้จากการจำลองพลวัตข้างต้นแล้ว ต่อมาเพื่อพยายามปรับปรุงให้โปรตีนที่ออกแบบได้จาก ZF Tools มีความชอบจับกับลำดับเบสเป้าหมายมากขึ้น จึงได้ทำการประเมินพลังงานในการสลายพันธะยึดจับ (decompose energy) ของกรดอะมิโนแต่ละหน่วยของ recognition helix ทั้ง 6 โมดูลของโปรตีน ZFPv1 กับลำดับเบสเป้าหมาย 2LTR ว่ากรดอะมิโนในส่วน recognition helix ใดที่มีความสามารถในการยึดจับกับเป้าหมายยังไม่ดีพอ หรือมีแนวโน้มที่สามารถนำมาปรับเปลี่ยนในทางแบบจำลองแล้วน่าจะเพิ่มความสามารถในการยึดจับได้ดีขึ้น โดยไม่จำเป็นต้องทำการทดลองในหลอดทดลองจริง ซึ่งผลการวิเคราะห์แสดงไว้ในรูป 3.6



รูป 3.6 พลังงานการสลายพันธะยึดจับ (decompose energy) ระหว่างกรดอะมิโนในตำแหน่ง -2 ถึง +6 ของ recognition helix ทั้ง 6 โมดูลของโปรตีน ZFPv1 กับลำดับเบสเป้าหมาย 2LTR

ในแต่ละโมดูลแสดงพลังงานในการสลายพันธะการยึดจับระหว่างกรดอะมิโนในตำแหน่ง -2 ถึง +6 ของ recognition helix ได้แก่ โมดูลที่ 1 (F1) ประกอบด้วยกรดอะมิโนในตำแหน่งที่ 15 ถึง 22

(SRSDKLVR) โมดูลที่ 2 (F2) ตำแหน่งที่ 43 ถึง 50 (STSGNLVR) โมดูลที่ 3 (F3) ตำแหน่งที่ 71 ถึง 78 (SRNDALTE) โมดูลที่ 4 (F4) ตำแหน่งที่ 99 ถึง 106 (SQSSSLVR) โมดูลที่ 5 (F5) ตำแหน่งที่ 127 ถึง 134 (SQSGDLRR) และโมดูลที่ 6 (F6) ตำแหน่งที่ 155 ถึง 162 (SQNSTLTE) จากรูป 3.6 จะพบว่าพลังงานการสลายพันธะยึดจับที่มีค่าบวก จะเป็นอันตรกิริยาที่ไม่เหมาะสม (unfavourable interactions) ในการเกิดเป็นสารประกอบ และกรดอะมิโนกับเบสในตำแหน่งที่เกิดอันตรกิริยาที่ไม่เหมาะสมพร้อมทั้งพลังงานที่เกี่ยวข้องเหล่านี้ แสดงไว้ในตาราง 3.5

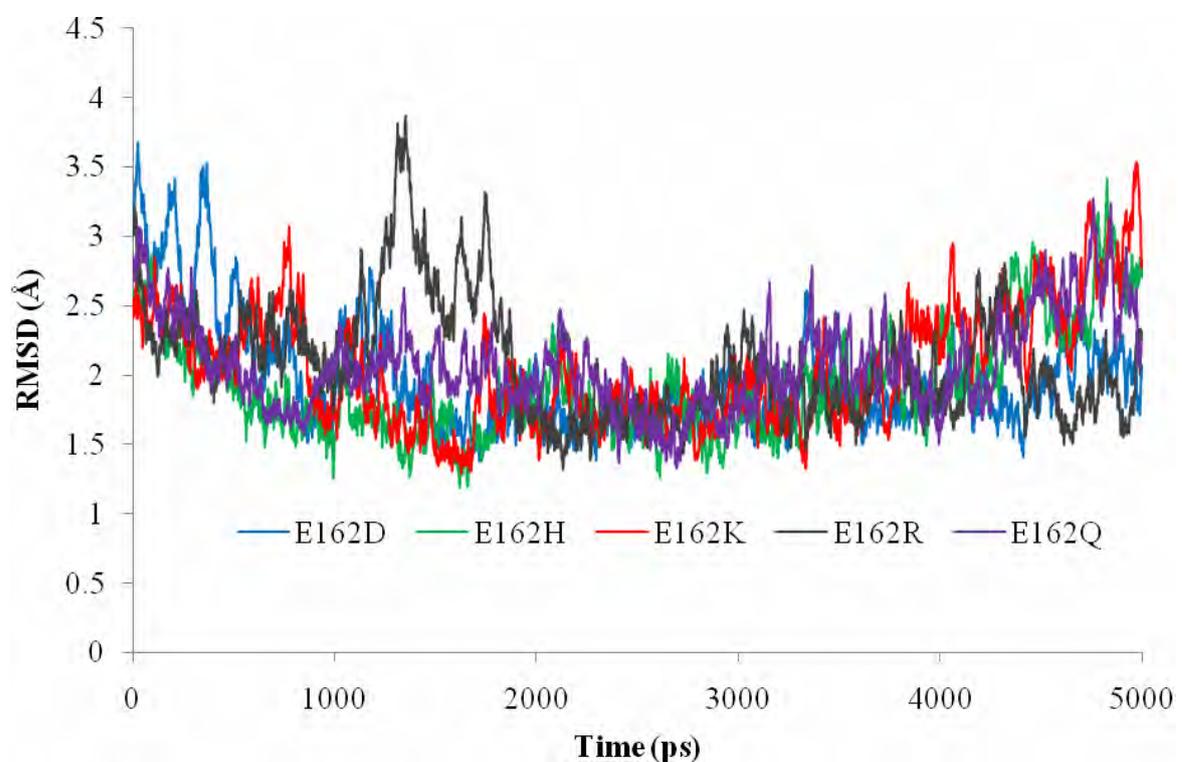
ตาราง 3.5 กรดอะมิโนบน recognition helix กับเบสในตำแหน่งที่เกิดอันตรกิริยาที่ไม่เหมาะสมในแบบจำลอง ZFPv1-GAT

Zinc finger modules	Amino acid-Base Interactions	Relative decompose energy (kcal/mol)
F1	Asp18(pos +2)---DG189	2.5
F3	Asp74(pos +2)---DC195	4.7
	Glu78(pos +6)---DC179	7.3
F5	Asp131(pos +3)---DC174	3.8
F6	Glu162(pos +6)---DC170	5.8

จากตาราง 3.5 พบว่ากรดอะมิโนที่เกิดอันตรกิริยาที่ไม่เหมาะสมทั้งหมดเป็นกรดอะมิโนที่มีประจุลบ ได้แก่ Glu และ Asp ซึ่งอาจเกิดการผลักกันระหว่างประจุลบของกรดอะมิโนทั้งสองกับหมู่ฟอสเฟตของโครงสร้างแกนของสายดีเอ็นเอ และสังเกตได้ว่าเบสที่เกี่ยวข้องในการเกิดอันตรกิริยาที่ไม่เหมาะสมนี้จะมีเพียงเบสไซโตซีนและกัวนีนเท่านั้น อาจเนื่องจากความหนาแน่นของอิเล็กตรอนของเบสทั้งสองชนิดนี้ที่มีมากกว่าเบสอะดีนีนและไทมีน ดังนั้นในการปรับเปลี่ยนชนิดของกรดอะมิโนจึงให้ความสนใจแก่ค่าพลังงานที่ไม่เหมาะสมที่มีค่าสูงที่สุด ได้แก่ กรดอะมิโน Glu78 ที่ตำแหน่ง +6 ของโมดูล 3 ซึ่งมีค่าพลังงานเท่ากับ +7.3 kcal/mol แต่ในโมดูล 3 นี้ ยังมีกรดอะมิโน Asp74 ในตำแหน่ง +2 ที่มีค่าพลังงานเท่ากับ +4.7 kcal/mol เช่นกัน ดังนั้นถ้าหากต้องการจะปรับเปลี่ยนชนิดของกรดอะมิโนทั้งสองในคราวเดียวกันอาจทำให้การแปลผลการทดลองมีความยุ่งยาก ว่าพลังงานการยึดจับที่ดีขึ้นหรือแย่ลงเป็นผลมาจากการปรับเปลี่ยนกรดอะมิโนใด ดังนั้นจึงเลือกจะปรับเปลี่ยนกรดอะมิโน Glu162 ที่ตำแหน่ง +6 ของโมดูล 6 ซึ่งมีค่าพลังงานเท่ากับ +5.8 kcal/mol ซึ่งมีค่าพลังงานสูงเป็นลำดับรองลงมา

และอยู่ในตำแหน่ง +6 เพียงตำแหน่งเดียว เพื่อทดสอบสมมติฐานที่คาดว่ากรดอะมิโนที่มีประจุลบบนสายโซ่ข้างอาจมีผลลดความสามารถในการเข้าจับของโปรตีน ZF ได้

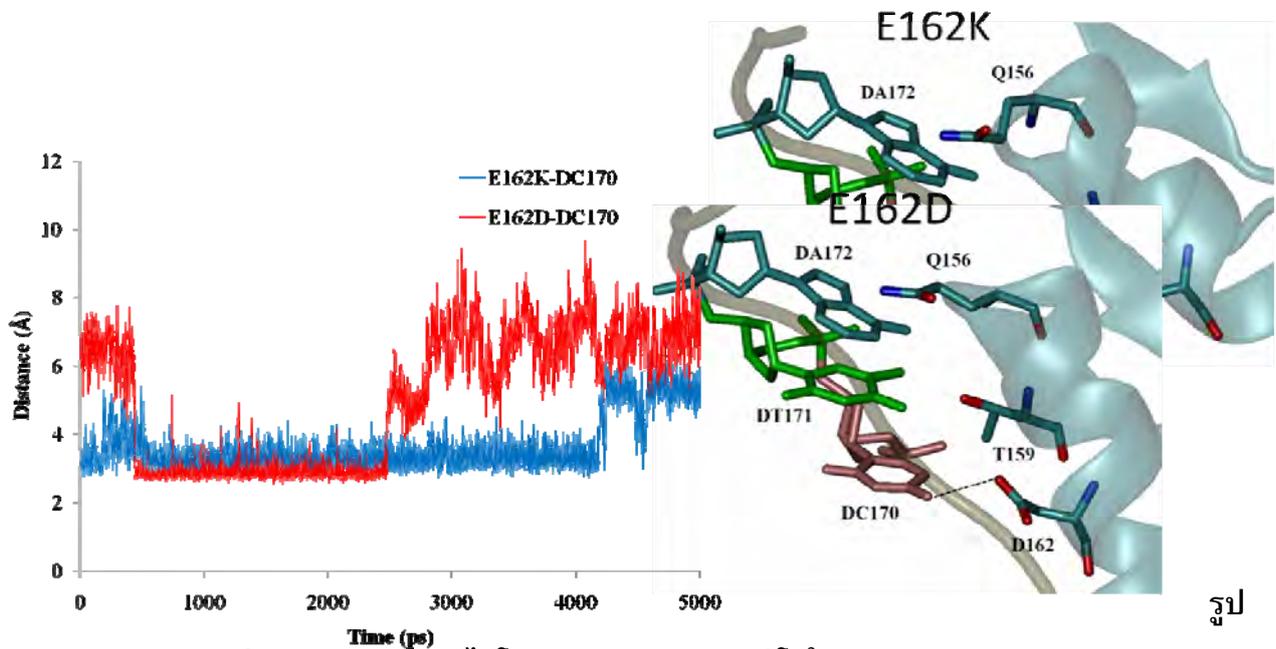
ต่อมา จึงทำการสร้างแบบจำลองโมเลกุลที่มีการปรับเปลี่ยนกรดอะมิโน Glu162 ขึ้นมาใหม่อีกจำนวน 5 ชุด ได้แก่ E162D E162H E162K E162R และ E162Q หลังจากนั้นจึงทำการจำลองพลวัตเป็นเวลา 5 ns จากนั้นจึงคำนวณหาค่า RMSD ของโครงสร้างทุกๆ 2 ps จากช่วงเวลา 5 ns เทียบกับโครงสร้างเฉลี่ยที่ได้จากการจำลองพลวัตแต่ละชุดแบบจำลอง (รูป 3.7) พบว่ามีค่า RMSD อยู่ในช่วงประมาณ 1.5 – 3.5 Å สามารถยอมรับได้ว่าแบบจำลองทั้งหมดอยู่ในสถานะที่เสถียร ไม่มีการแตกของพันธะในโครงสร้าง สามารถใช้เป็นตัวอย่างในการวิเคราะห์พลังงานในขั้นต่อไปได้



รูป 3.7 ค่า RMSD ของโครงสร้างทุกๆ 2 ps จากช่วงเวลา 5 ns เทียบกับโครงสร้างเฉลี่ยที่ได้จากการจำลองพลวัตของชุดแบบจำลอง E162D E162H E162K E162R และ E162Q

การวิเคราะห์รูปแบบการเกิดอันตรกิริยาในส่วนโมดูล 6 กับลำดับเบสเป้าหมาย พบว่ารูปแบบการเข้าจับของกรดอะมิโนกับลำดับเบสของทั้ง 5 ชุดแบบจำลอง (E162D E162H E162K E162R และ E162Q) และแบบจำลองต้นแบบ (ZFPv1-GAT) มีความคล้ายคลึงกัน กล่าวคือ พบการเกิดพันธะไฮโดรเจนระหว่างหมู่คาร์บอนิลของ Gln156 (ตำแหน่ง -1) กับหมู่เอมีนของ DA172 และไม่พบการเกิด

พันธะไฮโดรเจนในตำแหน่ง +3 ของ Thr159 ดังนั้นในการวิเคราะห์แบบจำลองทั้ง 5 ชุดที่มีการปรับเปลี่ยนจึงเห็นพิจารณาความสามารถในการเกิดพันธะไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง 162 ที่ถูกปรับเปลี่ยนไป กับเบสไซโตซีนในตำแหน่งที่ 170 ของทั้ง 5 ชุดแบบจำลอง ซึ่งจากแบบจำลองต้นแบบ ZFPv1-GAT ไม่สามารถตรวจพบการเกิดอันตรกิริยาระหว่าง Glu162 และ DC170 ได้ตลอดระยะเวลาที่ทำการจำลองพลวัต และคาดว่าสามารถตรวจพบได้ในกลุ่มกรดอะมิโนที่เป็นกลางหรือประจุบวก แต่พบว่าสามารถพบการเกิดอันตรกิริยา (ระยะน้อยกว่า 3.5 Å) ระหว่าง Lys162 หรือ Asp162 กับ DC170 ได้เพียง 2 ชุดแบบจำลองเท่านั้น ดังแสดงในรูป 3.8 ในแบบจำลอง E162D พบว่าหมู่คาร์บอนิลของ Asp162 สามารถเกิดพันธะไฮโดรเจนกับหมู่เอมีนของ DC170 ได้ในช่วง 0.5 – 2.8 ns ของการจำลองพลวัต ส่วนในแบบจำลอง E162K หมู่เอมีนของ Lys162 สามารถเกิดอันตรกิริยากับหมู่เอมีนของ DC170 ได้ในช่วง 4 ns แรก แต่เป็นที่น่าสังเกตว่าในแบบจำลองอื่น ได้แก่ E162H E162R และ E162Q ไม่พบว่ากรดอะมิโน His162 หรือ Arg162 หรือ Gln162 มีการเคลื่อนที่เข้าใกล้ DC170 แต่อย่างใด อาจเนื่องจากการเกิดอันตรกิริยาระหว่างกรดอะมิโนเหล่านี้กับเบสเป้าหมายนั้นอาจต้องอาศัยโมเลกุลน้ำเป็นตัวเชื่อมให้เกิดแรงยึดเหนี่ยวระหว่างโปรตีน ZF กับเบสเป้าหมาย ซึ่งการวิเคราะห์ในเชิงอันตรกิริยาที่อาจเกิดขึ้นจากโครงสร้างของแบบจำลองอาจยังไม่เพียงพอ จึงต้องทำการคำนวณพลังงานการยึดจับระหว่างโปรตีน ZFPv1 ที่ถูกปรับเปลี่ยนและสายดีเอ็นเอในแบบจำลองต่อไป

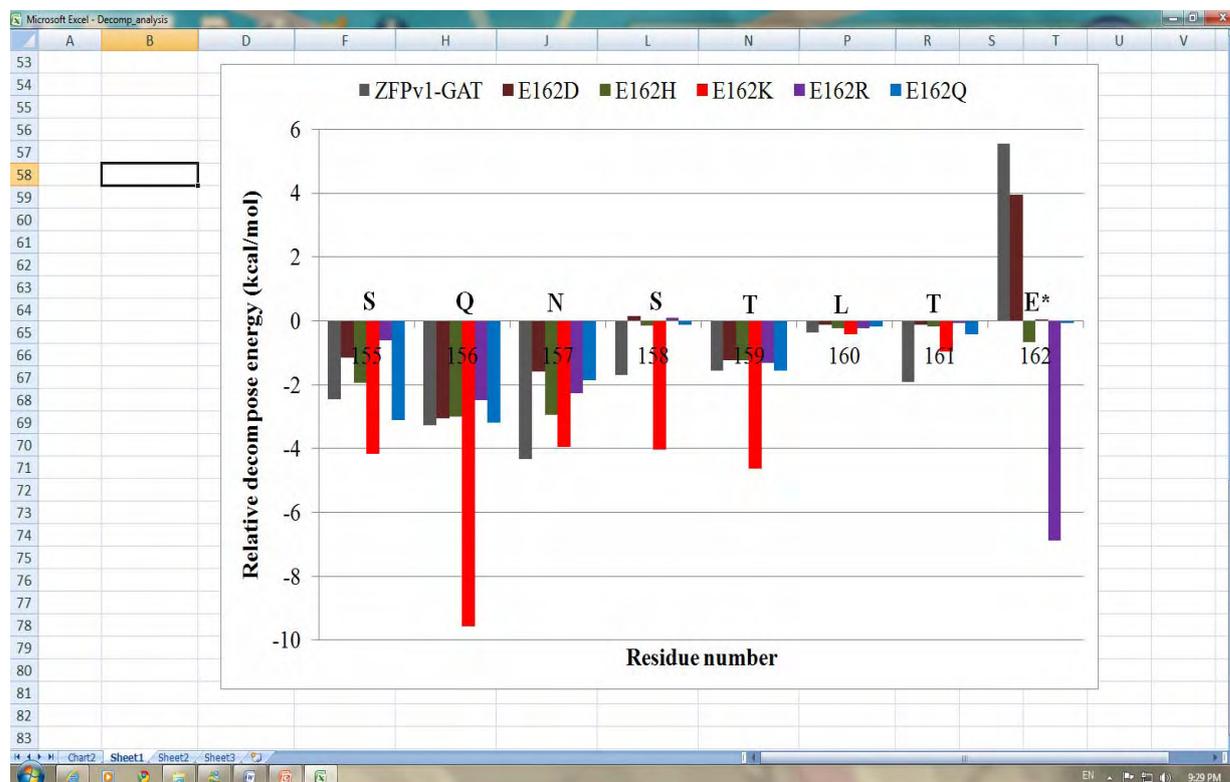


3.8 การเกิดพันธะ ไฮโดรเจนระหว่างกรดอะมิโนในตำแหน่ง +6 (Lys162 และ Asp162) ของ recognition helix ในโมเดล 6 ของแบบจำลอง E162K และ E162D

การคำนวณพลังงานในการยึดจับสัมพัทธ์ (relative binding energy) ของแต่ละชุดแบบจำลอง อาศัยหลักการ MM/GBSA ได้ผลดังแสดงในตาราง 3.6 พบว่าจากแบบจำลองที่มีการปรับเปลี่ยนจากกรดอะมิโน Glu162 (-323.2 kcal/mol) ไปเป็นกรดอะมิโนที่ไม่มีประจุ เช่น Gln (-349.3 kcal/mol) หรือกรดอะมิโนที่มีประจุบวก เช่น His (-359.8 kcal/mol) Lys (-372.8 kcal/mol) และ Arg (-342.5 kcal/mol) ต่างให้ค่าพลังงานในการยึดจับกับลำดับเบสเป้าหมายได้ดีกว่าโปรตีนต้นแบบ ZFPv1 ตั้งแต่ -19.3 ถึง -49.6 kcal/mol ส่วนการปรับเปลี่ยนเป็นกรดอะมิโน Asp (-308.6 kcal/mol) ซึ่งมีประจุลบและสายโซ่ข้างขนาดเล็กกว่า Glu กลับมีผลให้พลังงานในการยึดจับที่แยลง และจากการคำนวณพลังงานในการสลายพันธะยึดจับของกรดอะมิโนตำแหน่ง -2 ถึง +6 ของโมดูล 6 ในแบบจำลองที่มีการปรับเปลี่ยนชนิดกรดอะมิโน Glu162 ทั้ง 5 ชุดจากรูป 3.9 พบว่าแบบจำลอง E162K มีค่าพลังงานการยึดจับแข็งแรงที่สุด คือ -372.8 kcal/mol จากการคำนวณด้วยวิธี MM/GBSA ซึ่งมีค่าพลังงานที่สูงกว่าโปรตีนต้นแบบ ZFPv1 ถึง -49.6 kcal/mol เป็นผลจากการเพิ่มความแข็งแรงของอันตรกิริยาที่เกิดขึ้นของกรดอะมิโนในตำแหน่ง Ser155 Gln156 Ser158 Thr159 อย่างชัดเจน (กราฟแท่งสีแดง) ซึ่งได้จากการพลังงานการสลายพันธะยึดจับของ Lys162 แต่เป็นที่น่าสังเกตว่าการเปลี่ยนแปลงกรดอะมิโนจาก Glu162 ไปเป็น Lys162 นี้ ไม่มีผลเพิ่มการยึดจับมากขึ้นโดยตรงจาก Lys162 ตามที่คาดไว้จากการวิเคราะห์ระยะห่างระหว่างอะตอมที่สามารถเกิดพันธะไฮโดรเจนได้ แต่มีผลลดอันตรกิริยาที่ไม่เหมาะสมของ Glu162 หรือ Asp162 ที่เกิดขึ้น ส่วนกรดอะมิโน Arg162 มีผลเพิ่มความแข็งแรงในการยึดจับได้โดยตรงกับ DC170 ซึ่งคาดว่าเกิดผ่านโมเลกุลน้ำเป็นตัวกลาง แต่ไม่มีผลเสริมความแข็งแรงของการเกิดอันตรกิริยาของกรดอะมิโนอื่นในโมดูลเดียวกันเช่นในกรณีของ Lys162 อย่างไรก็ตามการประเมินความแข็งแรงหรือพลังงานในการยึดจับด้วยวิธี MM/GBSA เป็นการคำนวณจากทั้งโมเลกุลของโปรตีน ZF และสายดีเอ็นเอเป้าหมายในขณะที่เข้าจับกันเกิดเป็นสารประกอบ ส่วนการคำนวณพลังงานการสลายพันธะยึดจับเป็นการพิจารณาแต่ละหน่วยของกรดอะมิโนที่ศึกษาในรายละเอียด ดังนั้นการยืนยันถึงผลการศึกษาคงต้องอาศัยวิธีการทั้งสองในการพิสูจน์ ซึ่งจากผลการทดลองนี้พบว่าการปรับเปลี่ยนกรดอะมิโน Glu162 ที่อยู่ในตำแหน่ง +6 ของ recognition helix ในโมดูล 6 จากกรดอะมิโนที่มีประจุลบไปเป็นกรดอะมิโนอื่นที่มีประจุบวกหรือไม่มีประจุ คาดว่าจะสามารถเพิ่มความสามารถในการเข้าจับของโปรตีน ZFPv1 ต่อดีเอ็นเอเป้าหมายได้ดียิ่งขึ้น ซึ่งสามารถนำไปประยุกต์ในการพัฒนาการออกแบบโปรตีน ZF ให้มีความสามารถในการเข้าจับกับดีเอ็นเอเป้าหมายได้ โดยเฉพาะกรณีที่ต้องมีการแข่งขันการเข้าจับกับเอนไซม์ในธรรมชาติที่มีความจำเพาะต่อสายดีเอ็นเอเช่นกัน เช่น ในกรณีของเอนไซม์อินทิเกรส เป็นต้น

ตาราง 3.6 พลังงานการยึดจับสัมพัทธ์ของแบบจำลองที่มีการปรับเปลี่ยนชนิดกรดอะมิโนเทียบกับแบบจำลองต้นแบบ ZFPv1-GAT โดยวิธี MM/GBSA

ZFP-DNA models	Relative binding energy (kcal/mol)
ZFPv1-GAT	-323.2 ±19.1
E162D	-308.6 ±15.0
E162H	-359.8 ±26.4
E162K	-372.8 ±33.5
E162R	-342.5 ±20.6
E162Q	-349.3 ±14.6



รูป 3.9 พลังงานการสลายพันธะยึดจับของกรดอะมิโนตำแหน่ง -2 ถึง +6 ของโมดูล 6 ในแบบจำลอง ZFPv1-GAT E162D E162H E162K E162R และ E162Q

4. วิจารณ์ผลการทดลอง

โปรตีน ZF ชนิด Cys₂His₂ เป็นโครงสร้างตติยภูมิที่ถูกค้นพบว่ามีความสามารถจดจำลำดับเบสที่จำเพาะได้^[9-11, 34, 35] จึงมีความพยายามในการสังเคราะห์โปรตีน ZF เลียนแบบธรรมชาติให้มีความสามารถในการจับกับลำดับเบสเป้าหมายตามต้องการ ในการศึกษานี้อาศัยลำดับเบสในส่วน 2LTR ซึ่งมีความยาว 18 คู่เบสเป็นเป้าหมายสำหรับการออกแบบโปรตีน ZF โดยอาศัยโปรแกรมสำเร็จรูปจากเว็บไซต์ ZF Tools ได้โปรตีน ZF จำนวน 6 โมดูลที่มีโครงสร้างหลักเป็นชนิด Cys₂His₂ ได้แก่ ZFPv1 และได้อาศัยเทคนิคทางพันธุวิศวกรรมในการสังเคราะห์โปรตีน ZFPv1 นี้ขึ้นมาภายในเซลล์และพิสูจน์ความสามารถในการเข้าจับกับสายดีเอ็นเอเป้าหมายเพื่อประโยชน์ในการเป็นสารยับยั้งสำหรับเอนไซม์อินทิเกรสจากรายงานวิจัยก่อนหน้านี้^[28] โปรตีน ZFPv1 ถูกออกแบบโดยหลักการ Molecular Assembly ที่เป็นการเชื่อมต่อแต่ละโมดูลในส่วน recognition helix ที่จำเพาะต่อแต่ละ 3 คู่เบส (triplet) บนสายดีเอ็นเอเป้าหมาย 2LTR ซึ่งในแต่ละ recognition helix จะประกอบด้วยกรดอะมิโนที่สำคัญ 3 ตำแหน่ง ได้แก่ ตำแหน่ง -1 +3 และ +6 ที่สายโซ่ข้างของกรดอะมิโนเหล่านี้อยู่ในตำแหน่งที่สามารถเกิดอันตรกิริยากับ 3 คู่เบสบนสายดีเอ็นเอได้ และคาดว่าจะเป็นส่วนสำคัญที่เกี่ยวข้องกับรูปแบบการจดจำ (recognition pattern) ลำดับเบสของโปรตีน ZF และในตำแหน่ง +2 ที่สามารถเกิดอันตรกิริยากับเบสบนสายดีเอ็นเอตรงข้ามได้

จากการเปรียบเทียบผลการทำนายลำดับกรดอะมิโนที่ได้จาก ZF Tools และ ZiFDB ในตาราง 3.1 พบว่าข้อมูลของลำดับกรดอะมิโนบน recognition helix สำหรับเบสเป้าหมาย 5'-GNN-3' ได้แก่ 5'-GGG-3'(F1) 5'-GAT-3'(F2) 5'-GTA-3'(F4) และ 5'-GCA-3'(F5) มีความหลากหลายสูง และมีหลายกลุ่มวิจัยที่ทำการศึกษาทั้งผลที่ได้จากการทดลองและจากการสืบค้นข้อมูลจากโปรตีน ZF ที่พบบ่อยในธรรมชาติ พบว่าส่วน 5'-CTG-3'(F3) และ 5'-CTA-3'(F6) ความหลากหลายต่ำ ดังนั้นการทำนายลำดับกรดอะมิโนของโปรตีน ZF สำหรับ 3 คู่เบสเป้าหมาย 5'-GNN-3' คาดว่าจะมีโอกาสความถูกต้องสูงกว่า 3 คู่เบสอื่นๆ เนื่องจากมีข้อมูลทั้งในเชิงการทดลองและทางทฤษฎีรองรับเป็นจำนวนมาก โดยเฉพาะการทำนายชนิดของกรดอะมิโนในตำแหน่งที่สำคัญ ได้แก่ ตำแหน่ง +6 ที่เกิดอันตรกิริยากับเบส DG ในตำแหน่ง 5' พบว่ากรดอะมิโนที่เหมาะสมควรเป็น Arg และมีโอกาสพบได้มากถึง 64 – 100% และตำแหน่ง -1 ที่เกิดอันตรกิริยากับเบสตำแหน่ง 3' พบว่าถ้าเป็นเบส DG สามารถพบกรดอะมิโน Arg ในตำแหน่งที่เกิดอันตรกิริยาได้มากถึง 87 – 100% ส่วนเบสเป้าหมาย DA สามารถพบกรดอะมิโน Gln

ได้มากถึง 87 – 100% และถ้าเป็นเบสเป้าหมาย DT สามารถพบกรดอะมิโน Thr ได้เพียง 33% ส่วนตำแหน่ง +2 และ +3 ที่มีรายงานความสำคัญต่อการเข้าจับกับเบสเป้าหมาย พบว่ายังมีความหลากหลายของข้อมูลที่พบ แต่อย่างไรก็ตามผลทำนายที่ได้จาก ZF Tools ก็ยังให้ผลทำนายกรดอะมิโนที่มีโอกาสพบสูงเช่นเดียวกับจากฐานข้อมูล ซึ่งอาจเนื่องจากเป็นตำแหน่ง +2 และ +3 ที่อยู่ในตอนกลางของ recognition helix ซึ่งรับผิดชอบเกิดอันตรกิริยากับเบสตัวกลางของ 3 คู่เบสเป้าหมาย จึงอาจมีความสำคัญน้อยกว่าตำแหน่ง -1 และ +6 ที่อยู่ตอนต้นและปลายของ recognition helix ที่รับผิดชอบเบสในตำแหน่ง 5' และ 3' ดังนั้นจึงคาดว่าผลการทำนายของโปรตีน ZFPv1 ที่ได้จะมีความเหมาะสมต่อลำดับเบสเป้าหมาย 2LTR สูง

ในการศึกษานี้ได้สร้างแบบจำลองโมเลกุลของสารประกอบของโปรตีน ZFPv1 และลำดับเบสเป้าหมาย 2LTR ขึ้นจากโครงสร้างผลึกของโปรตีน ZF ความยาว 6 โมดูลที่ถูกจัดเก็บอยู่ในฐานข้อมูล PDB เป็นต้นแบบ เริ่มทำการศึกษาคำเฉพาะในการเข้าจับของโปรตีน ZFPv1 จากลำดับเบสเป้าหมายและที่ไม่ใช่เป้าหมาย คือ Zif1 (ZFPv1-GAT) และ Zif2 ตามลำดับ พบว่าการเกิดการเข้าจับระหว่างโปรตีน ZFPv1 กับสายดีเอ็นเอเป้าหมายในแบบจำลอง Zif1 นั้น มีความแข็งแรงมากกว่าการเข้าจับกับสายดีเอ็นเอที่ไม่ใช่เป้าหมายในแบบจำลอง Zif2 อย่างมาก คาดว่าโปรตีน ZFPv1 อาจจับกับดีเอ็นเอที่ไม่จำเพาะได้โดยอาศัยลักษณะโครงสร้างการจัดเรียงตัวของโปรตีน ZF ชนิด Cys₂His₂ ที่มีความเหมาะสมในการเข้าจับกับ major groove ของสายดีเอ็นเอโดยอาศัยการเกิดอันตรกิริยาระหว่างกรดอะมิโนที่ถูกสงวนไว้ ได้แก่ His ที่มีวงอิมิดาโซลในสายโซ่ข้างซึ่งอะตอม N_δ ซึ่งมีสภาพขั้วบวกจะสามารถเกิดแรงยึดเหนี่ยวไฟฟ้าสถิตย์กับหมู่ฟอสเฟตที่มีประจุลบได้ ทำให้โปรตีน ZF มาอยู่ใกล้กับสายดีเอ็นเอมากขึ้น ส่วน N_ε จะเกิดพันธะโคออดิเนตกับโลหะสังกะสีได้ รวมถึงกรดอะมิโน Lys หรือ Arg ที่อยู่ถัดจาก Cys ในส่วน β-sheet และ Lys ในส่วน linker (TGEKP) สามารถเกิดแรงยึดเหนี่ยวไฟฟ้าสถิตย์กับหมู่ฟอสเฟตที่มีประจุลบได้เช่นกัน^[3] รวมถึงอาจเหนี่ยวนำให้สายดีเอ็นเอมีการบิดรูป เช่น major groove ที่กว้างขึ้น ทำให้เกิดการจับกันระหว่างโปรตีน ZF และสายดีเอ็นเอได้ดียิ่งขึ้น^[36] เหตุการณ์เหล่านี้ช่วยให้โปรตีน ZF ยึดจับกับสายดีเอ็นเอได้แต่ไม่มีความจำเพาะหรือจับกันแบบไม่แข็งแรงพอ ส่วนชนิดของกรดอะมิโนในตำแหน่ง -1 +2 +3 และ +6 บน recognition helix จะสัมพันธ์กับอันตรกิริยาที่เกิดขึ้นจำเพาะต่อลำดับเบสเป้าหมายซึ่งคาดว่าเป็นปัจจัยสำคัญที่ทำให้โปรตีน ZF ซึ่งสามารถเข้าจับกับสายดีเอ็นเอทั่วไปได้อยู่แล้ว มีความสามารถในการจดจำลำดับเบสที่จำเพาะและกับชนิดของกรดอะมิโนในตำแหน่งข้างต้น และเกิดเป็นสารประกอบโปรตีน-ดีเอ็นเอที่เสถียรมากขึ้น ดังนั้นจึงคาดว่าการทำงานของโปรตีน ZF จะสามารถจับกับสายดีเอ็นเอทั่วไปได้อย่างไม่แข็งแรงเกิดเป็นสารประกอบที่ไม่จำเพาะ และสามารถหลุดออกไปจนกระทั่งเคลื่อนที่ไปจับกับสายดีเอ็นเอเป้าหมายจึง

เกิดการจับกันอย่างแข็งแรง คล้ายการทำงานของโปรตีนซ่อมแซมสายดีเอ็นเอ (DNA repair proteins) ที่สามารถตรวจพบความผิดปกติของเบสในสายดีเอ็นเอได้เพียงเล็กน้อยจากการเคลื่อนที่ไปตามสายดีเอ็นเอจนกระทั่งพบจุดที่ผิดปกติบนสายดีเอ็นเอจึงเกิดการซ่อมแซม ณ จุดนั้น

หลังจากนั้นจึงทำการทดลองถึงความสามารถของโปรตีน ZF ที่ออกแบบในการแยกความแตกต่างของเบสที่ต่างกันเพียงตำแหน่งเดียวบนสายดีเอ็นเอเป้าหมาย คือ ZFPv1 และ ZFPv2 ต่อลำดับเบสเป้าหมาย GAT และ GAA จากการคำนวณพลังงานการยึดจับระหว่างทั้ง 2 โมเลกุล ในกรณีของ ZFPv1 สามารถเข้าจับกับดีเอ็นเอเป้าหมาย GAT ได้แข็งแรงกว่า GAA ด้วยพลังงาน 8.5 kcal/mol ซึ่งสามารถบอกได้ว่า ZFPv1 มีความชอบจับ (affinity) กับเป้าหมายได้ดีกว่า แต่ไม่สามารถบอกได้ถึงความจำเพาะ (specificity) ที่มากกว่าหรือไม่ เนื่องจากไม่สามารถตรวจพบความแตกต่างในการเกิดแรงยึดจับที่แตกต่างกันระหว่างโปรตีน ZFPv1 กับ DT184 หรือ DA184 รวมถึง ZFPv2 ที่ออกแบบให้มีความจำเพาะต่อลำดับเบส GAA กลับพบว่าสามารถจับกับเบส GAT ได้แข็งแรงกว่า และไม่พบความแตกต่างในการเกิดแรงยึดจับที่แตกต่างกันระหว่างโปรตีน ZFPv2 กับ DT184 หรือ DA184 เช่นกัน ดังนั้นในกรณีที่ลำดับเบสเป้าหมายมีความแตกต่างกันเพียงเล็กน้อยจึงยังไม่สามารถสรุปได้ถึงความสามารถในการจำแนกความแตกต่างหรือความจำเพาะของโปรตีน ZF ที่ออกแบบได้ แต่อย่างไรก็ตาม ในการศึกษาที่ใช้ลำดับเบสเป้าหมายที่มีความยาว 18 คู่เบส ซึ่งมีโอกาสซ้ำกันน้อยมากในจีโนมของมนุษย์ เพื่อเป็นการป้องกันการจับกับลำดับเบสที่ไม่ใช่เป้าหมาย และอาจเกิดผลเสียที่ติดตามมาได้

การออกแบบโปรตีน ZF โดยอาศัยหลักการ Molecular Assembly มีรายงานความสำเร็จในการสังเคราะห์โปรตีน ZF เลียนแบบธรรมชาติ เพื่อเข้าจับกับดีเอ็นเอเป้าหมายได้อย่างมีประสิทธิภาพ เช่น การสังเคราะห์โปรตีน ZF เพื่อให้เป็น transcription factors^[37] และการสังเคราะห์โปรตีน ZF nucleases เพื่อให้จับกับลำดับเบสต่างๆในยีน CCR5 ของมนุษย์^[38] เป็นต้น แต่ก็พบรายงานความล้มเหลวในการทำนายและสังเคราะห์โปรตีน ZF^[13] ได้เช่นกัน ซึ่งปัจจัยที่ต้องพิจารณาหนึ่งคือความสามารถในการเข้าจับหรือความแข็งแรงในการเข้าจับกับสายดีเอ็นเอเป้าหมาย (binding affinity) ของโปรตีน ZF ที่ทำนายได้จากหลักการ Molecular Assembly ในการศึกษาพยายามสร้างแบบจำลองโมเลกุลที่มีการปรับเปลี่ยนชนิดของกรดอะมิโนในส่วน recognition helix เพื่อเพิ่มความสามารถในการเข้าจับกับสายดีเอ็นเอเป้าหมายของโปรตีน ZFPv1 โดยการศึกษาเชิงพลังงานการยึดจับสัมพัทธ์ (relative binding energy) ร่วมกับพลังงานในการสลายพันธะยึดจับ (decompose energy) ของกรดอะมิโน พบว่าในลำดับเบสเป้าหมาย 18 คู่เบสหรือ 6 โมดูลของโปรตีน ZFPv1 ที่ถูกทำนายด้วยโปรแกรม ZF Tools นั้น มีจำนวน 4 โมดูลที่มีเบสเป้าหมายเป็น 5'-GNN-3' ได้แก่

- โมดูล 1 RSDKLVR สำหรับ 5'-GGG-3'

- โมดูล 2 TSGNLVR สำหรับ 5'-GAT-3'
- โมดูล 4 QSSSLVR สำหรับ 5'-GTA-3' และ
- โมดูล 5 QSGDLRR สำหรับ 5'-GCA-3'

ซึ่งโมดูลสำหรับ 5'-GNN-3' คาดว่าจะมีโอกาสความถูกต้องสูงกว่าการทำนายโมดูลสำหรับ 5'-CNN-3' ในโมดูล 3 และ 6 ซึ่งมีเป้าหมายเป็น 5'-CTG-3' และ 5'-CTA-3' ตามลำดับ แม้ว่าการศึกษาสำหรับ 5'-CNN-3' จะมีรายงานการทำนายโมดูลถึง 15 แบบใน 16 แบบที่เป็นไปได้ทั้งหมดของ 5'-CNN-3' ยกเว้น 5'-CTC-3'^[10] ซึ่งทั้ง 5'-CTG-3' และ 5'-CTA-3' ที่เป็นเป้าหมายมีค่า K_D ในระดับ nM เช่นกัน ดังนั้นจึงคาดว่าน่าจะมีคามเหมาะสมในการทำนายโมดูลของ recognition helix แต่อย่างไรก็ตาม ข้อมูลการทำนาย recognition helix สำหรับ 5'-CNN-3' ที่ได้จากการทดลองยังมีความหลากหลายต่ำ มีเพียงรายงานจากกลุ่มวิจัยของ Barbas เป็นส่วนใหญ่ ดังนั้นในการทดลองเพื่อปรับเปลี่ยนชนิดของกรดอะมิโนจึงให้ความสนใจในส่วนของ 5'-CNN-3' เป็นพิเศษ รวมถึงผลการคำนวณพลังงานการสลายพันธะยึดจับของกรดอะมิโนแต่ละตำแหน่งของ recognition helix ทั้ง 6 โมดูล (รูป 3.6) พบว่ามีอันตรกิริยาที่ไม่เหมาะสมในการเกิดการยึดจับกันระหว่างโปรตีน ZFPv1 กับเบสเป้าหมาย 2LTR ในโมดูลที่ 1 (5'-GGG-3') 3 (5'-CTG-3') 5 (5'-GCA-3') และ 6 (5'-CTA-3') โดยกรดอะมิโนที่สำคัญที่เกิดพลังงานที่ไม่เหมาะสมสูงที่สุดในสองลำดับแรก คือ Glu78 ในตำแหน่ง +6 กับ DC179 ในโมดูลที่ 3 ด้วยพลังงาน 7.3 kcal/mol และ Glu162 ในตำแหน่ง +6 กับ DC170 ในโมดูลที่ 6 ด้วยพลังงาน 5.8 kcal/mol

จากฐานข้อมูลการทำนายชนิดของกรดอะมิโนในตำแหน่ง +6 ของ recognition helix ซึ่งรับผิดชอบในการจดจำสำหรับเบสไซโตซีนที่ปลาย 5' ของ 3 คู่เบสเป้าหมาย 5'-CNN-3' จะให้ผลการทำนายเป็นกรดอะมิโน Glu เสมอ ซึ่ง Glu เป็นกรดอะมิโนที่มีประจุลบบริเวณสายโซ่ข้าง ที่มีความเป็นไปได้ว่าอันตรกิริยาที่ไม่เหมาะสมที่เกิดขึ้นนั้น อาจเป็นผลจากการเกิดการผลักระหว่างความเป็นประจุลบของสายโซ่ข้างกับสภาพขั้วที่เป็นลบของอะตอมไนโตรเจนของหมู่เอมีนของเบส DC ได้ ดังนั้นจึงทดลองปรับเปลี่ยนกรดอะมิโน Glu162 ไปเป็น Asp ซึ่งเป็นกรดอะมิโนที่มีประจุลบบริเวณสายโซ่ข้างเช่นเดียวกับ Glu ให้ผลการยึดจับที่มีความแข็งแรงลดลงจากต้นแบบ จึงสันนิษฐานว่ากรดอะมิโนกลุ่มที่มีประจุลบบริเวณสายโซ่ข้างที่ตำแหน่ง +6 ไม่เหมาะสมในการเกิดแรงยึดจับกับเบสไซโตซีนที่ปลาย 5' หลังจากนั้นจึงทดลองปรับเปลี่ยนไปเป็นกรดอะมิโนกลุ่มที่มีประจุบวกหรือไม่มีประจุบริเวณสายโซ่ข้าง ได้แก่ His Lys Arg และ Gln พบว่าสามารถเพิ่มความแข็งแรงในการยึดจับระหว่างโปรตีน ZFPv1 กับเบสเป้าหมาย GAT ได้ด้วยพลังงานตั้งแต่ -19.3 ถึง -49.6 kcal/mol จากผลการคำนวณ

พลังงานการยึดจับสัมพัทธ์ (ตาราง 3.6) ซึ่งคาดว่าเป็นผลจากการเพิ่มแรงยึดจับระหว่างกรดอะมิโนที่ปรับเปลี่ยนกับเบส DC โดยตรง แต่จากการคำนวณพลังงานการสลายพันธะยึดจับของกรดอะมิโนที่ปรับเปลี่ยนไปนั้น (รูป 3.9) พบว่ามีเพียงกรดอะมิโน Arg เท่านั้นที่พบว่ามีแรงยึดจับกับ DC ได้ดียิ่งขึ้นโดยตรง ส่วนกรดอะมิโนอื่น เช่น His Lys และ Gln มีผลลดอันตรกิริยาที่ไม่เหมาะสมระหว่างกรดอะมิโนประจุลบกับประจุลบของหมู่ฟอสเฟตบนสายเบสเป้าหมาย และยังช่วยเพิ่มแรงยึดจับของกรดอะมิโนที่ปรับเปลี่ยนร่วมกับการเพิ่มความแข็งแรงของกรดอะมิโนตำแหน่งอื่นบน recognition helix ของโมดูล 6 นี้ด้วย โดยเฉพาะในกรณีของ E162K มีความน่าสนใจในการทดลองปรับเปลี่ยนชนิดของกรดอะมิโนในการทดลองจริง เนื่องจากให้ค่าพลังงานในการยึดจับได้สูงที่สุด และยังสามารถเพิ่มความแข็งแรงของการยึดจับของกรดอะมิโนในตำแหน่ง +3 ต่อเบสได้ดียิ่งขึ้น ซึ่งอาจเป็นการเพิ่มความจำเพาะในการเข้าจับของโปรตีน ZFPv1 ได้อีกด้วย ดังนั้นจากการวิเคราะห์เชิงโครงสร้างและเชิงพลังงานในการยึดจับที่ได้จากแบบจำลองโมเลกุลจะเห็นว่ามีความโน้มเอียงที่สามารถเพิ่มความแข็งแรงในการยึดจับของโปรตีน ZFPv1 ที่ออกแบบต่อลำดับเบสเป้าหมาย เพื่อเอาชนะการเข้าจับของเอนไซม์อินทิเกรส และอาจเพิ่มความจำเพาะในการยึดจับกับลำดับเบสเป้าหมายได้ด้วย

5. สรุปผลการทดลอง

เทคนิคการออกแบบโปรตีน ZF ชนิด Cys₂His₂ โดยวิธีทางคอมพิวเตอร์ สามารถทำนายลำดับกรดอะมิโนของโปรตีน ZFPv1 ที่จำเพาะต่อลำดับเบสเป้าหมาย 2LTR ซึ่งเป็นลำดับเบสเป้าหมายตามธรรมชาติของเอนไซม์อินทิเกรสได้อย่างมีประสิทธิภาพ แบบจำลองโมเลกุลระหว่างโปรตีน ZFPv1 ซึ่งเข้าจับกับลำดับเบสบนสายดีเอ็นเอที่แตกต่างกันได้ถูกพัฒนาขึ้นและทำการจำลองพลวัต เพื่อศึกษาถึงปัจจัยในด้านความจำเพาะและความแข็งแรงในการเข้าจับกับสายดีเอ็นเอเป้าหมาย พบว่าโปรตีน ZFPv1 นั้น มีความสามารถในการจำแนกลำดับเบสเป้าหมาย 2LTR ได้เป็นอย่างดีเยี่ยม ส่วนในกรณีที่สายดีเอ็นเอเป้าหมายมีความแตกต่างกันเพียงลำดับเบส 1 ตำแหน่ง อาจยังไม่สามารถสรุปได้แน่ชัดจากการทดลองนี้ แต่อย่างไรก็ตามเป็นโอกาสที่เกิดขึ้นได้น้อยมากในธรรมชาติ ส่วนในด้านความแข็งแรงในการเข้าจับของโปรตีน ZFPv1 กับสายดีเอ็นเอของ 2LTR จากการวิเคราะห์โครงสร้างแบบจำลองโมเลกุลและพลังงานการยึดจับระหว่างกันจากการจำลองพลวัต พบว่าการปรับเปลี่ยนกรดอะมิโนในตำแหน่ง +6 ของ recognition helix ที่เป็นกรดอะมิโนที่มีประจุลบบริเวณสายโซ่ข้าง ไปเป็นกรดอะมิโนที่ไม่มีประจุหรือมีประจุบวกมีแนวโน้มที่สามารถเพิ่มความแข็งแรงในการยึดจับของโปรตีน ZFPv1 กับสายดีเอ็นเอ 2LTR ได้เป็นอย่างดี เพื่อให้สามารถเอาชนะการเข้าจับของเอนไซม์อินทิเกรส และยังสามารถเพิ่มความจำเพาะในการยึดจับกับสายดีเอ็นเอเป้าหมายได้อีกด้วย ซึ่งมีความน่าสนใจเป็นอย่างยิ่งในการทดสอบในหลอดทดลองต่อไปในอนาคต และอาจเป็นข้อแนะนำเพิ่มเติมหนึ่ง สำหรับนักวิจัยที่สนใจการออกแบบโปรตีน ZF ให้มีความจำเพาะและสามารถจับกับลำดับเบสเป้าหมายได้อย่างแข็งแรง

เอกสารอ้างอิง

1. Krishna, S.S., Majumdar, I., Grishin, N. V., *Structural classification of zinc fingers: survey and summary*. Nucleic Acids Res, 2003. **31**(2): p. 532-50.
2. Tupler, R., Perini, G., Green, M. R., *Expressing the human genome*. Nature, 2001. **409**(6822): p. 832-3.
3. Wolfe, S.A., Nekludova, L., Pabo, C. O., *DNA recognition by Cys2His2 zinc finger proteins*. Annu Rev Biophys Biomol Struct, 2000. **29**: p. 183-212.
4. Pavletich, N.P., Pabo, C. O., *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å*. Science, 1991. **252**(5007): p. 809-17.
5. Swirnoff, A.H., Milbrandt, J., *DNA-binding specificity of NGFI-A and related zinc finger transcription factors*. Mol Cell Biol, 1995. **15**(4): p. 2275-87.
6. Christy, B., Nathans, D., *DNA binding site of the growth factor-inducible protein Zif268*. Proc Natl Acad Sci U S A, 1989. **86**(22): p. 8737-41.
7. Jamieson, A.C., Miller, J. C., Pabo, C. O., *Drug discovery with engineered zinc-finger proteins*. Nat Rev Drug Discov, 2003. **2**(5): p. 361-8.
8. Blancafort, P., Magnenat, L., Barbas, C. F., 3rd, *Scanning the human genome with combinatorial transcription factor libraries*. Nat Biotechnol, 2003. **21**(3): p. 269-74.
9. Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D., Barbas, C. F., 3rd, *Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors*. J Biol Chem, 2001. **276**(31): p. 29466-78.
10. Dreier, B., Fuller, R. P., Segal, D. J., Lund, C. V., Blancafort, P., Huber, A., Koksche, B., Barbas, C. F., 3rd, *Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors*. J Biol Chem, 2005. **280**(42): p. 35588-97.
11. Dreier, B., Segal, D. J., Barbas, C. F., 3rd, *Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains*. J Mol Biol, 2000. **303**(4): p. 489-502.
12. Segal, D.J., Dreier, B., Beerli, R. R., Barbas, C. F., 3rd, *Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences*. Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2758-63.
13. Ramirez, C.L., Foley, J. E., Wright, D. A., Muller-Lerch, F., Rahman, S. H., Cornu, T. I., Winfrey, R. J., Sander, J. D., Fu, F., Townsend, J. A., Cathomen, T., Voytas, D. F., Joung, J. K., *Unexpected failure rates for modular assembly of engineered zinc fingers*. Nat Methods, 2008. **5**(5): p. 374-5.
14. Sander, J.D., Maeder, M. L., Reyon, D., Voytas, D. F., Joung, J. K., Dobbs, D., *ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool*. Nucleic Acids Res, 2010. **38**(Web Server issue): p. W462-8.
15. Tan, S., Guschin, D., Davalos, A., Lee, Y. L., Snowden, A. W., Jouvenot, Y., Zhang, H. S., Howes, K., McNamara, A. R., Lai, A., Ullman, C., Reynolds, L., Moore, M., Isalan, M., Berg, L. P., Campos, B., Qi,

- H., Spratt, S. K., Case, C. C., Pabo, C. O., Campisi, J., Gregory, P. D., *Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity*. Proc Natl Acad Sci U S A, 2003. **100**(21): p. 11997-2002.
16. Persikov, A.V., Osada, R., Singh, M., *Predicting DNA recognition by Cys2His2 zinc finger proteins*. Bioinformatics, 2009. **25**(1): p. 22-9.
 17. Segal, D.J., Beerli, R. R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Kokscho, B., Lund, C. V., Magnenat, L., Valente, D., Barbas, C. F., 3rd, *Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins*. Biochemistry, 2003. **42**(7): p. 2137-48.
 18. Pensak, D.A., *Molecular modelling: scientific and technological boundaries*. Pure Appl Chem, 1989. **61**(3): p. 601-603.
 19. Humphrey, W., Dalke, A., Schulten, K., *VMD - Visual Molecular Dynamics*. J Mol Graphics, 1996. **14**: p. 33-38.
 20. Weiner, S.J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Weiner, P., *A new force field for molecular mechanical simulation of nucleic acids and proteins*. J Am Chem Soc 1984. **106**(3): p. 765-784.
 21. van Gunsteren, W.F., Berendsen, H. J. C., *Algorithms for macromolecular dynamics and constraint dynamics*. Mol Phys 1977. **34**(5): p. 1311-1327.
 22. Verlet, L., *Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules*. Phys Rev, 1967. **159**(1): p. 98-103.
 23. Manetti, M.E., Rossi, M., Costa, A. P., Clausen, A. M., Van Sluys, M. A., *Radiation of the Tnt1 retrotransposon superfamily in three Solanaceae genera*. BMC Evol Biol, 2007. **7**: p. 34.
 24. Li, L., Olvera, J. M., Yoder, K. E., Mitchell, R. S., Butler, S. L., Lieber, M., Martin, S. L., Bushman, F. D., *Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection*. EMBO J, 2001. **20**(12): p. 3272-81.
 25. Butler, S.L., Hansen, M. S., Bushman, F. D., *A quantitative assay for HIV DNA integration in vivo*. Nat Med, 2001. **7**(5): p. 631-4.
 26. Pierson, T.C., Kieffer, T. L., Ruff, C. T., Buck, C., Gange, S. J., Siliciano, R. F., *Intrinsic stability of episomal circles formed during human immunodeficiency virus type 1 replication*. J Virol, 2002. **76**(8): p. 4138-44.
 27. Butler, S.L., Johnson, E. P., Bushman, F. D., *Human immunodeficiency virus cDNA metabolism: notable stability of two-long terminal repeat circles*. J Virol, 2002. **76**(8): p. 3739-47.
 28. Sakhachornphop, S., Jiranusornkul, S., Kodchakorn, K., Nangola, S., Sirisanthana, T., Tayapiwatana, C., *Designed zinc finger protein interacting with the HIV-1 integrase recognition sequence at 2-LTR-circle junctions*. Protein Sci, 2009. **18**(11): p. 2219-30.
 29. Segal, D.J., Crotty, J. W., Bhakta, M. S., Barbas, C. F., 3rd, Horton, N. C., *Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA*. J Mol Biol, 2006. **363**(2): p. 405-21.
 30. Stote, R.H., Karplus, M., *Zinc binding in proteins and solution: a simple but accurate nonbonded representation*. Proteins, 1995. **23**(1): p. 12-31.

31. Jorgensen, W.L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., *Comparison of simple potential functions for simulating liquid water*. J Chem Phys, 1983. **79**(2): p. 926-935.
32. Onufriev, A., Bashford, D., Case, D. A., *Exploring protein native states and large-scale conformational changes with a modified generalized born model*. Proteins, 2004. **55**(2): p. 383-94.
33. Fu, F., Sander, J. D., Maeder, M., Thibodeau-Beganny, S., Joung, J. K., Dobbs, D., Miller, L., Voytas, D. F., *Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays*. Nucleic Acids Res, 2009. **37**(Database issue): p. D279-83.
34. Blancafort, P., Segal, D. J., Barbas, C. F., 3rd, *Designing transcription factor architectures for drug discovery*. Mol Pharmacol, 2004. **66**(6): p. 1361-71.
35. Zimmerman, K.A., Fischer, K. P., Joyce, M. A., Tyrrell, D. L., *Zinc finger proteins designed to specifically target duck hepatitis B virus covalently closed circular DNA inhibit viral transcription in tissue culture*. J Virol, 2008. **82**(16): p. 8013-21.
36. Lee, J., Kim, J. S., Seok, C., *Cooperativity and specificity of Cys2His2 zinc finger protein-DNA interactions: a molecular dynamics simulation study*. J Phys Chem B, 2010. **114**(22): p. 7662-71.
37. Beerli, R.R., Barbas, C. F., 3rd, *Engineering polydactyl zinc-finger transcription factors*. Nat Biotechnol, 2002. **20**(2): p. 135-41.
38. Kim, H.J., Lee, H. J., Kim, H., Cho, S. W., Kim, J. S., *Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly*. Genome Research, 2009. **19**(7): p. 1279-1288.

Designed zinc finger protein interacting with the HIV-1 integrase recognition sequence at 2-LTR-circle junctions

Supachai Sakkhachornphop,^{1,2} Supat Jiranusornkul,³ Kanchanok Kodchakorn,⁴ Sawitree Nangola,¹ Thira Sirisanthana,² and Chatchai Tayapiwatana^{1,5*}

¹Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand

²Research Institute for Health Sciences, Chiang Mai University, Chiang Mai 50200, Thailand

³Department of Pharmaceutical Sciences, Faculty of Pharmacy, Chiang Mai University, Chiang Mai 50200, Thailand

⁴Thailand Excellence Center for Tissue Engineering, Department of Biochemistry, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand

⁵Biomedical 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

Received 27 June 2009; Accepted 13 August 2009

DOI: 10.1002/pro.233

Published online 21 August 2009 proteinscience.org

Abstract: Integration of HIV-1 cDNA into the host genome is a crucial step for viral propagation. Two nucleotides, cytosine and adenine (CA), conserved at the 3' end of the viral cDNA genome, are cleaved by the viral integrase (IN) enzyme. As IN plays a crucial role in the early stages of the HIV-1 life cycle, substrate blockage of IN is an attractive strategy for therapeutic interference. In this study, we used the 2-LTR-circle junctions of HIV-1 DNA as a model to design zinc finger protein (ZFP) targeting at the end terminal portion of HIV-1 LTR. A six-contiguous ZFP, namely 2LTRZFP was designed using zinc finger tools. The designed motif was expressed and purified from *E. coli* to determine its binding properties. Surface plasmon resonance (SPR) was used to determine the binding affinity of 2LTRZFP to its target DNA. The level of dissociation constant (K_d) was 12.0 nM. The competitive SPR confirmed that 2LTRZFP specifically interacted with its target DNA. The qualitative binding activity was subsequently determined by EMSA and demonstrated the aforementioned correlation. In addition, molecular modeling and binding energy analyses were carried out to provide structural insight into the binding of 2LTRZFP to the specific and nonspecific DNA target. It is suggested that hydrogen-bonding interactions play a key role in the DNA recognition mechanisms of the designed ZFP. Our study suggested an alternative HIV therapeutic strategy using ZFP interference of the HIV integration process.

Keywords: HIV-1; AIDS; zinc finger protein; 2-LTR-circle junctions; molecular modeling

Abbreviations: AIDS, acquired immune deficiency syndrome; EMSA, electrophoretic mobility shift assay; GFP, green fluorescent protein; HIV-1, human immunodeficiency virus type 1; IN, integrase; LTR, long terminal repeat; MM/GBSA, molecular mechanics/generalized Born surface area; SPR, surface plasmon resonance; ZFP, zinc finger protein.

Grant sponsor: Royal Golden Jubilee Ph.D Program; Grant number: PHD/0112/2550; Grant sponsors: The Fogarty AIDS International for Research and Training Program, Johns Hopkins University, and the National Center for Genetic Engineering and Biotechnology (BIOTEC) of the National Science and Technology Development Agency, Thailand.

*Correspondence to: Chatchai Tayapiwatana, Division of Clinical Immunology, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai 50200, Thailand. E-mail: asimi002@chiangmai.ac.th

Introduction

Human immunodeficiency virus type 1 (HIV-1) is the etiologic agent of acquired immune deficiency syndrome (AIDS). It is one of the greatest health problems facing the world today.¹ According to UNAIDS and WHO, global estimates of the number of people living with HIV/AIDS as of the end of 2007 were 33 million adults and children.² HIV is a member of the family Retroviridae and subfamily lentivirinae. Soon after viral infection, the viral reverse transcriptase synthesizes a double-stranded DNA (ds-DNA). The subsequent insertion of viral genome into the host genome by integrase (IN) is a vital step in the HIV life cycle.³⁻⁵ Three forms of unintegrated viral DNA have been identified: a linear form which is flanked by long terminal repeat (LTR) copies at each terminus, and two circular molecules containing either one or two LTRs. The linear form is thought to be the substrate for integration.⁶⁻¹² Integrase is involved in removing 2 bp from the 3' ends of viral DNA, the so-called 3' processing. The second reaction catalyzed by IN, i.e., strand transfer, inserts both 3' ends of HIV-1 DNA into the host genome.^{13,14}

Current antiretroviral drugs therapy for AIDS involves the use of a multidrug cocktail referred to as highly active antiretroviral therapy (HAART).¹⁵ IN inhibitors belong to a novel class of drugs that function by inhibiting integration of proviral DNA into the host cell genome. IN inhibitors can be divided into two groups: 3' processing and selective strand transfer inhibitors. These drugs are currently undergoing clinical trials.^{16,17} However, several questions concerning the interaction of IN with its inhibitors remain unclear, and resistant strains have been found.¹⁸⁻²⁰ Therefore, a new approach targeting HIV-1 DNA at the level of the end terminal sequence of LTR by gene therapy may be a useful approach for inhibiting viral replication.

Our strategy herein is based on the specific DNA recognition by Cys₂His₂ zinc finger protein (ZFP). The Cys₂His₂ ZFP is the most common DNA-binding domain in the human and metazoan genome and plays a key role as a class of transcription factors. Its structure consists of a simple $\beta\beta\alpha$ fold of 30 amino acids in length, which is stabilized by hydrophobic interactions and chelation of a single zinc ion between two histidine residues and two cysteine residues.^{21,22} Within the residue at the N-terminus of the α -helix, the positions -1, 3, and 6 typically bind to three contiguous base pairs in a major groove of the target DNA.^{23,24} Therefore, several artificial ZFPs have been designed to target DNA sequences with high specificity and affinity.²⁵⁻²⁹

The ZF Tools server³⁰ implements an effective algorithm to generate ZFPs for *in silico* or *in vivo* applications. The program facilitates web-based design of ZFPs that specifically recognize DNA sequence motifs. In this study, we used the 2-LTR-circle junction

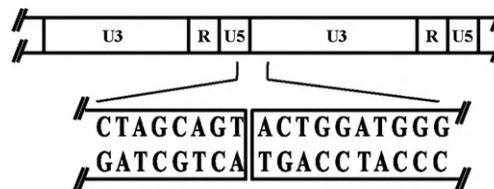


Figure 1. Schematic representation of the circle junction region of HIV-1.⁴⁵ Predicted organization of U3, R, and U5 in 2-LTR-circle junctions shown above, whereas sequence in box was selected for designing the ZFP.

tions of HIV-1 DNA as a model to design a ZFP targeted at the end terminal part of HIV-1 LTR, which is a region for 3' processing by IN of HIV-1. We designed a six-contiguous ZFP. Binding affinity was carried out by surface plasmon resonance (SPR) and an electrophoretic mobility shift assay (EMSA). Molecular modeling was also used to gain structural insight into the DNA recognition patterns of the designed ZFP, in the cases of being bound to either a specific or a nonspecific DNA sequence. Relative binding energies of the specific and nonspecific complexes were also calculated.

Results

Procurement of the target sequence and designing the 2LTRZFP

To target the HIV-1 DNA sequence of 2-LTR-circle junctions, the DNA sequence was submitted to the zinc finger tools using "Search DNA Sequence for Contiguous Target Sites" mode. Eight candidate DNA sequences were obtained and the binding prediction scores of these sequences were selected for designing the ZFP. The sequence that had the highest score was 5' CTAGCAGTACTGGATGGG 3' (Fig. 1).

The design tool generated the full-length amino acid sequence of 2LTRZFP. This protein was composed of 176 amino acids [Fig. 2(A)] for a six zinc finger motif. Each finger was linked with the canonical TGEKP linker.

The amino acid sequence of 2LTRZFP was reverse-translated into a nucleotide sequence, and all codons in this sequence were optimized [Fig. 2(B)]. The full-length optimized sequence was sent for full-gene synthesis by Blue Heron Biotechnology. Predicted properties of the resulting proteins were computed using ExPaSy proteomics tools. The resulting theoretical pI/M_w was 9.46/20.1 kDa. Prediction of the protein subcellular localization using the LOctree program indicated that 2LTRZFP is a nuclear protein and is found in the nucleus.

Construction of pTriEx-4- 2LTRZFP-GFP

The restriction endonucleases XcmI and SmaI in the plasmid vector pTriEx-4-GFP was used as a cloning sites for the 2LTRZFP gene fragment to construct the pTriEx-4-2LTRZFP-GFP containing His6 at N-

A

LEPGKEP	YKCECGKS	<u>FSRSDKLV</u> R	HQRTH	TGEKP
	YKCECGKS	<u>FSTSGNLV</u> R	HQRTH	TGEKP
	YKCECGKS	<u>FSRNDALTE</u>	HQRTH	TGEKP
	YKCECGKS	<u>FSQSSSLV</u> R	HQRTH	TGEKP
	YKCECGKS	<u>FSQSGDLRR</u>	HQRTH	TGEKP
	YKCECGKS	<u>FSQNSTLTE</u>	HQRTH	TGKKTG

B

```

5' ctggaaccgg gcaaaaacc gtacaaatgc cgggaatgcg gtaaaagctt
cagccgcagc gacaaatggt tgcgccacca ggcacccac accggcgaaa
aacggtacaa atgcccgtaa tgcggcaaaa gcttcagcac ctctggcaac
ctgggtgcgc accagcgcac ccacaccggc gaaaaccgt acaaatgccc
ggaatgctgc aaaagcttca gccgcaacga cgcctgacc gaacaccagc
gcacccacac cggcgaaaaa cgtacaaat gcccggaatg cggcaaaagc
ttcagccaga gcagcagcct ggtgcgccac cagcgcaccc acaccggcga
aaaaccgtac aatgcccggt aatgcggcaa aagcttcagc cagagcggcg
acctgcgcgc ccaccagcgc acccacaccg gcaaaaaacc gtacaaatgc
ccggaatgcg gcaaaaagctt cagccagaac agcacctga cgaacacca
gcgcacccac accggcaaaa aaaccagc 3' (528 bp)

```

C

			-1	3	6		
Aart	GEKP	YACPECGKS	FSRSDHLAE	HQRTH	TGEKP	32	
2LTRZFP	GEKP	<u>YKCECGKS</u>	<u>FSRSDKLV</u> R	HQRTH	TGEKP		
		YKCECGKS	FSDKIDLTR	HQRTH	TGEKP	60	
		YKCECGKS	<u>FSTSGNLV</u> R	HQRTH	TGEKP		
		YKCECGKS	FSQRANLRA	HQRTH	TGEKP	88	
		YKCECGKS	<u>FSRNDALTE</u>	HQRTH	TGEKP		
		YKCECGKS	FSQLAHLRA	HQRTH	TGEKP	116	
		YKCECGKS	<u>FSQSSSLV</u> R	HQRTH	TGEKP		
		YKCECGKS	FSREDNLHT	HQRTH	TGEKP	144	
		YKCECGKS	<u>FSQSGDLRR</u>	HQRTH	TGEKP		
		YKCECGKS	FSRRDALNV	HQRTH	TGEKP	167	
		YKCECGKS	<u>FSQNSTLTE</u>	HQRTH	TGEKP		

N-term backbone Recognition helix C-term backbone ZF linker

Figure 2. (A) Full-length amino acid sequence of 2LTRZFP. The amino acids shown by underline were located at the specific positions of -1, 3, and 6 from the N-terminal of the α -helix, respectively. (B) Full-length optimized nucleotide sequence of 2LTRZFP. (C) Primary sequence alignments of Aart (PDB code 2I13) and 2LTRZFP with 80% sequence identity. The underlined letters represent amino acid residues of the ZFP those are different from the Aart peptide. Positions -1, 3, and 6 above the recognition helix column represent amino acid residues often involved in DNA recognition. Numbers at the end of each line show the running number of the last amino acid residue.

terminal and fused-GFP at C-terminal under the control of the CMV, T7, or p10 promoters. The plasmid vector pTriEx-4-2LTRZFP-GFP was digested by the restriction endonucleases XcmI and SmaI and the product was subjected to 1% agarose gel electrophoresis and visualized by ethidium bromide staining. Restriction digest analysis showed that the pTriEx-4-2LTRZFP-GFP was constructed successfully at a 547 bp band (data not shown).

Protein expression and purification

2LTRZFP-GFP was expressed in its recombinant forms in *E. coli* Origami B (DE3) Novagen (Madison, WI).

Protein purification was performed using His-Bind column chromatography. The entire expression and purification processes were monitored by SDS-PAGE [Fig. 3(A)]. Comparison between lane 2 [total lysate of *E. coli* Origami B (DE3)] and lane 3 [Total lysate of *E. coli* Origami B (DE3) with pTriEx-4-2LTRZFP-GFP was induced by isopropyl β -D-thiogalactopyranoside (IPTG) after OD₆₀₀ reach 1.0 at 30°C for overnight] indicates that the expression of fusion protein His6-2LTRZFP-GFP (~50 kDa) was successful. Lane 6 shows that this fusion protein bound to the His-Bind column chromatography, and it can be eluted by elution buffer containing 1M imidazole, 0.5M NaCl, 20 mM Tris-HCl, pH 7.9 (lane 7).

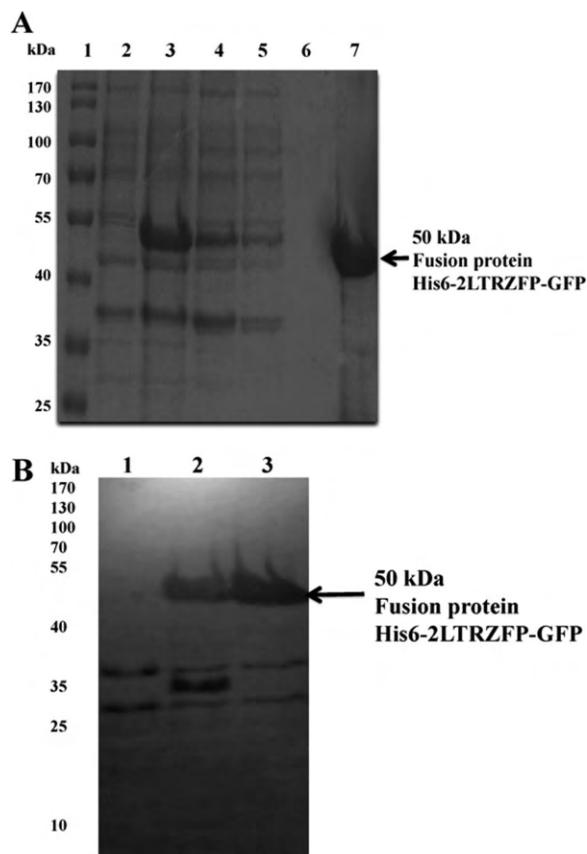


Figure 3. (A) SDS-PAGE analysis of His6-2LTRZFP-GFP at different and purification steps. Lane 1: protein marker, lane 2: total bacterial lysate, lane 3: total bacterial lysate with pTriEx-4-2LTRZFP-GFP after IPTG induction for overnight, lane 4: pass through lysate, lanes 5, 6: solution after washing the column with binding buffer and washing buffer, respectively, lane 7: elution of purified His6-2LTRZFP-GFP. The arrow indicates the size of 50 kDa of His6-2LTRZFP-GFP. Numbers on left show size in kilo Dalton of protein ladder. (B) Western blot analysis of His6-2LTRZFP-GFP at different condition of protein expression. Lane 1: total bacterial lysate, lanes 2, 3: total bacterial lysate with pTriEx-4-2LTRZFP-GFP after IPTG induction for 4 h, and overnight, respectively. The arrow indicates the size of 50 kDa of His6-2LTRZFP-GFP. Numbers on left show size in kilo Dalton of protein ladder.

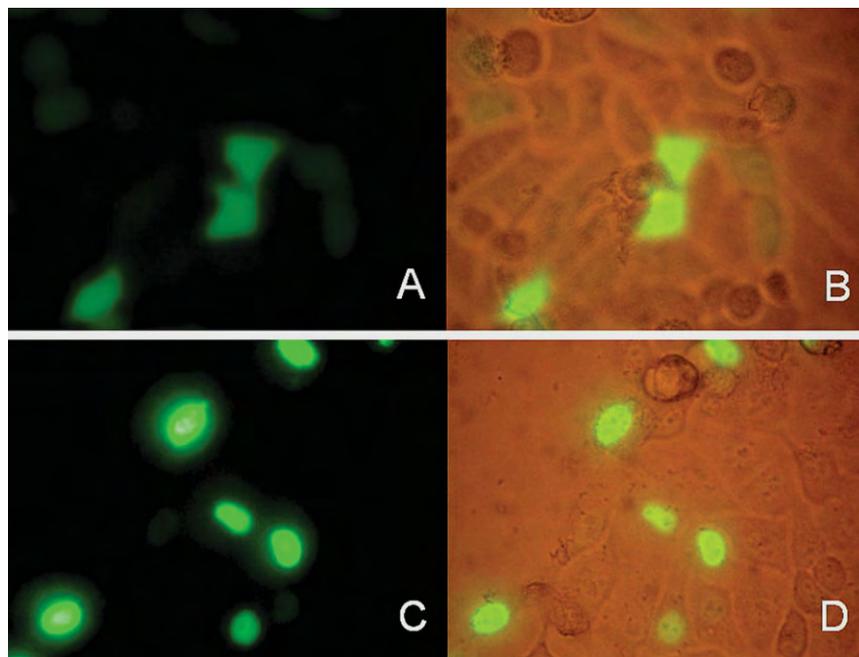


Figure 4. Expression of green fluorescent protein in HeLa cells 24 h after transfection. (A, B) Transfected with recombinant plasmid pTriEx-4-GFP. (C, D) Transfected with recombinant plasmid pTriEx-4-2LTRZFP-GFP.

Western blotting was performed by using anti-His mAb as primary antibody [Fig. 3(B)]; the analysis demonstrated that His6-2LTRZFP-GFP (~50 KDa) was successfully expressed. To separate GFP and 2LTRZFP, the purified recombinant 2LTRZFP-GFP containing the thrombin recognition sequence was cleaved via thrombin, and 2LTRZFP was purified by using His-Bind column chromatography. (Data not shown).

Expression of 2LTRZFP-GFP in HeLa cells

To test the expression of the 2LTRZFP-GFP recombinant protein in a mammalian system, pTriEx-4-2LTRZFP-GFP was transfected into a HeLa cell line. The emitted green fluorescence was observed under the microscope 24 and 48 h after transfection. The green fluorescence was observed in the nucleus in the pTriEx-4-2LTRZFP-GFP transfection group and in the cytoplasm for pTriEx-4-GFP control group (Fig. 4). This result indicated that 2LTRZFP-GFP recombinant protein is a nuclear protein and confirmed the successful folding of the protein.

Evaluating of dissociation constants and competitive DNA binding activity by SPR

To determine real-time binding kinetics between 2LTRZFP-GFP and its target DNA sequence, we used SPR as a tool for qualitative analysis. Different concentrations of 2LTRZFP-GFP were injected into the immobilized chip with specific ds-DNA. The 2LTRZFP-GFP could bind to its target ds-DNA on a nanomolar scale, with $K_d = 12.0$ nM [Fig. 5(A)]. We also performed a competitive SPR to determine the specificity

of binding between 2LTRZFP-GFP and its target DNA. The concentration of 2LTRZFP-GFP (1.2 μ M) was used to compete with 26.5 μ M of nonbiotinylated target ds-DNA and nonspecific ds-DNA before injection onto the immobilized chip with specific ds-DNA. The 2LTRZFP-GFP could bind to its target DNA sequence [result compared with nonspecific ds-DNA, Fig. 5(B)]. In addition, the specificity of binding activity was also assayed by injection of GFP and bovine serum albumin (BSA) as controls. These results demonstrated that the binding activity was involved with ZFP only [Fig. 5(A)]. To obtain the binding properties, we also tested the binding activity of 2LTRZFP-GFP and its target DNA sequence in zinc buffer containing 1 μ M, 1 mM, and 10 mM of EDTA. The binding activity was decreased when the concentration of EDTA was high. These findings supported the conclusion that the binding between 2LTRZFP-GFP and its target DNA sequence depended on zinc ions, and that this metal ion is important for stabilizing ZFP [Fig. 5(C)].

DNA binding activity by EMSA

To confirm the binding activity, EMSA was performed using an EMSA kit as described in the section Materials and Methods. A complete binding complex of 1 μ M of 2LTRZFP-GFP and 250 nM of its specific DNA duplex could be observed [see Fig. 6(A), lane 8] until no band of free DNA duplex was seen. At the same time, a faint band of binding complex was observed in the reaction of 2LTRZFP-GFP and nonspecific DNA duplex [Fig. 6(A), lane 9], whereas the intensity of free DNA duplex was not changed [Fig. 6(A), lanes 3 and 6]. This experiment also demonstrated that GFP in

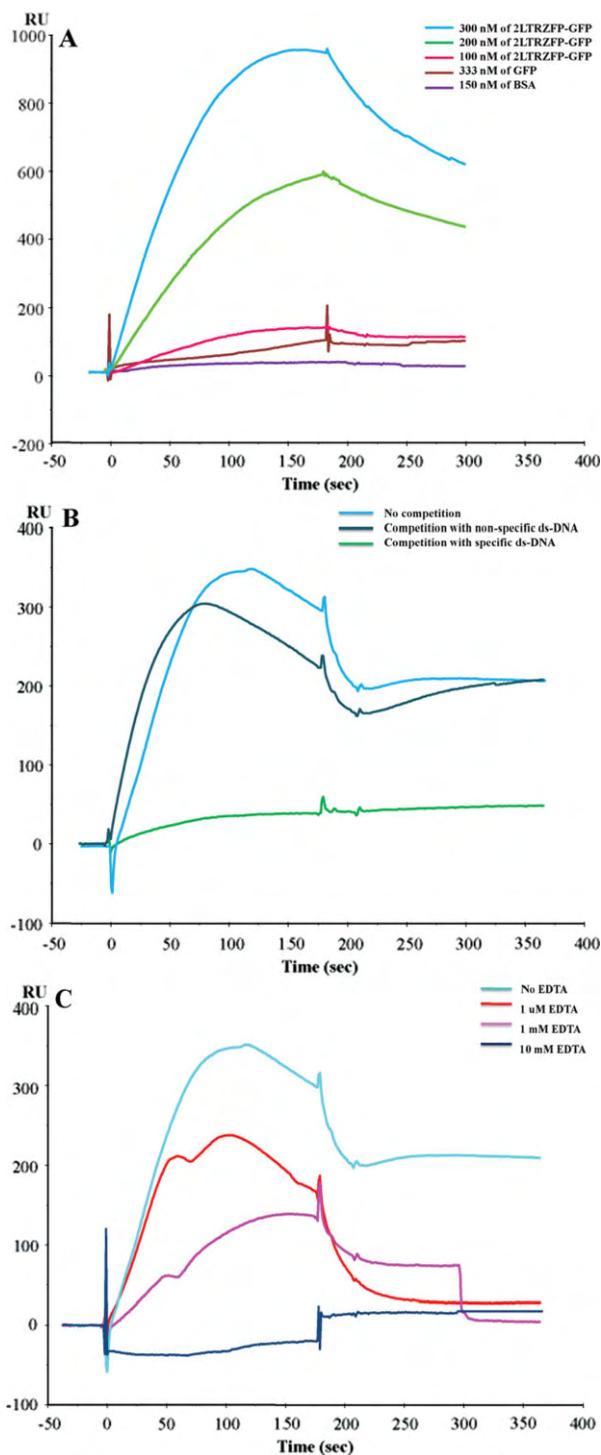


Figure 5. Kinetic analysis between 2LTRZFP-GFP and its target DNA sequence. (A) Binding kinetic of 2LTRZFP-GFP (300, 200, and 100 nM), GFP, and BSA to the immobilized chip with specific ds-DNA. (B) For the competitive SPR, 2LTRZFP-GFP was incubated with nonbiotinylated ds-DNA of its target ds-DNA, and nonspecific ds-DNA in zinc buffer for 15 min before injection. (C) 2LTRZFP-GFP was injected to determine the binding activity with its target DNA sequence in zinc buffer containing 1 μ M, 1 mM, and 10 mM of EDTA.

2LTRZFP-GFP recombinant protein was not involved in the binding [lanes 4, 5, and 6 of Fig. 6(A)]. The dilution effect of DNA binding complex can be observed in Figure 6(B,C).

2LTRZFP-DNA models and their binding energy analyses

Molecular models of the designed 2LTRZFP-DNA complexes, Zif1 and Zif2, were remodeled based on the crystal structure of Aart bound to DNA (PDB code 2I13, see more details in the Materials and Methods), where all backbone atoms were maintained in the same position as found in the crystal structure. Amino acid residues within helical helices were modified solely according to our 2LTRZFP sequence [Fig. 2(C)]. Pair-wise similarity comparison between Aart and 2LTRZFP proteins was also performed using the blastp algorithm of the BLAST tool.³¹ The result showed a relatively high sequence identity (80%).

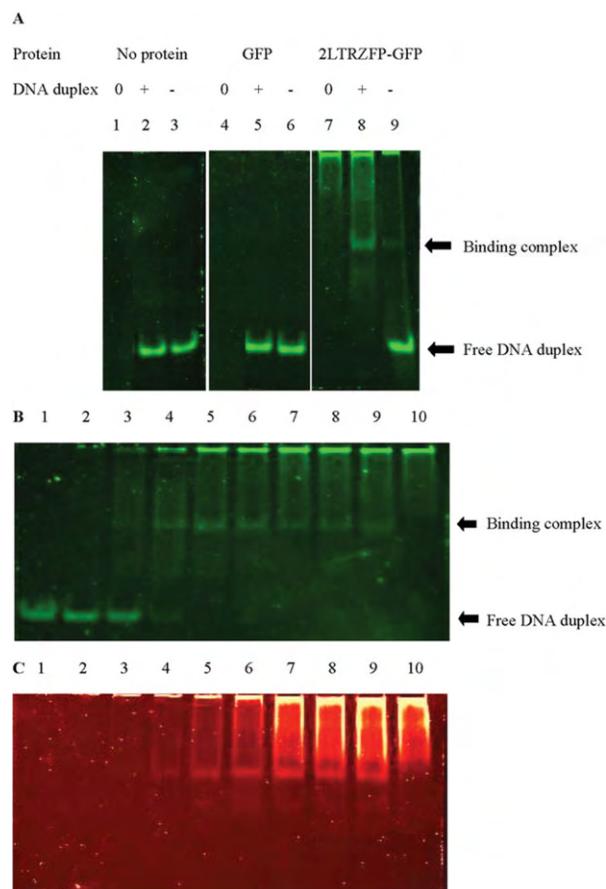


Figure 6. Electrophoretic mobility shift assay (EMSA). (A) The protein used for each reaction is 1 μ M, whereas [0], no DNA duplex; [+], specific DNA duplex; [-], nonspecific DNA duplex. The gel was stained with SYBR[®] Green EMSA stain (B) or with SYPRO[®] Ruby EMSA stain (C). In panels B and C: lanes 1 to 9, the specific DNA duplex used for each reaction was 250 nM, whereas 2LTRZFP-GFP was varied, the concentration from 0, 0.125, 0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0 μ M, respectively. Lane 10, 2LTRZFP-GFP was used 3.0 μ M, but no DNA duplex.

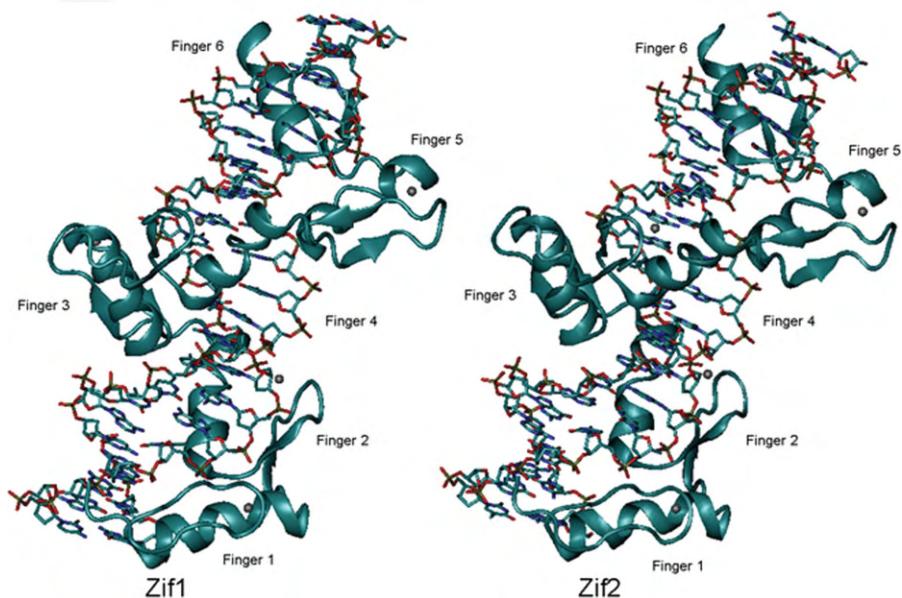


Figure 7. Comparisons of conformation of the recognition binding region in Zif1 (in left panel) and Zif2 (in right panel). The nucleotides are shown in a licorice model and the protein backbone in a cartoon diagram.

After a 2500-cycle minimization for each of Zif1 and Zif2 complex, the root mean square deviation (RMSD) of the minimized structure of each model was measured. The RMSDs of 2LTRZFP and DNA backbones were 0.51 and 0.94 Å, respectively, suggesting that Zif1 and Zif2 are relatively similar in conformation (Fig. 7). As the first crystal structure of Zif268 bound to DNA was obtained, it has been suggested that amino acid positions -1, 3, and 6, with respect to the start of the recognition helix, play specific roles for binding to three contiguous base pairs in each DNA triplet.^{23,24} The residue-by-residue analysis of 2LTRZFP-DNA interaction for each residue in the recognition helix to each DNA triplet compared to its nonspecific triplet was subsequently explored to understand a DNA recognition pattern for 2LTRZFP.

Figure 8 demonstrates all possible hydrogen-bonding interaction between each zinc finger motif and DNA residues in both specific (Zif1) and nonspecific (Zif2) complexes, and the distances are listed in Table I. In general, a major contribution to stabilizing the 2LTRZFP-DNA complexes is the electrostatic interactions between amino acid residues in the zinc finger motif and nucleotides in the DNA target, where there is no water-mediated interaction apparent in these complexes.

In finger 1, although the hydrogen bonding network among ARG16, LYS19, and ARG22 to its DNA target in Zif1 and Zif2 is similar, the hydrogen bonds in Zif2 are weaker than those of Zif1 with respect to its longer bonding distances. The hydrogen-bond distances in finger 2 are very similar in Zif1 and Zif2, whereas they are completely different in finger 3. The amino N atom of ARG72 is bonded to the carbonyl group of DG181 in

the Zif1 complex; however, there is no hydrogen bonding between ARG72 and DC181 in Zif2 because of repulsive interaction of the positively charged amino group. The O atom of ASP74 is hydrogen bonded to the N atom of DC195 in Zif1, whereas, in Zif2, the steric effect of the methyl group in DT195 could lead to an improper formation of hydrogen bonds. A hydrogen bond donor at the amino N4 atom of DC179 (Zif1) efficiently establishes a hydrogen bond with GLU78, where the amino group of DA179 (Zif2) is located away from GLU78. Hydrogen-bond patterns in fingers 4, 5, and 6 in Zif1 and Zif2 are similar. The negatively charged amino acids, aspartic acid and glutamic acid, may play an important role in DNA recognition as being involved in selective contacts with cytosine only. Moreover, serine residues in position -2 of the recognition helices are likely to selectively bind to the phosphate backbone of the target DNA as well as arginine residues in position 6 that can bind to all bases of DNA. Finally, from the minimized structure, 2LTRZFP is more probable to establish hydrogen bonds within Zif1 than Zif2, resulting in the more selective binding to the target DNA.

In addition to the structural analysis, the molecular mechanics/generalized Born surface area (MM/GBSA) approach was used to further calculate relative binding energies between the designed peptide and the target DNA compared to its nonspecific sequence (Zif2). The $\Delta\Delta G_{\text{binding}}$ reveals that 2LTRZFP is more favorable for binding to the target DNA in Zif1 than that of Zif2 by -35.0 ± 0.5 kcal/mol (see Table II). This suggests that the Zif1 complex is mainly stabilized by the electrostatic interaction (E_{es}) as the better hydrogen-bonding network is presented.

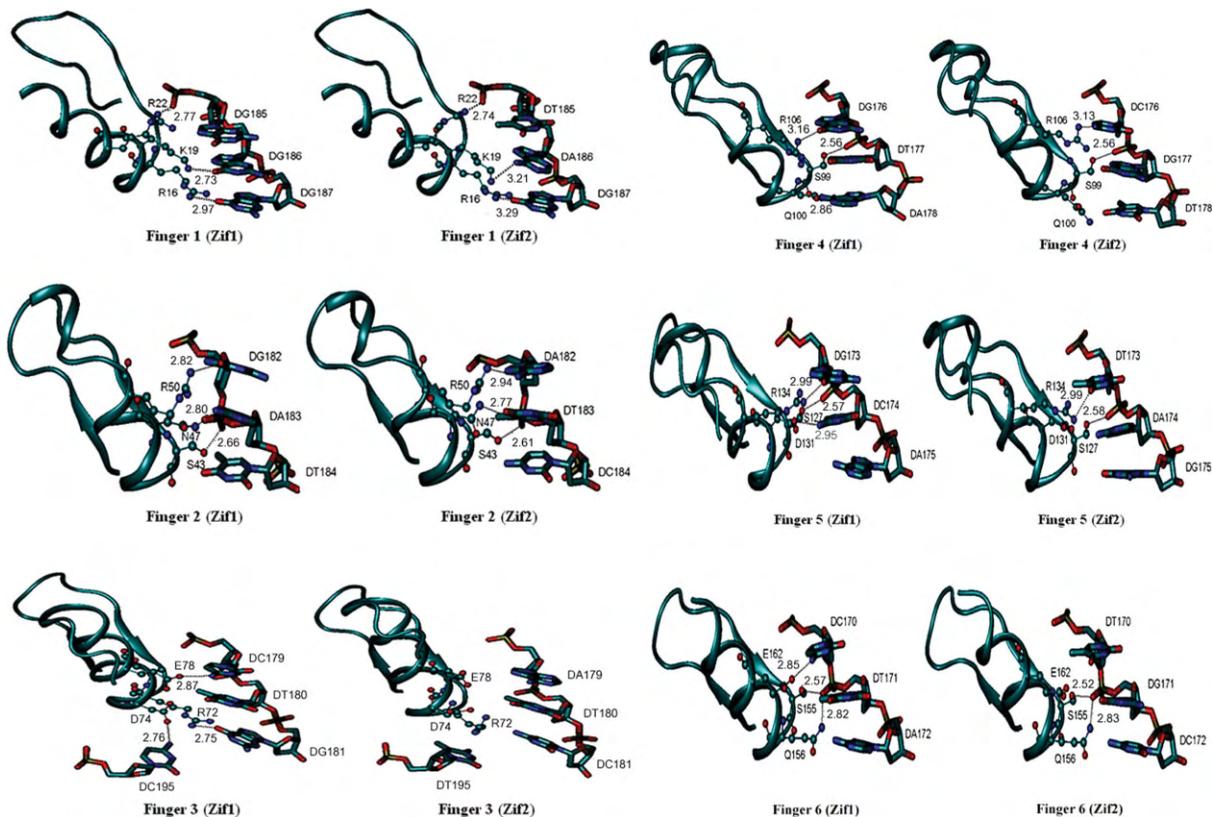


Figure 8. Possible hydrogen bonding between the zinc finger residues and the target DNA triplet within distance $< 3.5 \text{ \AA}$ between H-bond donor and H-bond acceptor associated with the distance.

Discussion

The Cys₂His₂ ZFPs have been demonstrated to be specific binding proteins to the recognition sequences.^{25–29} In this study, a six-finger protein was constructed specifically to 18 bp of 2-LTR-circle junctions. The first, second, and third fingers can target the terminal sequence of 5' LTR, whereas the terminal sequence of

3' LTR can be targeted by the other three fingers. The 2LTRZFP-GFP was expressed and purified from *E. coli* to investigate its binding properties. The binding activity was on the nanomolar scale, which is similar to the affinity level of HIV-1 IN.^{32,33} This finding suggested that the HIV-1 IN might be interfered with by the 2LTRZFP as well. The specificity of this binding was

Table I. Hydrogen-Bonding Distances between 2LTRZFP Residues and DNA Residues

	2LTRZFP residue	DNA residue	Distance (Å)		2LTRZFP residue	DNA residue	Distance (Å)	
Finger 1	ARG16(NH)	Zif1: DG187(O6)	2.97	Finger 4	SER99(OH)	Zif1: DT177(O2P)	2.56	
		Zif2: DG187(O6)	3.29				Zif2: DT177(O2P)	2.56
	LYS19(NZ)	Zif1: DG186(O6)	2.73			GLN100(O)	Zif1: DA178(N6)	2.86
Finger 2	ARG22(NH)	Zif1: DG185(O2P)	2.77			Zif2: DT178	–	
		Zif2: DT185(O2P)	2.74	Finger 5	ARG106(NH)	Zif1: DG176(O6)	3.16	
	SER43(OH)	Zif1: DA183(O1P)	2.66				Zif2: DC176(N4)	3.13
		Zif2: DT183(O1P)	2.61			SER127(OH)	Zif1: DC174(O2P)	2.57
	ASN47(ND)	Zif1: DA183(O2P)	2.80				Zif2: DA174(O2P)	2.58
	Zif2: DT183(O2P)	2.77			ASP131(OH)	Zif1: DC174(N4)	2.95	
Finger 3	ARG50(NH)	Zif1: DG182(O6)	2.82			Zif2: DA174	–	
		Zif2: DA182(N6)	2.94	Finger 6	ARG134(NH)	Zif1: DG173(O6)	2.99	
	ARG72(NH)	Zif1: DG181(O6)	2.75				Zif2: DT173(O4)	2.99
		Zif2: DC181	–			SER155(OH)	Zif1: DT171(O2P)	2.57
	ASP74(OH)	Zif1: DC195(N4)	2.76				Zif2: DG171(O2P)	2.52
	Zif2: DT195	–			GLN156(NH)	Zif1: DT171(O2P)	2.82	
	GLU78(OH)	Zif1: DC179(N4)	2.87			Zif2: DG171(O2P)	2.83	
		Zif2: DA179	–		GLU162(OH)	Zif1: DC170(N4)	2.85	
			–			Zif2: DT170	–	

Atom names are indicated in parentheses.

Table II. Estimated Relative Binding Energies of Specific (Zif1) and Nonspecific (Zif2) Complexes from the MM/GBSA Approach

Sequence	$\Delta\Delta E_{vdW}$ (kcal mol ⁻¹)	$\Delta\Delta E_{es}$ (kcal mol ⁻¹)	$\Delta\Delta G_{pol}$ (kcal mol ⁻¹)	$\Delta\Delta G_{nonpol}$ (kcal mol ⁻¹)	$\Delta\Delta G_{binding}$ (kcal mol ⁻¹)
Zif1-Zif2	-3.1 ± 1.7	-255.9 ± 9.0	226.3 ± 8.2	-2.3 ± 0.0	-35.0 ± 0.5

obtained by competitive SPR, which indicated that there was specific binding of 2LTRZFP-GFP and its target DNA. However, this binding affinity was decreased when the concentration of EDTA in binding buffer was increased. This data demonstrated the influence of zinc ion on correct folding of 2LTRZFP domains in recognizing a specific DNA sequence. Moreover, 2LTRZFP can specifically bind to its target ds-DNA, whereas binding of GFP in the C-terminal part was not facilitated.

Additionally, theoretical studies were used to investigate at the atomic level the DNA recognition of 2LTRZFP. Although it is well known that amino acid positions -1, 3, and 6 within the recognition helix directly contact the DNA triplet, positions -2, 1, and 5 are also described as playing a role in binding to the phosphate backbone.³⁴ Our study suggests that serine at position -2 and arginine at position 6 are very important in binding to the phosphate DNA backbone and 5'-nucleotide, respectively. However, the roles of amino acid positions -1 and 3 remain unclear. The results herein are particularly interesting because single or multiple mutations of amino acid position -1 and 3 would be able to enhance selectivity of the ZFP motif to the 2-LTR-circle junctions.

Mutation of the IN enzyme encoded by *pol* gene can occur,¹⁸⁻²⁰ while numerous studies have reported that the important 2-7 base pairs upstream and the well-conserved CA dinucleotide at the 3' end of viral cDNA play key roles in processing and strand transfer.³⁵⁻⁴⁰ Therefore, the advantage of targeting the 3'-end terminal part of HIV-1 LTR to interfere with the IN enzyme is an attractive strategy for a new approach to gene therapy.

Many investigations have targeted viral replication by ZFP recognition of the DNA of viruses. The engineered ZFPs targeting the Sp1 binding site in promoter region of HIV-1 have been reported, and one of them was found to inhibit HIV-1 replication by 75%.⁴¹ In 2004, Segal *et al.* demonstrated that a transcriptional repressor protein, namely KRAB-HLTR3, was able to achieve 100-fold repression of transcription from the HIV-1 promoter. This transcription factor also repressed the replication of several HIV-1 strains 10- to 100-fold in T-cell line and primary blood mononuclear cells with no significant cytotoxicity.⁴² Recently, ZFPs were designed to bind DNA sequences in the duck hepatitis B virus to inhibit viral transcription in tissue culture. Two candidate ZFPs decreased production of viral products and progeny viral genomes.²⁵ These studies

supported the use of ZFPs to inhibit viral replication, opening new avenues in gene therapy. There are a various methods available to deliver a protein for gene therapy. The *in vivo* adenoviral gene transfer of ZFPs has been performed to induce angiogenesis in a mouse model,⁴³ whereas retroviral and lentiviral gene transfer have been delivered the artificial zinc finger transcription factors to bind sites in the HIV-1 promoter to repress the replication of several HIV-1 strains.⁴² Recently, various nonbiological and biological carrier systems have been developed for HIV-1 gene therapy, such as nanoparticles, liposomes, or synthetic polymers, which can be taken up by many cell types.⁴⁴

Our study has demonstrated that 2LTRZFP can specifically bind to the HIV-1 integrase recognition sequence at 2-LTR-circle junctions. The findings may be applied for limiting viral integration to maximize the impact of HIV gene therapy in the future.

Materials and Methods

Zinc finger protein design

Studying order to identify target sites within the DNA sequence, the HIV-1 DNA sequence of 2-LTR-circle junctions⁴⁵ was submitted to the ZF Tools server of the Barbas Laboratory of the Scripps Research Institute.³⁰ The minimum target size of 18 base pairs was set to obtain the output DNA target sequences and amino acid sequences for the six fingers of the ZFP. This amino acid sequence was then reverse-translated into a nucleotide sequence. Codon usage of the amino acid sequences of the selected ZFP was optimized by using the JAVA Codon Adaptation Tool (JCat)⁴⁶ and by manual optimization. The designed ZFP was further modified with flanking XcmI and SmaI restriction endonuclease sites. The full-length optimized DNA sequence was sent for full-gene synthesis to Blue Heron Biotechnology (Bothell, WA). Predicted properties of the resulting proteins were computed using ExPaSy proteomics tools.⁴⁷ Prediction of protein subcellular localization was done using the LOctree program.⁴⁸

Plasmid construction

The full-length optimized DNA sequence of 2LTRZFP was synthesized and cloned into pUC19 vectors by Blue Heron Biotechnology. The 2LTRZFP gene fragment was ligated to pTriEx-4-GFP using the flanking XcmI and SmaI sites to construct pTriEx-4-2LTRZFP-

GFP, creating an N-terminal His6-fusion protein with GFP in the C-terminal part. This plasmid has a multiple expression system driven by the CMV, T7, or p10 promoters. The ligation product was transformed into the competent *E. coli* XL-1 Blue cells and plated on Luria-Bertani (LB) agar containing 100 µg/mL of ampicillin. The plasmid miniprep was performed using a QIAGEN Miniprep Kit (Qiagen, Hilden, Germany). The constructed plasmid was preliminarily identified by the restriction endonucleases XcmI and SmaI. PCR and DNA sequencing were performed for confirmation.

Expression and purification of His6-2LTRZFP-GFP

We used *E. coli* Origami B (DE3) as the expression strain. A 10 mL Terrific broth preculture, supplemented with 100 µg/mL ampicillin, 12.5 µg/mL kanamycin, and 12.5 µg/mL tetracycline was incubated at 37°C until reaching an absorbance at 600 nm of about 1.0–1.8. The preculture was diluted 1:100 into 100 mL Terrific broth medium with 100 µg/mL ampicillin, 12.5 µg/mL kanamycin, 12.5 µg/mL tetracycline and supplemented with 100 µM Zn₂SO₄. The culture was incubated at 37°C. When an absorbance about 1.0 at 600 nm was reached, the bacterial culture was induced by 0.1 mM IPTG at 30°C for overnight. Bacteria were then harvested by centrifugation (7500g at 4°C for 15 min). The bacterial pellets were resuspended in 6.25 mL of B-PER II Bacterial Protein Extraction Reagent (Pierce, Rockford, IL) and lysed by ultrasonication in an ice bath. The lysate was centrifuged at 15,000g for 20 min at 4°C. The supernatant was collected and filtered with microfiltration membranes (0.22-µm pore size). The clear solution containing His6-2LTRZFP-GFP was applied to His-bind column chromatography (Novagen, San Diego, CA) for protein purification. The eluted fraction containing His6-2LTRZFP-GFP was observed using Coomassie brilliant blue R250 stain (Bio-Rad, Hercules, CA). The purified protein was kept in final concentration 25% (W/V) glycerol for long-term storage at –80°C. The protein concentration was quantified by using the Micro-BCA protein assay (Pierce). All the samples were analyzed by SDS-PAGE. Western blot analysis was performed on a Hybond-P polyvinylidene fluoride membrane. (Amersham Bioscience, Piscataway, NJ). After being blocked with 5% skim milk in PBS, the proteins were probed with anti-his-tag monoclonal antibody (Genscript, Piscataway, NJ) as a primary antibody and horseradish peroxidase-labeled goat anti-mouse (Sigma, St Louis, MO) as a secondary antibody, using the ECL system (GE Healthcare, Buckinghamshire, UK).

Cell culture and transfection of HeLa cells

HeLa cervical carcinoma cells were kindly obtained from Dr. A. Lieber, University of Washington, Seattle,

WA. HeLa cells (1×10^5) were seeded onto 24-well plate in humidified atmosphere under 5% CO₂ at 37°C in Dulbecco's Modified Eagle's medium (DMEM) (Gibco, Grand Island, NY) containing penicillin (100 U/mL), streptomycin (100 µg/mL), and 2 mM L-glutamine, supplemented with 10% fetal bovine serum (FBS) (Gibco) for 24 h before transfection. Transfection mixture was prepared by adding 1 µg of the pTriEx-4-2LTRZFP-GFP or control vector pTriEx-4-GFP and 2 µL of GeneJamer (Stratagene, La Jolla, CA) into DMEM up to 200 µL. Then, the mixture was incubated at room temperature for 10 min. The culture supernatant was discarded and 300 µL of fresh DMEM containing 10% FBS and antibiotics was added. The transfection mixture was added to the cells and incubated at 37°C in 5% CO₂ for 5 h. A total of 500 µL of DMEM containing 10% FBS and antibiotics was added and further cultured at 37°C in 5% CO₂ for 24–48 h. Green fluorescent protein (GFP) was observed under a fluorescent microscope.

Double-stranded DNA preparation

ds-DNA of specific and nonspecific target DNA was prepared for testing the binding activity of 2LTRZFP-GFP. A pair of specific ds-DNA (sense) was designed as follows: 5'-AAA TCT CTA GCA GTA CTG GAT GGG CTA ATT-3' and a pair of nonspecific ds-DNA (sense) was also designed as follows 5'-TGA CAG TGC TAG CGT ATC ATC TAG TCG ACG-3'. The specific sequences to 2LTRZFP-GFP are shown in italic. The reaction mixture (100 µL) for annealing was composed of each 600 pmol single-stranded DNA (ss-DNA) and complementary strand in 50 mM NaCl. The mixture was heated at 95°C for 5 min and then slowly cooled to room temperature for 90 min. For immobilization of ds-DNA on a sensor chip SA (Biacore AB, Uppsala, Sweden) for SPR, the 3' end of the antisense strand of specific target DNA was labeled with biotin.

Surface plasmon resonance

SPR was carried out on a BIACORE 2000™ biosensor (Biacore AB). The target DNA duplexes were immobilized by injecting the biotinylated specific ds-DNA of 0.5 µg/mL in 0.3M NaCl on a Sensor Chip SA (Biacore AB) at a flow rate 5 µL/min in running buffer (HBS-EP), containing 10 mM HEPES (pH 7.4), 150 mM NaCl, 3 mM EDTA, 0.005% (v/v) surfactant P20 (Biacore AB). Typically 800–1000 RUs of the target ds-DNA were immobilized. After DNA immobilization, the chip was washed with 50 mM NaOH/1M NaCl and then primed with zinc buffer [10 mM Tris-HCl, pH 7.5, 90 mM KCl, 1 mM MgCl₂, 90 µM ZnSO₄, 5 mM dithiothreitol, and 0.5 mM phenylmethylsulfonylfluoride), which was used as a running buffer for a period of binding analysis. ZFP was diluted in zinc buffer and 60 µL was injected at a flow rate of 20 µL/min, followed by a dissociation phase of 180 s. Before each injection, the baseline stability was achieved by

injecting 1M NaCl for 2 min. To calculate the binding constants (K_d), the kinetic parameters were evaluated with BIA evaluation software 3.1 (Biacore AB) using a 1:1 binding model with mass transfer. For competitive SPR, 2LTRZFP-GFP was incubated with different concentrations of nonbiotinylated ds-DNA of its target ds-DNA, and nonspecific ds-DNA in zinc buffer for 15 min before injection.

Electrophoretic mobility shift assay

2LTRZFP-GFP or GFP was incubated with 250 nM of DNA duplex at room temperature for 1 h in zinc buffer. Total volumes were 10 μ L per reaction. The reaction mixtures were mixed with 2 μ L of 6 \times EMSA gel-loading solution (component D) of EMSA kit [E33075] (Invitrogen, Paisley, UK) before being loaded on 5% nondenaturing polyacrylamide gels using 100 V, 30 min. Gels were stained by using two fluorescent dyes for detection—SYBR[®] green EMSA nucleic acid gel stain (component A) and SYPRO[®] ruby EMSA protein gel stain (component B) by following the protocol from the same kit. The stained gels were imaged at an excited state of 488 nm using a Typhoon Trio phosphorImager (GE Healthcare Biosciences, Piscataway, NJ).

Molecular modeling of 2LTRZFP-DNA complexes

Construction of the 2LTRZFP-DNA complexes was initiated using an X-ray crystallographic structure of a designed six zinc finger motif (Aart) bound to an oligonucleotide from PDB entry 2I13³⁴ as a template. The sequence of the Aart peptide was then modified (using the Discovery Studio version 1.7 program package) in recognition helices according to our designed ZFP resulting from ZF Tools.³⁰ The DNA duplex from the crystal structure of the Aart-DNA complex was further remodeled to 21-mers oligonucleotides of 2-LTR-circle junctions, as follows:

5'-CTCTAGCAGTACTGGATGGGC-3'

3'-GAGATCGTCATGACCTACCCG-5' termed Zif1

as a specific complex

and

5'-AGTGCTAGCGTATCATCTAGT-3'

3'-TCACGATCGCATAGTAGATCA-5' termed Zif2

as a nonspecific complex.

All backbone atoms of the protein and DNA sequences for both modified models were maintained as in the crystal structure. The ZFP was thoroughly prepared before performing the energy calculations. Protonation states of ZF residues were decided by reassignments of all histidine residues to δ -histidine (HID) and all cysteine to CYM (cysteine with a negative charge). Finally, a longer range electrostatic model ($\sigma = 1.70$ Å, $\epsilon = 0.67$ kcal/mol) of the zinc ion (Zn^{2+})⁴⁹ was used instead of the original nonbonded zinc parameters from the AMBER force field, avoiding a disorder of the ZFP during minimizations. The Zn^{2+}

was coordinated within the ZFP by two CYM and two HID residues. A geometry with a tetrahedrally coordinated Zn^{2+} in the center of the ZFP was subsequently maintained throughout the minimizations. The remaining protein and nucleic acid parameters were determined using the AMBER ff03 force field.⁵⁰

Both Zif1 and Zif2 models were then solvated with an explicit TIP3P water model⁵¹ in a truncated octahedral box with a minimum 10 Å distance between any solute atom and a box edge. Periodic boundary conditions were applied for avoiding the impacts from the atoms on the outer surfaces drifting away from the simulation box. The systems were neutralized with a minimum number of sodium ions and were then energy minimized in two stages (only solvent and ions at first, then all atoms). Each minimization stage was performed using the first 100 steps of a steepest descent algorithm followed up with 2400 steps of conjugate gradient minimization to get closer to an energy minimum. The minimized conformations of Zif1 and Zif2 were visualized and structurally analyzed.

Analysis of the ZFP-DNA interaction energies

To investigate the relative binding energies between the designed ZFP and the target oligonucleotides, a short trajectory of each protein-DNA complex was produced. The system was treated as a ligand/receptor complex, in which the designed ZF motif was the ligand and the target DNA was the receptor. The MM/GBSA approach was then used to calculate the free energies of molecules (G_x) in solution as described in previous work⁵² [see Eqs. (1) and (2)]. The molecular mechanical energies, the van der Waals interaction (E_{vdw}), electrostatic interaction (E_{es}), and internal energy (E_{in}), as well as the free energy of the GB solvated system (G_{pol}) and the hydrophobic contribution to the solvation free energy (G_{nonpol}) were calculated with the SANDER program. The GB^{OBC} solvent model⁵³ with $igb = 2$ was used in this study. The binding energies of ZF motifs to their DNA and the relative free energies between Zif1 and Zif2 are calculated as in Eqs. (3) and (4), averaging over trajectories from the minimizations.

$$G_x = E_{vdw} + E_{es} + E_{in} + G_{pol} + G_{nonpol} \quad (1)$$

$$G_{nonpol} = 0.0072 \text{ kcal mol}^{-1} \text{Å}^{-2} \times SA \quad (2)$$

$$\Delta G_{binding} = G_{complex} - G_{receptor} - G_{ligand} \quad (3)$$

$$\Delta \Delta G_{binding} = \Delta G_{binding}(Zif1) - \Delta G_{binding}(Zif2) \quad (4)$$

Acknowledgments

The authors thank Professor Carlos F. Barbas III for providing them with EMSA kit and Associate Professor Dr. Prachya Kongtawelert for his assistance to provide the BIACORE 2000TM biosensor and NANOTEC, Thailand for the access to Discovery Studio version 1.7 program

package. They thank Dr. Dale Taneyhill for proofreading the manuscript.

References

1. Mann JM, Chin J, Piot P, Quinn T (1988) The international epidemiology of AIDS. *Sci Am* 259:82–89.
2. UNAIDS/World Health Organization (2008) Report on the Global HIV/AIDS Epidemic. August, Mexico City.
3. Englund G, Theodore TS, Freed EO, Engelman A, Martin MA (1995) Integration is required for productive infection of monocyte-derived macrophages by human immunodeficiency virus type 1. *J Virol* 69:3216–3219.
4. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* 69:2729–2736.
5. Sakai H, Kawamura M, Sakuragi J, Sakuragi S, Shibata R, Ishimoto A, Ono N, Ueda S, Adachi A (1993) Integration is essential for efficient gene expression of human immunodeficiency virus type 1. *J Virol* 67:1169–1174.
6. Swanstrom R, Delorbe WJ, Bishop JM, Varmus HE (1981) Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. *Proc Natl Acad Sci USA* 78:124–128.
7. Shoemaker C, Goff S, Gilboa E, Paskind M, Mitra SW, Baltimore D (1980) Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc Natl Acad Sci USA* 77:3932–3936.
8. Shank PR, Varmus HE (1978) Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. *J Virol* 25:104–114.
9. Yoshimura FK, Weinberg RA (1979) Restriction endonuclease cleavage of linear and closed circular murine leukemia viral DNAs: discovery of a smaller circular form. *Cell* 16:323–332.
10. Fritsch E, Temin HM (1977) Formation and structure of infectious DNA of spleen necrosis virus. *J Virol* 21:119–130.
11. Guntaka RV, Richards OC, Shank PR, Kung HJ, Davidson N (1976) Covalently closed circular DNA of avian sarcoma virus: purification from nuclei of infected quail tumor cells and measurement by electron microscopy and gel electrophoresis. *J Mol Biol* 106:337–357.
12. Gianni AM, Smotkin D, Weinberg RA (1975) Murine leukemia virus: detection of unintegrated double-stranded DNA forms of the provirus. *Proc Natl Acad Sci USA* 72:447–451.
13. Vink C, Plasterk RH (1993) The human immunodeficiency virus integrase protein. *Trends Genet* 9:433–438.
14. Katz RA, Skalka AM (1994) The retroviral enzymes. *Annu Rev Biochem* 63:133–173.
15. De Clercq E (2001) New developments in anti-HIV chemotherapy. *Curr Med Chem* 8:1543–1572.
16. Evering TH, Markowitz M (2008) Raltegravir: an integrase inhibitor for HIV-1. *Expert Opin Invest Drugs* 17:413–422.
17. Grant P, Zolopa A (2008) Integrase inhibitors: a clinical review of raltegravir and elvitegravir. *J HIV Ther* 13:36–39.
18. Delelis O, Malet I, Na L, Tchertanov L, Calvez V, Marcelin AG, Subra F, Deprez E, Mouscadet JF (2009) The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. *Nucleic Acids Res* 37:1193–1201.
19. Hicks C, Gulick RM, Roquebert B, Blum L, Collin G, Damond F, Peytavin G, Leleu J, Matheron S, Chene G, Brun-Vezinet F, Descamps D (2009) Raltegravir: the first HIV type 1 integrase inhibitor selection of the Q148R integrase inhibitor resistance mutation in a failing raltegravir containing regimen. *Clin Infect Dis* 48:931–939.
20. Charpentier C, Karmochkine M, Laureillard D, Tisserand P, Belec L, Weiss L, Si-Mohamed A, Piketty C (2008) Drug resistance profiles for the HIV integrase gene in patients failing raltegravir salvage therapy. *HIV Med* 9:765–770.
21. Miller J, Mclachlan AD, Klug A (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J* 4:1609–1614.
22. Lee MS, Gippert GP, Soman KV, Case DA, Wright PE (1989) Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 245:635–637.
23. Elrod-Erickson M, Rould MA, Nekudova L, Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* 4:1171–1180.
24. Pavletich NP, Pabo CO (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252:809–817.
25. Zimmerman KA, Fischer KP, Joyce MA, Tyrrell DL (2008) Zinc finger proteins designed to specifically target duck hepatitis B virus covalently closed circular DNA inhibit viral transcription in tissue culture. *J Virol* 82:8013–8021.
26. Dreier B, Segal DJ, Barbas CF (2000) Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* 303:489–502.
27. Dreier B, Fuller RP, Segal DJ, Lund CV, Blancafort P, Huber A, Kokscho B, Barbas CF (2005) Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 280:35588–35597.
28. Dreier B, Beerli RR, Segal DJ, Flippin JD, Barbas CF (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* 276:29466–29478.
29. Blancafort P, Segal DJ, Barbas CF (2004) Designing transcription factor architectures for drug discovery. *Mol Pharmacol* 66:1361–1371.
30. Mandell JG, Barbas CF (2006) Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* 34:W516–W523.
31. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
32. Deprez E, Barbe S, Kolaski M, Leh H, Zouhiri F, Auclair C, Brochon JC, Le Bret M, Mouscadet JF (2004) Mechanism of HIV-1 integrase inhibition by styrylquinoline derivatives in vitro. *Mol Pharmacol* 65:85–98.
33. Bugreev DV, Baranova S, Zakharova OD, Parrisi V, Desjobert C, Sottofattori E, Balbi A, Litvak S, Tarrago-Litvak L, Nevinsky GA (2003) Dynamic, thermodynamic, and kinetic basis for recognition and transformation of DNA by human immunodeficiency virus type 1 integrase. *Biochemistry* 42:9235–9247.
34. Segal DJ, Crotty JW, Bhakta MS, Barbas CF, Horton NC, Mandell JG, Eberhardy SR, Goncalves J, Coelho S, Berkhout B, Dreier B, Fuller R, Lund CV, Blancafort P, Huber A, Kokscho B (2006) Structure of Aart, a

- designed six-finger zinc finger peptide, bound to DNA. *J Mol Biol* 363:405–421.
35. Balakrishnan M, Jonsson CB (1997) Functional identification of nucleotides conferring substrate specificity to retroviral integrase reactions. *J Virol* 71:1025–1035.
 36. Katzman M, Sudol M (1996) Influence of subterminal viral DNA nucleotides on differential susceptibility to cleavage by human immunodeficiency virus type 1 and visna virus integrases. *J Virol* 70:9069–9073.
 37. Yoshinaga T, Fujiwara T (1995) Different roles of bases within the integration signal sequence of human immunodeficiency virus type 1 in vitro. *J Virol* 69:3233–3236.
 38. Van den Ent FM, Vink C, Plasterk RH (1994) DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. *J Virol* 68:7825–7832.
 39. Vink C, Van Gent DC, Elgersma Y, Plasterk RH (1991) Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J Virol* 65:4636–4644.
 40. Lafemina RL, Callahan PL, Cordingley MG (1991) Substrate specificity of recombinant human immunodeficiency virus integrase protein. *J Virol* 65:5624–5630.
 41. Reynolds L, Ullman C, Moore M, Isalan M, West MJ, Clapham P, Klug A, Choo Y (2003) Repression of the HIV-1 5' LTR promoter and inhibition of HIV-1 replication by using engineered zinc-finger transcription factors. *Proc Natl Acad Sci USA* 100:1615–1620.
 42. Segal DJ, Goncalves J, Eberhardy S, Swan CH, Torbett BE, Li X, Barbas CF (2004) Attenuation of HIV-1 replication in primary human cells with a designed zinc finger transcription factor. *J Biol Chem* 279:14509–14519.
 43. Rebar EJ, Huang Y, Hickey R, Nath AK, Meoli D, Nath S, Chen B, Xu L, Liang Y, Jamieson AC, Zhang L, Spratt SK, Case CC, Wolffe A, Giordano FJ (2002) Induction of angiogenesis in a mouse model using engineered transcription factors. *Nat Med* 8:1427–1432.
 44. Lanao JM, Briones E, Colino CI (2007) Recent advances in delivery systems for anti-HIV1 therapy. *J Drug Target* 15:21–36.
 45. Whitcomb JM, Kumar R, Hughes SH (1990) Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. *J Virol* 64:4903–4906.
 46. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res* 33:W526–W531.
 47. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* 112:531–552.
 48. Nair R, Rost B (2005) Mimicking cellular sorting improves prediction of subcellular localization. *J Mol Biol* 348:85–100.
 49. Stote RH, Karplus M (1995) Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins* 23:12–31.
 50. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem* 24:1999–2012.
 51. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
 52. Jiranusornkul S, Laughton CA (2008) Destabilization of DNA duplexes by oxidative damage at guanine: implications for lesion recognition and repair. *J R Soc Interface* 5 (Suppl 3):S191–S198.
 53. Onufriev A, Bashford D, Case DA (2000) Modification of the generalized Born model suitable for macromolecules. *J Phys Chem B* 104:3712–3720.