



รายงานวิจัยฉบับสมบูรณ์

โครงการ ScFv และนาโนบอดีมนุษย์ผ่านเข้าเซลล์ที่ยับยั้งไทโรซีนไคเนสของ
Epidermal Growth Factor Receptor

โดย รศ.ดร.เกียรติทวี ชวงศ์โกมล และคณะ

12 พฤษภาคม 2560

โครงการ ScFv และนาโนบอดีมนุษย์ผ่านเข้าเซลล์ที่ยับยั้งไทโรซีนไคเนสของ
Epidermal Growth Factor Receptor

รศ.ดร.เกียรติทวี ชูวงศ์โกมล
ภาควิชาชีวเคมี คณะวิทยาศาสตร์
มหาวิทยาลัยเกษตรศาสตร์

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย
และมหาวิทยาลัยเกษตรศาสตร์

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกว. และมหาวิทยาลัยเกษตรศาสตร์ไม่จำเป็นต้อง
เห็นด้วยเสมอไป)

ชื่อโครงการ ScFv และนาโนบอดีมนุษย์ผ่านเข้าเซลล์ที่ยับยั้งไทโรซีนไคเนสของ Epidermal Growth Factor Receptor
ระยะเวลาโครงการ 3 ปี

ชื่อหัวหน้าโครงการวิจัยผู้รับทุน รศ.ดร.เกียรติทิพย์ ชูวงศ์โกลม

รายงานในช่วงตั้งแต่วันที่ 16 กรกฎาคม 2557 ถึง 15 กรกฎาคม 2560

ABSTRACT

Aberration of epidermal growth factor receptor (EGFR) mostly found in various solid cancers such as breast, cervical, head and neck, and including lung cancer. 85% of lung cancer often have been found into non-small cell lung cancer (NSCLC) which mostly cause patient death. Therefore, the first strategy is chemotherapy for attenuating EGFR activity. Monoclonal antibodies (mAb) and small molecule tyrosine kinase inhibitors (TKI) have been used as anti-cancer agents. Unfortunately, they remain many disadvantage such non-specific effective, side effect, lack of penetrable property. However, decades of biology research on cancer therapy have been revealed many strategies for improving efficiency of anticancer drug. We used the computational studies to understand the protein-protein interaction that activate kinase activity and MIG6 peptide that inhibit activity of EGFR. The computational studies revealed the key residue in the interaction and inhibiting mechanism. Furthermore we aim to find nanobodies as inhibitors which were selected from humanized VH/V_HH phage display library. The nanobody clone containing EGFR-high affinity and EGFR-inhibitory were introduced cell-penetrable peptide (Arg-9) for making cell-penetrable nanobodies. Cell-based experiment revealed that VH18, VH36 modulate cell viability of NSCLC cell lines (A549) with nanomolar range (181nM and 9.61nM, respectively) compare to approved drug. Confocal microscope suggested that all cell-penetrable nanobodies able to penetrate cell. Moreover, combination of molecular docking and epitope searching reveal that VH36 can bind to C-lope of EGFR which essential for EGFR-dimerized activation. This antibody fragment can be developed as a therapeutic agent for cancer therapy.

บทคัดย่อ

ความผิดปกติของโปรตีนตัวรับอีพิเดอมอลโกรทแฟคเตอร์ (EGFR) มักพบส่วนมากในมะเร็งหลายชนิด เช่น มะเร็งเต้านม มะเร็งปากมดลูก มะเร็งในระบบสืบพันธุ์และลำคอ รวมไปถึงมะเร็งปอด มากไปกว่านั้น 85% ของผู้ป่วยโรคมะเร็งปอดมักพบว่าเป็นเซลล์มะเร็งปอดชนิดไม่เล็ก (NSCLC) ดังนั้น EGFR จึงเป็นโปรตีนเป้าหมายในการพัฒนายาในการรักษาโรคมะเร็ง ในปัจจุบันตัวยับยั้งการทำงานของโปรตีน EGFR นั้นแบ่งออก 2 ชนิดคือ สารโมเลกุลเล็กและแอนติบอดี อย่างไรก็ตาม

ตามของจำกัดในการใช้สำหรับการรักษาโรคมะเร็งยังคงเป็นอุปสรรค ตัวอย่างเช่น การขาดความจำเพาะเจาะจงในการรักษาที่โปรตีนเป้าหมาย การเกิดผลข้างเคียงในการรักษา การขาดความสามารถในการแทรกเข้าเซลล์ ในการศึกษาขั้นนี้ได้ทำการศึกษากับคอมพิวเตอร์เพื่อที่จะเข้าใจการจับกันระหว่างโปรตีนกับโปรตีนเพื่อกระตุ้นการทำงานของไคเนส และการจับกับ MIG6 เปปไทด์ ที่ยับยั้งกิจกรรมของไคเนส การศึกษาด้วยคอมพิวเตอร์แสดงให้เห็นถึงกรดอะมิโนที่สำคัญต่อการกลไกการจับกันและการยับยั้ง นอกจากนี้ ในงานวิจัยนี้ผู้วิจัยได้คัดเลือก VH/V_H ที่จำเพาะต่อไทโรซีนไคเนสโดเมนของ EGFR (TK-EGFR) จากเทคโนโลยีคลังผาจ หลังจากนั้นคัดเลือกโคลนที่มีความสามารถในการจับและความสามารถในการยับยั้งการทำงานของ TK-EGFR มาต่อเข้ากับเปปไทด์แทรกเข้าเซลล์ หรือ Arg-9 เพื่อสร้างให้นาโนบอดีสามารถผ่านเข้าเซลล์ได้ เพื่อทดสอบความเป็นพิษต่อเซลล์มะเร็งปอดชนิดไม่เล็ก (A549) พบว่า นาโนบอดีโคลน VH18, V_HH35 และ VH36 มีความสามารถในการยับยั้งการเจริญเติบโตของเซลล์มะเร็งได้ในระดับนาโนโมลาร์ โดยมีค่า IC₅₀ 181 nM, 9.16nM และ 9.61nM ตามลำดับ เมื่อเปรียบเทียบกับยาที่ใช้ในการรักษาในปัจจุบัน ผลจากกล้องฟลูออเรสเซนซ์แบบตัดเงื่อนไขพบว่านาโนบอดีดังกล่าวมีความสามารถในการแทรกผ่านเข้าเซลล์ได้ มากไปกว่านั้นทำผสมผสานผลจากการหาบริเวณจับรวมกับกับการทำนายโครงสร้างโดยใช้คอมพิวเตอร์บ่งบอกว่า VH36 จับที่บริเวณ C-lope ของ EGFR ซึ่งเป็นบริเวณที่สำคัญในการเกิดการกระตุ้นให้เกิดการทำงาน ดังนั้นนาโนบอดีจึงเป็นโมเลกุลที่น่าสนใจในการพัฒนาเป็นยาในการรักษาโรคมะเร็ง

ความสำคัญและที่มาของปัญหา

ในปัจจุบันโรคมะเร็งรับว่าเป็นสาเหตุหลักของการเสียชีวิตของคนทั่วโลก ตัวอย่างชนิดของมะเร็งที่พบได้บ่อย เช่น มะเร็งเต้านม มะเร็งปากมดลูก มะเร็งตับ และมะเร็งปอดเป็นต้น โดยเฉพาะมะเร็งปอดพบว่าสามารถเกิดได้ทั้งในเพศชายและเพศหญิงในอัตราส่วนที่เท่ากัน รวมทั้งยังสามารถเกิดในได้ทั้งในคนที่สูบบุหรี่และไม่สูบบุหรี่ด้วย จากการสำรวจพบว่า 85% ของผู้ป่วยมะเร็งปอดมักพบว่าเป็นชนิด เซลล์มะเร็งปอดชนิดไม่เล็ก (Non-small cell lung cancer; NSCLC) และอีก 15% จะอยู่ในกลุ่ม เซลล์มะเร็งปอดชนิดเล็ก (Small cell lung cancer; SCLC) อย่างไรก็ตามผู้ป่วยโรคมะเร็งปอดชนิด SCLC มักจะเสียชีวิตภายในเวลาอันสั้นและไม่สามารถคาดการณ์ระยะของโรคได้ สำหรับผู้ป่วยโรคมะเร็งปอดชนิด NSCLC นั้นสามารถแบ่งระยะของโรคได้ทำให้สามารถหาวิธีการในการรักษาหรือควบคุมความรุนแรงของโรคได้ ดังนั้นโรคมะเร็งปอดชนิด NSCLC จึงเป็นความท้าทายในการพัฒนาตัวยับยั้งโดยใช้โปรตีนเป้าหมาย จากงานวิจัยที่ผ่านมาได้มีการรายงานไว้ว่า มะเร็งปอดชนิด NSCLC มักจะเกี่ยวกับการแสดงออกของโปรตีนตัวรับ Epidermal Growth Factor Receptor (EGFR) ซึ่งเป็นโปรตีนตัวรับบนผิวเซลล์ชนิดหนึ่ง ซึ่งจัดอยู่ในกลุ่มของ โปรตีนไทโรซีนไคเนส มีบทบาทสำคัญในการควบคุมการส่งสัญญาณในระดับเซลล์ โดยมีผลต่อการควบคุม การแบ่งเซลล์ โปรแกรมการตายของเซลล์ (Apoptosis) การยึดเกาะหรือ การเคลื่อนที่ของเซลล์ (Cell adhesion and migration) เป็นต้น การควบคุมดังกล่าวนี้มักจะอาศัยตัวกระตุ้นจากภายนอก (External ligand) เช่น EGF (Epidermal Growth Factor), TGF- α (Transforming Growth Factor -alpha), Neuregulin เป็นต้น เมื่อพิจารณาในโครงสร้างพบว่า โปรตีนตัวรับในกลุ่มของไทโรซีนไคเนส (ErbB family) ประกอบด้วยส่วนที่สำคัญ 3 ส่วน คือ Extracellular domain หรือ Ligand binding domain, Transmembrane helix และ ไทโรซีนไคเนสโดเมน (Tyrosine kinase domain) ทั้งการแ่งของการแสดงออกมากกว่าปกติ (Overexpression) และการกลายพันธุ์ (Mutation) ในปัจจุบันพบว่ายาที่ใช้ในการรักษามะเร็งปอดชนิด NSCLC โดยมุ่งเป้าไปที่โปรตีนตัวรับ Epidermal Growth Factor Receptor (EGFR) นั้นแบ่งออกเป็น 2 กลุ่ม คือ ตัวยับยั้งในกลุ่มแอนติบอดี (Antibody-based inhibitor) ซึ่งจะจำเพาะเป็นในบริเวณ Extracellular domain โดยมีหน้าในการแย่งจับกับลิแกนด์ (Competitive binding) ตัวยับยั้งในกลุ่มสารโมเลกุลเล็ก (Small molecule inhibitor) จะจำเพาะในบริเวณไทโรซีนไคเนสโดเมน โดยทำหน้าที่เป็นสารโมเลกุลเหมือน ATP (ATP-

analogue) โดยเข้าจับบริเวณ Active site ของบริเวณ Enzymatic domain แต่อย่างไรก็ตามข้อเสียของยาทั้งสองกลุ่มนี้คือ ถ้าเกิดการกลายพันธุ์ของไทโรซีนไคเนสโดเมนจะทำให้การยับยั้งด้วยตัวยับยั้งในกลุ่มแอนติบอดีไม่ได้ผลดีมากนักเนื่องจากจะเกิดการกระตุ้นโดยไม่ใช้ลิแกนด์ (Ligand-independent activation) แต่การยับยั้งด้วยตัวยับยั้งในกลุ่มสารโมเลกุลเล็กนั้นจะขาดความจำเพาะในการยับยั้งเนื่องจากโปรตีนส่วนใหญ่ในเซลล์นั้นมันจะเป็นเอนไซม์ในกลุ่มไคเนสซึ่งใช้ ATP เป็นสับสเตรท ด้วยเหตุผลเหล่านี้จึงนำไปสู่วัตถุประสงค์ในงานวิจัยนี้

ในการทดลองผู้วิจัยในได้ใช้ Phage display technology ในการคัดเลือกแอนติบอดี ที่สามารถยับยั้ง การทำงานของ tyrosine kinase (TK) โดเมนจาก EGFR โดยอาศัยสมบัติของ Phagemid ในการเป็นพาหะ ในการลำเลียงยีนของแอนติบอดีในตัวพาหะมาทำการแสดงออกของโปรตีนในเซลล์แบคทีเรีย โดยการใช้คลังฟาจแอนติบอดีสายเดี่ยว (Human Single chain Fragment variable; HuScFv) และ ชิ้นส่วนโดเมนเดี่ยว (Humanized Single domain; VH/V_HH) ที่ได้จากสัตว์เลี้ยงลูกด้วยนมประเภทหนูที่แสดงบนโปรตีนโครงสร้างของฟาจ M13 ในคลังฟาจในขั้นแรกผู้วิจัยได้ทำการคัดเลือกแอนติบอดีที่ความจำเพาะต่อ TK โดยใช้ เทคนิค Single round Biopanning จากผลการทดลองพบว่า นาโนบอดี (VH18, V_HH35, VH36) มีความสามารถในการจับกับไทโรซีนไคเนสโดเมนได้ดี และมีความสามารถในการยับยั้งการทำงานของไทโรซีนไคเนสโดเมนในการลดการเกิดเติมหมู่ฟอสเฟต (Phosphorylation) มากไปกว่านั้นการเติม Cell penetrating peptide (CCP) ที่บริเวณปลาย N ของนาโนบอดีทำให้มีความสามารถในการเข้าเซลล์โดยใช้กล้องฟลูออเรสเซนซ์แบบตัดเฉือน (Confocal fluorescence microscope) ส่วนผลการทดลองในเซลล์มะเร็งปอดชนิด NSCLC พบว่ามีความสามารถในการยับยั้งในระดับนาโนโมลาร์ เมื่อเทียบกับเซลล์มะเร็งชนิดอื่น รวมทั้งยังมีความสามารถในการยับยั้งการเคลื่อนที่ของเซลล์มะเร็งอีกด้วย เป็นไปได้ว่านาโนบอดีมีความสามารถในการพัฒนาเป็นยาในการรักษาโรคมะเร็งปอดชนิด NSCLC ต่อไปในอนาคต

วัตถุประสงค์ของการดำเนินการ

1. เพื่อศึกษาจำลองการจับกันบริเวณที่สำคัญของไทโรซีนไคเนสโดเมนของ EGFR
2. คัดเลือกนาโนบอดีที่มีความสามารถในการยับยั้งการทำงานของ EGFR ที่บริเวณไทโรซีน ไคเนสโดเมน
3. ผลิต Cell-permeable nanobodies สำหรับนำไปทดสอบในเซลล์ (Cell line)
4. ทดสอบความสามารถในการยับยั้งการเจริญเติบโตของเซลล์มะเร็งปอดชนิด NSCLC

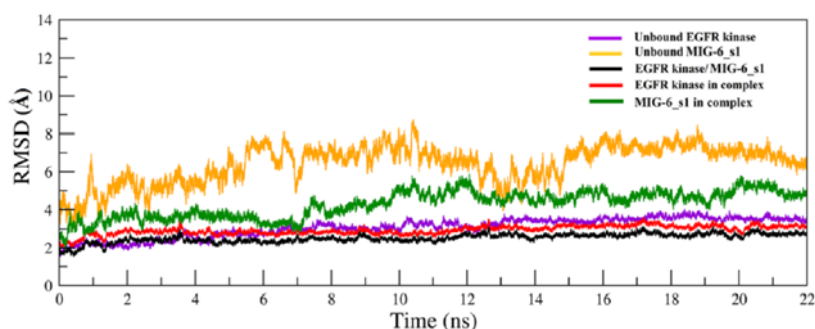
ผลการทดลอง

1. การจำลองการจับกันระหว่าง TK กับเปปไทด์ MIG6

การเตรียมโมเดลและศึกษาค่าพลังงานและพฤติกรรมการเคลื่อนที่แบบไดนามิกส์ (MD) เตรียมโครงสร้างของ TK โดเมน ของ EGFR และ MIG6_s1 แบบ asymmetric kinase dimer จากฐานข้อมูล PDB โมเดลโมโนเมอร์ของ TK โดเมน ของ EGFR แบบ activators (residues 678-959) และ MIG6_s1 (residues 337-362) โดยใช้ PDB code 2RFE [1] ส่วน PDB code 2GS6 จะใช้เป็นโมเดลสำหรับศึกษา TK โดเมน ของ EGFR แบบ receiver จากนั้นเตรียมไฟล์เพื่อใช้สำหรับการทำ MD ด้วย AMBER 12 [2] ที่ปลาย N และ C terminus ของ TK โดเมน ของ EGFR และ MIG-6 เติม acetyl group และปลาย N เติม N-methyl group ตามลำดับ ในการทำศึกษาการเคลื่อนที่ของโปรตีนโดยใช้ ff03 force field และเติมน้ำด้วยชนิด TIP3P [3] จากนั้นคำนวณการจับกันด้วยพลังงานอิสระด้วย MM-PBSA [4] ระหว่างโปรตีน-โปรตีน ด้วยการเลือกศึกษา 300

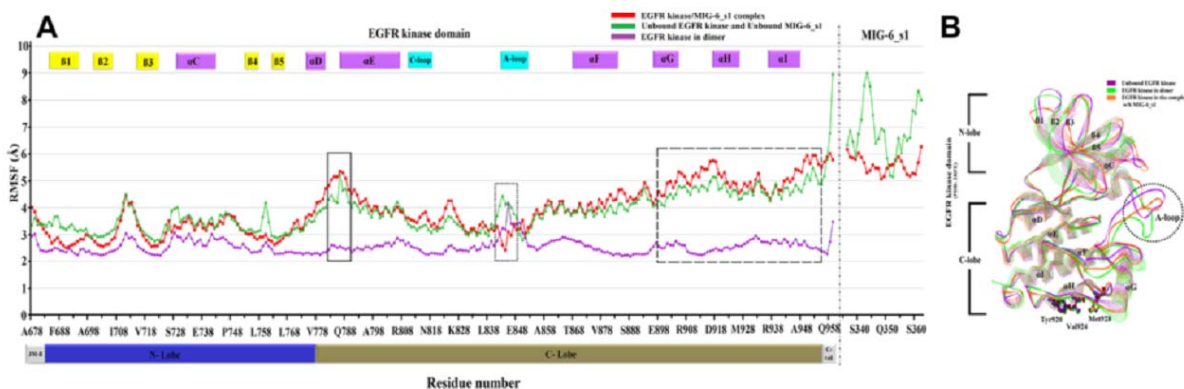
snapshot จากช่วง 3 ns สุดท้ายของทำ MD stimulation โดยคำนวณทุกช่วงของการ snapshot และทำการคำนวณ alanine scanning mutagenesis [5] คือการค้นหาหรือระบุบริเวณ “hot spot” ที่อินเทอร์เฟซของการจับกัน ซึ่งเป็นบริเวณที่สำคัญต่อการจับกันของโปรตีน ค้นหา key residues ของ TK โดเมน ของ EGFR/ MIG-6_s1 ที่บริเวณอินเทอร์เฟซของโปรตีน ด้วยการนำเอา alanine มาแทนที่กรดอะมิโนเดิมที่มีอยู่ โดยไม่ทำให้โครงสร้างของโปรตีนเปลี่ยนแปลงไป จากนั้นทำการคำนวณด้วยค่าพลังงานของการเปลี่ยนแปลงด้วย MM-PBSA

การศึกษาความคงที่และความยืดหยุ่นของโครงสร้างในระบบ 3 ระบบ ที่ผ่านการทำ MD stimulations ได้แก่ 1. Unbound EGFR 2. Unbound MIG-6_s1 และ 3. EGFR kinase/MIG-6_s1 ดังภาพที่ 1 ถูกนำมาสร้างกราฟ RMSD เพื่อดูความเสถียรของระบบที่ใช้ในการศึกษาจะเห็นได้ว่าทั้งสามระบบมีความเสถียรในช่วง 22 ns ของการศึกษา



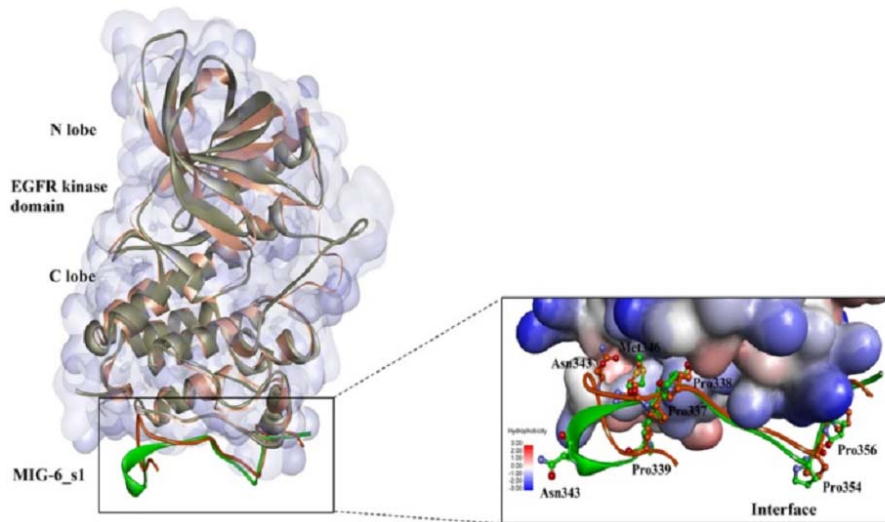
ภาพที่ 1 ค่า RMSD ของโครงสร้างหลักของ EGFR kinase/MIG-6_s1 สถานะ bound และ unbound ในช่วง 22 ns ของการทำ MD

ของเมื่อเปรียบเทียบการเปลี่ยนแปลงโครงสร้างในระบบ bound และ unbound ดังภาพที่ 2 พบว่าค่าเฉลี่ยของกราฟ RMSD ของทั้งสองระบบมีความใกล้เคียงกัน คือ 3.0 Å ของระบบ unbound และ 2.9 Å ของระบบ bound เมื่อเปรียบเทียบของ bound และ unbound ของ MIG-6_s1 พบว่าระบบ bound มีความเสถียรมากกว่า unbound โดยเฉลี่ย 4.2 Å ของ ระบบ bound และ 6.4 Å ของระบบ unbound ซึ่งแสดงให้เห็นว่าระบบ bound ของ MIG-6_s1 มีความเสถียรมากกว่า unbound อย่างมีนัยสำคัญ ซึ่งโดยทั่วไประบบ unbound MIG_6-s1 มีค่า RMSD สูง เนื่องจาก MIG-6_s1 มีจำนวนกรดอะมิโนที่สั้นกว่า (27 กรดอะมิโน) และเมื่อเปรียบเทียบทั้งสองระบบของ unbound EGFR และ EGFR kinase/MIG-6_S1 ด้วยกราฟ RMSF หรือการวิเคราะห์การเคลื่อนไหวของอะตอมเฉลี่ย โปรไฟล์ RMSF ของ unbound TK โดเมน ของ EGFR (สีเขียว) ดังภาพที่ 2A



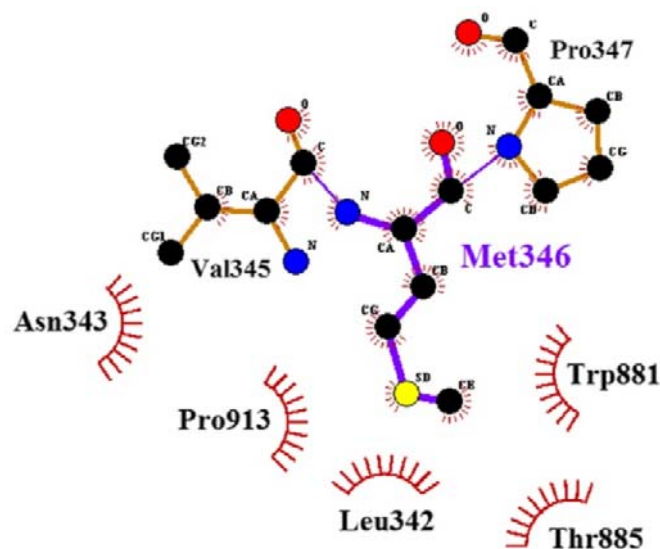
ภาพที่ 2 ค่า RMSF และภาพการซ้อนทับกันของโครงสร้างTK โดเมน ของ EGFR แสดงให้เห็นพฤติกรรมเคลื่อนไหวที่ของโปรตีนคล้ายกับระบบของ EGFR (สีแดง) ในทางตรงกันข้ามโปรไฟล์ RMSF ของระบบ unbound MIG_6-S1 มีค่าความผันผวนสูงกว่าการจับกันเป็น complex ของ MIG-6_s1 ซึ่งทำให้เห็นรูปแบบของความแตกต่างในหลายๆส่วนของบริเวณอินเทอร์เฟซ EGFR กับ MIG-6_s1 ส่วนการวิเคราะห์ค่า RMSF ของ EGFR N-lobe พบว่ามีโปรไฟล์ RMSF คล้ายกันในทุกสามระบบ นอกจากนี้ยังพบความไม่แน่นอนของความผันผวนในส่วน of Ala840-Gly850 (A-

loop) [6] ซึ่งเกี่ยวข้องกับการเกิด missing residues ใน EGFR X-ray crystal structure จากข้อสังเกตนี้แสดงให้เห็นว่าการจับกันของ TK โดเมน ของ EGFRสามารถช่วยให้การจับของ MIG-6_s1 เป็นคอมเพล็กซ์ได้ดีขึ้น การเปลี่ยนแปลงโครงสร้างของ EGFR kinase/MIG-6_s1 ในช่วง 3 ns สุดท้ายของการทำ MD stimulation พบว่าโครงสร้างในช่วงแรกและโครงสร้างในช่วงของระบบเข้าสู่สมดุลนั้นโครงสร้างส่วนใหญ่คล้ายคลึงกัน ยกเว้นการเปลี่ยนแปลงในส่วนขงลูปของ MIG-6_s1 โดยเฉพาะ Asn343 ของ MIG-6_s1 ที่บริเวณลูปของโครงสร้างมีการเลื่อนไปของโซ่ข้างไปใกล้ชิดกับอินเทอร์เฟซของ TK โดเมน ของ EGFRมากกว่าการลดพลังงานของโครงสร้างเริ่มต้นดังภาพที่ 3

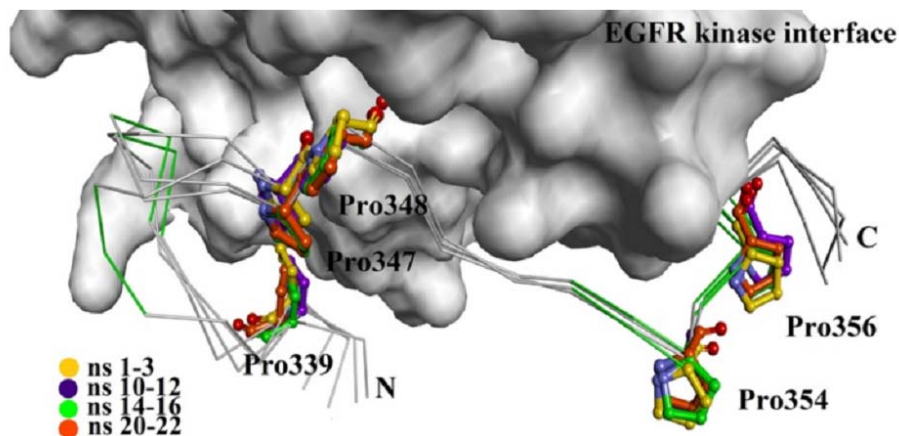


ภาพที่ 3 การซ้อนทับของโครงสร้างจากช่วง 3 ns สุดท้ายของการทำ MD (สีชมพู) และช่วงการลดพลังงานของโครงสร้างเริ่มต้น (สีเขียว) ของ EGFR kinase / MIG-6_s1

ส่วนของปฏิสัมพันธ์ของไฮโดรโฟบิกบริเวณผิวของTK โดเมน ของ EGFRประกอบด้วย Leu343, Trp881, Thr885, Pro913 และ Met346 ของโซ่ข้างของ MIG-6_s1 ซึ่งอยู่มั่นคงในบริเวณของโพรงของโปรตีนด้วยโปรแกรม Ligplot ดังภาพที่ 4 กรดอะมิโนโพลีเมอร์อาจมีผลต่อความแข็งแรงของโครงสร้างเปปไทด์ นอกจากนี้ยังมีการวิเคราะห์โครงสร้างของโพลีเมอร์ โดยกรดอะมิโนโพลีเมอร์ทั้ง 5 ชนิดของ MIG-6_s1 (Pro347, Pro348, Pro354, Pro356, Pro339) อยู่ในรูปแบบ cis และฟอร์มโครงสร้างแบบ turn ส่วน 4 ชนิดของกรดอะมิโน (Pro347, Pro348, Pro354, Pro356) ยังคงอยู่ในโครงสร้างแบบ bent ที่บริเวณอินเทอร์เฟซดังภาพที่ 5

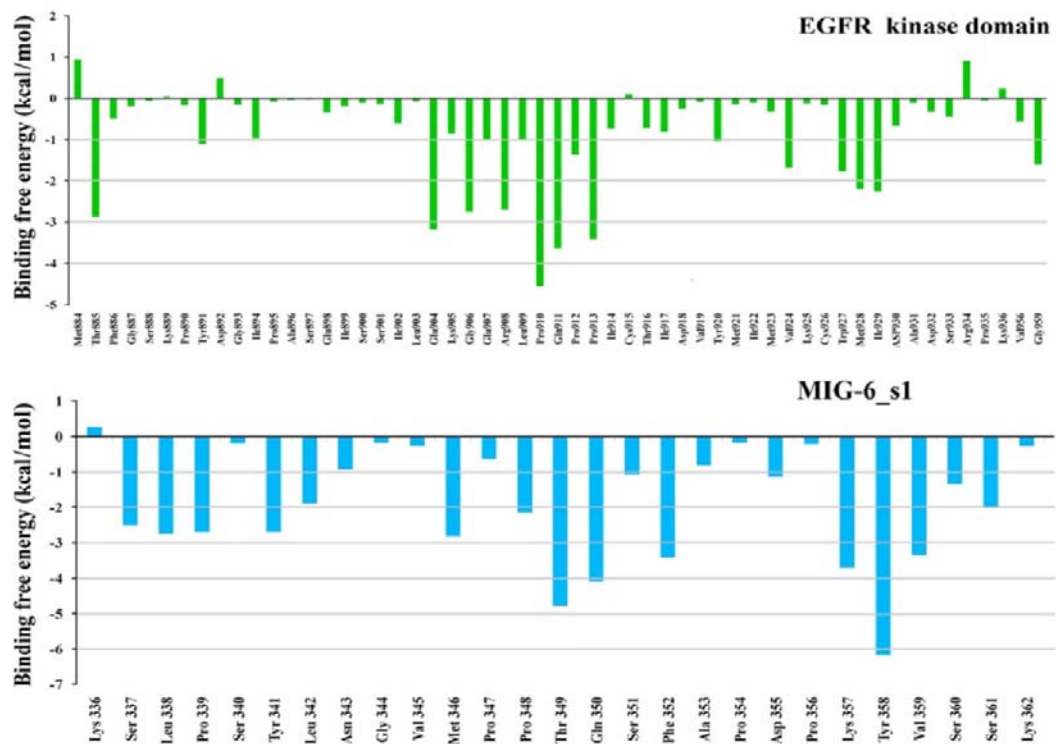


ภาพที่ 4 ปฏิสัมพันธ์ระหว่าง Met346 กับเรซิดิวบริเวณข้างเคียงจากโปรแกรม LIGPLOT

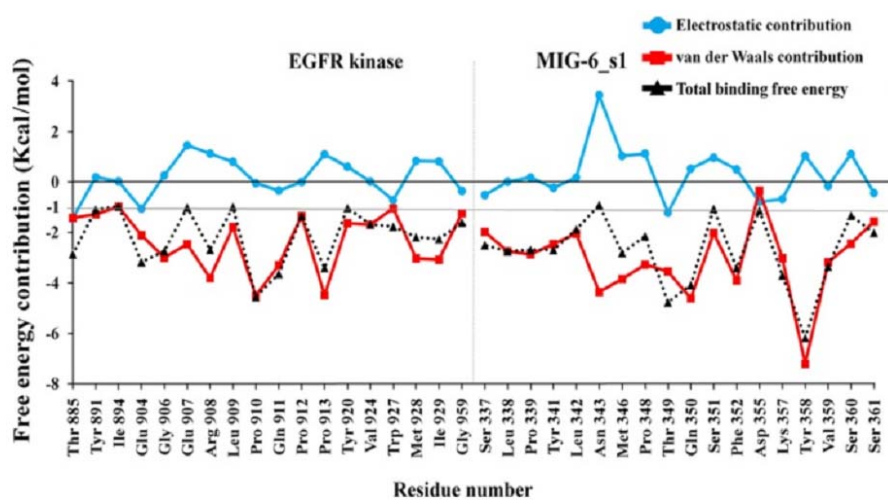


ภาพที่ 5 การจัดเรียงโครงสร้างเฉลี่ยระหว่างช่วง 1-3 ns (สีเหลือง), 10-12 ns (สีม่วง), 14-16 ns (สีม่วง) และ 20-22 ns (สีส้ม) แสดงให้เห็นความคงที่ของ prolines ในระหว่างการจำลอง

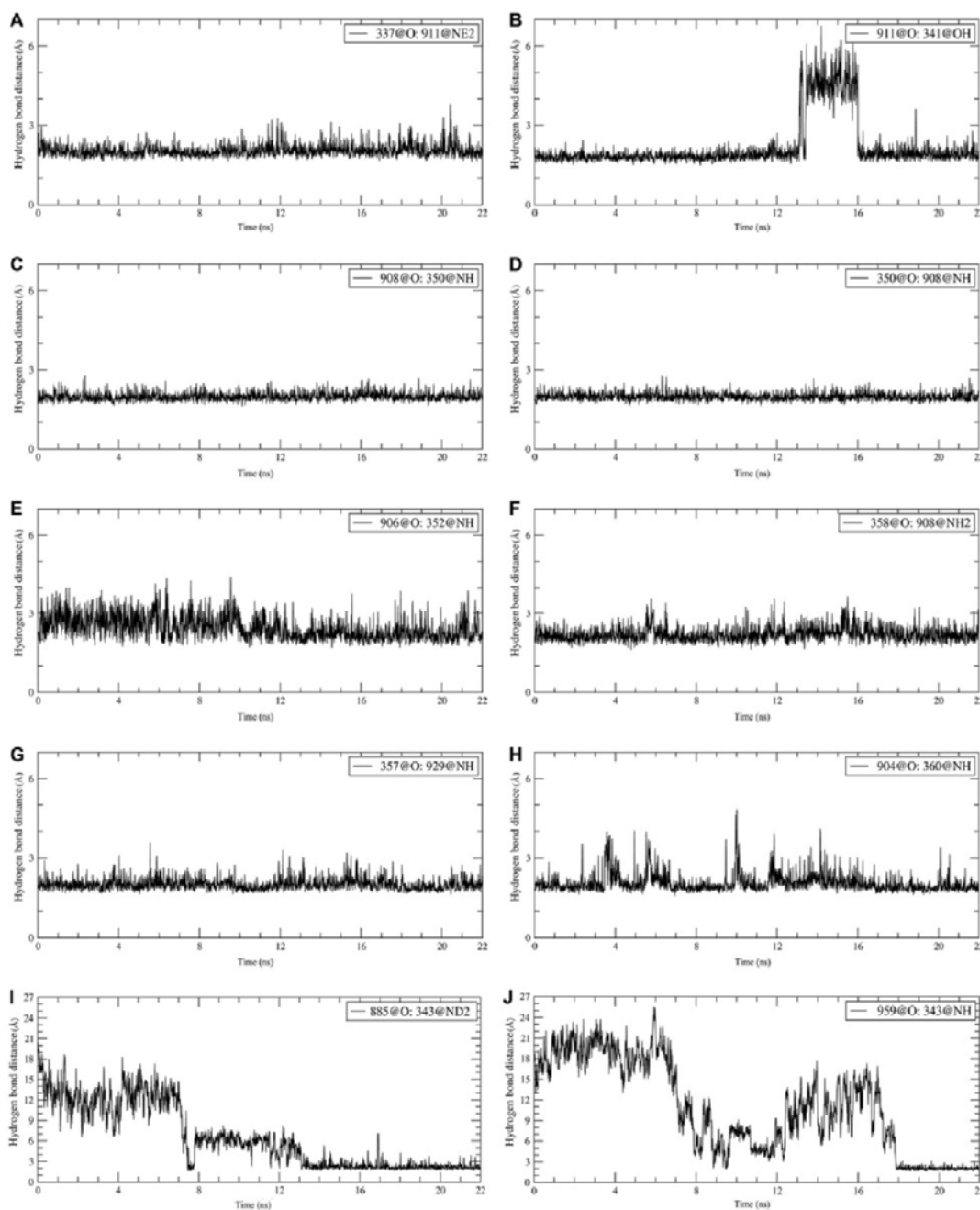
ค่าพลังงานการสลายตัวของเรซิดิวที่อยู่บนอินเทอร์เฟซ เรซิดิวที่สำคัญของการจับของ EGFR kinase/MIG-6_s1 ที่อินเทอร์เฟซคัดเลือกจากค่า cut-off ที่ ΔG_{tot} , PB $\leq$ กิโลแคลอรี / โมล 0.1(เหมาะแก่การจับ) ดังภาพที่ 6 ประกอบไปด้วย 13 เรซิดิวที่สำคัญของ MIG-6_s1 และ 9 เรซิดิวที่สำคัญต่อการจับกัน TK โดเมน ของ EGFRซึ่งมีค่าพลังงานอิสระต่อการจับน้อยกว่า -1 kcal/mol โดย 13 เรซิดิวเหล่านั้นได้แก่ Ser337, Leu338, Pro339, Tyr341, Met346, Pro348, Thr349, Gln350, Phe352, Lys357, Tyr358, Val และ 359Ser ส่วนเรซิดิวที่สำคัญของ 361TK โดเมน ของ EGFRได้แก่ Thr885, Glu904, Gly906, Arg908, Pro910, Gln911, Pro913, Met และ 928Ile929 ดังภาพที่7 ข้อมูลการจับกันด้วยพันธะไฮโดรเจนระหว่าง MIG-6_s1 peptide และ TK โดเมน ของ EGFRใน 3 ns สุดท้ายระหว่างการทำ MD stimulation โดยกรดอะมิโนเหล่านี้เกิดการฟอร์มพันธะไฮโดรเจน Ser337, Tyr341, Asn350, Phe352, Lys357, Tyr และ 358Ser จาก 360 MIG-6_s และ 1 ส่วนที่เหลือ Glu904, Gly906, Arg908, Gln และ 911Ile929 ของ TK โดเมน ของ EGFRซึ่งรูปแบบการจับของพันธะไฮโดรเจนแบบ strong, medium และ weak กำหนดด้วยอัตราการเกิดพันธะดังนี้ $> 75\%$, $75\%-50\%$ และ $< 50\%$ ตามลำดับ ระยะห่างของพันธะไฮโดรเจนของปฏิกิริยาระหว่างพันธะไฮโดรเจนมากที่สุดในช่วง 22ns ของการทำ MD stimulation ดังภาพที่ 6B



ภาพที่ 6 ค่าพลังงานอิสระที่ลดลงของแต่ละเรซิดิวของ EGFR และ MIG-6_s1



ภาพที่ 7 ค่าพลังงานอิสระที่ลดลงของแต่ละเรซิดิวที่บริเวณอินเทอร์เฟซของ EGFR kinase และ MIG-6_s1

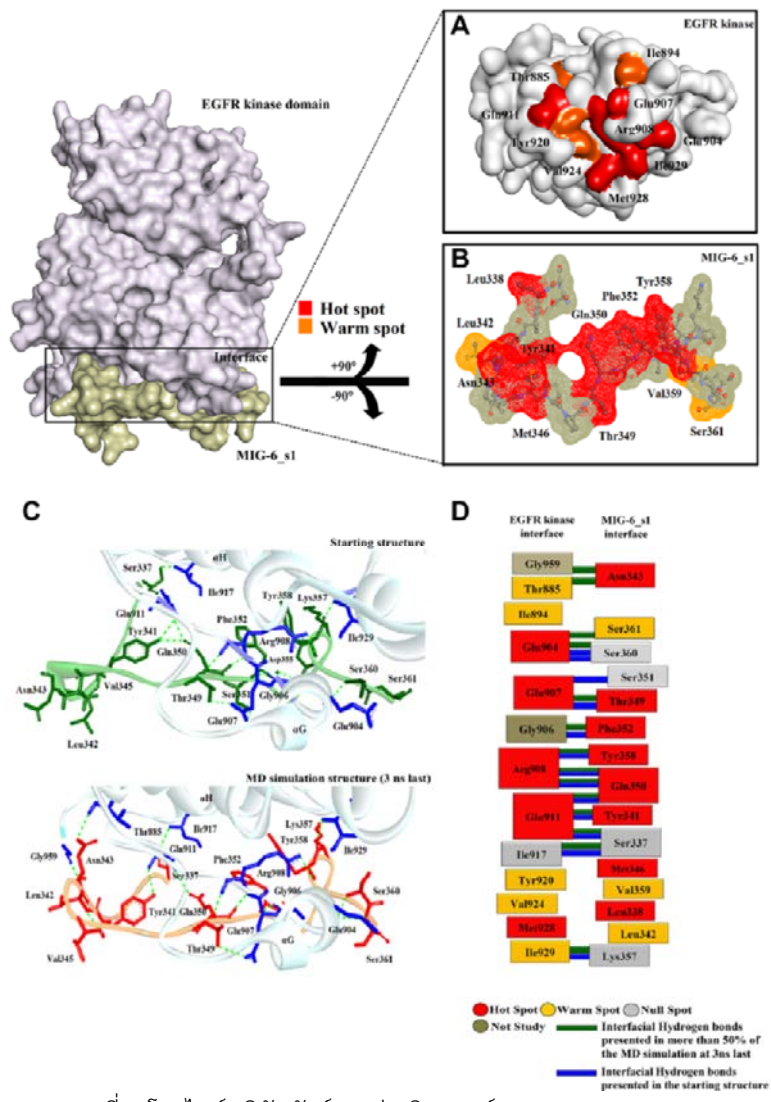


ภาพที่ 8 โปรไฟล์ระยะทางของพันธะไฮโดรเจน)A) ระหว่าง Gln-911NE-2HE และ 21Ser-337O (B) ระหว่าง Tyr-341OH และ Gln-911O (C) ระหว่างGln-350NH และ Arg-908O (D) ระหว่าง Arg-908NH และ Gln-350O (E) ระหว่าง Phe-352 NH และ Gly-906O (F) ระหว่าง Arg-908NH-2HHและ 22Tyr-358O (G) ระหว่าง Ile-929NH และ Lys-357O (H) ระหว่าง ser-360NH และ Glu-904O (I) ระหว่าง Asn-343ND-2HD และ 21Thr-885O (J) ระหว่างAsn-343NH และ Gly-959O

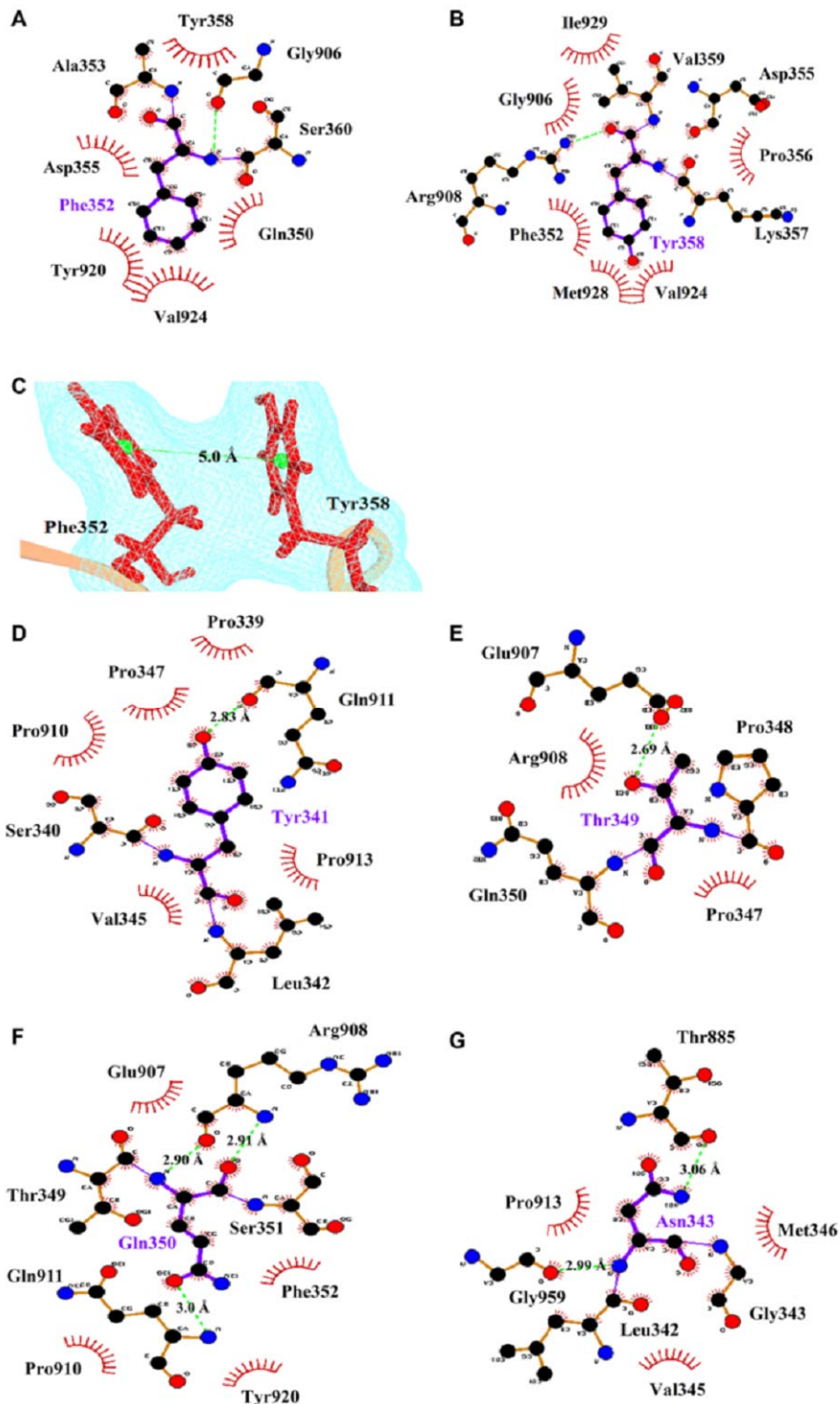
ผลจากการทำ alanine scanning และการคำนวณค่าพลังงานอิสระของแต่ละเรซิดิวดังภาพที่ 9 สามารถแบ่งออกเป็น 6 เรซิดิวที่เป็น “hot spot” บนอินเทอร์เฟซของ TK โดเมน ของ EGFR ได้แก่ Glu904, Glu907, Arg908, Gln911, Met 928and Ile และ 9298 เรซิดิวที่เป็น “hot spot” บนอินเทอร์เฟซของ MIG-6_s1 ได้แก่ Leu338, Try341, Asn343, Met346, Thr349, Gln350, Phe 352and Tryส่วนของเรซิดิวที่เป็น 358 “warm spot” บนอินเทอร์เฟซของ TK โดเมน ของ EGFR 4 เรซิดิว ได้แก่ Thr885, Ile94, Tyr920 และ Val924 ส่วนเรซิดิวที่เป็น warm spot บนอินเทอร์เฟซของ MIG-6_s1 ได้แก่ Leu342, Val359, and Ser361 อย่างไรก็ตามการคำนวณ alanine scanning มีข้อจำกัดโดย Gly906,

Pro910, Pro912, Pro913, Gly959 บนอินเทอร์เฟซ TK โดเมน ของ EGFR และ Pro339, Pro บนอินเทอร์เฟซ 348 MIG-6_{s1} เเรซิดิวเหล่านี้จะไม่ถูกนำไปคำนวณ 1 ซึ่งหมายความว่าค่าพลังงานอิสระที่ลดลง จะพบ 8 เเรซิดิวที่สำคัญ แต่ถ้าใช้วิธีการคำนวณแบบ alanine scanning จะไม่พบเเรซิดิวเหล่านี้ (Tyr891, Leu909, Trp927, Ser337, Ser351, Asp355, Lys357 and Ser360) โดยทั่วไปเเรซิดิวที่สำคัญที่พบโดยใช้วิธีการคำนวณทั้งสองวิธี เช่น Val924, Met346, Phe352 และ Try ซึ่งได้จากการทำการกลายพันธุ์ด้วยการแทนที่ด้วย 358alanine ทำให้เเรซิดิวเหล่านี้ไม่สามารถจับกับเปปไทด์ได้ [1]

การศึกษาการกลายพันธุ์ของ MIG-6_{s1} ที่เเรซิดิว Met346, Phe 352 and Tyrต่อการจับกับ 358TK โดเมน ของ EGFR ซึ่ง Met346 อาจจะเป็นเเรซิดิวที่สำคัญของ MID-6_{s1} มีค่าพลังงานอิสระประมาณ -8.2kcal/mol ที่ก่อให้เกิดปฏิสัมพันธ์กับไฮโดรฟอบิกที่บริเวณอินเทอร์เฟซของ EGFR kinase/MIG-6_{s1} complex โซ่ข้างของ Met346 อยู่ในโพรงส่วนที่ไม่ชอบน้ำประกอบด้วย Leu343, Trp881 และ Thr85 (warm spot) และ Pro913 ที่บริเวณอินเทอร์เฟซของ TK โดเมน ของ EGFR การกลายพันธุ์ของ Met346 เป็น Ala346 มีค่าพลังงานประมาณ -8.1 กิโลแคลอรี โมล ทำให้โซ่ข้างมีความสำคัญใน / โพรงของ TK โดเมน ของ EGFR ผลดังกล่าวนี้สอดคล้องกับโครงสร้างผลึกที่แสดงให้เห็นว่า Met346 ที่อยู่ในโพรงที่ไม่ชอบน้ำของ EGFR kinase และจากผลการทดลองที่ทำให้เกิดการกลายพันธุ์ของ Met346 ทำให้ไม่สามารถจับกับ EGFR ได้ ส่วน Phe352 และ Tyr358 ของ MIG-6_{s1} ได้ถูกระบุไว้ในงานวิจัยนี้ว่าเป็นเเรซิดิวที่จำเป็นสำหรับการจับกันที่มีพลังงาน และ 4.3 โมล ตามลำดับ โดยโครงสร้างหลัก / กิโกลแคลอรี 2.6(backbone) สร้างพันธะไฮโดรเจนกับเเรซิดิว “hot spot” Gly906 และ Arg908 ส่วนโซ่ข้างจะอยู่ในส่วนที่ไม่ชอบน้ำ นอกจากนี้ยังมีปฏิสัมพันธ์ระหว่างไฮโดรเจนและอิเล็กโตรอสแตติก ซึ่งเป็นปัจจัยสำคัญสำหรับการจับกันของ MIG-6_{s1} และ TK โดเมน ของ EGFR โดยมีเเรซิดิว Tyr341 ของ MIG-6_{s1} เป็นเเรซิดิวที่สำคัญของการจับกันระหว่าง MIG-6_{s1} และ TK โดเมน ของ EGFR มีค่าพลังงานเท่ากับ -1.88 กิโลแคลอรี โมล ต่อเเรซิดิว / ในระหว่างการทำ MD, Tyr341 สร้างพันธะไฮโดรเจนกับ Gln911 (hot spot) และอยู่ในโพรงส่วนที่ไม่ชอบน้ำของ Pro339, Pro347, Pro910 และ Pro913 นอกจากนี้ค่า $\Delta\Delta G$ ของการแทนอะลานีนของ Tyr ประมาณ 341-กิโลแคลอรี / โมล ซึ่ง 9.7 ได้รับผลจากการสูญเสียแรงวัลเดอวาลและอิเล็กโตรอสแตติก เเรซิดิวที่สำคัญอีกอย่างหนึ่งสำหรับการจับกันเป็นคอมเพล็กซ์ คือ Thr349 ของ MIG-6_{s1} ซึ่งฟอร์มพันธะไฮโดรเจนกับ Glu907 (hot spot) กับ TK โดเมน ของ EGFR ค่า $\Delta\Delta G$ ของการแทนอะลานีนของ Thr ประมาณ 349-กิโลแคลอรีต่อโมล ซึ่งเป็นผลมาจากการสูญเสียแรงอิเล็กโตรอสแตติก นอกจากนี้ 4.7 โครงสร้างหลักของ Gln พันธะกับ 2 ยังสร้างพันธะไฮโดรเจนอีก 350Arg) 908hot spot ในขณะที่โซ่ข้างไม่เพียงแต่สร้าง (พันธะไฮโดรเจนกับ Gln) 911hot spot แต่ยังก่อให้เกิดแรง วนเดอร์วาลส์กับ (Pro และ 910Try) 920warm spot ใน (ระหว่างการทำ MD stimulation ไม่ได้เพียงแต่สร้างพันธะไฮโดรเจนระหว่าง Ser361 และ Glu904 แต่ยังสร้างพันธะไฮโดรเจนที่เกี่ยวข้องกับ Asn343 ที่บริเวณลูปของ MIG-6_{s1} และโซ่ข้างของ MIG-6_{s1} เคลื่อนไปทางอินเทอร์เฟซของ TK โดเมน ของ EGFR และสร้างพันธะไฮโดรเจนกับ Thr885 (warm spot) และ Gly959 ดังภาพที่ 10



ภาพที่ 9 โปรไฟล์ปฏิสัมพันธ์ระหว่างอินเทอร์เฟซของ EGFR/MIG-6_s1

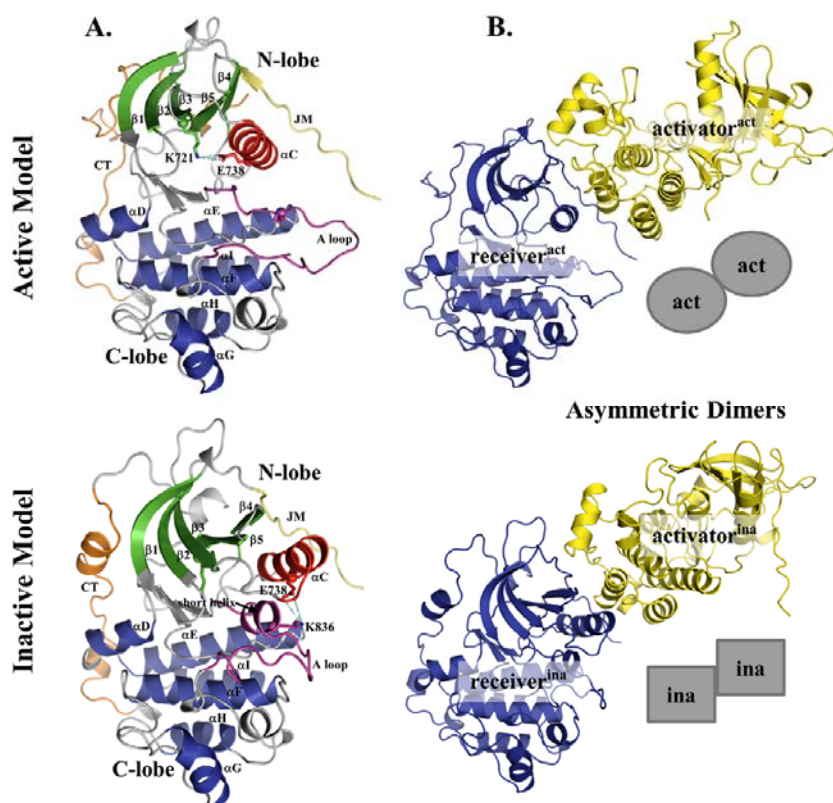


ภาพที่ 10 ปฏิสัมพันธ์ของเรซิดิวที่สำคัญของ Mlg-6_s1 กับเรซิดิวในบริเวณใกล้เคียง

2. การจำลองแบบทางพลวัตเชิงโมเลกุลของ TK domain ของ EGFR ในรูปแบบ asymmetric dimer : การจำลองโครงสร้างในรูปแบบแอคทีฟและอินแอคทีฟของไคเนสโดเมน

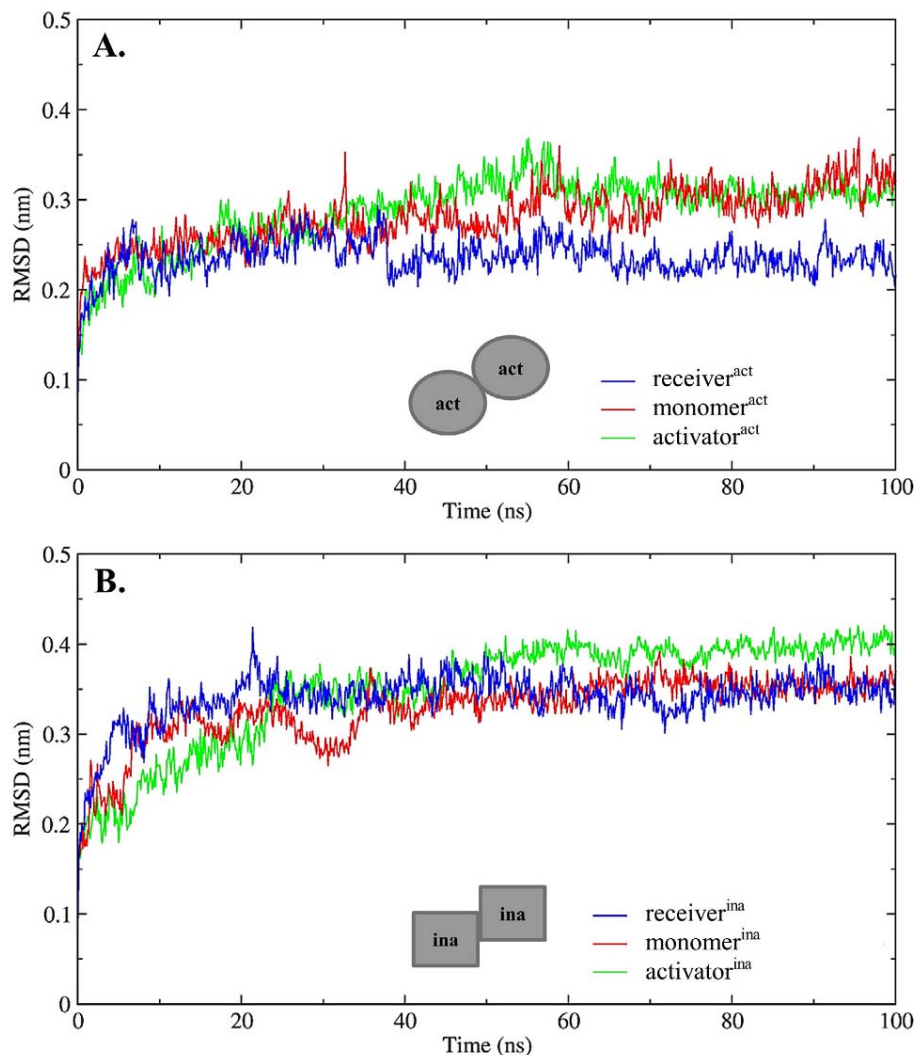
การใช้ homology-based เพื่อจำลองโมเดลโครงสร้าง 3 มิติของโครงสร้างแบบแอคทีฟและอินแอคทีฟของ TK โดเมน ของ EGFR เตรียมโครงสร้างของโปรตีน TK โดเมน ของ EGFR โดเมนทั้งในรูปแบบแอคทีฟและอินแอคทีฟซึ่งได้จากฐานข้อมูล PDB จากนั้นใช้โปรแกรม MODELLER สร้างโมเดลฮีเอพอาร์ในส่วนที่จะศึกษา C-terminal ส่วนของ JM (G672 to I682) และขยายส่วนของ CT segments (V 956 to I994) เพื่อใช้ในการศึกษารูปแอคทีฟฟอร์มใช้ PDB code 1m17 และ 2GS6 รูปอินแอคทีฟใช้ PDB code 1XKK, 2GS 7 and 2R4B จากนั้นนำโครงสร้างสามมิติไปทำ MD stimulation แบ่งออกเป็น โมโนเมอร์ ตัวรับสัญญาณและตัวกระตุ้น ทั้งในรูปแบบแอคทีฟและอินแอคทีฟฟอร์มด้วย GROMACS 4.5 และวิเคราะห์ผลการทำ MD ด้วย GROMACS utilities [7]

แบบจำลองเชิงโมเดล homology ผ่านการวิเคราะห์สเตอริโอเคมีด้วย Ramachandran plot พบว่าไม่มีเรซิดิวอยู่ในพื้นที่ที่ไม่เหมาะสมของกราฟทั้งแอคทีฟและอินแอคทีฟ 90% ของเรซิดิวตกอยู่ในพื้นที่ที่ยอมรับแสดงให้เห็นถึงที่น่าเชื่อถือของทั้งสองโมเดล โครงสร้างแบบแอคทีฟและอินแอคทีฟถูกสร้างขึ้นตามการศึกษาของ Zhang et al. [8] ดังภาพที่ 11B



ภาพที่ 11 โมเดลโครงสร้างโมโนเมอร์ (A) และโครงสร้างอะซิมเมตริก (B) ในรูปแบบแอคทีฟและอินแอคทีฟ TK โดเมน ของ EGFR โดเมน ไคเนสโดเมนรูปแบบโมโนเมอร์ในส่วน JM และ CT แสดงสีตามโครงสร้างดังนี้ N-lobe, β -strands (สีเขียว): C-lobe, α -helices (สีน้ำเงิน): α C-helix, สีแดง: activation loop, สีม่วงแดง เรซิดิว K721-E738 และ E738-K836 ซึ่งเป็น salt bridge แสดงด้วยเส้นประสีฟ้า และผลจากการทำ MD stimulation 100 ns กับอะซิมและโมโนเมอร์ของ TK โดเมน ของ EGFR โดเมน เพื่อตรวจสอบผลของการจับกันเป็นไดเมอร์มีผลต่อความยืดหยุ่นและความเป็นไดนามิก โดยใช้ค่า

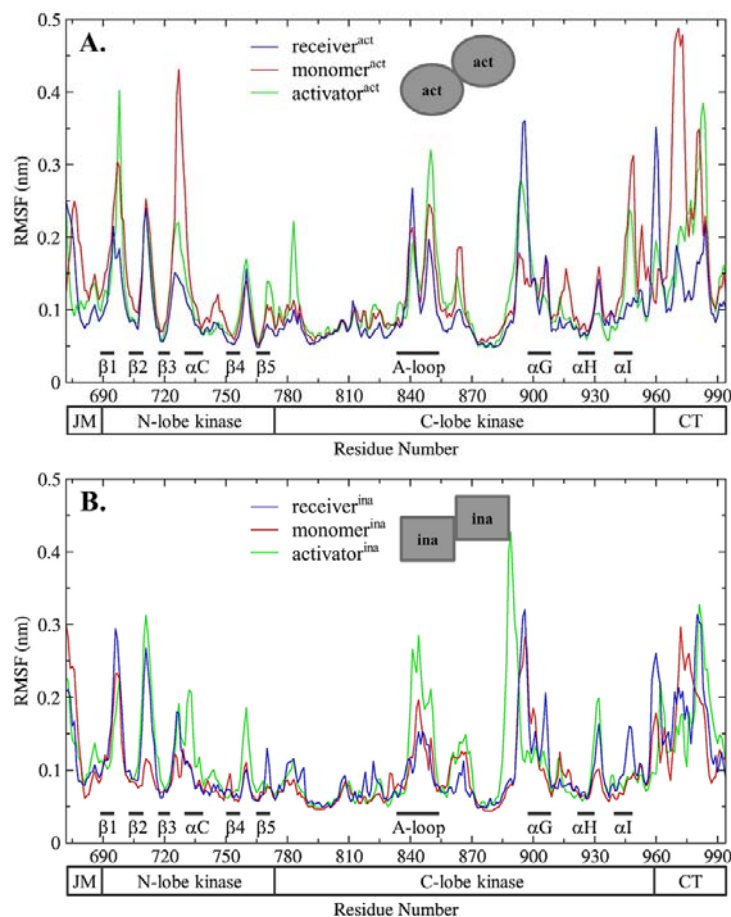
RMSD ในการสังเกตความเสถียรของระบบและความใกล้เคียงกันระหว่างโครงสร้างที่จำลองขึ้นกับโครงสร้างอ้างอิงในช่วงเวลาของการทำ MD ดังภาพที่ 12



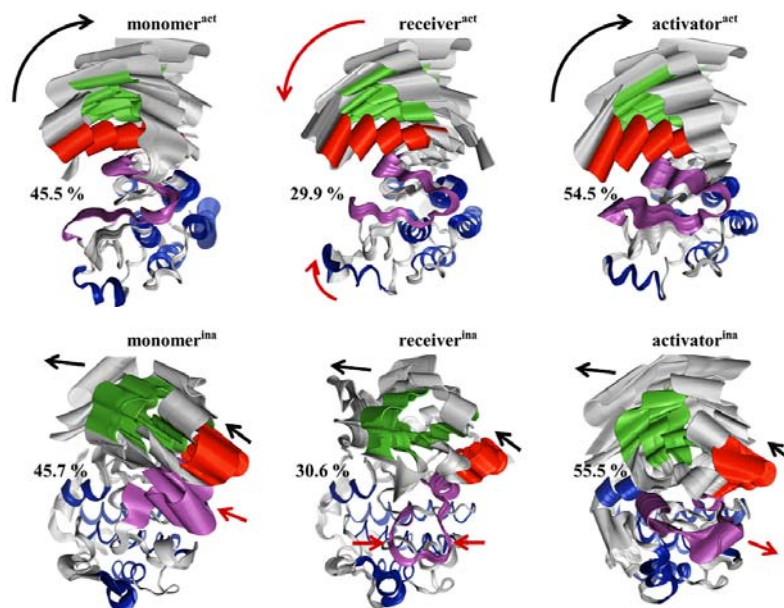
ภาพที่ 12 ค่า RMSD ของโครงสร้างหลักเป็นฟังก์ชันของเวลาในรูปแบบของแอคทีฟ (A) อินแอคทีฟ (B) จากโครงสร้างเริ่มต้นเพื่อให้เข้าใจถึงความยืดหยุ่นของโครงสร้างในแต่ละส่วนของ TK โดเมน ของ EGFR โดเมน จึงใช้ค่า RMSF เพื่อคำนวณค่าพลังงานในแต่ละเรซิดิวดังภาพที่ 3 การวิเคราะห์องค์ประกอบหลักของโครงสร้างแบบแอคทีฟ ด้วย MD trajectories ด้วยการดูในแต่ละ MD trajectories เพื่อระบุลักษณะเด่นของการเปลี่ยนแปลง เรียกว่า PC ในรูปแอคทีฟ 1 PC1 จะกระจายอย่างมากในส่วน N-lobe ของไคเนสโดเมนทั้ง 3 แอคทีฟโมเดลดังภาพที่ 4

อย่างไรก็ตามทิศทางของการเคลื่อนที่ของตัวรับ (receiver) ดูเหมือนตรงกันข้ามกับที่สังเกตเห็นได้ในตัวกระตุ้น (activator) และ โมโนเมอร์ ซึ่งไปมีส่วนร่วมกับการเคลื่อนที่ประมาณ %9.29, %5.54 และ %5.45 ของการเคลื่อนไหวทั้งหมดในโปรตีนตามลำดับ ผลที่เกิดขึ้นเหล่านี้ชี้ให้เห็นว่าอะซิมเมตริกไดเมอร์สามารถเปลี่ยนแปลงพลศาสตร์และลดความผันผวนของ TK โดเมน ของ EGFR โดเมน การเคลื่อนไหวระหว่างโครงสร้างแบบแอคทีฟมีแนวโน้มที่จะเกี่ยวข้องกับการเปิดและปิดบริเวณของ hinge region ซึ่งเป็นบริเวณที่มีความสำคัญของการจับกับ ATP และเร่งปฏิกิริยากับโปรตีนที่เป็น ไคเนสตัวอื่นๆ เช่น PKA [9] การเปิดและปิดของโครงสร้างอีจีเอฟอาร์แบบแอคทีฟเกิดจากการศึกษาด้วยการทำ MD ซึ่งเป็นการเคลื่อนที่ของไคเนสโดเมนอย่างใกล้ชิดควบคู่กับการเปลี่ยนแปลงโครงสร้างส่วนของ JM และ CT [10] ซึ่งหมายความว่า

เปิดและปิดของไคเนสโดเมนเป็นจุดสำคัญสำหรับของโปรตีนที่เป็นไคเนสในสภาวะที่เป็นแอคทีฟฟอร์ม ส่วนการศึกษาโครงสร้างแบบอินแอคทีฟ จากการทำ MD พบว่า PC1 มีการเคลื่อนที่คล้ายๆกันเกี่ยวกับการกระจายของ N-lobe แบบ rigid ในส่วนของ C-lobe ของไคเนสโดเมน การเคลื่อนที่ของ N-lobe มีลักษณะเป็นมุมฉากเช่นเดียวกับโมเดลแบบแอคทีฟ ความผันผวนของการเคลื่อนที่ในตัวรับ (receiver) ลดลงกว่าในตัวกระตุ้น (activator) และโมโนเมอร์มีส่วนประมาณ %6.30, %5.55 and %7.45 จากการเคลื่อนที่ทั้งหมดในโปรตีน ผลการทดลองเหล่านี้ชี้ให้เห็นว่าไม่เหมือนกับโครงสร้างแบบแอคทีฟ การฟอร์มเป็นอะซิมเมตริกไดเมอร์ในโครงสร้างแบบอินแอคทีฟไม่ได้เปลี่ยนไดนามิกส์ส่วนของ N-lobe ของ TK โดเมนของ EGFR โดเมน แต่ลดการผันผวนของโปรตีนแทนจะสังเกตได้ว่าการเคลื่อนที่ของ loop ในทั้ง 3 อินแอคทีฟฟอร์มมีความแตกต่างกันอย่างมีนัยสำคัญ loop A ของตัวรับเคลื่อนที่ได้น้อยลง ในขณะที่ loop A ของตัวกระตุ้นสามารถเคลื่อนที่ไปในทิศทางตรงกันข้ามกับ N-lobe การเคลื่อนที่ของ N-lobe นี้ C-helix และ loop A ในโมโนเมอร์มีความสัมพันธ์กันอย่างมาก แต่ทำงานของ PC1 ในรูปอินแอคทีฟ ยังไม่ชัดเจน การเคลื่อนที่ของ N-lobe กับการเคลื่อนที่ของเกลียว αC แบบปกติในสามรูปอินแอคทีฟ ทั้งสามโมเดล และชี้ให้เห็นว่าจะไม่ตอบสนองเฉพาะต่อการเปลี่ยนแปลงตามโครงสร้างของตัวรับ (receiver) ในการ dimerization นอกจากนี้การเคลื่อนที่ที่แตกต่างกันใน loop a ของโมเลกุลที่อินแอคทีฟอาจทำให้มีความยืดหยุ่นของ loop a ข้อสันนิษฐานนี้เป็นผลมาจากข้อมูลจำนวนมากจากการสุ่มตัวอย่างการทำ MD รวมทั้งโครงสร้างผลึกซึ่งทั้งหมดนี้มีความผิดปกติในส่วนปลายของ loop A (เรซิดิว E-842K (851) ดังภาพ 13B)

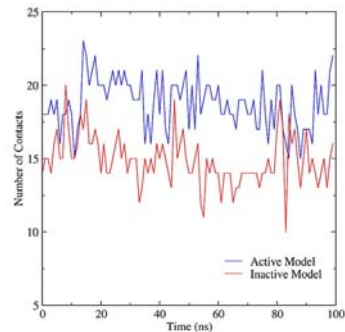
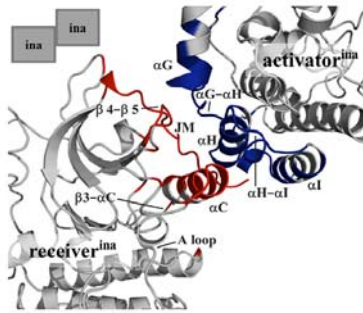


ภาพที่ 13 กราฟ RMSF ของโครงสร้างหลักแสดงฟังก์ชันต่อเรซิดิวจากโมเดลในรูปแอคทีฟและอินแอคทีฟ ส่วนของโครงสร้าง TK โดเมน ของ EGFR โดเมนและส่วนของ JM และ CT ในช่วงแต่ละ 50 ns

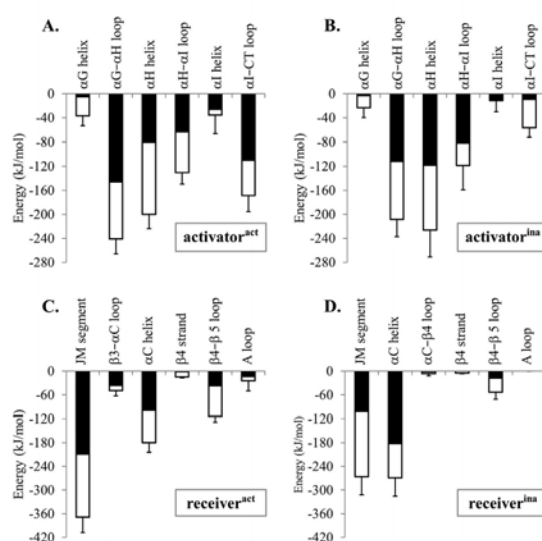


ภาพที่ 14 เปรียบเทียบการเคลื่อนที่ของโดเมนอธิบายด้วย PC1 จากการทำแบบจำลองการเคลื่อนที่ของรูปแบบแอคทีฟและอินแอคทีฟ ส่วนของโครงสร้างแบ่งออกตามสีดังนี้ N-lobe β -strands, สีเขียว; C-lobe α -helix, สีนํ้าเงิน; α C-helix, สีแดง; activation loop สีม่วงแดง และลูกศรชี้ทิศทางการเคลื่อนที่

การศึกษาปัจจัยที่ไม่มีส่วนเกี่ยวข้องกับการรักษาเสถียรภาพของปฏิสัมพันธ์ตัวกระตุ้นและตัวรับเป็นอะซิมเมตริกโดเมน TK โดเมน ของ EGFR โดเมน พบว่าพลังงานของปฏิสัมพันธ์ทั้งหมดของโมเดลในรูปอินแอคทีฟมีค่าประมาณ 4/5 น้อยกว่าโมเดลในรูปแอคทีฟ จากผลลัพธ์นี้อาจเกิดจากจำนวนของอะตอมที่มีปฏิสัมพันธ์กัน เช่น H-bond และแรงดึงดูดระหว่างโมเลกุลที่ไม่ชอบน้ำระหว่างตัวรับและตัวกระตุ้นของโมเดลในรูปแอคทีฟมากกว่าโมเดลในรูปอินแอคทีฟ ผลการทำ MD นอกจากนี้ยังพบว่าพลังงานวันเดอร์วาลส์มีความอ่อนกว่าอิเล็กโตรสแตติกในทั้งสองรูปแบบ อาจเกิดจากปฏิกิริยาพันธะโคเวเลนต์แบบมีขั้ว (พันธะไฮโดรเจน) การวิเคราะห์ตำแหน่งของเรซิดิวที่อยู่ในช่วง ≤ 50 นาโนเมตรในช่วง 5.0ns พบว่าบริเวณส่วนโครงสร้างของตัวกระตุ้นเกี่ยวข้องกับ อินเทอร์เฟซของโดเมน มีความคล้ายคลึงกันระหว่างโมเดล 2 โดเมน ตั้งแต่ αF - αG loop, αG helix, αG - αH loop, αH helix, αH - αI loop, αI helix, I-CT loop and the N-terminal ส่วนของ of CT ดังภาพที่ 15

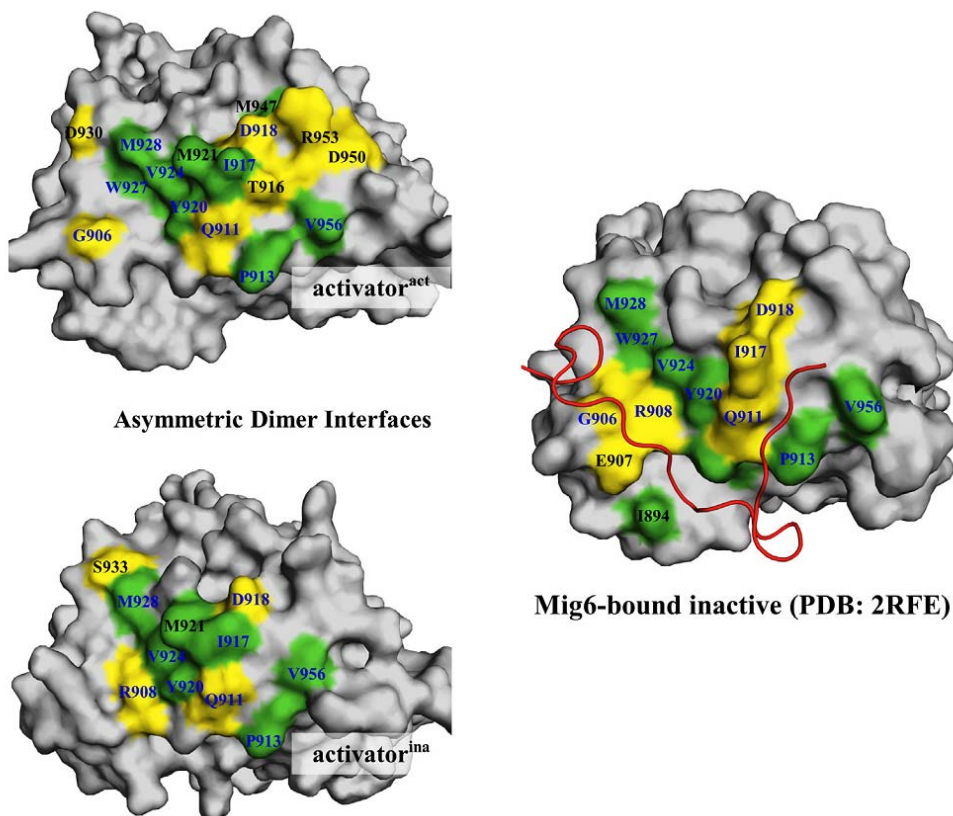


ในทางตรงกันข้ามโครงสร้างของตัวรับตอบสนองอินเทอร์เฟซของโดเมอร์รวมถึง ส่วน JM α C helix, α C- β 4 loop, β 4 strand, β 4- β 5 loop and β 5 strand อินเทอร์เฟซของโดเมอร์ของรูปแบบแอกทิฟจะเห็นว่า loop ระหว่าง α G และ α H (α G- α H loop) ของตัวกระตุ้นแสดงค่าพลังงานติดลบสูงมากจะเห็นได้ว่าเรซิดิวเหล่านี้มีผลต่อการฟอร์มเป็นโครงสร้างเป็นอะซิมเมตริกโดเมอร์และการกลายพันธ์ทำให้แอกทิวิตีของอีจีเอฟอาร์ในเซลล์ลดลง ดังภาพที่ 16พลังงานมีค่าติดลบสูงมากในส่วนของJM ของตัวรับ และยังสามารถบอกได้ถึงความสำคัญของส่วนนี้ในการฟอร์มตัวเป็นเป็นอะซิมเมตริกโดเมอร์ ส่วนอินเทอร์เฟซของโดเมอร์ของโครงสร้างแบบอินแอคทีฟ แตกต่างจาก อินเทอร์เฟซของโดเมอร์ของโครงสร้างแบบแอกทิฟ ปฏิสัมพันธ์ระหว่างเกลียว α H ของตัวกระตุ้น และ α C ของตัวรับอาจเป็นส่วนสำคัญของการฟอร์มเป็นเป็นอะซิมเมตริกโดเมอร์ของโครงสร้างแบบอินแอคทีฟ



ภาพที่16 พลังงานศักย์ระหว่างตัวรับและตัวกระตุ้นสังเกตได้จากการทำ MD ของแบบแอคทีฟ (A,C) แบบอินแอคทีฟ (B,D) พลังงานของการมีปฏิสัมพันธ์แสดงเป็นพลังงานที่ลดลงของ C-lobe ของตัวกระตุ้น และ N-lobe ของตัวรับ อิเล็กโตรสแตติกและแรงวนเดอร์วาลส์แสดงเป็นแถบสีดำและสีขาวตามลำดับ แถบแสดงข้อผิดพลาดแสดงค่าเบี่ยงเบนมาตรฐานของปฏิสัมพันธ์ทั้งหมด

เมื่อเปรียบเทียบกับกรจับที่อินเทอร์เฟซของ MIG-6_s1 โครงสร้างผลึกของTK โดเมน ของ EGFRโดเมน และส่วนของ Mig6 ซึ่งเป็นตัวยับยั้งแบบ feedback inhibitor ของอีจีเอฟอาร์ แสดงให้เห็นว่าเรซิดิว K336-K362 ของโปรตีน Mig6 สามารถจับกับ C-lobe ของTK โดเมน ของ EGFRโดเมน โครงสร้างผลึกของโครงสร้างดังกล่าวแสดงให้เห็นว่า Mig6 peptide ใช้โครงสร้างที่ยาวและวางอยู่บน อินเทอร์เฟซของ C-lobe ของโคเนส อินเทอร์เฟซถูกปกคลุมด้วยด้วย αG และ αH helices และ αF - αG , αG - αH , αH - αI loop, ซึ่งฝังอยู่ในพื้นที่ประมาณ 1 นาโนเมตร 0.187มีเรซิดิวไม่กี่ชนิดที่ [อยู่บนเกลียว αH ซึ่งช่วยในการเชื่อมต่อกับ อินเทอร์เฟซ และเมื่อเปรียบเทียบปฏิกิริยาของเรซิดิวที่อยู่บน C-lobe โคเนส โดเมนระหว่างอินเทอร์เฟซโดเมอร์แบบเป็นอะซิมเมตริกและที่จับกับ Mig พบว่าส่วนของการจับกับ 6MIG-6 ทับซ้อนกับ อินเทอร์เฟซของเป็นอะซิมเมตริกโดเมอร์เรซิดิวของกรดอะมิโนที่ใช้อินเทอร์เฟซร่วมกันรวมถึงมีพันธะไฮโดรเจน ได้แก่ G906, R908, Qใน911 αG - αH loop; I917, D ในเกลียว 918 αH และ ii) ส่วนที่ไม่ชอบน้ำปฏิสัมพันธ์ผ่าน P ใน 913 αG - αH loop; Y920, V921, V924, W บนเกลียว 927 αH ; M928 on the αH - αI loop เช่นเดียวกับ V956 บนส่วนของ CT การศึกษาทางชีวเคมีและชีวฟิสิกส์ของการจับ Mig6 และการยับยั้งอีจีเอฟอาร์ แสดงให้เห็นว่าการเปลี่ยนให้เกิดการกลายพันธุ์ของ V924R ที่โคเนสโดเมนทำให้ไม่สามารถจับกับเปปไทด์ได้ ในทางกลับกันโคเนสโดเมนที่มีการกลายพันธุ์ I682Q มีความสามารถในการจับเช่นเดียวกับโคเนสโดเมนปกติ ดังภาพที่ 17



ภาพที่17 การเปรียบเทียบการจับกันบริเวณอินเทอร์เฟซที่ C-lobe ของตัวกระตุ้นระหว่างโมเดลแอคทีฟและอินแอคทีฟ(ภาพซ้าย) เช่นเดียวกับโครงสร้างของ Mig6 จับกับโครงสร้างอินแอคทีฟ (PDB code:2RFE) Mig6 peptide แสดงด้วยริบบอนสีแดง บริเวณที่จับของอินเทอร์เฟซแสดงพื้นผิวและกรดอะมิโนเรซิดิวที่ก่อให้เกิดพันธะไฮโดรเจนและปฏิกิริยาไม่ชอบน้ำแสดงด้วยสีเหลืองและสีเขียวตามลำดับ เรซิดิวที่ปฏิกิริยาร่วมกันแสดงด้วยตัวอักษรสีน้ำเงิน

3. การคัดเลือกแอนติบอดีจาก HuScFv และ VH/V_HH libraries

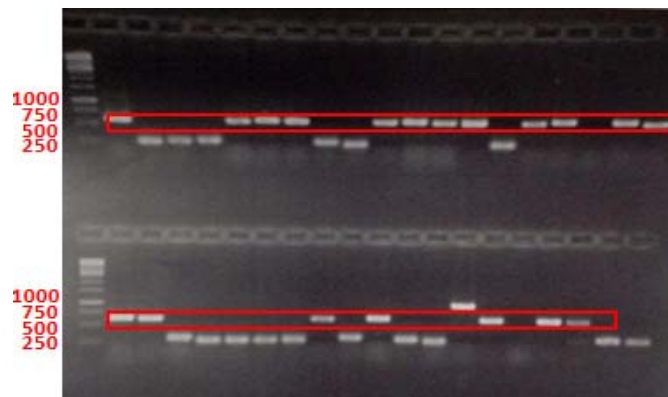
เทคนิค Single round biopanning ถูกนำมาใช้เพื่อคัดเลือกแอนติบอดีที่มีความสามารถกับ TK โดยขั้นแรกเริ่มจากการนำ TK มาตรึงบน ELISA Plate ด้วย Coating buffer (Bicarbonate buffer pH 9.6) โดยกำหนดความเข้มข้นของ TK ที่ 2 μ g ต่อ well จากนั้นทำการ blocking ด้วย 3%BSA แล้วนำไปบ่มเป็นเวลา 2 ชั่วโมงที่อุณหภูมิ 37 องศา จากนั้นล้าง BSA ส่วนเกินออกให้หมด จึงเติม library ลงไป ขั้นตอนเป็นขั้นตอนที่เรียกว่า Phage enrichment นำไปบ่มเป็นเวลา 1 ชั่วโมงที่อุณหภูมิ 37 องศา จากนั้นล้าง unbound phage ออกให้หมด หลังจากเติม Mid-log Phase HB2151 *E.coli* เพื่อให้ฟาจ infect ดังกล่าว ทำให้เกิดการถ่ายถอดข้อมูล ทางพันธุกรรมหรือ phagemid จากฟาจสู่แบคทีเรีย จากนั้นนำไปบ่มเป็นเวลา 1 ชั่วโมงที่อุณหภูมิ 37 องศา ขั้นตอนสุดท้าย ทำแบคทีเรียที่ได้ มาเลี้ยง บนอาหาร แข็ง เพื่อให้เกิดการสร้างโคโลนีสำหรับตรวจสอบเทคนิค PCR ต่อไป

3.1 การคัดเลือกแอนติบอดีโดยใช้ PCR

เนื่องจาก HuScFv และ VH/V_HH นั้นเป็นส่วนส่วนของยีนที่อยู่ใน vector ที่เป็น phagemid ซึ่งต่างกับ plasmid ตรงที่ phagemid นั้นมีจุดเริ่มต้นในการสังเคราะห์ DNA ของแบคทีเรีย และฟาจอยู่ร่วมกัน ส่วน primer ที่ใช้ในการตรวจสอบยีนนั้นจะจำเพาะกับยีนที่เราต้องการ โดยขนาดของ amplicon ที่ได้จาก HuScFv ประมาณ 1000 bp (ภาพที่18) ในส่วนของ VH/V_HH จะมีขนาดของ amplicon ประมาณ 750 bp (ภาพที่19)โดยสภาวะที่เหมาะสม ในการเพิ่มจำนวนยีน เพื่อตรวจยีนของแอนติบอดี 95 องศา 1 นาที 55 องศา 30 วินาที และ 72 องศา 1 นาที



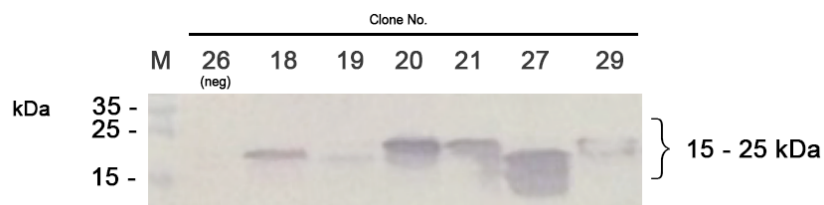
ภาพที่ 18 แสดงแถบ amplicon HuScFv ที่พบใน HuScFv library 1% Agarose gel



ภาพที่ 19 แสดงแถบ amplicon VH/V_HH ที่พบใน VH/V_HH library 1% Agarose gel

3.2 การทดสอบการแสดงออกของใน Phagemid ในเซลล์แบคทีเรีย (HB 2151 E.coli)

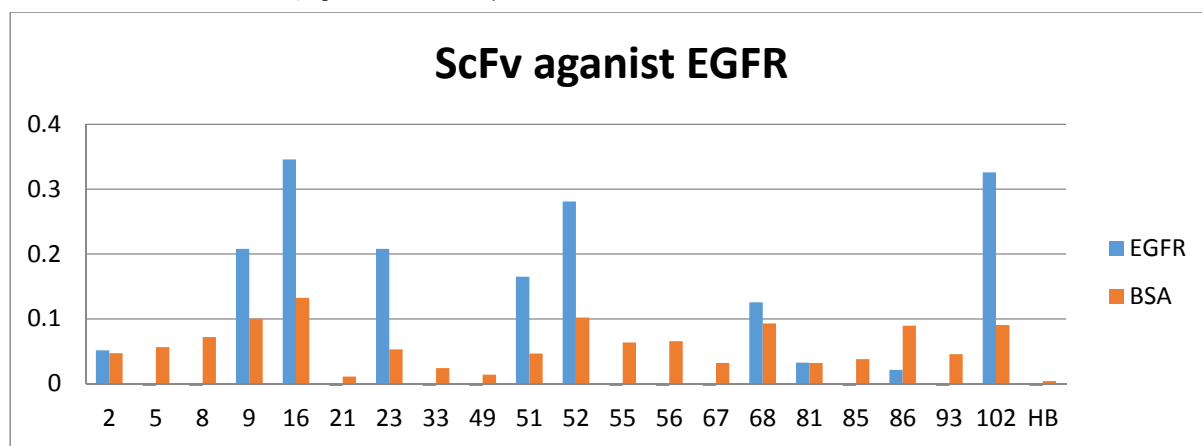
จากการทดลอง ผู้วิจัยได้นำโคลนในข้อ 3.3 มาเลี้ยงในอาหารเลี้ยงเชื้อ จนมีค่าการดูดกลืน ที่ความยาว คลื่น 600 nm (OD_{600}) ประมาณ 0.5-0.6 จากนั้นทำการเหนี่ยวนำให้มีการแสดงออกแอนติบอดี โดยใช้ 1mM IPTG จากนั้นทำไปบ่มที่อุณหภูมิ 37 องศา เป็นเวลา 3 ชั่วโมง หลังจากนั้นนำมาปั่นตกตะกอนเซลล์ เพื่อตรวจสอบการแสดงออกด้วยการแยกโดยใช้ 14% acrylamide SDS-PAGE หลังจากนั้นตรวจสอบ การแสดงออกของแอนติบอดีด้วยเทคนิค western blot analysis (1:3000 Mouse anti E-tag primary antibodies, 1:5000 Goat anti mouse AP-conjugated secondary antibodies) ขนาดของ HuScFv ประมาณ 23-30 kDa ขนาดของ VH/VHH ประมาณ 15-20 kDa



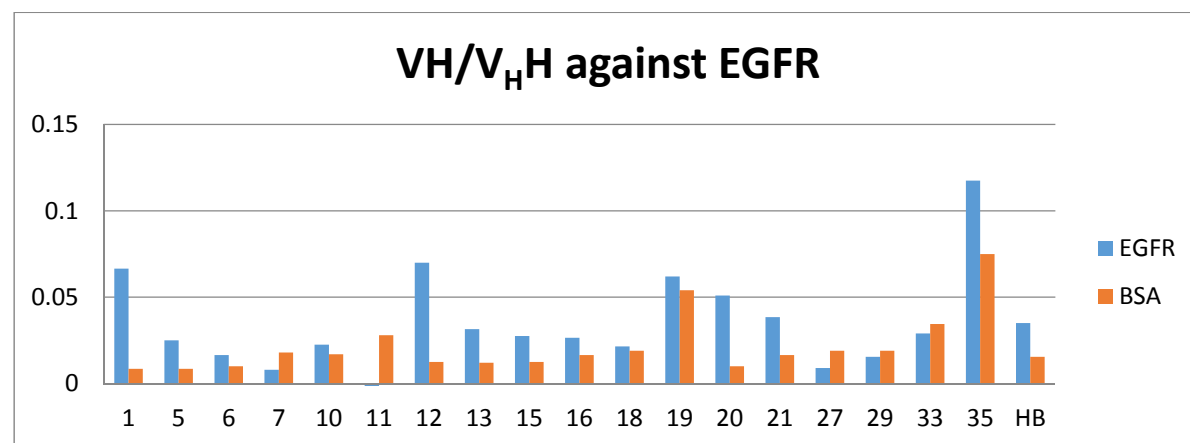
ภาพที่ 20 แสดงแถบการแสดงออกของ VH/V_HH

3.3 การทดสอบความสามารถในการจับระหว่างแอนติบอดีกับ TK โดยใช้เทคนิค Enzyme-linked immunosorbent assay และ western binding

ในการทดลอง ELISA ขั้นตอนแรกเริ่มจากการตรึง TK ลงใน ELISA plate โดยเจือจางโปรตีน ให้มีความเข้มข้น 2 μg ใน 100 μl ต่อ well และในส่วนของชุดควบคุมให้ตรึงด้วย bovine serum albumin (BSA) ในความเข้มข้นที่เท่ากัน นำไปบ่มที่อุณหภูมิ 37 องศา เป็นเวลา 16-18 ชั่วโมง จนกระทั่ง ELISA plate แห้ง จากนั้นล้างโปรตีนส่วนเกินออก แล้ว Blocking ด้วย 3% BSA เป็นเวลา 2 ชั่วโมง หลังจากนั้นล้าง BSA ส่วนเกิน แล้วเติม lysate ของ HB 2151 E.coli ที่มีการแสดงออกของแอนติบอดี (ทำ Normalization ด้วยโปรแกรม digidoc จากผลของ Western blot) นำไปบ่มที่อุณหภูมิ 37 องศา เป็นเวลา 2 ชั่วโมง หลังจากนั้นล้าง lysate ของ HB 2151 E.coli ที่มีการแสดงออกของแอนติบอดีส่วนเกินออก เตรียม detection complex โดยใช้ 1:3000 mouse anti E-tag primary antibodies mouse anti E-tag และ 1:5000 Goat anti mouse HRP-conjugated secondary antibodies วัดที่ความยาวคลื่น 450 nm



ภาพที่ 21 แสดงผล ELISA ของ HuScFv ต่อ EGFR

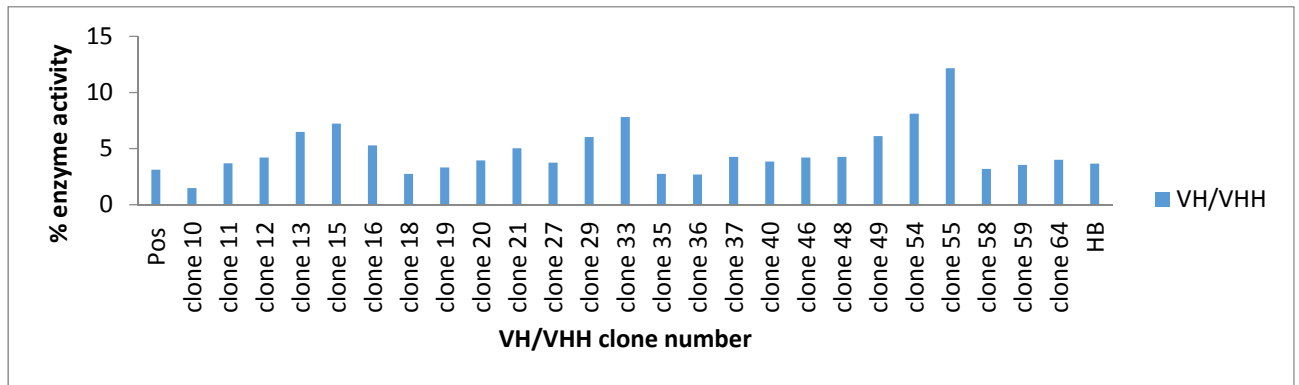


ภาพที่ 22 แสดงผล ELISA ของ VH/VH ต่อ EGFR

จากผลการทดลองพบว่า HuScFv มีความสามารถในการจับกับได้ดีกว่า VH/VH ประมาณ 5 เท่าเมื่อเปรียบเทียบตัวที่มีความสามารถในการจับกับ TK ทั้งสองกลุ่มการทดลองเปรียบเทียบกับชุดควบคุมแต่อย่างไรก็ตามผู้วิจัยพบว่าความสามารถในการจับไม่ได้บ่งบอกความสามารถในการยับยั้งการทำงานของได้

3.4 การทดสอบความสามารถของแอนติบอดีในการยับยั้งการทำงานของ TK

ผู้วิจัยได้ใช้ Kinase Glo Max Assay kit การตรวจสอบการยับยั้งการทำงานของ TK โดยแอนติบอดีที่ได้จาก Libraries โดยหลักการของ Kinase Glo assay kit จะวัดปริมาณ ATP ที่คงเหลือโดยใช้เอนไซม์ Luciferase เป็นตัวบ่งชี้ (Chemiluminascescence Indicator) โดยเอนไซม์ดังกล่าวจะใช้ ATP แล้วใช้สัญญาณ Chemiluminascescence ออกมา ดังนั้นถ้าแอนติบอดีสามารถยับยั้งการทำงานของ TK ได้ สัญญาณ Chemiluminascescence ต่ำ จากภาพที่ 23 แสดงเปอร์เซ็นต์การยับยั้งการทำงานของ TK โดย VH/V_HH (HuScFvที่ได้จาก library เกิดความความผิดพลาดทางความหลากหลายใน library)

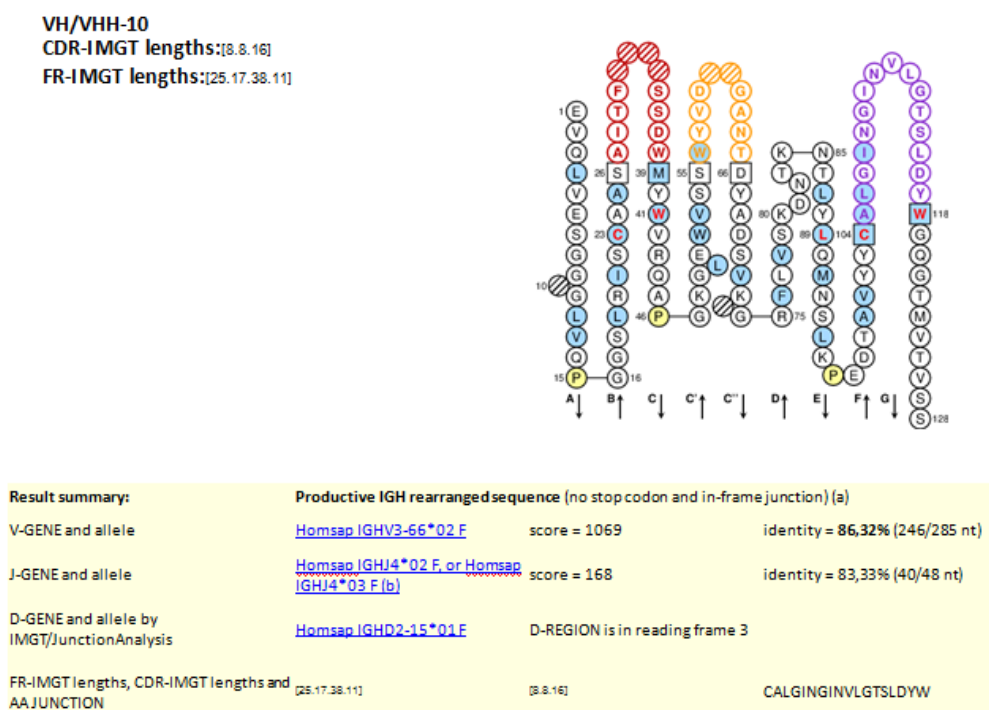


ภาพที่ 23 แผนภูมิ แสดงความสามารถในการยับยั้งการทำงานของ TK โดย VH/V_HH โดยใช้ Kinase Glo Max assay kit

จากผลการทดลองพบว่า VH/VHH 10, 18, 35 และ 36 มีความสามารถในการยับยั้งการทำงานของ TK แต่ 55 มีความสามารถในการกระตุ้นการทำงานของ TK ได้ดีเมื่อเปรียบเทียบกับ Positive control ผู้วิจัยจึงเลือกทั้ง 5 โคลนนี้นำมาลำดับนิวคลีโอไทด์ต่อไปเพื่อทำนายโครงสร้าง

เมื่อทราบลำดับนิวคลีโอไทด์แล้วผู้วิจัยได้แปลรหัสเป็นกรดอะมิโนเมื่อนำไปเปรียบเทียบกับลำดับนิวคลีโอไทด์ของมนุษย์โดยใช้ฐานข้อมูลทางด้านอิมมูโนโกลูลิน และพบ VH ที่ได้มีความคล้ายคลึงกับยีนของมนุษย์มากกว่า 85% (ภาพที่ 24)

ภาพที่ 24 แสดงตัวอย่างลำดับกรดอะมิโนของ VH/V_HH



3.5 การผลิต Cell-permeable antibodies

จากผลการทดลองขั้นต้นพบว่า แอนติบอดีจาก VH/V_HH library จำนวน 4 โคลน มีความสามารถในการยับยั้งการทำงานของของ TK เพื่อประสิทธิภาพในการเข้าเซลล์ ไปสู่โปรตีนเป้าหมายภายในเซลล์ ยีนของแอนติบอดีถูกเพิ่มจำนวนด้วยไพรเมอร์จำเพาะ

F-R9 ScFv-LIC

5' GGT TGG GAA TTG CAA CGT **CGT CGT CGT CGT CGT CGT CGT CGT CGT** GCG GCC CAG CCG GCC 3'

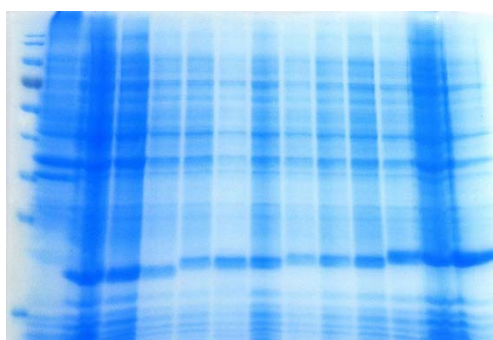
R-E ScFv-LIC

5' GGA GAT GGG AAG TCA TTAA ACG CGG TTC CAG CGG ATC 3'

ขนาดของ PCR product ที่ได้ประมาณ 850 bp หลังจากนั้นชิ้นส่วนยีนถูกทำให้บริสุทธิ์โดยวิธี Gel Extraction ชิ้นส่วนยีนที่ผ่านการทำให้บริสุทธิ์จะถูกนำเข้าสู่พลาสมิด pLATE52 โดยวิธี Ligation Independent Cloning pLATE52-VH/V_HH-R9 ถูก Transform เข้าสู่ JM109 *E.Coli* เพื่อทำการ Screen แบคทีเรียที่ได้รับ pLATE52-VH/V_HH-R9 โดยวิธี Colonies PCR หลังจากนั้นนำ Positive ที่ได้มาสกัด Plasmid เพื่อ Sequencing และ Transform เข้าสู่ Rosetta α (DE3) *E.Coli* สำหรับการตรวจสอบการแสดงออกของโปรตีน



ภาพที่ 25 ชิ้นส่วนยีน pLATE52-VH/V_HH-R9



ภาพที่ 26 การแสดงออกของโปรตีนใน pLATE52-VH/V_HH-R9
(1mM IPTG, 37c, 4hr)

3.6 การเปรียบเทียบลำดับนิวคลีโอไทด์ของโปรตีนลูกผสมนาโนบอดี

หลังจากที่ผู้วิจัยได้ย้ายยีนที่ผลิตนาโนบอดี ที่สามารถเข้าสู่เซลล์ได้สู่ pLATE52 ทั้ง 5 โคลน พบว่าโคลนที่น่าสนใจคือ VH10, VH18, V_HH35, VH36 และ VH55 ตามลำดับ

การทดลองเริ่มจากการนำโคลนของแต่ละโคลนมาเลี้ยงใน 5 mL LB broth + 10 µg/mL Ampicillin นำไปเลี้ยงที่อุณหภูมิ 37°C ประมาณ 16-18 ชั่วโมง หลังจากนั้นนำเชื้อที่ได้มาสกัดพลาสมิดด้วยชุดสกัด วัดความเข้มข้นด้วย nanodrop spectrometer ที่ความยาวคลื่น 260 nm โดยไพรเมอร์ที่ใช้ในการหาลำดับนิวคลีโอไทด์คือ LIC-FW และ LIC-RW

	6xHis	T7 tag	WELQ	Arg-9	FR-1
VH18	HHHHHH	GMASMTGGQQMG	RSGWELQ	RRRRRRRRR	AAQPAEVQLVESGGGLVQPGGSLRLS
VH55	HHHHHH	GMASMTGGQQMG	RSGWELQ	RRRRRRRRR	AAQPAEVQLVESGGGLVQPGGSLRLS
VH36	HHHHHH	GMASMTGGQQMG	RSGWELQ	RRRRRRRRR	AAQPAQVQLVESGGGLVQPGGSLRLS
VHH35	HHHHHH	GMASMTGGQQMG	RSGWELQ	RRRRRRRRR	AAQPAEVQLVESGGGSVQAGGSLRPS
	*****	*****	*****	*****	*****:*****.*****

	FR-1	CDR-1	FR-2	CDR-2	FR-3
VH18	CAAS	GFTFSHYW	MSWVRQAPGKGLEWVAN	IKQDGSEK	YFVDSVKGRFTISRDNAKNSLYL
VH55	CAAS	GFTFSHYW	MSWVRQAPGKGLEWVAN	IKQDGSEK	YFVDSVKGRFTISRDNAKSSLYL
VH36	CAAS	GFTFNYYL	MYWVRQAPRLGLEWVS-	--TTGGIA	AYADSVKGRFTISRDNAKNTMYL
VHH35	CAAS	GYTYCSYD	MSWYRQAPGKEREFVSA	IDSDGS-T	SYADSVKGRFTISQDNAKNTVYL
	****	*:*:	* * *	***:	*. :.*****:****.::**

	FR-3	CDR-3	FR-4	E-tag
VH18	QMNSLRAEDTAVYYC	ARVPETTVTTGPLLYYYYGMDV	WGQGTIVTVSS	AAAGAPVPYPDP
VH55	QMNSLRAEDTAVYYC	ARVPETTVTTGPLPYYYYYGMDV	WSQGTIVIVSS	AAAGAPVPYPDP
VH36	QMNSLEPVDATVYYC	VRGGGEYVSP-----ID	WGQGTIVTVSS	AAAGAPVPYPDP
VHH35	QMNSLKPEDTAVYYC	KTDG-----YGCWSFY	NGQGTMTVTVSS	AAAGAPVPYPDP
	*****.	***:***	.*** * ***	*****

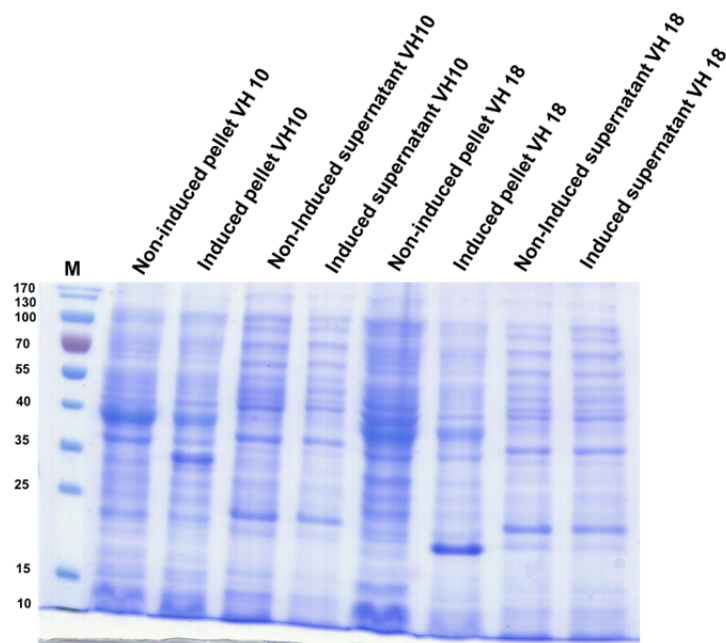
	E-tag
VH18	LEPR
VH55	LEPR
VH36	LEPR
VHH35	LEPR

ภาพที่ 27 แผนภาพแสดงลำดับนิวคลีโอไทด์ของโปรตีนลูกผสมนาโนบอดี VH18, V_HH35, VH36, และ VH36 (6xHis; 6xHistidine tag, WELQ; Protease cleavage site, FR;Framework, CDR;Complementarity Determining Region)

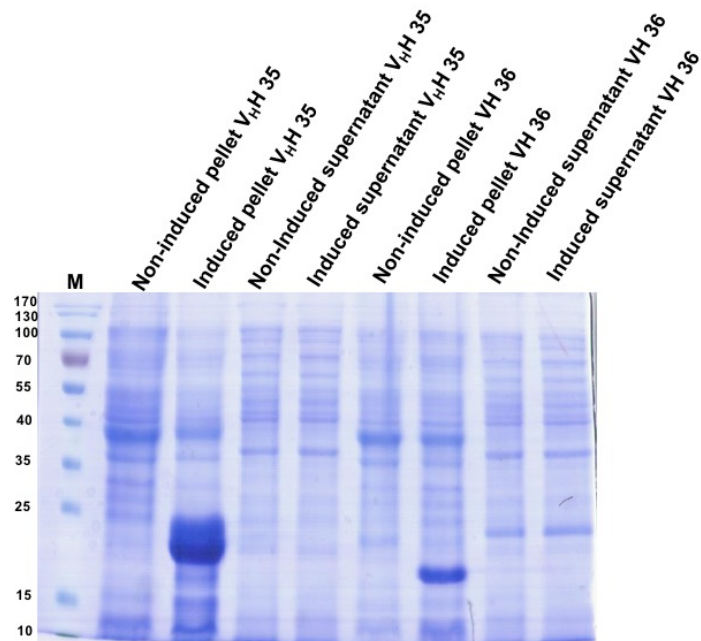
ผลการทดลองที่ได้จะถูกระบุโดยใช้โปรแกรม Clustal W2 ในการเปรียบเทียบความแตกต่างของโปรตีนลูกผสมนาโนบอดี จากภาพที่ 18 พบว่า 6xHis tag ได้ถูกเติมที่ปลาย N เพื่อใช้ในการทำให้บริสุทธิ์ ตามด้วย T7 tag บริเวณตัดของเอนไซม์โปรตีเอส และ Arg-9 (Cell-penetrating peptide) สำหรับการนำของโปรตีนลูกผสมนาโนบอดีเข้าสู่เซลล์ ส่วนในบริเวณปลาย C ยังคงมีลำดับกรดอะมิโนอนุรักษ์ของ E-tag ในส่วนของโครงสร้างหลักนั้นพบว่าบริเวณที่เป็นบริเวณกรดอะมิโนอนุรักษ์ของนาโนบอดีตามธรรมชาติยังคงอยู่ ซึ่งจะอยู่ในส่วนของ FR-1-FR-4 ในส่วนบริเวณที่มีความหลากหลายของกรดอะมิโนนั้นพบว่าจะอยู่ในช่วง CDR-3 ของโปรตีนลูกผสมนาโนบอดี แต่อย่างไรก็ตามพบว่าโปรตีนลูกผสมนาโนบอดี VH18, VH55 มีความคล้ายคลึงในส่วนของลำดับกรดอะมิโน >95%

3.7 การหาสภาวะที่เหมาะสมในผลิตโปรตีนลูกผสมนาโนบอดี

เพื่อทำการผลิตโปรตีนลูกผสมนาโนบอดีในปริมาณมากพอสำหรับการทดลองในขั้นต่อไป ผู้วิจัยได้เริ่มจากการหาสภาวะที่เหมาะสมในเหนี่ยวนำให้เกิดการแสดงออกของโปรตีนลูกผสมนาโนบอดีในแบคทีเรีย ชนิด E.coli สายพันธุ์ Rosetta เริ่มต้นการทดลองผู้วิจัยได้เลี้ยงแบคทีเรียที่มีพลาสมิดของนาโนบอดีลงบนอาหารแข็งเพื่อให้ได้โคโลนีเดี่ยว หลังจากนั้นนำไปเลี้ยงต่อในอาหารเหลวเป็นเวลา 16-18 ชั่วโมง จากนั้น เติม 1%ของเชื้อดังกล่าวลงในอาหารเหลว เลี้ยงที่อุณหภูมิ 37°C ประมาณ 3 ชั่วโมง หรือจนกระทั่ง $OD_{600} \approx 0.5-0.6$ จากนั้นทำการเหนี่ยวนำให้มีการผลิตโปรตีนโดยการเติม โดยให้มี IPTG ความเข้มข้นสุดท้ายที่ 0.5 mmol/L แล้วนำไปเลี้ยงต่อที่อุณหภูมิ 37°C เป็นเวลา 3 ชั่วโมง ทำการเก็บตะกอนเซลล์โดยเครื่องปั่นเหวี่ยง ทำการแตกเซลล์โดยใช้คลื่นเสียงที่ 25% นำไปปั่นเหวี่ยงแยกตะกอนเซลล์อีกครั้ง เก็บส่วนใสและตะกอนมาแยกโดยวิธี SDS-PAGE



ภาพที่ 28 12% SDS-PAGE แสดงการแสดงออกของโปรตีนลูกผสมนาโนบอดี VH10 และ VH18 ในสภาวะที่ถูกเหนี่ยวนำด้วย 0.5 mmol/L IPTG ที่อุณหภูมิ 37°C



ภาพที่ 29 12% SDS-PAGE แสดงการแสดงออกของโปรตีนลูกผสมนาโนบอดี V_HH35 และ V_H36 ในสภาวะที่ถูกเหนี่ยวนำด้วย 0.5 mmol/L IPTG ที่อุณหภูมิ 37°C

ตารางที่ 1 ตารางแสดงน้ำหนักโมเลกุล จำนวนกรดอะมิโน น้ำหนักโมเลกุล และค่า pI ของโปรตีนลูกผสมนาโนบอดี

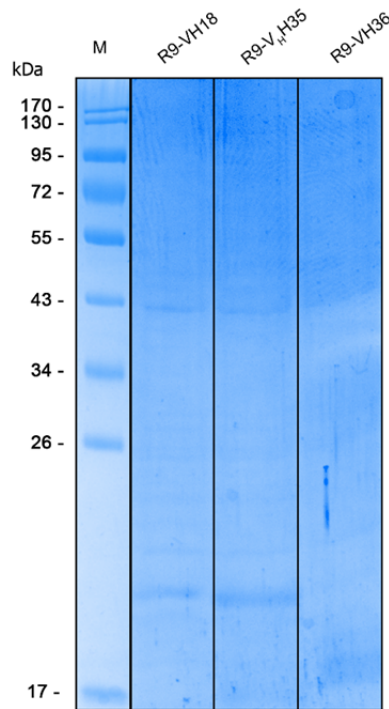
Sample	aa.	MW	pI
VH10	296	32	9.36
VH18	184	20	9.73
V _H H35	173	19	9.10
VH36	171	18	10
VH55	184	20	9.73

ผลการทดลองพบว่าการแสดงออกของโปรตีนลูกผสมนาโนบอดีนั้นจะแสดงออกในรูปแบบที่ไม่ละลายน้ำมากกว่าในสภาวะที่ถูกเหนี่ยวนำด้วย 0.5 mmol/L IPTG ที่อุณหภูมิ 37°C โดยเฉพาะอย่างยิ่งโปรตีนลูกผสม V_HH35 ซึ่งมีการแสดงออกมากเมื่อเปรียบเทียบกับโคลนอื่นในสภาวะเดียวกัน เนื่องจากความเหมือนกันของโปรตีนลูกผสมนาโนบอดี VH18 และ VH55 ผู้วิจัยจึงได้เลือกทำเฉพาะโคลน VH18 ในส่วนของ VH10 ที่มีน้ำหนักโมเลกุลมากกว่าโคลนอื่นเนื่องจาก VH10 มีโครงสร้างทั้งในส่วนของ Heavy chain (VH) และ Light chain ของ variable domain หรือเรียกอีกอย่างหนึ่งว่า แอนทิบอดีสายเดี่ยว (Single-Chain Fragment Variable domain; ScFv) จึงมีน้ำหนักมากกว่าโคลนอื่น น้ำหนักโมเลกุลของโปรตีนลูกผสมนาโนบอดีแต่ละโคลนสามารถเปรียบเทียบได้จากตารางที่ 1

3.8 การม้วนพับโปรตีนลูกผสมนาโนบอดี (Protein Refolding)

หลังจากที่ได้ Denatured protein หรือ Solubilized Protein จากกระบวนการผลิตโปรตีนจากแบคทีเรีย หลังจากนั้นนำมาทำให้เข้มข้นโดยใช้ 3kDa cut off amicon ให้มีปริมาตรเหลือประมาณ 500 μ L และกำหนดความเข้มข้นโดย

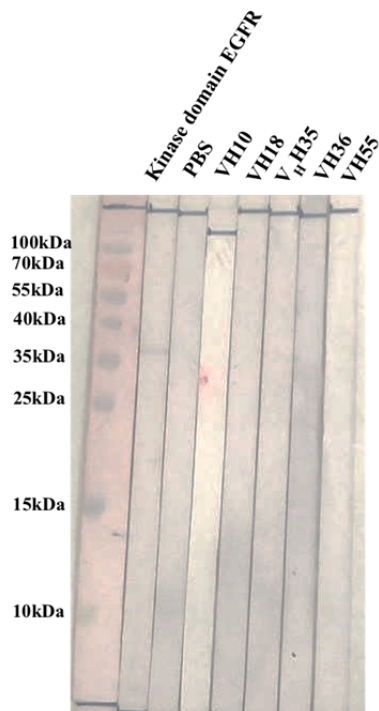
Bradford assay หลังจากนั้นนำไปเจือจางใน Refolding Buffer โดยอาศัยหลักการ ลดความเข้มข้นของสารที่ทำให้โปรตีนเสียสภาพ ซึ่งในการทดลองนี้ผู้วิจัยใช้ Urea ที่มีความเข้มข้น 8 mol/L โดยอาศัยการเจือจาง (Dilution Refolding) แบบ Drop to Drop ใน Refolding buffer ให้มีความเข้มข้นสุดท้ายของโปรตีนเท่ากับ 0.1 mg/mL นำมาที่อุณหภูมิ 4⁰c เป็นเวลา 1 ชั่วโมง หลังจากนั้นนำมากรองด้วย 0.22μm nitrocellulose filter สุดท้ายเปลี่ยนเป็นบัฟเฟอร์ที่เหมาะสมในการทำการทดลองต่อไป



ภาพที่ 30 12% SDS-PAGE แสดงแถบโปรตีนนาโนบอดีย้อมด้วย comassie blue

3.9 การทดสอบความสามารถในการจับกันของนาโนบอดีกับโปรตีน EGFR

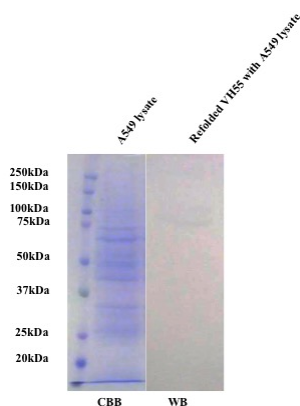
เพื่อทดสอบความสามารถของ Nanobody ทั้ง 5 โคลน ผู้วิจัยได้ใช้วิธี Western blot binding โดยเริ่มจากการเตรียม โปรตีนลูกผสม ไคเนสโดเมน ของ EGFR ที่ผ่านการ ≈) Refold 40kDa (ให้มีความเข้มข้น 20μg/mL) นำไปแยกผ่านกระบวนการอิเล็กโตรโฟรีซิส และย้ายลงสู่ NC membrane หลังจากนั้นนำ ที่ผ่านกระบวนการ NC membrane Blot มาตัดเป็นแท่ง ต่อมานำไป Blocking ด้วย Blocking buffer ที่อุณหภูมิ 37⁰ เป็นเวลา 2 ชั่วโมง แล้วนำบ่มกับ cell lysate ของแบคทีเรียที่มีการผลิตนาโนบอดีในแต่ละโคลน ที่อุณหภูมิ 37⁰ เป็นเวลา 2 ชั่วโมง หลังจากนั้นทำการติดตามด้วย กระบวนการ western blotting (1:3000 Mouse anti His-tag primary antibodies, 1:5000 Goat anti mouse AP-conjugated secondary antibodies) สับสเตรทที่ใช้ คือ BCIP/NBT



ภาพที่ 31 Western blot binding strip

3.10 การทดสอบความสามารถในการจับกันของนาโนบอดีกับโปรตีน EGFR กับเซลล์มะเร็ง

เพื่อทดสอบความสามารถของ Refolded nanobody กับโปรตีนในสภาวะจริงภายในเซลล์มะเร็ง การทดลองเริ่มจากการเลี้ยงเซลล์มะเร็งปอดชนิด A549 ในอาหารเลี้ยงเซลล์ชนิด DMEM ที่มีส่วนผสมของ 10% Fetal Bovine Serum (FBS) 100U penicillin-streptomycin ที่อุณหภูมิ 37°C , 5% CO_2 ในฟลasks T75 cm^2 จนกระทั่งเซลล์เจริญเติบโต 90% ของพื้นที่ฟลasks ผู้วิจัยได้ทำการย่อยเซลล์ (trypsinization) โดยใช้ 0.25% trypsin/EDTA solution ปั่นแยกตะกอนแล้วสกัดโปรตีนออกจากเซลล์โดยใช้ Extraction buffer จากนั้นทำการแยกด้วย PAGE-SDS และ Western blotting โดยใช้ Refolded Nanobody 5 โคลน (1:3000 Mouse anti His-tag primary antibodies, 1:5000 Goat anti mouse AP-conjugated secondary antibodies) สับสเตรทที่ใช้ คือ BCIP/NBT



ภาพที่ 32 นาโนบอดีสามารถจับกับ EGFR ในเซลล์มะเร็งชนิด A549

3.11 การเลี้ยงเซลล์มะเร็ง

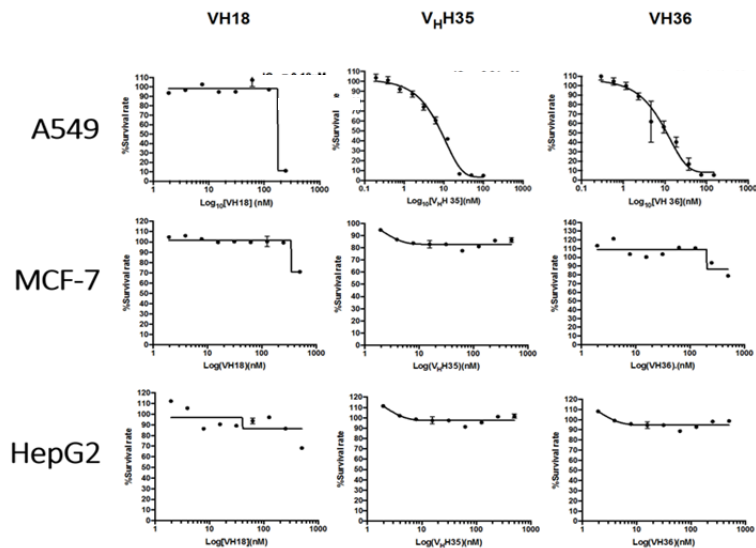
การทดลองเริ่มต้นจากเซลล์มะเร็งปอดชนิด non-small cell lung cancer cell (A549) ถูกนำมาเลี้ยงในอาหารเลี้ยงเซลล์ชนิด Dulbecco's Modified Eagle Medium (DMEM) โดยมีเติม 10% Fetal Bovine Serum (FBS) และ 100U/mL Penicillin/Streptomycin เซลล์ถูกเลี้ยงในตู้บ่มที่มีการควบคุมก๊าซคาร์บอนไดออกไซด์ (CO₂ Incubator) ที่อุณหภูมิ 37 °C, 5%CO₂ เซลล์ถูกเลี้ยงใน T25 cm³ cell culture flask จนกระทั่งเซลล์แผ่กระจายตัวจนทั่วฟลasks (90% confluence) ใช้เวลาประมาณ 3-7 วัน หลังจากนั้นเซลล์จะถูก Sub-culture โดยการ Trypsinize ด้วย 10x Trypsin/EDTA solution หลังจากนั้น trypsinized cells จะถูกทำให้เป็นเซลล์เดี่ยวก่อนที่จะนำไปเลี้ยงต่อใน T75cm³ cell culture flask เซลล์ถูกเลี้ยงต่อในสภาวะที่เหมาะสมตามที่กล่าวข้างต้น ประมาณ 3-5 วัน เซลล์จะพร้อมสำหรับการทดลองในขั้นต่อไป

3.12 กำหนดค่า IC₅₀ โดย Cell-based Experiment

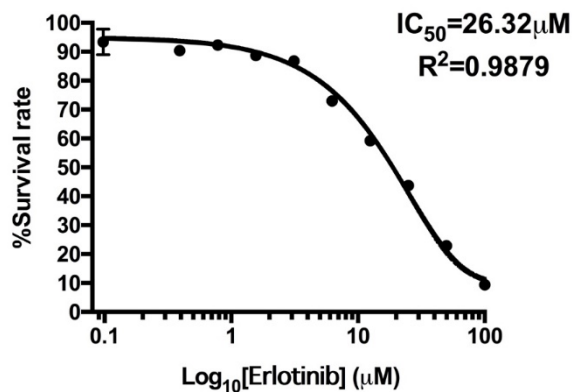
การทดลองเริ่มต้นการเพาะเลี้ยงเซลล์มะเร็งชนิด A549, HepG2 และ MCF-7 จำนวน 5x10³ เซลล์ ต่อ well ใน 96-well plate ด้วยอาหารเลี้ยงเซลล์ชนิด complete DMEM (10%FBS, 100U/mL penicillin/streptomycin) จากนั้นนำไปบ่มในสภาวะที่เหมาะสม ที่อุณหภูมิ 37 °C, 5%CO₂ เป็นเวลา (16-18 ชั่วโมง หลังจากนั้นเซลล์จะถูกเปลี่ยนเป็นอาหารที่มีส่วนผสมของนาโนโบดี้ที่มีความเข้มข้นตามที่กำหนดไว้ (100nM, 50nM, 25nM, 12.5nM, 6.25nM, 3.125nM, 1.5nM, 0.75nM, 0.3125nM) ทุกความเข้มข้นของนาโนโบดี้แต่ละตัวจะถูกเจือจางใน Serum-free DMEM (SFM) ในส่วนของชุดควบคุมนั้น ผู้วิจัยได้ใช้ Phosphate Buffer Saline (PBS) เป็น Blank สำหรับการทดลองนี้ เพื่อเปรียบเทียบความสามารถของนาโนโบดี้กับยารักษามะเร็งที่ใช้ในปัจจุบัน ในการทดลองนี้ผู้วิจัยเลือก Erlotinib (Tarceva) (100µM, 50µM, 25µM, 12.5µM, 6.25µM, 3.125µM, 1.5µM, 0.75µM, 0.3125µM) เป็นตัวยาที่มุ่งเป้าในการรักษามะเร็งปอดที่มีการแสดงออกของ EGFR โดยในชุดการทดลองนี้ได้ใช้ DMSO เป็น Blank เป็นชุดควบคุม โดยกำหนดให้มีความเข้มข้นไม่เกิน 0.1%DMSO ในแต่ละ well หลังจากที่ยบ่มเซลล์กับนาโนโบดี้เป็นเวลา 72 ชั่วโมง ให้ดูดอาหารเก่าทิ้งแล้วเติมด้วย Complete DMEM ที่มีส่วนผสมของสารละลาย MTT โดยให้มีความเข้มข้นสุดท้ายอยู่ที่ 0.5mg/mL นำไปบ่มต่อที่ตู้ CO₂ Incubator เป็นเวลา 3 ชั่วโมง หลังจากนั้นจะสังเกตเห็นว่าเกิดผลึกสีม่วงที่บริเวณด้านล่างของ well เนื่องจากเซลล์ที่ยังมีชีวิต (Cell viability) จะสามารถเปลี่ยน MTT เป็นผลึก Formazan ซึ่งเป็นผลึกสีม่วงไม่ละลายน้ำ ซึ่งเกิดจากการใช้เอนไซม์ mitochondria reductase เป็นตัวบ่งชี้ว่าเซลล์ยังมีชีวิตอยู่ จากนั้น เติม 50 DMSOµL แล้วนำไปวัดค่าความดูดกลืนแสงที่ความยาวคลื่น 570 nm นำค่าที่ได้มาคำนวณหา %viability โดยใช้สูตร

$$\% \text{viability} = \left(\frac{\text{ค่า OD ของเซลล์ที่เลี้ยงในอาหารเลี้ยงเซลล์ที่มีส่วนผสมของนาโนโบดี้}}{\text{ค่า OD ของเซลล์ที่เลี้ยงในอาหารเลี้ยงเซลล์ที่เติม DMSO เป็น Blank}} \right) \times 100$$

สุดท้ายกำหนดค่า IC₅₀ จาก Graphpad Prism 6.0



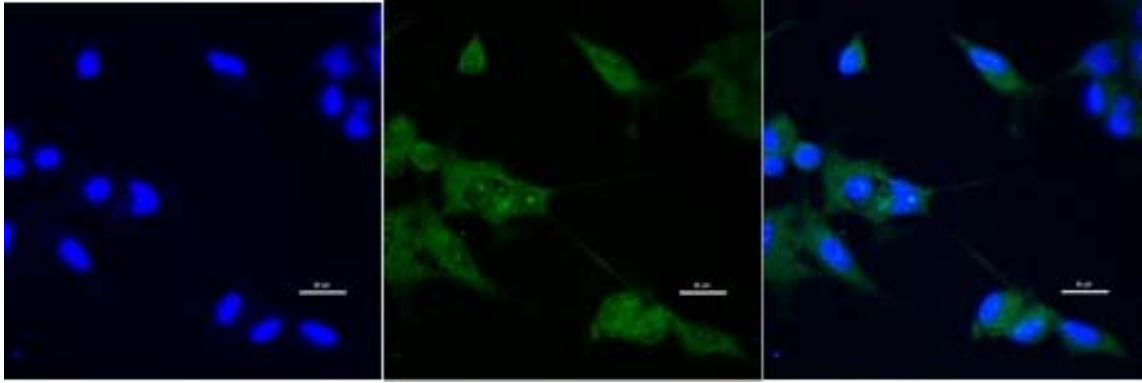
ภาพที่ 33 กราฟแสดงค่าความสามารถในการยับยั้ง (IC_{50}) ของนาโนบอดี



ภาพที่ 34 กราฟแสดงค่าความสามารถในการยับยั้ง (IC_{50}) ของ Erlotinib

3.13 การทดสอบความสามารถในการเข้าเซลล์ (Cell penetrating) โดยใช้กล้องฟลูออเรสเซนซ์แบบตัดเฉือน (Confocal fluorescence microscope)

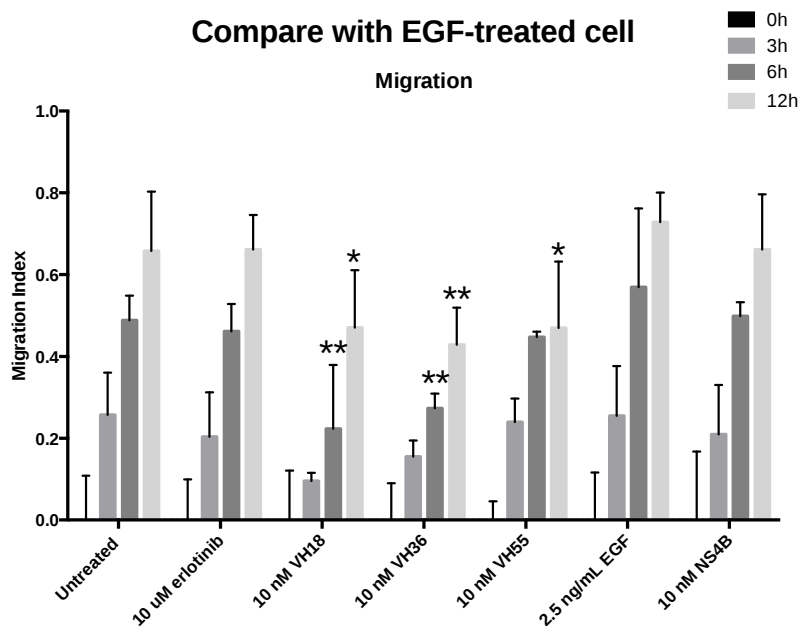
เซลล์ถูกเลี้ยงในโดบนแผ่นสไลด์ใน 24 well plate เป็นเวลา 16-18 ชั่วโมง ใน complete DMEM ที่สภาวะที่เหมาะสม ($5\%CO_2$, $37^{\circ}C$) หลังจากดูดอาหารเก่าทิ้งและเลี้ยงเซลล์ต่อด้วย SFM เป็นเวลา 16-18 ชั่วโมง เมื่อครบกำหนดเวลาเซลล์จะถูกล้าง 1XPBS ที่อุ่น ก่อนที่จะบ่มด้วย $40\ \mu g$ นาโนบอดีใน SFM เป็นเวลา 1 ชั่วโมง หลังจากนั้นล้างนาโนบอดีส่วนเกินออกด้วย 1XPBS ที่เย็นจัด หลังจากนั้นเซลล์จะถูกตรึง (Fixation) ด้วย 100% methanol ที่เย็นจัด เป็นเวลา 5 นาที ที่อุณหภูมิ $-20^{\circ}C$ ล้าง methanol ส่วนเกินด้วย 1XPBS ที่เย็นจัด หลังจากนั้น permeabilize เซลล์ด้วย 1% Trion X-100 เป็นเวลา 15 นาทีที่อุณหภูมิห้อง หลังจากนั้นเติม $200\ \mu l$ Blocking buffer (PBST- 0.1% , 1% BSA, $0.3M$ Glycine, 10% FBS) บ่มต่อเป็นเวลา 1 ชั่วโมง ที่อุณหภูมิ $4^{\circ}C$ หลังจากนั้นล้างด้วย 1XPBS ที่เย็นจัด และบ่มด้วย primary antibody (Mouse anti-Histidine tag) อัตราส่วน (1:1000) ใน 1XPBS เป็นเวลา 1 ชั่วโมง และ secondary antibody (Goat anti-mouse Alexa flour 488) อัตราส่วน (1:1000) เป็นเวลา 1 ชั่วโมง สุดท้ายเซลล์ถูกย้อมด้วย DAPI ก่อนนำไปตรวจสอบด้วยกล้องฟลูออเรสเซนซ์แบบตัดเฉือน



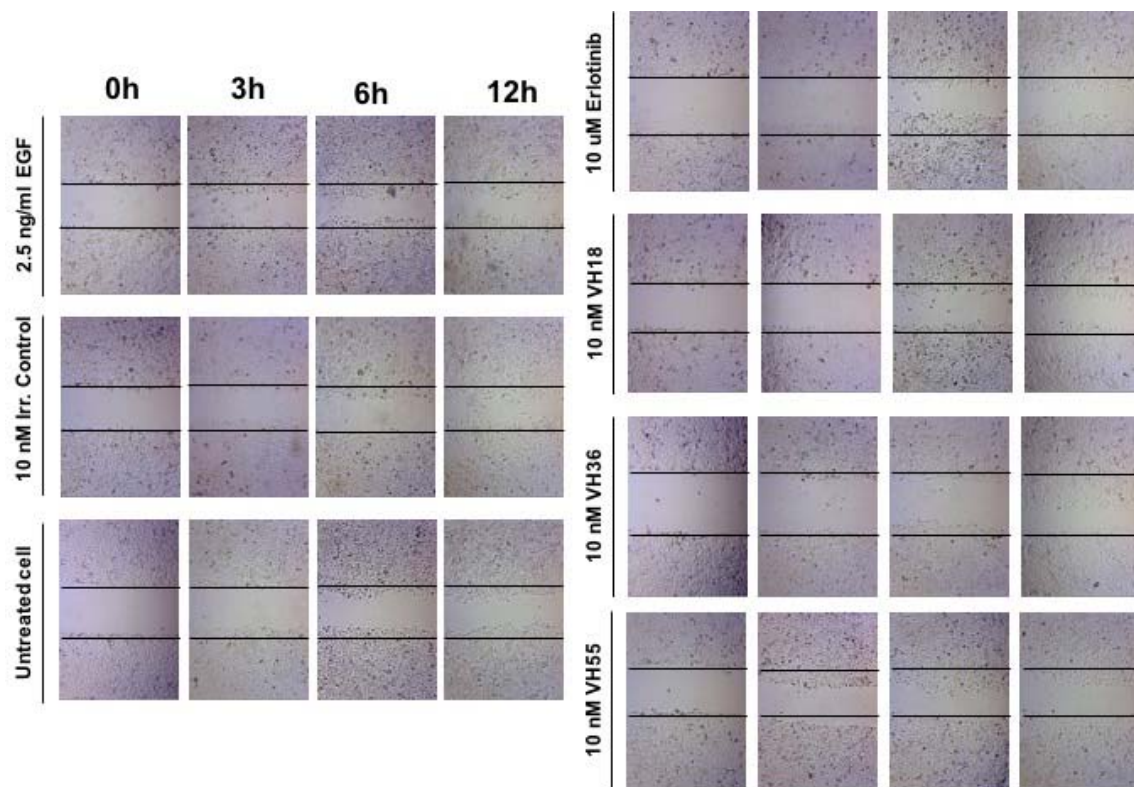
ภาพที่ 35 ภาพจากกล้องฟลูออเรสเซนซ์แบบตัดเฉือนของนาโนบอดี No. 18 แสดงให้เห็นว่านาโนบอดีสามารถเข้าสู่เซลล์ได้ ภาพซ้าย การย้อมเซลล์ด้วย DAPI, ภาพกลาง Alexa fluor 488, ภาพขวา DAPI และ Alexa fluor 488

3.14 การทดสอบความสามารถของนาโนบอดีในการยับยั้งการเคลื่อนที่ของเซลล์มะเร็ง

เซลล์ (5×10^5 cell/well) ถูกเลี้ยงใน 24 well plate เป็นเวลา 16-18 ชั่วโมง โดยสังเกตให้เซลล์เจริญเติบโตจนเต็มประมาณ 90 % ของพื้นที่ หลังจากนั้นเลี้ยงด้วย SFM เป็นเวลา 24 ชั่วโมง ล้างเซลล์ด้วย 1XPBS เติม 10 nM ของนาโนบอดีในแต่ละโคลนซึ่งเตรียมใน SFM ลงไปบ่มเป็นเวลา 1 ชั่วโมงที่สภาวะที่เหมาะสม ในตู้ CO₂ Incubator หลังจากนั้นนำเซลล์มาล้างด้วย 1XPBS เพื่อทดสอบประสิทธิภาพของนาโนบอดี 100 ng/mL Epidermal Growth Factor (EGF) ใน SFM เป็นเวลา 15 นาทีที่สภาวะที่เหมาะสม ในตู้ CO₂ Incubator หลังจากนั้นดูดอาหารทิ้ง ใช้ ทิป 200 μ L ชีดยบน monolayer ให้เป็นรอยตามยาว จากนั้นเลี้ยงเซลล์ที่หลุดออกด้วย 1XPBS แล้วเลี้ยงเซลล์ต่อด้วย complete DMEM การเคลื่อนที่ของเซลล์จะถูกเก็บภาพทุก 3, 6, 12 ชั่วโมง ระยะทางในการเคลื่อนที่จะถูกกำหนดโดย ImageJ software เพื่อนำมาคำนวณหาดัชนีในการเคลื่อนที่ (Migration Index, MI) $MI = (1 - (\text{ระยะทางในการเคลื่อนที่ที่ชั่วโมงใดๆ} / \text{ระยะในการเคลื่อนที่ที่ชั่วโมงที่ 0})) \times 100$



ภาพที่ 36 แผนภูมิแท่งแสดงดัชนีการเคลื่อนที่ของเซลล์มะเร็งเมื่อถูกรักษาด้วยนาโนบอดีเปรียบเทียบกับยาด้านมะเร็ง (Erlotinib) โดยกำหนดให้ EGF-treated cell เป็นชุดควบคุม



ภาพที่ 37 ภาพแสดงการเคลื่อนที่ของเซลล์มะเร็งเมื่อถูกรักษาด้วยนาโนบอดีเปรียบเทียบกับยาด้านมะเร็ง (Erlotinib) โดยกำหนดให้ EGR-treated cell เป็นชุดควบคุม

สรุปผลการทดลอง

จากผลการทดลองพบว่าผู้วิจัยได้ใช้คอมพิวเตอร์ในการศึกษาการจับกันระหว่างไทโรซีนไคเนสโดเมนของ EGFR เพื่อกระตุ้นให้ทำงาน และการจับกับเปปไทด์ MIG6 เพื่อยับยั้งการทำงานของไทโรซีนไคเนส นอกจากนี้ได้คัดเลือกนาโนบอดีที่จำเพาะกับไทโรซีนไคเนสของ EGFR ผ่านกระบวนการที่เรียกว่า Biopanning หลังจากนั้นได้ทำการทดสอบคุณสมบัติในการจับกับโปรตีนเป้าหมายโดยใช้เทคนิค ELISA พบว่ามี นาโนบอดีโคลนที่ 10, 18, 35, 36, 55 ที่มีความสามารถในการจับกับโปรตีนเป้าหมายได้ดี รวมทั้งยังมีความสามารถในการยับยั้งกิจกรรมการทำงานของโปรตีนไทโรซีนไคเนสได้อีกด้วยเพื่อเพิ่มความสามารถในการเข้าเซลล์ผู้วิจัยได้เพิ่ม Arg-9 ที่ N-terminal ของนาโนบอดี ผลจากกล้องฟลูออเรสเซนซ์แบบตัดเฉือนพบว่า นาโนบอดีที่มี Arg-9 มีความสามารถในการเข้าเซลล์ รวมทั้งยังมีความสามารถในการยับยั้งการเจริญเติบโตของเซลล์มะเร็งปอดชนิด A549 อย่างจำเพาะเจาะจง โดยมีค่า IC_{50} อยู่ในระดับ nM ซึ่งดีกว่าค่า IC_{50} ของยาด้านมะเร็งที่ใช้อยู่ในปัจจุบัน มากไปกว่านั้นนาโนบอดียังมีความสามารถในการยับยั้งการเคลื่อนที่ของเซลล์อีกด้วย

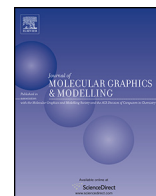
เอกสารอ้างอิง

- [1] Zhang X, Pickin KA, Bose R, Jura N, Cole PA, Kuriyan J. Inhibition of the EGF receptor by binding of MIG6 to an activating kinase domain interface. *Nature*. 2007;450 (7170) :741–4.
- [2] Case DA, Cheatham 3rd TE, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. *J Comput Chem*. 2005;26 (16) :1668–88.
- [3] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Physics*. 1983;79(2) :926–35.
- [4] Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*. 2000;33(12) :889–97.
- [5] Zhong H, Carlson HA. Computational studies and peptidomimetic design for the human p53-MDM2 complex. *Proteins*. 2005;58(1):222–34.
- [6] Red Brewer M, Choi SH, Alvarado D, Moravcevic K, Pozzi A, Lemmon MA, et al. The juxtamembrane region of the EGF receptor functions as an activation domain. *Mol Cell*. 2009;34 (6):641–51.
- [7] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, et al., *Gromacs User Manual Version 4.5.4*, 2010 www.gromacs.org
- [8] X. Zhang, J. Gureasko, K. Shen, P.A. Cole, J. Kuriyan, An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor, *Cell* 125 (2006) 1137–1149.
- [9] N. Narayana, S. Cox, X. Nguyen-huu, L.F. Ten Eyck, S.S. Taylor, A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility, *Structure* 5 (1997) 921–935.
- [10] M. Mustafa, A. Mirza, N. Kannan, Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: a molecular dynamics study, *Proteins* 79 (2011) 99–114.

Output ที่ได้จากโครงการ

1. Songtawee, N.; Bevan, D.R.; **Choowongkamon**, K. Molecular dynamics of the asymmetric dimers of EGFR: Simulations on the active and inactive conformations of the kinase domain. *J. Mol. Graph. Model.*, **2015**, 58, 16-29.
2. Moonrin N; Songtawee N; Rattanabunyong S; Chunsrivirod S; Mokmak W; Tongsimma S, **Choowongkamon** K. Understanding the molecular basis of EGFR kinase domain/MIG-6 peptide recognition complex using computational analyses. *BMC Bioinform.*, **2015**, 16(1):103.
3. Lueach Tabtimai, Praphasri Suphakun, Pimonwan Srisook, Trin Athigapanich, Duangnapa Dirivan, Pichamon Kiatwuthinon, Wanpen Chicumpa, Kiattawee **Choowongkamon**, Cell-Penetrable Nanobodies Specific to Tyrosine Kinase Domain of Epidermal Growth Factor (EGFR) Able to Modulate Cell Viability of Human Lung Adenocarcinoma A549 Cells. Poster presentation at *The 5th International Biochemistry and Molecular Biology Conference 2016*, B.P. Samila Beach Hotel, Songkhla (26-27 May, 2016)

ภาคผนวก



Molecular dynamics of the asymmetric dimers of EGFR: Simulations on the active and inactive conformations of the kinase domain



Napat Songtawe^a, David R. Bevan^b, Kiattawee Choowongkamon^{a,c,*}

^a Department of Biochemistry, Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand

^b Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, United States

^c Center for Advanced Studies in Tropical Natural Resources, National Research University-Kasetsart University, Kasetsart University, Chatuchak, Bangkok 10900, Thailand

ARTICLE INFO

Article history:

Accepted 5 March 2015

Available online 14 March 2015

Keywords:

EGFR

Kinase domain

Asymmetric dimers

Molecular dynamics

Essential dynamics

Intermolecular interactions

ABSTRACT

Abnormal activation of EGFR is associated with human cancer, and thus it is a key target for inhibition in cancer therapy. There is evidence suggesting that the activation mechanism of EGFR is based upon the formation of the asymmetric dimer of the kinase domains. Here, we performed MD simulations on the asymmetric dimer for both active and inactive conformations of EGFR kinase domain to investigate flexibility and intrinsic motions of the proteins. Simulations of the active conformation showed that the formation of the asymmetric dimer changes the dynamics of EGFR kinase domain by suppressing fluctuation of the protein and altering the direction of motion of the protein. In contrast, the asymmetric dimerization of the inactive conformation does not alter the overall fluctuation of the kinase domain and does not initiate destabilizing of the inactive structure. We also investigated the intermolecular interactions in the EGFR asymmetric dimers and found that in the active conformation the interactions are dominated by loop–loop contacts rather than those from the helix–helix interactions. In contrast, helix–helix interaction seemed to be more significant for the inactive kinase structure. This work helps us to better understand the conformational flexibility and dynamics of the EGFR kinase domain, as well as provides information that may be useful to develop newer classes of inhibitors that can block allosteric sites rather than the more traditional catalytic site.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein receptor tyrosine kinase that plays crucial roles in cellular responses to environments mediated by its natural ligands and growth factors. Aberrant activation or mutations of EGFR can promote tumor cell growth and survival of non-small cell lung cancer (NSCLC) as well as other types of cancer, e.g. liver, stomach, colorectal, breast and esophageal cancer [1,2]. Much effort on cancer research has been focused on designing drugs that inhibit EGFR activity [3]. EGFR can be divided into three structural domains: (i) an extracellular ligand-binding domain, (ii) a single transmembrane segment, and (iii) an intracellular tyrosine kinase domain flanked by a juxtamembrane (JM) and an extended C-terminal tail (CT) segments [4]. It is known that ligand-induced dimerization of the extracellular domain results in the activation of the kinase

domain, catalyzing the γ -phosphate transfer reaction from ATP to the conserved tyrosine residue at its C-terminal tail [5,6]. Since autophosphorylation at the C-terminal tail initiates several downstream signaling pathways [6]. As a result, EGFR kinase activity requires careful regulation.

Numerous crystal structures of the human EGFR kinase domain revealed the characteristics common to almost all protein kinases in that it is comprised of two lobes: (i) a smaller N-terminal lobe consisting mainly of β -strands with a single large α C helix; and (ii) a larger C-terminal lobe, which is almost exclusively α -helical with a long activation (A) loop. The ATP-binding site is located at the hinge region between two lobes, and thus motions of the lobes can determine dynamics of the active site of the enzyme (Fig. 1). Among the known crystal structures, it has been observed that the EGFR kinase domain exhibits at least two fundamentally different conformations. One adopts a catalytically active conformation in the presence of the anti-cancer drug erlotinib [7], and the other form is in an autoinhibited/inactive conformation bound to the drug lapatinib [8]. The pronounced structural features of the active conformation compared to that in the inactive one include: (i) orientation of the α C helix closer to the active site, resulting in the salt bridge formation of E738 on the helix and K721 on the

* Corresponding author at: Department of Biochemistry, Faculty of Science, Kasetsart University, Chatuchak, Bangkok 10900, Thailand. Tel.: +66 2 5625555x2051; fax: +66 2 5614627.

E-mail address: fscikt@ku.ac.th (K. Choowongkamon).

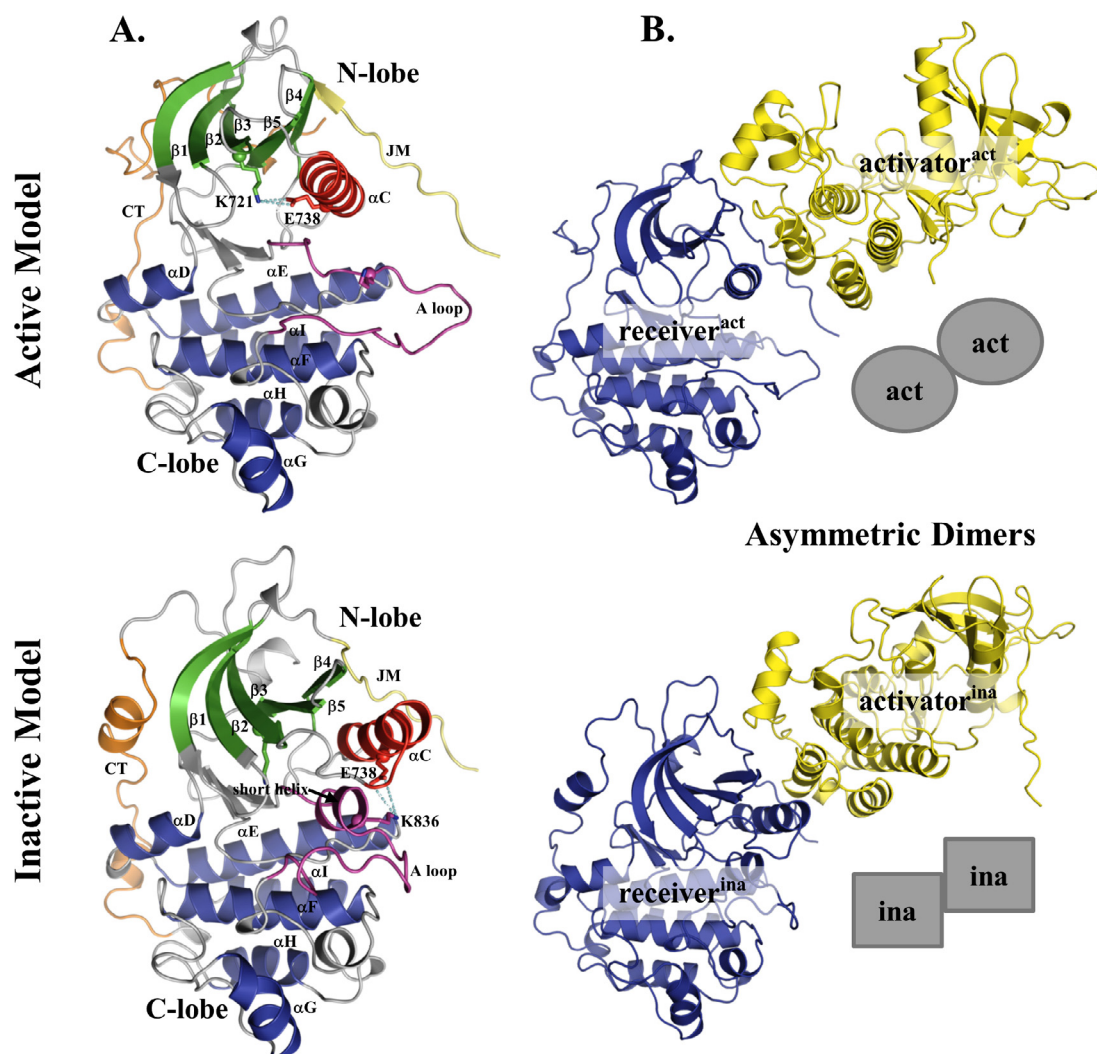


Fig. 1. Structural models for the monomers (A) and asymmetric dimers (B) of active and inactive EGFR kinase domain. The monomeric kinase domain with JM and CT segments are colored according to their structural regions; N-lobe β -strands, green; C-lobe α -helices, blue; α C-helix, red; activation loop, magenta. The K721–E738 and E738–K836 salt bridges are indicated in cyan dashed lines.

β 5-strand, which is essential for stabilizing the proper position of the ATP phosphates, and (ii) extension of the A loop away from the active site, which allows accessibility of a peptide substrate to the binding site. By contrast, the outward displacement of the α C helix coupled to the closed conformation of the A loop observed in the inactive conformation sterically hinder the substrate binding (Fig. 1A).

Knowledge of the autoinhibition, activation and feedback inhibition mechanisms of the EGFR kinase domain has been significantly advanced in recent years. It is thought that EGFR kinase domain is intrinsically autoinhibited, and its intermolecular interactions promote the kinase activation. Recent studies have demonstrated that activation of EGFR kinase results from the asymmetric dimer formation, in which the C-lobe of one kinase (so-called activator) interacts with the N-lobe of the other (so-called receiver), eventually leading to activation of the latter which is in the active-like conformation [9–11]. When the asymmetric dimer interface is disrupted, either by point mutations at the dimer interface, e.g. V924R of the activator and I682Q of the receiver, or by binding of cytoplasmic protein Mig6, that there is an intracellular negative feedback EGFR inhibitor, the kinase domain possess a low level of basal activity toward substrate peptides [9,10]. The crystal structure of Mig6-bound EGFR kinase domain

also demonstrated that the kinase structure adopts the autoinhibited/inactive conformation and Mig6 binds to the C-lobe of the kinase domain in which it contributes to EGFR inhibition through blocking the asymmetric dimer interface [10], emphasizing the important role of the asymmetric dimer formation in EGFR activation as well as a particularly intriguing perspective for the receptor inhibition.

A noticeable observation on the role of the asymmetric dimer in EGFR activation is that one crystal structure of EGFR and HER4 with their well-resolved juxtamembrane (JM) segment exhibits the inactive conformation but still possess the ability to form an asymmetric dimer that resembles the dimer of the active conformation [12,13]. Crystallographic analyses demonstrated that the C-terminal portion of the JM segment provides an additional interaction for the receiver to interact with the C-lobe of the activator in the context of the asymmetric dimer, and such interaction plays a critical role in stabilizing the dimerization. It could be hypothesized that the capability of the C-terminal portion of JM segment to dimerize kinase molecules even though they are in the inactive conformation might be essential for an early step of the receptor activation when two kinase molecules with inactive conformation are brought in the vicinity of each other and then interact in the context of the asymmetric dimer [14].

Protein conformations are actually expressed as a solution ensemble that is more dynamic rather than the single static snapshot of the actual state provided by crystallographic experiments, and thus it is limiting a broader understanding of the protein dynamics. Moreover, crystal contacts may stabilize non-native conformational states and thus introduce artifacts that may lead to inaccurate interpretations. Molecular dynamics (MD) simulation are an ideal method for analyzing conformational flexibility of protein molecules. The technique has been successfully used to study several biological systems, e.g. protein–protein, protein–lipid and protein–ligand interactions at the atomic-level detail. Several MD simulations have previously been applied for studying EGFR kinase domain and other family members in several contexts including the enzyme–drug interactions [15–19] and protein conformational studies [20–26] as well as the membrane-bound full length receptor [27]. In this study, we performed 100-ns MD simulations followed by principal component analysis (PCA) on the asymmetric dimer of both active and inactive conformations of EGFR kinase domain to investigate the structural flexibility and intrinsic motions of each kinase molecule in the dimers. We also investigated intermolecular interactions at the dimer interfaces and compared such interactions to those of the Mig6–EGFR complex observed in the crystal structure. Information gained from our simulations can give insights into better understanding in conformational flexibility and dynamics of EGFR in the context of the asymmetric dimer, as well as provide useful information for further development of a new class of inhibitors that will block the allosteric sites rather than targeting the catalytic sites of the enzyme.

2. Computational methods

2.1. Structural preparation

The atomic coordinates of human EGFR kinase domain determined for both active and inactive conformations were obtained from the Protein Data Bank (PDB). Missing and disordered residues in the retrieved structures were examined using Swiss-PDB-Viewer v4.1 [28]. The program was also used to reconstruct missing atoms of side chains and remove alternative conformations of some residues. The PDB codes that were used to create a crystallographic ensemble of human EGFR kinase domain include the active conformation as follows: 1M14, 1M17, 2GS6, 3VJO, 2GS2, 2ITW, 2ITX, 2J5F, 2J5E, 2J6M, 2ITY, 2ITN, 2ITV, 4I23 and inactive conformations: 3POZ, 2RGP, 3BEL, 1XKK, 2RF9, 3GOP, 2GS7, 3GT8, 3LZB, 4I22, 4I24. All of which the residues before L683 and after L955 were removed to reduce complexities. The residues from L683 to L955 are responsible for the actual kinase domain of human EGFR. Missing residues in those retrieved crystal structures were subsequently constructed using MODELLER 9v8 [29] offered by the ModLoop web server [30].

The MODELLER program was also used to construct the EGFR models in our study which include the C-terminal portion of JM (from G672 to I682), actual kinase domain (from L683 to L955), and extended CT segments (from V956 to I994). To construct the model of the active conformation, the PDB codes 1M17 and 2GS6 were used as the templates whereas 1XKK, 2GS7 and 2R4B (the crystal structure of human HER4) were used to construct that of the inactive conformation. Since evidence suggested that the JM and CT segments are essential for protein dynamics of the kinase domain, and biophysical studies showed that the deletion of these two segments significantly reduce the ability of EGFR kinase to form an asymmetric dimer [11,12,22]. For this reason, the 2GS6, 1XKK and 2GS7 were used to spatially restrain an orientation of the CT segments that should interact with both the C-lobe (via V956–P968) and the N-lobe (via E980–I994) of the kinase domain. Note that

the extended CT segment in the 1M17 distantly extends away from and loses any interaction with the kinase domain [7]. On the other hand, 2R4B was used to model the residues G672–N676 of the JM segment and residues E848–G850 of the A loop in the inactive conformation of EGFR kinase domain since those residues are not well resolved in all of the inactive EGFR structures. The stereochemical qualities of the resultant models were assessed using PROCHECK v3.5 [31].

A model of the asymmetric dimer of EGFR kinase domain for both active and inactive conformations was constructed according to the study of Zhang et al. [9] using the PDB codes 2GS6 and 2R4B as templates, respectively. A model of two EGFR monomers was created by removing the activator molecules from the dimeric models, and therefore they correspond to the receiver in the asymmetric dimer. We note that each molecule in our models is referred to here as the receiver^{act}, activator^{act}, monomer^{act} and the receiver^{ina}, activator^{ina}, monomer^{ina} for simulation of the active and inactive models, respectively. Prior to an MD simulation, the N- and C-termini of each amino acid chain were capped with the acetyl and amino groups, respectively, to give the neutral charged ends and mimic the full-length protein in which additional amino acids would be present. The GROMOS96-53A6 force field [32] was applied to all protein structures in simulated systems and the ionization state of amino acid residues was set according to the standard protocol. The SPC water model [33] was used for the solvent.

2.2. Molecular dynamics simulation

MD simulations for all EGFR models were carried out using explicit-solvent periodic boundary conditions using GROMACS v4.5 [34,35]. Each model was solvated in a rectangular box keeping a distance of 1.2 nm between the solutes and the sides of the solvent box. Each of the solvated systems was neutralized by adding enough sodium and chloride ions to give a concentration of 100 mM. All of the solvated systems were then energy-minimized using the steepest descent algorithm either until the maximum force was lower than 1000 kJ/mol/nm on any atom or until additional steps result in a potential energy change of less than 1 kJ/mol. All simulation systems were first equilibrated in three phases with position restraint (the force constant of 1000 kJ/mol/nm) applied to all heavy atoms of the proteins, allowing H atoms, and solvent and ion molecules to freely move. The first step was conducted under NVT conditions at 300 K of temperature using the modified Berendsen (velocity rescaling) thermostat [36,37]. The following step was conducted under NPT conditions at 1 bar of pressure using Parrinello–Rahman barostat [38,39]. In the last step, Nosé–Hoover thermostat was applied because it generates a correct kinetic ensemble and allows for fluctuations that produce more natural dynamics [40,41]. After equilibration, any position restraint of all heavy atoms in proteins was removed. The full dynamics production was subsequently performed under the NPT condition at 300 K and 1 bar of the system.

For all dynamics procedures, hydrogen bond lengths were constrained using LINCS algorithm which allows for a 2.0 fs-time step [42]. A cut-off distance for the short-range neighbor list was set to 1.0 and 1.4 nm for the electrostatic and van der Waals interactions, respectively. Long-range electrostatic interactions are approximated using the Particle Mesh Ewald (PME) method [43,44]. The atomic coordinates were recorded every 10 ps for the data collection.

2.3. Trajectory analysis

MD trajectories were analyzed using utilities in the GRO-MACS suite [45] that includes the backbone root mean square deviation, root mean square fluctuation, atomic distances and

residue contacts. Changes in secondary structure content during the simulations were assessed using DSSP [46]. An analysis of intermolecular interaction in the dimeric models (receiver vs. activator) includes interaction energies, hydrogen bonds and hydrophobic contacts. Hydrogen bonding is defined following the default values that are (i) donor–acceptor distance ≤ 0.35 nm, and (ii) acceptor–donor–hydrogen angle $\leq 30^\circ$. The Protein Interaction Calculator (PIC) web server [47] was used to determine hydrophobic interactions, by which each of the 100 dimeric structures extracted from the last 50 ns of simulation time was submitted one-by-one to the server. Hydrophobic interactions are identified using a distance cut-off of 0.5 nm between non-polar groups in the non-polar side chains of amino acid residues. It should be noted that no aromatic–aromatic, aromatic–sulfur and cation– π interaction was observed at the interface of both EGFR dimeric models. The Grace v5.1 [48] was used to plot the 2D data and the 3D images were created and rendered with PyMOL v1.3 [49].

2.4. Principal component analysis

The protein ensembles include (i) a set of twenty-five crystal structures of human EGFR kinase domain (as mentioned above) and (ii) six separate MD sets that differ in terms of both molecule type (monomer, receiver and activator) and protein conformation (active and inactive) of 5000 structures, which were extracted from the last 50 ns of simulation. Each protein ensemble was separately subjected to performing a principal component analysis (PCA) using the GROMACS utilities [45]. Mass-weighted covariance matrices were constructed based on backbone fluctuations from their average position after least-squares fitting to remove overall rotational and translational motions. Diagonalization of each covariance matrix then produced a set of eigenvectors and corresponding eigenvalues, which represent the direction and amplitude of a motion, respectively. Eigenvectors were subsequently ranked by decreasing eigenvalues, and such that the first eigenvector, which is also referred to as the first principal component (PC1), represents the largest contribution to the total fluctuation of the protein [50,51]. A motion represented by each PC was visualized by projecting the structures from ensembles (or trajectories) onto each eigenvector of interest and then transforming them back into atomic coordinates. Two extreme projections along that eigenvector were then be interpolated to create an illustration using PyMOL v1.3 [49].

3. Results and discussion

We used a homology-based approach to construct 3D models for both active and inactive conformations of the entire EGFR kinase with its flanked JM and extended CT segments (the residues G672 to I994). Our homology models were stereochemically evaluated according to Ramachandran plots. No residue is located in disallowed region of the plot for both active and inactive models, and more than 90% of the residues are in most favored regions (data not shown) suggesting the reliability of our resultant models. The structural model for the EGFR asymmetric dimer of both active and inactive conformations were then constructed according to the study of Zhang et al. [9] (Fig. 1B). It should be noted that although the relative orientation between receiver and activator in both dimeric models seems identical, the interface area of the active model was slightly larger than that of the inactive one (i.e. 10.4 vs. 9.6 nm², respectively) as estimated from PDBePISA server [52].

The 100-ns MD simulations were carried out on the asymmetric dimers as well as their separate monomer to investigate how dimerization affects the flexibility and dynamics of the EGFR kinase

domain. The backbone-atom root mean square deviation (RMSD) was first measured to observe stability of the simulations and the degree of similarity between simulated and reference structures over the simulation time (Fig. 2). With respect to their initial structure, the backbone RMSD values increased more than 0.3 nm and then fluctuated around 0.3–0.4 nm; nevertheless, the values of the receiver^{act} molecule fluctuated around 0.2–0.3 nm. The RMSD of the receiver in both of the dimeric models remained stable after 50 ns when compared to those of the monomer and activator models which slightly increased. In the following sections: we discuss the results of the MD simulations as follows: (i) the difference in conformational flexibility among the receiver, activator and monomeric molecules observed from the simulations, (ii) essential dynamics of the MD trajectories with and without respect to the crystallographic ensembles, and (iii) intermolecular interactions between receiver and activator from the simulations of the asymmetric dimers.

3.1. Conformational flexibility of EGFR kinase domains

To give insight into conformational flexibility with respect to each structural region of the EGFR kinase domain, we then measured a backbone root mean square fluctuation (RMSF) as a function of the residue number (Fig. 3). The reference structures were averaged over the last 50 ns simulation time. The RMSF was also qualitatively illustrated in the backbone traces and tubes of the 50 snapshots from the simulations (supplementary data S1 and S2, respectively).

3.1.1. Flexibility of the active conformation

As shown in Fig. 3A, in the active models the RMSF of the receiver were reduced in the N-lobe and CT segment when compared to those of the monomer and activator, in particular the loops between $\beta 1$ – $\beta 2$ and $\beta 3$ – αC and the CT residues D970–D990. This result indicated that the dimerization process suppresses the conformational flexibility of the N-lobe of the kinase domain as well as of the CT segment. The reduced fluctuation in the $\beta 3$ – αC loop is consistent with the other published work in which this region in the EGFR monomer is intrinsically disordered and becomes ordered when dimerization [24]. The $\beta 1$ – $\beta 2$ loop is important for organizing the ATP-binding since it makes contacts with the γ -phosphate of ATP. Slightly lower RMSF values were also observed in the hinge region of the receiver (the loop following $\beta 5$ strand), which makes contacts with the ATP-adenine ring. Furthermore, the K721–E738 salt bridge seemed to be more stable in the receiver than in the monomer and activator (supplementary data S3), and thus consistent with the others [20,24,53]. The salt bridge allows the side chain of K721 to bind to the α - and β -phosphates of ATP, and reorients the αC helix closer to the ATP binding site. These results indicated that the asymmetric dimerization helps stabilizing the active conformation, presumably involving the organization of the ATP-binding site.

The significant difference in the RMSF values was not observed in the JM segment (Fig. 3A). In the asymmetric dimer, this segment of the receiver obviously interacts with the C-lobe of the activator. The RMSF values of the receiver were slightly lower than those of the monomer and activator in contrast to the previous work, in which more fluctuations of the JM segment were observed the monomeric kinase [22]. It might be because in the monomer and activator the JM segment could make contacts with the A loop rather than left flexibly (Supplementary data S2). The reduced fluctuation of the receiver was also observed in various regions of the C-lobe of the kinase domain, for examples the distal part of the A loop, αG – αH loop, αI helix and the loop between αI and CT segment. These results indicated that the asymmetric dimerization could also alter the flexibility in regions that are distant from the dimerization interface.

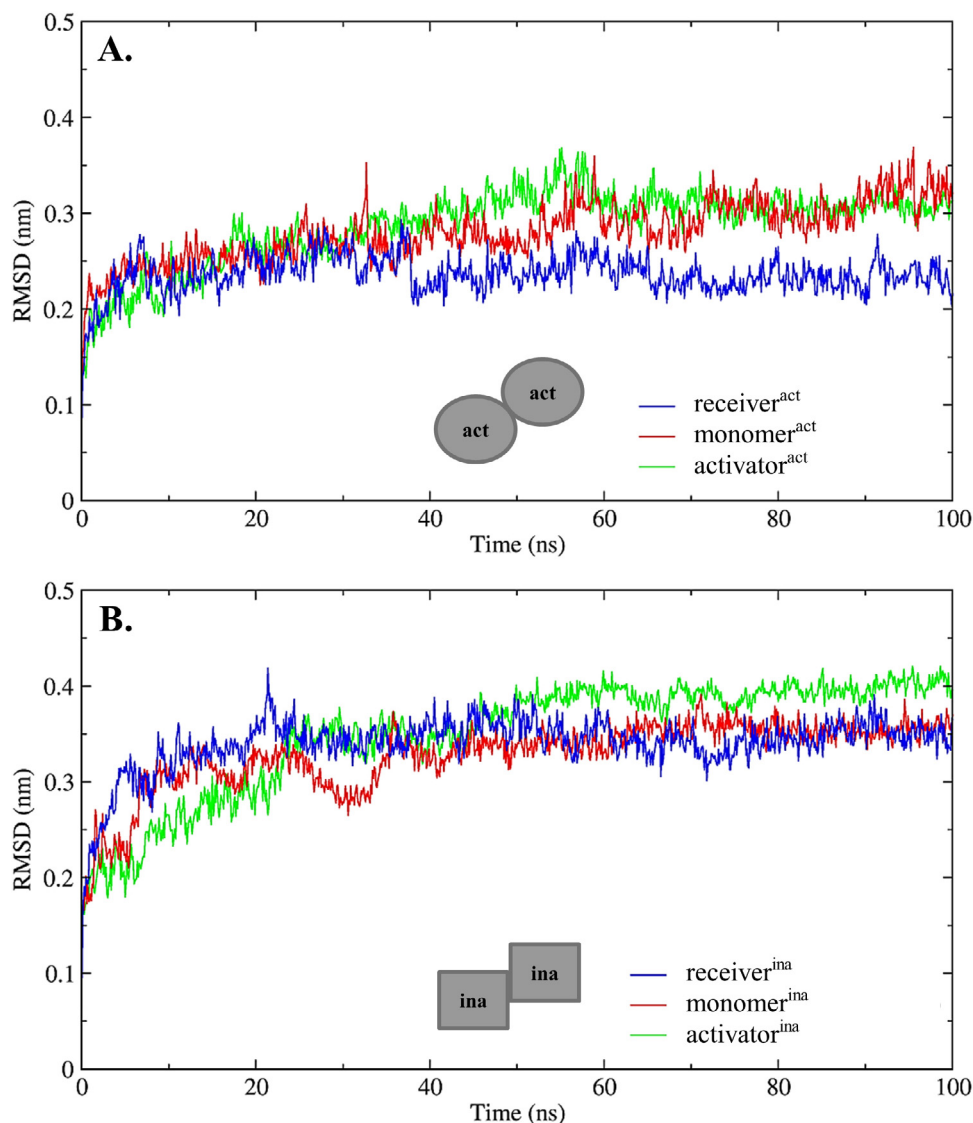


Fig. 2. Backbone RMSD plots as a function of time from the active (A) and inactive (B) models with respect to their initial structure.

3.1.2. Flexibility of the inactive conformation

In contrast to the active conformation, the flexibility of the receiver and activator in the inactive model was similar. The dimerization did not alter fluctuations in the N-lobe of the receiver; except for the α C helix and β 4– β 5 loop (Fig. 3B), which are located in the dimer interface. Interestingly, the RMSF of the β 2– β 3 loop in the monomer was significantly lower than those of both the receiver and the activator. This loop segment is not a part of the dimer interface and is disordered in at least four crystal structures of the EGFR kinase domain. The β 1– β 2 loop also showed more fluctuations in all three inactive molecules. In our previous study, the loop exhibited less fluctuation in the inactive structure, but became more flexible when binding to the drug lapatinib [25]. In addition to the N-lobe, the simulations also showed that the distal portion of the A loop (residues E844–K851) exhibited lower fluctuations in the receiver than that of activator and monomer, even slightly (Fig. 3B). This portion was disordered in all crystal structures of the inactive conformation of EGFR. One possible explanation for this is that this portion in the receiver seemed to make contacts with the C-lobe (the loops preceding α F and α G helix) whereas it was solvent-exposed in the activator (Supplementary data S2).

Since it has been thought that the asymmetric dimer of the inactive conformation might be an early step of the activation of EGFR kinase domain [14], we therefore followed the RMSD values of (i) the active models with respect to the inactive structures, and (ii) the inactive models with respect to the active structure during the 100-ns simulation time (supplementary data S4). A global switch from the active to inactive conformations and vice versa was not observed. In the inactive models, the RMSD at the end of simulation with respect to the active structure of the receiver^{ina}, which would be expected to close to the active conformation when formed the asymmetric dimer, is 0.66 nm. Neither active nor inactive monomer conformation closes to the inactive and active conformations at the end of simulation (RMSD values of 0.57 nm and 0.62 nm, respectively). These results indicated that the 100-ns time scale was still too short to observe either the transition between two distinct conformations or to sample an overlapping intermediate structure. A recent MD study has demonstrated that no active-to-inactive (and vice versa) conformational transition of the wild type monomeric EGFR kinase domain was observed during the 200-ns simulation time [23]. Nevertheless, another study of the other has demonstrated that an intermediate state between

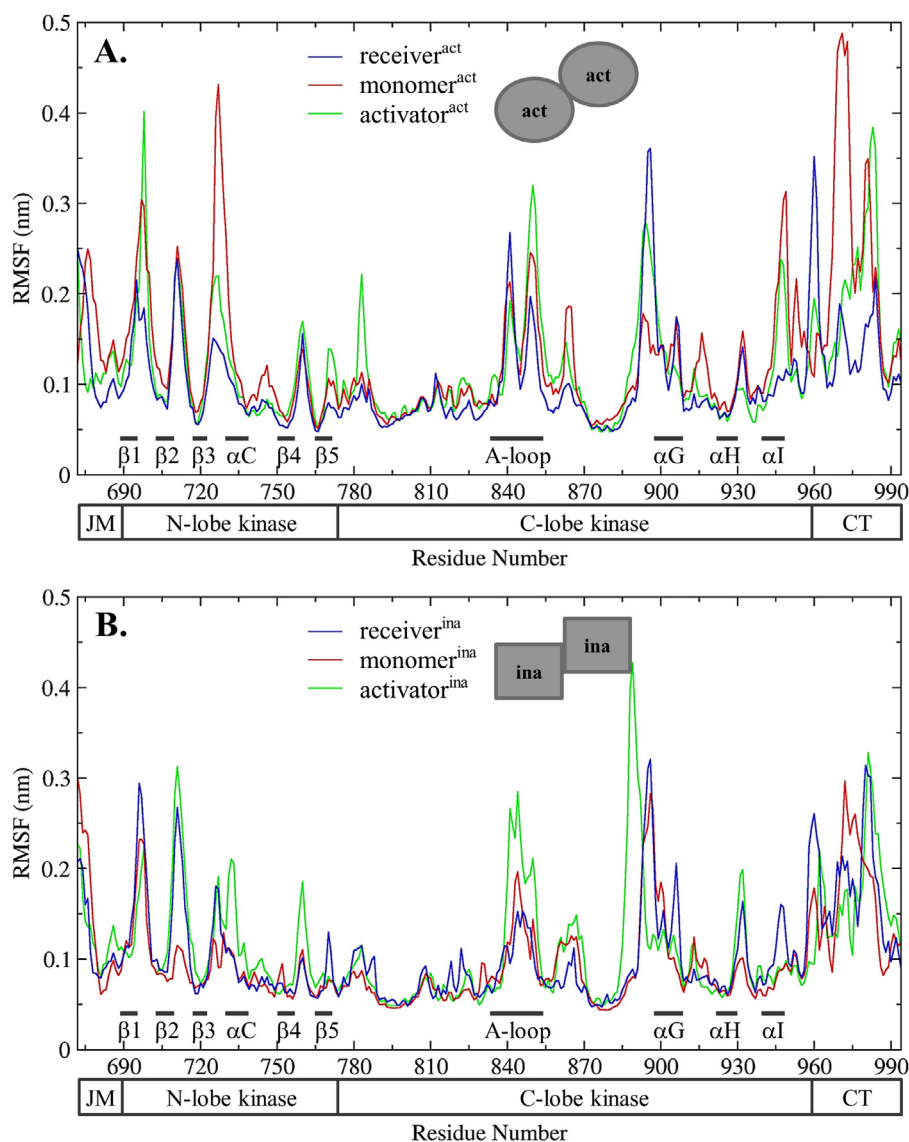


Fig. 3. Backbone RMSF plots as a function of residues from the active (A) and inactive (B) models. The various structural regions the EGFR kinase domain and the JM and CT segments are indicated. The last 50 ns in each MD trajectory were taken to calculate in all cases.

active and inactive conformations of EGFR kinase domain can be captured in a microsecond time scale MD simulation [24].

We were subsequently interested in examining the destabilization effects on the structure of the receiver^{ina} molecule upon the dimerization process. There are several factors stabilizing the inactive state of EGFR, e.g. a short helix of residues L834–L838 located on the A loop that makes hydrophobic contacts with the αC helix. The interaction results in the obstruction of the K721–E738 salt bridge, but the formation of the E738–K836 salt bridge instead [9,20]. Destabilization of this hydrophobic interaction has been thought to be one of the mechanisms for EGFR kinase activation, e.g. L834R mutation [20]. In addition, an α helix formed at the CT residues F973 to M978 partially blocks the ATP-binding site of the 1XKK structure [8,22]. In our inactive models, we found that the helix of L834 to L838 of the receiver remained stable, in contrast to the monomer of which the helix unwound (supplementary data S5). The results were also in accordance with the critical E738–K836 salt bridge that is more stable in the receiver than observed in the monomer (Supplementary data S3). In combination with the results of RMSD as mentioned above it would be suggested that the asymmetric dimerization of EGFR, at least 100-ns

time scale simulations could not even disrupt interactions stabilizing the inactive conformation, in which it has been thought to be an early step required for conformational changes to the activated kinase.

3.2. Essential dynamics of EGFR kinase domains

Principal component analysis (PCA) can be used to identify and compare the principal modes of motion of the EGFR kinase domain using either crystallographic or MD ensembles. The first few principal components (so-called essential dynamics, ED) describe the large-scale collective motions that represent the functionally important global movements, in contrast to the remaining principal components that capture the smaller, localized and often insignificant fluctuations that should not influence protein function [50,51]. In this study, PCA was applied to investigate the ability of our MD simulations to sample the experimentally known motions, which were identified from the crystallographic experimental data set of EGFR kinase domain. In addition, the quantity and direction represented by the first principal component (PC1), which

was individually generated from the MD simulations, were visually compared.

3.2.1. Principal components of the protein ensemble from crystal structures

We established the conformational changes captured in the crystallographic ensembles as an experimental reference. The crystallographic ensemble is a set of twenty-five crystal structures of

EGFR kinase domain (i.e. 12 wild type/active, 2 mutant/active, 5 wild type/inactive and 6 mutant/inactive structures), and can produce a common set of PCs for the active and inactive conformations (supplementary data S6A), and therefore making direct comparison possible among different protein models. This approach has been previously used on other proteins for which substantial crystallographic data exists [54–56]. A projection of the crystal structures onto the two-dimensional (2D) essential subspace defined by the

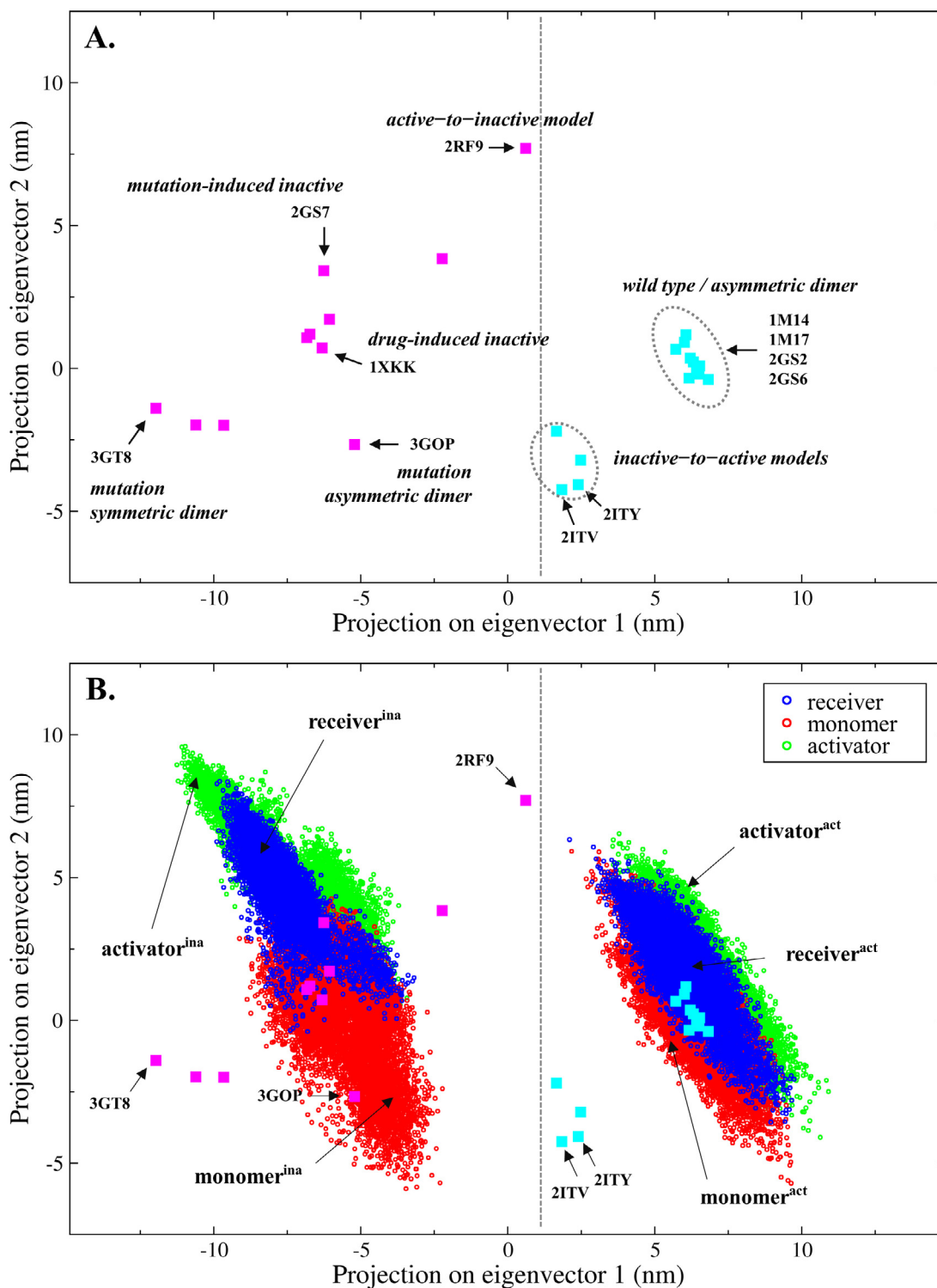


Fig. 4. Principal components (PC) identified from crystallographic ensemble of EGFR kinase domain. Crystal structures (A) and MD structures (B) were projected onto the plane defined by PC1 and PC2. Active and inactive conformations of the crystal structures are shown in cyan and magenta filled rectangles, respectively. Selected PDB codes (discussed in the text) as well as the areas for the simulated structures of the last 10 ns are indicated.

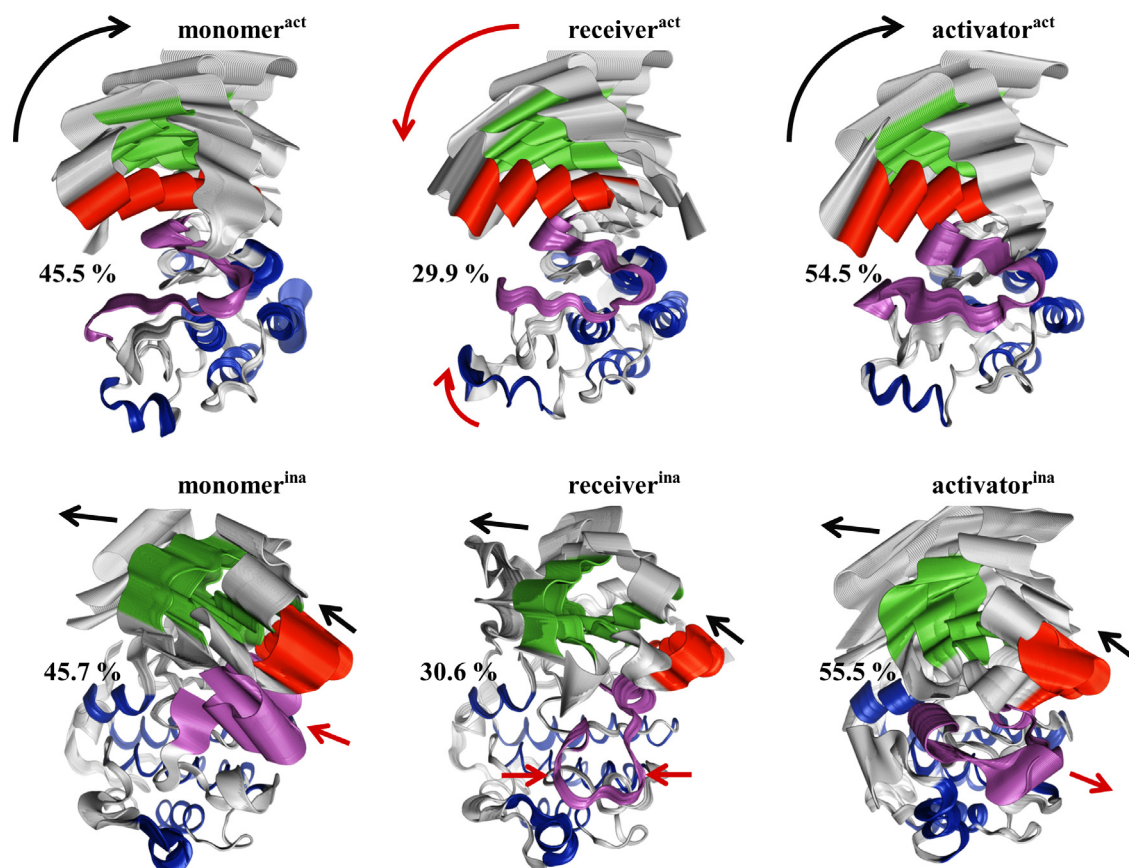


Fig. 5. Comparison of the domain motion described by PC1 from the simulations of the active and inactive models. The two extreme projections for PC1 of the monomer, receiver and activator are illustrated statically by showing all frames simultaneously as backbone trace representation. The structural regions are separately colored as follows: N-lobe β -strands, green; C-lobe α -helices, blue; α C-helix, red; activation loop, magenta. Arrows indicate the approximate direction of the motion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

first two principal components of the crystallographic ensemble was shown in Fig. 4A. The 2D projection could define the clusters that correlate well with two distinct conformations of EGFR kinase domain, in which the PC1 mode was able to differentiate between the active and inactive conformations with positive and negative values, respectively. Visualization of a protein motion was illustrated as multi-frames of the two extreme structures (supplementary S6B). The PC1, which represented approximately 71% of the total motion of the ensemble, involved the large movement of the N-lobe of the kinase domain, e.g. α C helix and almost β -strands, as well as extension of the A loop to be solvent-exposed. The functional significance of this motion presumably modulates the shape of the ATP binding site of EGFR by rearrangement of structural elements that are important for catalysis. The PC2, which comprised only 11% of the total motion of the ensemble, also involved the movement of the N-lobe of the kinase domain, but in the opposite direction observed in PC1. This N-lobe displacement might be related to the inactive-to-active conformational transition of EGFR kinase domain.

We then projected the protein ensemble obtained from all MD trajectories onto the 2D essential subspace of the crystallographic ensemble to investigate conformational sampling of the MD simulations with respect to the experiments. As shown in Fig. 4B, the simulations of the active and inactive models were remarkably distinguished using the PC1 defined from the experiments. Almost the active models largely occupied the unique group of the wild type active EGFR kinase structures (cyan rectangles) and none of them was sampled in the region of the minority. Unlike the crystal structures of inactive EGFR kinase domain (magenta rectangles), the

cluster of the inactive models was less distributed throughout the negative value of PC1. Note that only the monomer^{ina} could sample a region of the asymmetric dimerized-inactive EGFR kinase domain (PDB code 3G0P) [12]. The structure has been thought to be an early step of the kinase activation. Furthermore, none of all inactive models was sampled in the area of the symmetric dimerized-autoinhibited/inactive EGFR kinase domain (PDB code 3GT8) [11]. From the results of PCA, it could be implied that the asymmetric dimer constructed in our EGFR inactive model is unlikely to be in either the autoinhibition or activation form, and thus it would suggest an alternative role of the inactive conformation of EGFR in the context of the asymmetric dimer. In addition, since there was no overlap between the MD ensembles from the active and inactive models, in accordance with the result of RMSD analysis, this indicated that the conformational transitions of the EGFR kinase domain were not observed, at least during the 100-ns simulation time.

No crystal structures have been observed in the intermediate conformational region between active and inactive clusters to date, e.g. the Mig6-bound (PDB code 2RF9) and the L834R oncogenic mutant (2ITN) structures. A recent study suggested that the Mig6-bound inactive conformation of EGFR kinase domain [10] is highly consistent with the intrinsically disordered intermediate sampled from the active to inactive transition of EGFR kinase domain [24]. On the other hand, the structure of the L834R oncogenic mutant protein may represent the inactive-to-active transition since there is evidence that induction of this mutation in the EGFR kinase domain promotes either destabilization of the inactive conformation [20] or the shift the active-inactive equilibrium to prefer the active state

Table 1

List of H-bonds observed at the dimer interface from the simulations of the active and inactive models. The last 50 ns of the trajectory were taken in all cases.

Activator		Receiver		% Occupancy
Residue	Location	Residue	Location	
<i>Active model</i>				
G906-NH	α G- α H loop	L758-CO	β 4 strand	93
G906-CO	α G- α H loop	L758-NH	β 4 strand	58
Q911-OE	α G- α H loop	L680-NH	JM segment	98
Q911-NE	α G- α H loop	A678-CO	JM segment	97
T916-OG	α G- α H loop	N676-CO	JM segment	72
T916-OG	α G- α H loop	A678-NH	JM segment	79
I917-NH	α H helix	A678-CO	JM segment	21
D918-OD1	α H helix	Y740-OH	α C helix	37
D918-OD2	α H helix	Y740-OH	α C helix	81
D930-OD1	α H- α I loop	T727-OG	β 3- α C loop	44
D930-OD2	α H- α I loop	T727-OG	β 3- α C loop	54
D950-OD1	α I-CT loop	G672-NH	JM segment	68
D950-OD1	α I-CT loop	E673-NH	JM segment	36
D950-OD2	α I-CT loop	G672-NH	JM segment	56
D950-OD2	α I-CT loop	E673-NH	JM segment	58
R953-CO	α I-CT loop	N676-NH	JM segment	18
<i>Inactive model</i>				
R908-N1	α G- α H loop	L758-CO	β 4 strand	17
R908-N2	α G- α H loop	L758-CO	β 4 strand	12
Q911-OE	α G- α H loop	L680-NH	JM segment	93
Q911-NE	α G- α H loop	A678-CO	JM segment	88
I917-NH	α H helix	N676-CO	JM segment	37
D918-OD1	α H helix	Y740-OH	α C helix	46
D918-OD2	α H helix	Y740-OH	α C helix	47
S933-OG	α H- α I loop	K733-NH	α C helix	16

[23]. In addition, it should be noted that the drug gefitinib-bound structure (PDB code 2ITY) was sampled in the same area of the L834R mutant structure, suggesting the similar effect of gefitinib does to the inactive conformation. Our previous simulations have shown that gefitinib could destabilize both active and inactive conformations of EGFR kinase domain [25].

3.2.2. Principal component of the active conformation from MD trajectories

PCA was also performed on the individual MD trajectories to characterize the dominant motion captured by the first principal component (PC1). In the active models, PC1 involved the large displacement in the N-lobe with respect to the C-lobe of the kinase domain in all three active models (Fig. 5, upper panel). Nevertheless, the direction of this motion in the receiver seemed opposite to that observed in the activator and monomer, and the contribution of the motion represented approximately 29.9%, 54.5% and 45.5% of the total motion in the protein, respectively. These results suggested that the asymmetric dimer can alter the global dynamics and suppress fluctuations of the active EGFR kinase domain. The interlobal motion of the active conformation is likely to be related to the opening and the closing of the binding cleft around the hinge region that is known to be important for ATP binding and catalysis in other protein kinases, e.g. PKA [57]. This opening and closing motion of the active conformation of EGFR was also identified in the another MD study in which the motion in the kinase domain is closely coupled with the conformational changes of the JM and CT segments [22], implying that the opening and closing motion in the kinase domain is the hallmark for protein kinases in the active conformational state, presumably for the tight regulation of a catalytic activity.

3.2.3. Principal component of the inactive conformation from MD trajectories

In all three inactive models, PC1 exhibited a similar motion that involves the rigid-body displacement of the N-lobe with respect to the C-lobe of the kinase domain (Fig. 5, lower panel). This motion

Table 2

List of hydrophobic interactions observed at the dimer interface from the simulations of the active and inactive models. The last 50 ns of the trajectory were taken in all cases.

Activator		Receiver		% Occupancy
Residue	Location	Residue	Location	
Active model				
P913	α G- α H loop	L679	JM segment	10
I917	α H helix	A678	JM segment	100
I917	α H helix	L680	JM segment	100
I917	α H helix	Y740	α C helix	100
I917	α H helix	A743	α C helix	100
Y920	α H helix	L680	JM segment	100
Y920	α H helix	I682	JM segment	88
M921	α H helix	L680	JM segment	76
M921	α H helix	L736	α C helix	100
M921	α H helix	A739	α C helix	44
M921	α H helix	Y740	α C helix	100
V924	α H helix	L680	JM segment	84
V924	α H helix	I682	JM segment	90
V924	α H helix	L736	α C helix	92
V924	α H helix	I756	β 4 strand	48
V924	α H helix	L758	β 4 strand	100
W927	α H helix	I682	JM segment	18
W927	α H helix	L758	β 4 strand	68
M928	α H- α I loop	L736	α C helix	98
M928	α H- α I loop	L758	β 4 strand	100
M928	α H- α I loop	V762	β 5 strand	94
M947	α I helix	Y740	α C helix	12
V956	CT tail	P675	JM segment	98
Inactive model				
P913	α G- α H loop	L679	JM segment	88
I917	α H helix	A678	JM segment	86
I917	α H helix	L680	JM segment	98
I917	α H helix	Y740	α C helix	100
I917	α H helix	A743	α C helix	96
Y920	α H helix	L680	JM segment	84
Y920	α H helix	I682	JM segment	94
M921	α H helix	L680	JM segment	24
M921	α H helix	L736	α C helix	38
M921	α H helix	Y740	α C helix	100
V924	α H helix	L680	JM segment	42
V924	α H helix	I682	JM segment	90
V924	α H helix	L736	α C helix	100
V924	α H helix	I756	β 4 strand	20
V924	α H helix	L758	β 4 strand	58
M928	α H- α I loop	L736	α C helix	96
M928	α H- α I loop	L758	β 4 strand	92
V956	CT tail	A674	JM segment	12
V956	CT tail	P675	JM segment	92

of the N-lobe looked roughly orthogonal to that of identified in the active models. The fluctuation of the motion is weaker in the receiver than in the activator and monomer, which represented approximately 30.6%, 55.5% and 45.7% of the total motion in the protein, respectively. These results suggested that, unlike the active conformation, the asymmetric dimer formation in the inactive conformation did not drastically change the dynamics of the N-lobe of EGFR kinase domain, but decreased the fluctuation of the protein instead. It would be noted that the motion of the A loop among three inactive models was found to be significantly different. The A loop of the receiver was less mobile. While the A loop of the activator moved toward the opposite direction with respect to the N-lobe, the motion of the N-lobe, α C helix and A loop in the monomer were highly correlated. Functional significance of PC1 observed from the inactive models is unclear. The motion of the N-lobe with the α C helix movement is common in three inactive models, and these suggested that it would not be specifically responsible for conformational changes of the receiver^{ina} molecule upon the dimerization. Moreover, the different motions in the A loop of the three inactive molecules might suggest the intrinsic flexibility of the loop, presumably as a result of various populations of the A

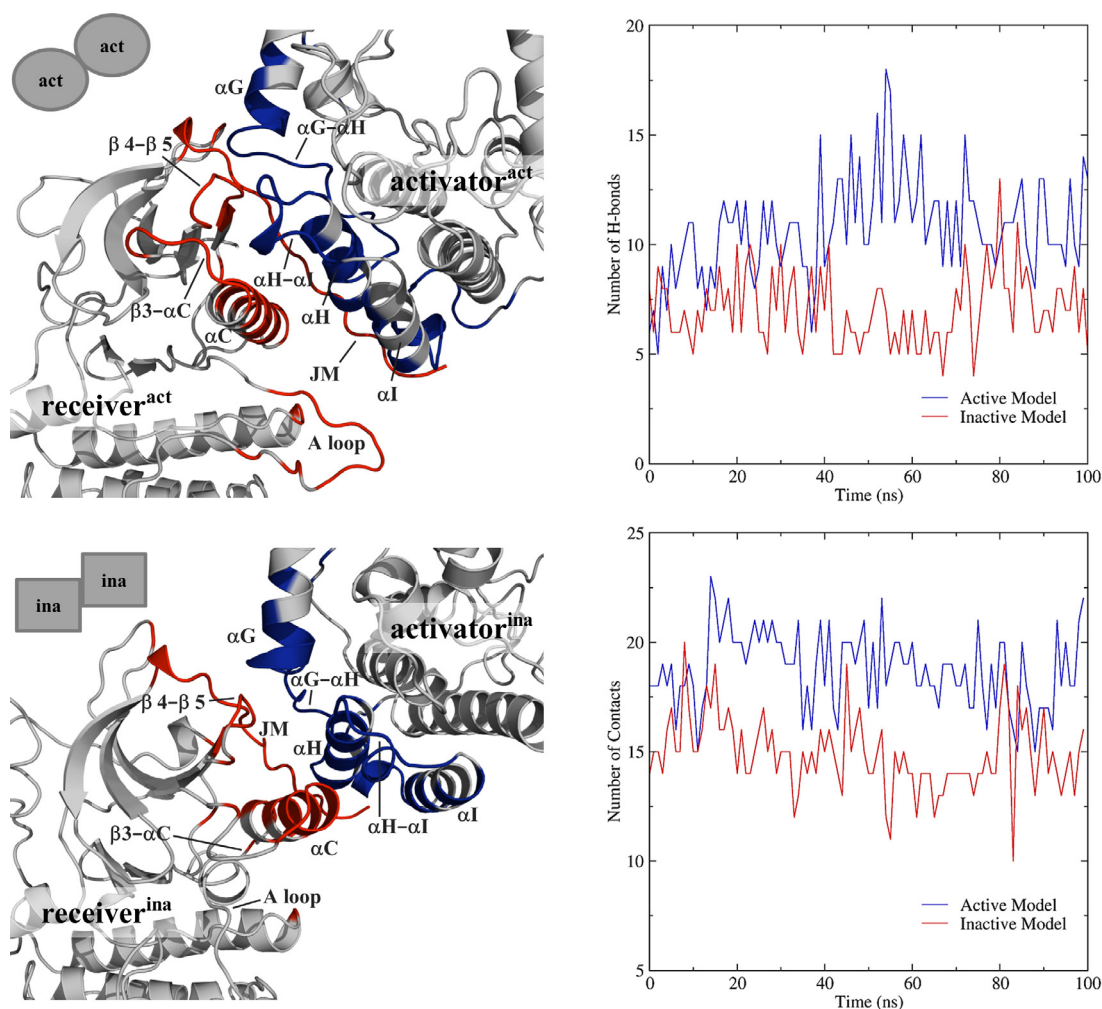


Fig. 6. Ribbon representation of the dimer interface (left images) and number of intermolecular contacts (H-bonds and hydrophobic interactions) as a function of time between the receiver and activator (right images) were observed from the simulations of the active and inactive models. Structural regions responsible for the dimer interface in the activator and receiver are colored in blue and red, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

loop conformation sampled in the MD simulations as well as the crystal structures, all of which are disordered in the distal portion of the A loop (residues E842–K851).

It has been stated that the formation of an asymmetric dimer is essential for the first step of EGFR activation when two inactive kinase molecules are brought into close proximity and interact with each other [14]. Our inactive model in the context of the asymmetric dimer would serve as a starting point to explore an early step of EGFR activation. However, the destabilizing effects on the inactive conformation upon the asymmetric dimerization could not be observed. These results suggested that the functional relevance of the asymmetric dimer formation of the inactive conformation of EGFR kinase domain would not be so evident; or it might play a subtle role involved in regulation rather than a direct mechanism on EGFR activation. Another mechanism involved in regulation of EGFR subpopulations could be suggested. A dynamic equilibrium has been proposed to exist between the two distinct EGFR kinase conformations, and it is maintained by their thermodynamic stabilities [58]. Biochemical experiments demonstrated that the isolated kinase has a low catalytic activity in solution, and when its local concentration was increased by attaching the isolated protein to the surface of lipid vesicles, the EGFR activity dramatically increased, indicating intermolecular interaction between isolated molecules [9]. It would be therefore possible

that the dimerization by the activator to the active monomeric kinase molecule may help stabilize the active conformation of such exiting monomeric populations, eventually leading to shifting the equilibrium between active and inactive conformations, and such intermolecular interactions preferably EGFR kinase populations of the more stable active conformational states.

3.3. Intermolecular interactions at EGFR dimer interfaces

We next tried to study the nonbonded factors stabilizing the activator–receiver interactions in the asymmetric dimers of EGFR kinase domain. The possibility of the asymmetric dimer formation for the inactive conformations was not excluded in these analyses, and therefore this can be compared to interactions observed from the active conformation. Pairs of hydrogen (H–) bonds and hydrophobic interactions observed between the activator and receiver with respect to the domains in which each amino acid belongs to are listed in Tables 1 and 2. Each of the interaction pairs are maintained at least 10% of the time during the last 50 ns of the simulations. In addition, the potential energies that contribute to such interactions (referred to here as interaction energies) were calculated as the sum of the electrostatic and van der Waals (vdW) energies over the last 50 ns of simulation. The interaction energies

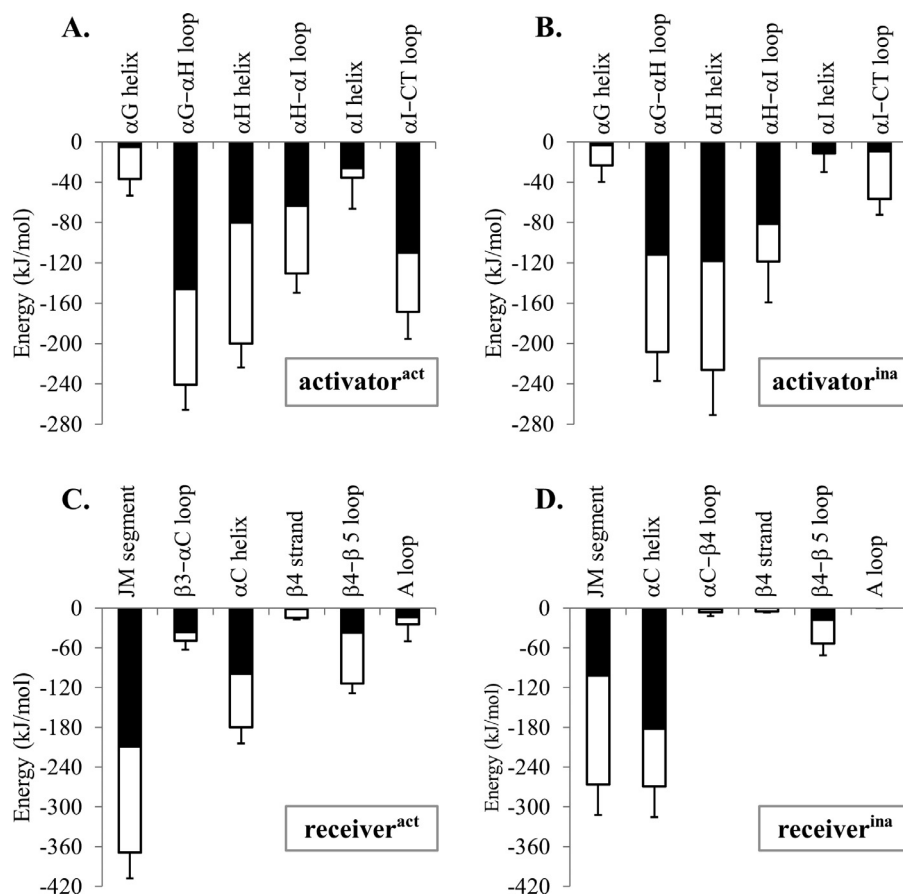


Fig. 7. Potential energies for interaction between the activator and receiver observed from the simulations of the active (A, C) and inactive (B, D) models. The interaction energies are represented as region-based decomposition of the C-lobe of the activator and the N-lobe of the receiver. The electrostatic and van der Waals contributions are shown in black and white bars, respectively. Error bars represent the standard deviation of the total interactions. The last 50 ns of each trajectory were taken in all cases.

with respect to various structural regions of EGFR kinase domain for each dimeric model could be quantitatively compared (Fig. 7).

The total interaction energy observed from the inactive model was estimated as being 4/5 weaker than that observed from the simulation of the active model. This result would be expected because the number of atomic contacts (i.e. H-bonds and hydrophobic interactions) between the receiver and activator of the active model were more than those observed in the inactive models throughout the simulations (Fig. 6, right panel). It should also be noted that vdW energies were found to be weaker than electrostatic interactions in both cases, presumably indicating the more contributions mediated by polar interactions (e.g. H-bonds). Analyzing the residue contacts within a distance of ≤ 0.5 nm during the last 50 ns revealed that the structural regions of the activator involving the dimer interface are similar between the two dimeric models, which range from the αF-αG loop, αG helix, αG-αH loop, αH helix, αH-αI loop, αI helix, αI-CT loop and the N-terminal portion of CT segment (Fig. 6, left panel). On the other hand, the structural regions of the receiver responsible for the dimer interfaces include the JM segment, αC helix, αC-β4 loop, β4 strand, β4-β5 loop and β5 strand (Fig. 6, left panel). As it is expected owing to the in-orientation of the αC helix in the active conformation, the loop preceding αC helix (β3-αC loop) of the receiver could make contacts with the αH-αI loop of the activator. Several polar residues located on the extended portion of the A loop of the receiver^{act} could also make interactions with the αI-CT loop of the activator. The A loop of the receiver^{ina} was found to move down to make contacts with the C-lobe surface of the same molecule instead.

3.3.1. Dimer interface of the active conformation

In the active model, it can be seen that the loop between αG and αH (αG-αH loop) of the activator exhibited the largest negative interaction energies (Fig. 7A). The loop made interactions with the JM segment and β4 strand of the receiver. Several H-bonds contributing such interactions include (i) the main chain-main chain H-bond of G906-L758, (ii) side chain-main chain H-bond of Q911-L680 and Q911-A678, and T916-N676 (Table 1). On the other hand, several residues in the αH helix of the activator (I917, Y920, M921, V924 and W927), as well as M928 on the αH-αI loop and V956 on the CT formed the hydrophobic interactions with a number of nonpolar residues on the JM segment, αC helix and β4 strand of the receiver (Table 2). Such hydrophobic contributions are consistent with experiments, which demonstrated that those residues are critical for the asymmetric dimer formation, and their mutations abolished the EGFR activity in cells [9,11]. The large negative interaction energies of the JM segment of the receiver also indicated the importance of this segment in the formation of the asymmetric dimer (Fig. 7B), in accord with several experiments [9–12].

From the crystal structures of active EGFR kinase domain (the PDB codes 2GS2 and 2GS6) [9], it can be seen that the hydrophobic residues located on the JM segment, αC helix, β4 and β5 strand (listed in Table 2) form a large hydrophobic surface area. The so-called hydrophobic patch is buried in the inactive structure, but becomes solvent-exposed in the active conformation. It has been stated that the asymmetric dimer stabilizes the active conformation by compensation for the thermodynamic penalty associated with the solvent exposure of the patch in the active conformation

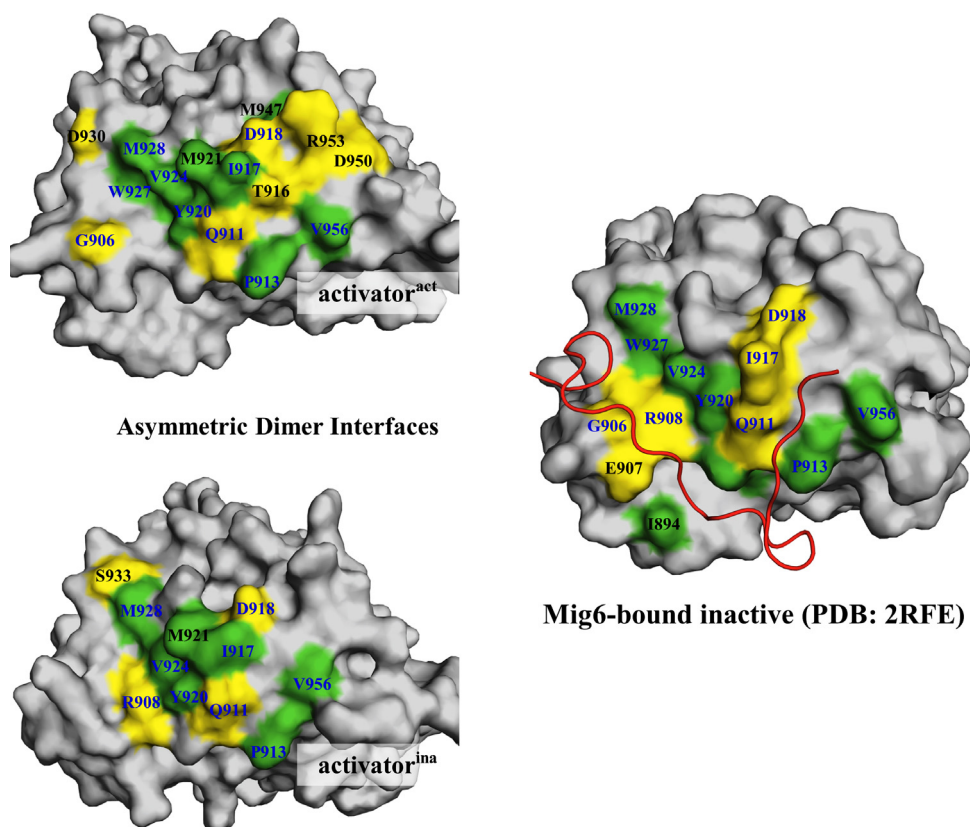


Fig. 8. Comparison of the binding interface at the C-lobe of the activator between the active and inactive models (left images) as well as the crystal structure of Mig6-bound inactive conformation (PDB code 2RFE). Mig6 peptide is shown in the red ribbon. The binding interfaces are shown in a surface representation and the amino acid residues contributing to H-bonds and hydrophobic interactions are colored in yellow and green, respectively. The shared interacting residues are labeled in blue letters. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[59]. Although it has been suggested that the core of the EGFR asymmetric dimer interface is dominated by hydrophobic interactions, our analyses on the estimated interaction energies suggested that in addition to the helix–helix packing (i.e. α H vs. α C), several interactions on the loop regions in particular the H-bond formation of both α G– α H and α I–CT loops of the activator with the JM segment of the receiver could be major contributing factors that stabilize the asymmetric dimer of EGFR kinase domain.

3.3.2. Dimer interface of the inactive conformation

Unlike the dimer interface of the active model, the helical interaction between the α H of the activator and α C of the receiver could be the major contribution in the asymmetric dimer of the inactive conformation (Fig. 7B and D). The α H helix exhibited more negative interaction energies than those from the α G– α H loop. Several hydrophobic interactions in the α H helix found in the active conformation were also observed in the dimer interface of the inactive conformation. These residues include I917, Y920, M921 and V924 on the α H helix, which made contacts with A678, L680 and I682 on the JM segment as well as L736, Y740 and A743 in the α C helix (Table 2). The strong hydrophobic interactions on the loop segments observed in the active conformation were also found (i.e. the M928–L758 and V956–P675). Although the H-bond interactions seemed to be contributing less in the asymmetric dimer of the inactive conformation since the number of interactions were less than those found in the active model, strong H-bond interactions between the α G– α H loop and the JM segment (i.e. Q911–L680 and Q911–A678) were observed (Table 1).

The precise location of the asymmetric dimer interface for the EGFR inactive conformation has not been proven to date. However, evidence from crystallographic analyses of PDB codes 2R4B and

3GOP suggested the possibility of the asymmetric dimer formation for the inactive conformation of EGFR kinase domain [12,13]. It should be noted that the crystal structure of the V924R-mutated EGFR kinase domain, which adopts the inactive conformation, cannot form the asymmetric dimer (PDB code 2GS7); nevertheless, it demonstrated extensive contacts in the context of the symmetric dimer instead (PDB code 3GT8 [11]). This finding indicates that the residue V924 of the activator is critical for the asymmetric dimer formation for both active and inactive conformations of EGFR kinase domain.

3.3.3. Comparison with the Mig6-kinase binding interface

The crystal structure of the complex between the EGFR kinase domain and a fragment of Mig6, which is a negative feedback inhibitor of EGFR, shows that residues K336–K362 of the Mig6 protein can bind to the C-lobe of EGFR kinase domain [10]. The crystal structure of such complex revealed that Mig6 peptide adopts an extended conformation and places on the superficial surface of the C-lobe of the kinase. The interface is dominated by the α G and α H helices, and the α F– α G, α G– α H, α H– α I loops, which buries the area of $\approx 18.00 \text{ nm}^2$ [10]. Such interactions are largely polar; nonetheless, a few hydrophobic residues located on the α H helix contribute to the interface.

We then compared the interacting residues located on the C-lobe kinase domain between the asymmetric dimer interfaces and the Mig6-binding interface (Fig. 8). It was found that the Mig6-bound interface overlaps with the asymmetric dimer interfaces. The amino acid residues that contribute to the shared interface include (i) the H-bond contributions via G906, R908, Q911 on the α G– α H loop; I917, D918 on the α H helix, and (ii) hydrophobic interactions via P913 on the α G– α H loop; Y920, V921, V924, W927

on the α H helix; M928 on the α H– α I loop as well as V956 on the CT segment. The biochemical and biophysical studies on the Mig6 binding and EGFR inhibition suggested that introducing V924R mutation to the kinase domain can abolish the peptide binding. On the other hand, the kinase domain with the I682Q mutation has the same binding affinity as for the wild type protein [10]. Residue I682 is located on the JM segment on the N-lobe of EGFR kinase domain, and this mutant form is not activated because it is unable to form the asymmetric dimer, indicating that the Mig6 peptide can inhibit the EGFR activity in the context of the asymmetric dimer formation, but not in the basal activity of the monomeric kinase domain. The blockage of the asymmetric dimer interface for EGFR inhibition by the cytoplasmic protein Mig6 suggests one of the alternative approaches for development of a new class of inhibitors that can bind to the dimer interface on the C-lobe of EGFR kinase domain.

4. Conclusions

We have performed explicit-solvent MD simulations on two different conformations of EGFR kinase domain in the context of the asymmetric dimer. Although the active-to-inactive conformational transition (and vice versa) in both receiver and monomeric molecules was not observed within the 100-ns simulations, the long MD simulations in our study do allow us to assess relevant protein dynamics observed from the multiple ensemble simulations [23,56]. Our simulations on the active conformation showed that the asymmetric dimerization changed the global motion as well as suppressed the fluctuation of the receiver molecule. The reduced fluctuation in the intrinsically disordered regions and the dynamics alteration of EGFR kinase domain are consistent with the other works [20,22,24]. On the other hand, the asymmetric dimerization of the inactive conformation did not promote a destabilizing effect on the receiver molecule. The dimerization process suppressed its fluctuation to a greater extent and significantly changed its global dynamics; nevertheless, different kinds of motion in the A loop of all three inactive molecules would reflect intrinsic flexibility of the loop of EGFR inactive conformation.

We have investigated intermolecular interactions (i.e. H-bonds and hydrophobic interactions) at the asymmetric dimer interface of the active and inactive conformations of EGFR. The structural regions and key amino acid residues on the C-lobe of the activator as well as the N-lobe of the receiver responsible for such interactions in the asymmetric dimers were identified. The larger contributions by electrostatics rather than vdW interactions may correspond to presence of multiple H-bonds between several loops on the C-lobe of the activator and the JM segment of the N-lobe of the receiver. Our simulations also suggests that the interactions between the loop segments could play a role in the stabilization of the asymmetric dimers in addition to the hydrophobic contact of the helix–helix interaction, as previously demonstrated by crystallographic structures [9]. Furthermore, when compared with the interface of the Mig6–EGFR complex [10], several residues on the C-lobe of the kinase domain responsible for such peptide–protein interactions are also common to the asymmetric dimer interfaces. Since obstruction of the asymmetric dimer interface leads to the inhibition of EGFR activity by Mig6, this suggests another approach for development of novel inhibitors that bind to the C-lobe of the kinase domain. Information gained from our simulations should help us to locate key interacting motifs at the dimer interface on the C-lobe of EGFR kinase domain. This is the first step of a multi-step journey toward the design of peptide inhibitors to EGFR that can prevent the activator from either activating the isolated inactive receiver at an early stage of activation or to help stabilize the active monomeric populations in equilibrium.

Acknowledgments

This work was supported by the Higher Education Research Promotion and National Research University Project of Thailand, Office of the Higher Education Commission and the Thailand Research Fund grants RMU5180032 (KC). The authors are grateful for the support of the Faculty of Science, Kasetsart University and Professor Dr. Supa Hannongbua in particular. NS was supported by a grant under the program Strategic Scholarships for Frontier Research Network for the Joint PhD Program Thai Doctoral degree from the Office of the Higher Education Commission, Ministry of Education, Thailand. KC is supported by Faculty of Science, Kasetsart University, and Thailand Research Fund (RSA5780066). All MD simulations were performed on the Virginia Tech's System X Supercomputer of the Terascale Computing Facility at Virginia Tech, United States. The authors also thank Dr. Justin A. Lemkul for his useful discussions about the results of this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmglm.2015.03.002>.

References

- [1] F.R. Hirsch, M. Varella-Garcia, P.A. Bunn Jr., M.V. Di Maria, R. Veve, R.M. Bremmes, et al., Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis, *J. Clin. Oncol.* 21 (2003) 3798–3807.
- [2] M.A. Lemmon, J. Schlessinger, Cell signaling by receptor tyrosine kinases, *Cell* 141 (2010) 1117–1134.
- [3] N.E. Hynes, H.A. Lane, ERBB receptors and cancer: the complexity of targeted inhibitors, *Nat. Rev. Cancer* 5 (2005) 341–354.
- [4] K.M. Ferguson, Structure-based view of epidermal growth factor receptor regulation, *Annu. Rev. Biophys.* 37 (2008) 353–373.
- [5] G.M. Walton, W.S. Chen, M.G. Rosenfeld, G.N. Gill, Analysis of deletions of the carboxyl terminus of the epidermal growth factor receptor reveals self-phosphorylation at tyrosine 992 and enhanced *in vivo* tyrosine phosphorylation of cell substrates, *J. Biol. Chem.* 265 (1990) 1750–1754.
- [6] Y. Yarden, M.X. Sliwkowski, Untangling the ErbB signalling network, *Nat. Rev. Mol. Cell Biol.* 2 (2001) 127–137.
- [7] J. Stamos, M.X. Sliwkowski, C. Eigenbrot, Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor, *J. Biol. Chem.* 277 (2002) 46265–46272.
- [8] E.R. Wood, A.T. Truesdale, O.B. McDonald, D. Yuan, A. Hassell, S.H. Dickerson, et al., A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells, *Cancer Res.* 64 (2004) 6652–6659.
- [9] X. Zhang, J. Gureasko, K. Shen, P.A. Cole, J. Kuriyan, An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor, *Cell* 125 (2006) 1137–1149.
- [10] X. Zhang, K.A. Pickin, R. Bose, N. Jura, P.A. Cole, J. Kuriyan, Inhibition of the EGF receptor by binding of MIG6 to an activating kinase domain interface, *Nature* 450 (2007) 741–744.
- [11] N. Jura, N.F. Endres, K. Engel, S. Deindl, R. Das, M.H. Lamers, et al., Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment, *Cell* 137 (2009) 1293–1307.
- [12] M. Red Brewer, S.H. Choi, D. Alvarado, K. Moravcevic, A. Pozzi, M.A. Lemmon, et al., The juxtamembrane region of the EGF receptor functions as an activation domain, *Mol. Cell* 34 (2009) 641–651.
- [13] E.R. Wood, L.M. Shewchuk, B. Ellis, P. Brignola, R.L. Brashear, T.R. Caferro, et al., 6-Ethynylthieno[3,2-d]- and 6-ethynylthieno[2,3-d]pyrimidin-4-anilines as tunable covalent modifiers of ErbB kinases, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 2773–2778.
- [14] N. Jura, X. Zhang, N.F. Endres, M.A. Seeliger, T. Schindler, J. Kuriyan, Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms, *Mol. Cell* 42 (2011) 9–22.
- [15] B. Liu, B. Bernard, J.H. Wu, Impact of EGFR point mutations on the sensitivity to gefitinib: insights from comparative structural analyses and molecular dynamics simulations, *Proteins* 65 (2006) 331–346.
- [16] T.E. Balius, R.C. Rizzo, Quantitative prediction of fold resistance for inhibitors of EGFR, *Biochemistry* 48 (2009) 8435–8448.
- [17] A. Dixit, G.M. Verkhivker, Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations, *PLoS Comput. Biol.* 5 (2009), e1000487.

- [18] S. Wan, P.V. Coveney, Rapid and accurate ranking of binding affinities of epidermal growth factor receptor sequences with selected lung cancer drugs, *J. R. Soc. Interface* 8 (2011) 1114–1127.
- [19] O. Sawatdichaikul, S. Hannongbua, C. Sangma, P. Wolschann, K. Choowongkamon, In silico screening of epidermal growth factor receptor (EGFR) in the tyrosine kinase domain through a medicinal plant compound database, *J. Mol. Model.* 18 (2012) 1241–1254.
- [20] A. Papakyriakou, D. Vourloumis, F. Tzortzatos-Stathopoulou, M. Karpus, Conformational dynamics of the EGFR kinase domain reveals structural features involved in activation, *Proteins* 76 (2009) 375–386.
- [21] S.E. Telesco, R. Radhakrishnan, Atomistic insights into regulatory mechanisms of the HER2 tyrosine kinase domain: a molecular dynamics study, *Biophys. J.* 96 (2009) 2321–2334.
- [22] M. Mustafa, A. Mirza, N. Kannan, Conformational regulation of the EGFR kinase core by the juxtamembrane and C-terminal tail: a molecular dynamics study, *Proteins* 79 (2011) 99–114.
- [23] S. Wan, P.V. Coveney, Molecular dynamics simulation reveals structural and thermodynamic features of kinase activation by cancer mutations within the epidermal growth factor receptor, *J. Comput. Chem.* 32 (2011) 2843–2852.
- [24] Y. Shan, M.P. Eastwood, X. Zhang, E.T. Kim, A. Arkhipov, R.O. Dror, et al., Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization, *Cell* 149 (2012) 860–870.
- [25] N. Songtawee, M.P. Gleeson, K. Choowongkamon, Computational study of EGFR inhibition: molecular dynamics studies on the active and inactive protein conformations, *J. Mol. Model.* 19 (2013) 497–509.
- [26] Y. Shan, A. Arkhipov, E.T. Kim, A.C. Pan, D.E. Shaw, Transitions to catalytically inactive conformations in EGFR kinase, *Proc. Natl. Acad. Sci. U.S.A.* 110 (2013) 7270–7275.
- [27] A. Arkhipov, Y. Shan, R. Das, N.F. Endres, M.P. Eastwood, D.E. Wemmer, et al., Architecture and membrane interactions of the EGF receptor, *Cell* 152 (2013) 557–569.
- [28] N. Guex, M.C. Peitsch, SWISS-MODEL the Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18 (1997) 2714–2723.
- [29] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (1993) 779–815.
- [30] A. Fiser, A. Sali, ModLoop: automated modeling of loops in protein structures, *Bioinformatics* 19 (2003) 2500–2501.
- [31] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [32] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, *J. Comput. Chem.* 25 (2004) 1656–1676.
- [33] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, in: B. Pullman (Ed.), *Intermolecular Forces*, Reidel, Dordrecht, The Netherlands, 1981, pp. 331–342.
- [34] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J. Berendsen, GROMACS: fast, flexible, and free, *J. Comput. Chem.* 26 (2005) 1701–1718.
- [35] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, et al., GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, *Bioinformatics* 29 (2013) 845–854.
- [36] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [37] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (2007) 014101.
- [38] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (1981) 7182–7190.
- [39] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, *Mol. Phys.* 50 (1983) 1055–1076.
- [40] S. Nosé, A molecular dynamics method for simulations in the canonical ensemble, *Mol. Phys.* 52 (1984) 255–268.
- [41] W.G. Hoover, Canonical dynamics: equilibrium phase-space distributions, *Phys. Rev. A* 31 (1985) 1695–1697.
- [42] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [43] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [44] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [45] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, et al., *Gromacs User Manual Version 4.5.4*, 2010 www.gromacs.org
- [46] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (1983) 2577–2637.
- [47] K.G. Tina, R. Bhadra, N. Srinivasan, PIC: protein interactions calculator, *Nucleic Acids Res.* 35 (2007) W473–W476.
- [48] P.J. Turner, Grace, Center for Coastal and Land-Margin Research Oregon Graduate Institute of Science and Technology, Beaverton, OR, USA, 2008.
- [49] W.L. DeLano, The PyMOL Molecular Graphics System, Version 1.3r1, Schrödinger, LLC, 2010.
- [50] A. Amadei, A.B. Linssen, H.J. Berendsen, Essential dynamics of proteins, *Proteins* 17 (1993) 412–425.
- [51] S. Hayward, B.L. de Groot, Normal modes and essential dynamics, *Methods Mol. Biol.* 443 (2008) 89–106.
- [52] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, *J. Mol. Biol.* 372 (2007) 774–797.
- [53] A.J. Shih, S.E. Telesco, S.H. Choi, M.A. Lemmon, R. Radhakrishnan, Molecular dynamics analysis of conserved hydrophobic and hydrophilic bond-interaction networks in ErbB family kinases, *Biochem. J.* 436 (2011) 241–251.
- [54] D.M. van Aalten, D.A. Conn, B.L. de Groot, H.J. Berendsen, J.B. Findlay, A. Amadei, Protein dynamics derived from clusters of crystal structures, *Biophys. J.* 73 (1997) 2891–2896.
- [55] B.L. de Groot, S. Hayward, D.M. van Aalten, A. Amadei, H.J. Berendsen, Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data, *Proteins* 31 (1998) 116–127.
- [56] A. Ivetic, J.A. McCammon, Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations, *J. Mol. Biol.* 388 (2009) 644–658.
- [57] N. Narayana, S. Cox, X. Nguyen-huu, L.F. Ten Eyck, S.S. Taylor, A binary complex of the catalytic subunit of cAMP-dependent protein kinase and adenosine further defines conformational flexibility, *Structure* 5 (1997) 921–935.
- [58] M. Landau, N. Ben-Tal, Dynamic equilibrium between multiple active and inactive conformations explains regulation and oncogenic mutations in ErbB receptors, *Biochim. Biophys. Acta* 1785 (2008) 12–31.
- [59] R. Bose, X. Zhang, The ErbB kinase domain: structural perspectives into kinase activation and inhibition, *Exp. Cell Res.* 315 (2009) 649–658.

RESEARCH ARTICLE

Open Access

Understanding the molecular basis of EGFR kinase domain/MIG-6 peptide recognition complex using computational analyses

Ninnutt Moonrin¹, Napat Songtawee², Siriluk Rattanabunyong¹, Surasuk Chunsriviro³, Wanwimon Mokmak^{4,5}, Sissades Tongsima^{5*} and Kiattawee Choowongkamon^{1,5,6*}

Abstract

Background: Epidermal growth factor receptor (EGFR) signalling plays a major role in biological processes, including cell proliferation, differentiation and survival. Since the over-expression of EGFR causes human cancers, EGFR is an attractive drug target. A tumor suppressor endogenous protein, MIG-6, is known to suppress EGFR over-expression by binding to the C-lobe of EGFR kinase. Thus, this C-lobe of the EGFR kinase is a potential new target for EGFR kinase activity inhibition. In this study, molecular dynamics (MD) simulations and binding free energy calculations were used to investigate the protein-peptide interactions between EGFR kinase and a 27-residue peptide derived from MIG-6_s1 segment (residues 336–362).

Results: These 27 residues of MIG-6_s1 were modeled from the published MIG-6 X-ray structure. The binding dynamics were detailed by applying the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) method to predict the binding free energy. Both van der Waals interactions and non-polar solvation were favorable driving forces for binding process. Six residues of EGFR kinase and eight residues of MIG-6_s1 residues were shown to be responsible for interface binding in which we investigated per residue free energy decomposition and the results from the computational alanine scanning approach. These residues also had higher hydrogen bond occupancies than other residues at the binding interface. The results from the aforementioned calculations reasonably agreed with the previous experimental mutagenesis studies.

Conclusions: Molecular dynamics simulations were used to investigate the interactions of MIG-6_s1 to EGFR kinase domain. Our study provides an insight into such interactions that is useful in guiding the design of novel anticancer therapeutics. The information on our modelled peptide interface with EGFR kinase could be a possible candidate for an EGFR dimerization inhibitor.

Keywords: EGFR, Cancer, Inhibitor, Tyrosine kinase, MIG-6 segment1, Molecular dynamics simulations

Background

EGFR, also known as ErbB1 or HER, is a receptor tyrosine kinase that mediates in biological process of normal physiology [1]. An over-expression of EGFR has often been associated with a variety of human cancers that

can provide the basic properties required for cancer growth, including cell proliferation, anti-apoptosis, metastasis and angiogenesis [2]. These properties render EGFR an attractive target for cancer therapy. Nonetheless, several reports show that the secondary mutations in EGFR cause anticancer drug resistances [3–5], instigating a requirement of efficient treatments to overcome these resistances. Zhang et al. [6] showed that a new target therapy could be done by the suppression of EGFR activation through an allosteric mechanism. This approach was supported by an association of the endogenous negative-feedback-inhibitor protein, called MIG-6, that regulates

* Correspondence: sissades@biotec.or.th; fscitc@ku.ac.th

⁵Genome Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency (NSTDA), 113 Thailand Science Park, Phahonyothin Road, Khlong Nueng, Pathum Thani, Khlong Luang 12120, Thailand

¹Department of Biochemistry, Faculty of Science, Kasetsart University, 50 Ngam, Wong Wan Rd, Bangkok, Chatuchak 10900, Thailand

Full list of author information is available at the end of the article

the level of EGFR activation [7,8]. The crystal structure of the EGFR kinase/MIG-6 complex showed that a segment of MIG-6 (residues 337–361), called MIG-6_s1 interacts with the C-lobe of EGFR kinase domain (Figure 1) [6]. The binding of MIG-6_s1 to EGFR kinase prevents an asymmetric dimer formation, leading to the inhibition of EGFR activation at micro molar binding affinity level [6]. The detailed analysis of the protein-protein interactions (PPI) at an atomic level of EGFR kinase and MIG-6_s1 is yet limited. Molecular dynamics (MD) simulations and free energy calculation can be used to obtain the knowledge on the dynamics of the structures and the binding energy profiles of the protein complexes in solution [9].

In this study, MD simulations were performed on the EGFR kinase/MIG-6_s1 complex to obtain the knowledge on their binding dynamics and structural changes. The binding free energy calculation by the MM-PBSA method was applied to gain a better understanding on the binding energetics of the complex. The per residue free energy decomposition and alanine scanning methods were used to elucidate the residues important for interface binding and to obtain their energetic contributions. Our results provided a better understanding of the binding interactions between EGFR kinase and MIG-6_s1 peptide. Moreover, such information could be used to improve or develop more efficient anti-cancer drugs.

Methods

Model setup and molecular dynamics simulations

The starting structures of EGFR kinase/MIG-6_s1 and the asymmetric kinase dimer complexes were based on

the X-ray crystal structure from the protein data bank (PDB). The monomer structures of the EGFR kinase activators (residues 678–959) and MIG-6_s1 (residues 337–362) were used (PDB code: 2RFE) [6], while PDB code: 2GS6 represents the EGFR kinase receiver model. The N- and C-termini of EGFR kinase and MIG-6_s1 were capped with an acetyl group (ACE) and an N-methyl group (NME), respectively [10]. All simulations were performed using AMBER 12 and the ff03 force field [11]. Each system was solvated using the atomistic TIP3P water model [12] in an octahedron truncated box with a buffer distance of 7 Å. Hydrogen atoms were added, and the counter-ions of Na⁺ or Cl[−] were used to neutralize the system. To remove bad contacts in the crystal structure, we applied three-step minimization to each system before performing the simulations. Positional restraints were applied to the whole system in the first and the second steps with a force constant of 10 kcal/(mol Å²) and 2 kcal/(mol Å²), respectively. In the third step, all atoms were allowed to move freely without restraints. The energy minimization in each step was carried out using 2,000 cycles of the steepest descent algorithm, followed by 300 cycles of the conjugate gradient algorithm. Each system was heated from 0 to 300K during the 100 ps in an NVT ensemble and was further equilibrated for 500 ps in an NPT ensemble. Then it was simulated for 22 ns in an NPT ensemble at 1 atm and 300K, using the 2 fs integration time step. During the simulations, the particle mesh Ewald (PME) method [13] was used to treat the long-range electrostatics. Moreover, the SHAKE algorithm [14] was applied to

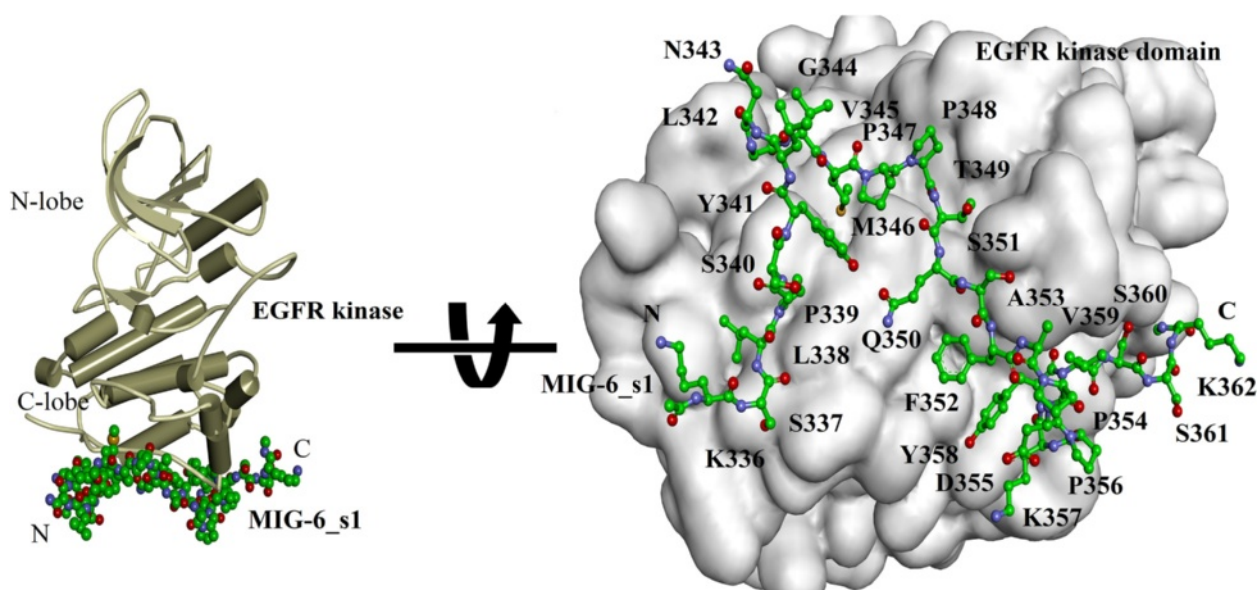


Figure 1 Cartoon representation of the crystal structure of the EGFR kinase/MIG-6_s1 complex. The positions of MIG-6_s1 residues were spanned on the EGFR kinase interface.

eliminate bond-stretching freedom for all bonds involving hydrogen, thereby allowing the use of a 2 fs time step.

Calculation of binding free energy using the MM-PBSA approach

The MM-PBSA approach [15] was used to calculate protein-protein binding free energies. 300 snapshots were extracted from the last 3 ns of the MD trajectories to perform MM-PBSA. The binding free energy of each snapshot was calculated as follows:

$$\Delta G_{\text{(bind)}} = G_{\text{(complex)}} - G_{\text{(protein)}} - G_{\text{(ligand)}} \quad (1)$$

The free energy (G) for each molecule can be computed from the following equations:

$$G = \Delta E_{\text{(gas)}} - \Delta G_{\text{(sol)}} - T\Delta S \quad (2)$$

$$\Delta E_{\text{(gas)}} = E_{\text{(int)}} + E_{\text{(vdw)}} + E_{\text{(ele)}} \quad (3)$$

$$E_{\text{(int)}} = E_{\text{(bond)}} + E_{\text{(angle)}} + E_{\text{(torsion)}} \quad (4)$$

$$G_{\text{(sol)}, \text{PB}} = G_{\text{(PB)}} + G_{\text{(nonpol, sol)}} \quad (5)$$

$$G_{\text{(nonpol, sol)}} = \gamma \text{SASA} \quad (6)$$

$\Delta E_{\text{(gas)}}$ is the gas-phase energy, which is the sum of the internal (E_{int}), van der Waals (E_{vdw}) and Coulomb energies (E_{ele}) (Eq. 3). Internal energy (E_{int}) comprises the bond (E_{bond}), angle (E_{angle}) and torsion (E_{torsion}) energies, respectively (Eq.4). The solvation free energy ($G_{\text{sol, PB}}$) comes from the combination of polar (G_{PB}) and non-polar contributions ($G_{\text{nonpol, sol}}$) (Eq.5). Note that $G_{\text{(PB)}}$ is the polar solvation calculated by the Poisson-Boltzmann (PB) equation with the dielectric constants of 1.0 for solute and 80.0 for solvent. The non-polar component ($G_{\text{nonpol, sol}}$) is defined in Eq. 6, where the constant γ was set to 0.0072 kcal/mol, while SASA stands for the solvent-accessible-surface area calculated by the linear combination of pairwise overlap [16] model. Finally, $-T\Delta S$ in Eq. 2 is an entropy term calculated from the sum of translational, rotational and vibrational components using normal mode analysis. To gain more insights into the residue contributions, the residues of MIG-6_s1 responsible for the binding to the EGFR kinase surface were computed by using the MM-PBSA per residue free energy decomposition method (*mm_pbsa* module) of AMBER12. These calculations were based on the same snapshots as those used in the binding free energy calculations.

Computational alanine scanning mutagenesis

To investigate the contributions of an individual residue to the binding interface and to identify the “hot-spot” residues that are very important for interface binding,

we used computational alanine scanning to identify these key residues of the EGFR kinase/MIG-6_s1 interface as well as the asymmetric dimer interface. Alanine scanning replaces an original residue with alanine; this substitution eliminates the side-chain but does not introduce large conformational changes to alter the mode of binding [17]. The binding free energy of the wild type and mutant were calculated using a post-processing treatment of the *mm_pbsa* module and the difference in their binding free energies was computed using the following equation [10,18]:

$$\Delta\Delta G_{\text{binding}} = \Delta G_{\text{binding of wildtype}} - \Delta G_{\text{binding of mutant}} \quad (7)$$

Based on the post-processing energy calculation [19,20], the alanine mutant structure was generated by eliminating the residue side-chain with a hydrogen atom of alanine prior to the calculation. Note that the smallest amino acid, glycine, was not included since it introduces conformational flexibility into the protein backbone. Proline was also not included because its backbone conformation differs from that of alanine [21-23]. The 300 snapshots extracted from the last 3 ns of the MD runs were used in these calculations. If the difference in binding free energy between the wild type and the mutant of a residue ($\Delta\Delta G_{\text{binding}}$) becomes positive, such an alanine substitution is favorable, i.e., this residue is probably not important in the interface binding. However, the negative value of $\Delta\Delta G_{\text{binding}}$ indicates the unfavorable alanine substitution, i.e., this residue plays significant role in the interface binding [18]. These key residues were clarified as the residues that significantly decrease the binding free energy for at least 2.0 kcal/mol. From literatures, residues with strong binding should reduce the binding free energy greater than or equal to 4 kcal/mol [21,24].

Results and discussion

Structural stabilities and flexibilities of the systems

To assess the dynamic stabilities of the systems, we performed MD simulations of EGFR kinase and MIG-6_s1 in both unbound and bound states with explicit water for 22 ns at 300K under 1 atm pressure. The equilibration of MD simulations can be observed from the convergence of the root-mean-square deviation (RMSD) plot, based on the backbone (C α , N and O) atoms of each snapshot as compared to the initial energy minimized structure.

The RMSD plots of three systems: 1) unbound EGFR kinase 2) unbound MIG-6_s1 and 3) EGFR kinase/MIG-6_s1 complex are shown in Figure 2. For the bound system, beside the plot for overall complex (black line), we also plotted the RMSD values for only EGFR kinase

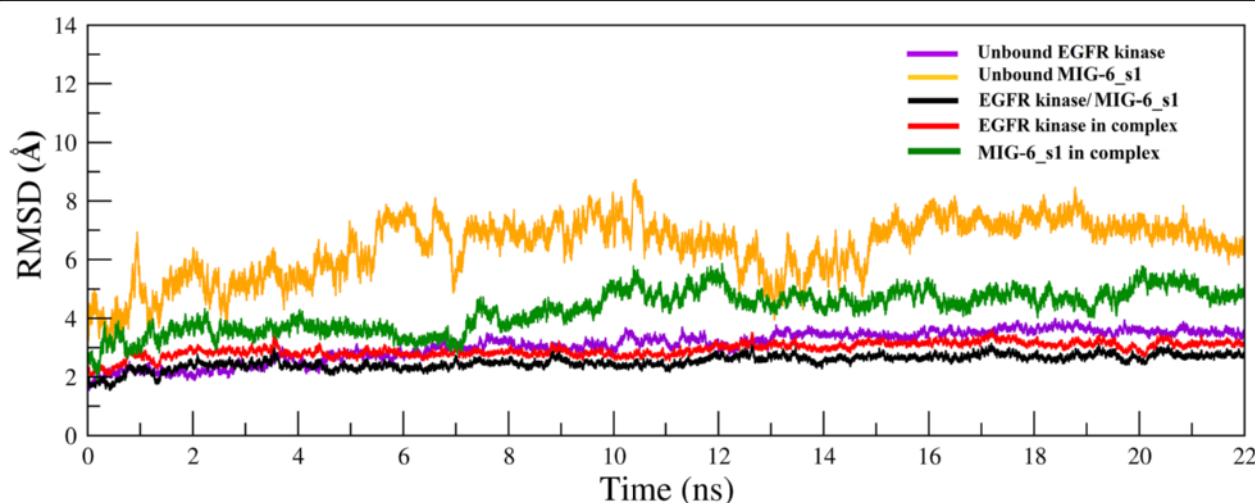


Figure 2 The Root-Mean-Square Deviation (RMSD) plots of the EGFR kinase /MIG-6_s1 backbones in the bound and unbound states of the 22-ns simulation. RMSD plots of unbound EGFR kinase and unbound MIG-6_s1 are shown in purple and yellow lines, respectively. For the complex of EGFR kinase/MIG-6_s1, three RMSD plots are presented. The black line represents the RMSD of the entire complex of EGFR kinase/MIG-6_s1. The RMSD plot for only EGFR kinase in the complex is shown in red. The RMSD plot for only MIG-6_s1 in the complex is shown in green.

(red line) and for only MIG-6_s1 (green line). The RMSD plots of “unbound EGFR kinase”, “EGFR kinase/MIG-6_s1 complex” and “only EGFR kinase in complex” reached their plateaus after about 2 ns. The RMSD plots of only MIG-6_s1 in the EGFR complex reached its plateau after about 7 ns. These three systems were equilibrated during the 22 ns simulations.

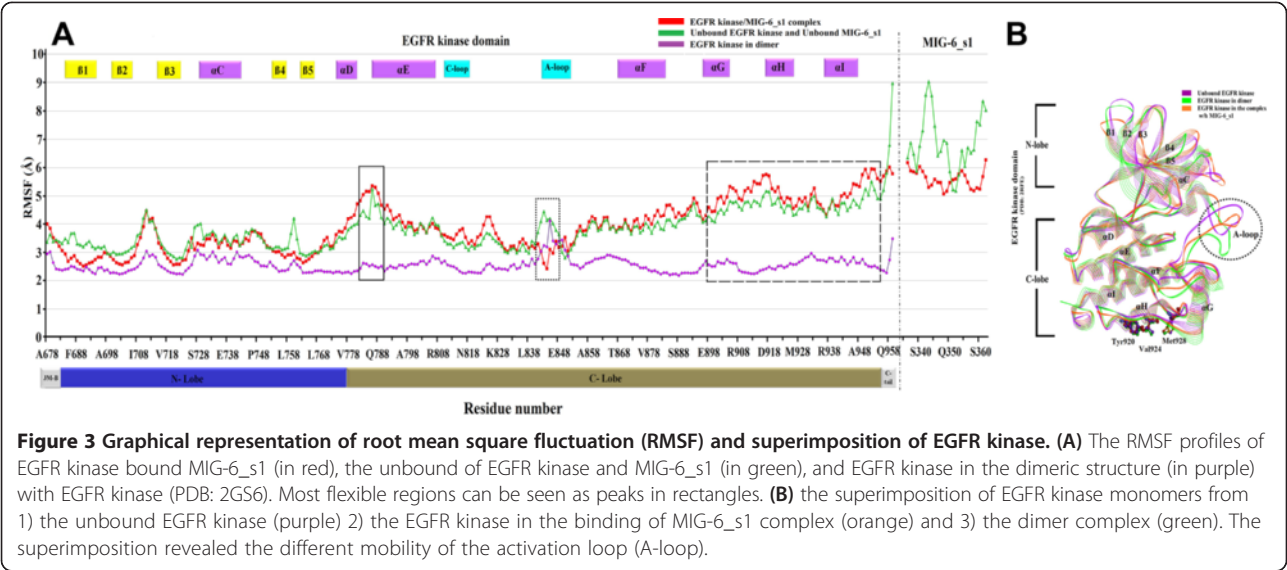
The conformational changes in the bound and unbound simulations were compared (Figure 2). Both bound and unbound EGFR kinase RMSD plots revealed similar patterns with average RMSD value of ~ 3.0 Å for unbound EGFR and ~ 2.9 Å for bound. For bound vs. unbound MIG-6_s1, the RMSD plot of bound MIG-6_s1 is more stable than that of unbound state with average RMSD of ~ 4.2 Å for bound MIG-6_s1 and ~ 6.4 Å for the unbound MIG-6_s1. It can be seen that the bound MIG-6_s1 was significantly more stable than the unbound state as seen in the lower RMSD fluctuation. Note that the unbound MIG-6_s1 has generally higher RMSD values because MIG-6_s1 is rather short peptide (27 amino acids) and becomes very flexible in solution. This flexible short peptide phenomenon were also presented in [17,25] where the unbound short peptide p53 could not be stabilized in solution but could be bound stably to the binding cleft of MDM2 protein. Thus, from the RMSD results, our MD simulations are reliable enough for further investigation.

The root-mean-square fluctuations (RMSF) of α C atom from the MD simulated structures against the starting structure were calculated to identify the most flexible regions of the protein-peptide structure. The larger RMSF value conveys more flexible region while the lower RMSF value entails the more constrained region.

The comparison of both unbound EGFR system and EGFR kinase/MIG-6_s1 complex system based on their RMSF profiles (Figure 3A). The RMSF profile of unbound EGFR kinase system (green) showed similar behavior to that of EGFR complex system (red). On the contrary, the RMSF profile of MIG-6_s1 in the unbound state showed much higher fluctuation than that of its complex state. When comparing the RMSF EGFR kinase profile with the dimer complex profile (purple), we observed distinguishable patterns in several regions of the EGFR interface to MIG-6_s1. Within the EGFR N-lobe region, similar RMSF profiles from all three systems were observed; however the corresponding RMSF profiles were lower than the other regions reflecting lower flexibility.

RMSF patterns and their values in the EGFR C-lobe region differ greatly from that of N-lobe region. Particularly, residues from both EGFR kinase/MIG-6_s1 complex and unbound EGFR kinase, especially in the junction α D– α E, and along the three α G, α H and α I helices revealed high RMSF values (reaching 5–6 Å); while the EGFR kinase in dimer complex shows stable RMSF value ~ 3 Å. We hypothesize that during EGFR dimerization these helices must interact with the juxtamembrane B portion of the EGFR receiver; hence losing the dimerization causes the EGFR activator to have higher RMSF (see unbound EGFR kinase). Furthermore, we observed uncertain fluctuations in the part of Ala840–Gly850, called A-loop, which correlates with the missing residues in the EGFR X-ray crystal structure (Figure 3B) [26].

Toward the end of the RMSF plots, the magnitude of unbound MIG-6 plot (green) differed greatly from that of complex of EGFR/MIG-6_s1 plot (red). In particular the RMSF of MIG-6_s1 in EGFR complex (red) was about

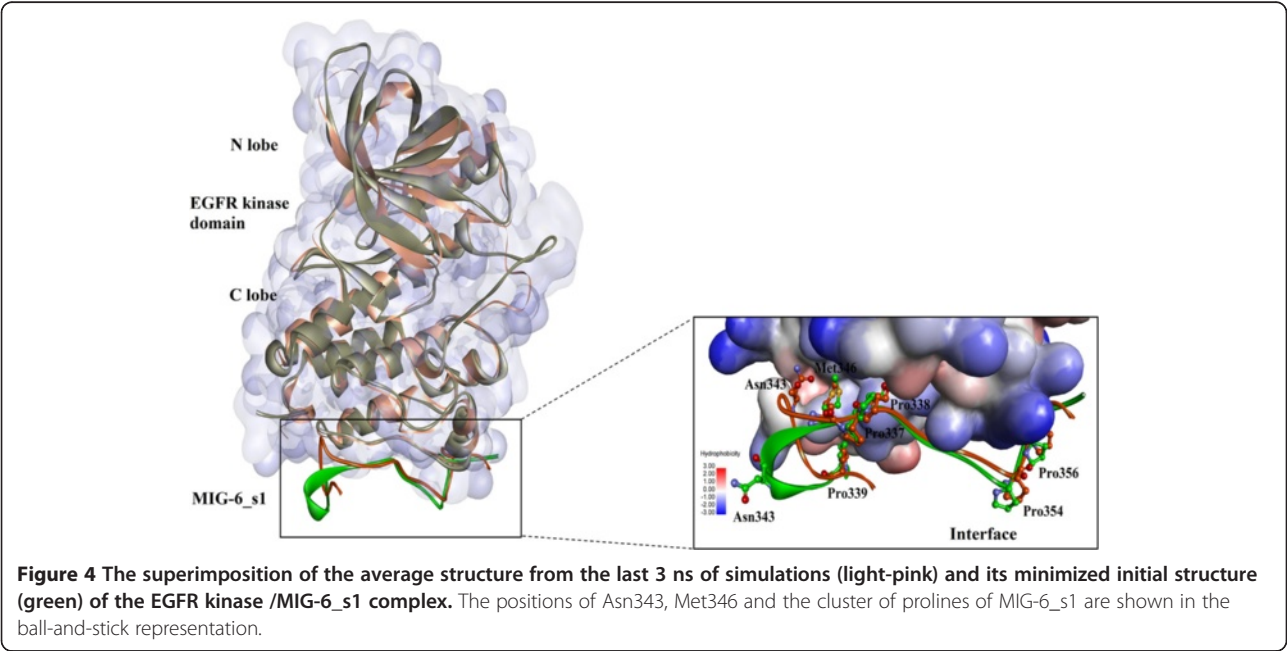


5–6 Å, while the RMSF of unbound MIG-6_s1 (green) swung much higher. These evidences suggest that the binding of EGFR kinase could help MIG-6_s1 structure to stabilize in complex.

The conformational changes of the EGFR kinase/MIG-6_s1 complex

Twenty-seven residues of MIG-6_s1 (residue 336–362) can bind to the distal surface of the C-lobe in the EGFR kinase domain to prevent the activation of EGFR kinase by asymmetric dimer (Figure 1). To gain a better understanding on the structural differences between the initial structure and the average simulated structure, the

superimposition of the minimized initial structure of the EGFR kinase/MIG-6_s1 complex and its average structure from the 300 snapshots of the last 3-ns simulations are shown in Figure 4. The conformations of the average structure and the minimized initial structure were mostly similar, except for the shift in the loop region of MIG-6_s1. Specifically, Asn343 of MIG-6_s1 at the loop region of the average structure shifted its side-chain up closer to the EGFR kinase surface region than that of the minimized initial structure (Figure 4). This contact was not observed previously in the initial structure possibly due to crystal packing contacts in the flexible loop segment [27].



In terms of hydrophobic interactions, the hydrophobic pocket of the EGFR kinase interface consists of Leu343, Trp881, Thr885 and Pro913, and the side-chain of Met346 of MIG-6_s1 resided stably in this pocket, as shown in Figure 5 (plotted by LIGPLOT program [28]). Since prolines can affect the rigidity of peptide structures, the conformations of prolines were also analyzed in this study. Five prolines of MIG-6_s1 (Pro347, Pro348, Pro354, Pro356 and Pro339) were in *cis*-configurations and located on the turn. Four of these residues (Pro347, Pro348, Pro354 and Pro356) remained in bent conformation at the interface (Figure 6) [6]. Prolines on the turn might also form interactions between the proline ring and nearby aromatic residues (CH- π) [29] and help stabilize the MIG-6_s1 complex. Similar to prolines in the SH3 domain, the positions of two prolines were relatively fixed to constrain the structural movements [30].

The calculations of binding free energies

To compute the binding free energies of MIG-6_s1 peptides to EGFR kinase and to gain insights into the peptide-protein binding interactions, the MM-PBSA approach was applied to the 300 snapshots taken from the last 3 ns of the production runs. The calculated results are presented in Table 1. The predicted binding free energy of EGFR kinase/MIG-6_s1 complex was -142.7 kcal/mol. The major favorable components of the MIG-6_s1 binding were the van der Waals term in gas-phase

($\Delta E_{\text{vdw}} = -120.8$ kcal/mol) and the non-polar part of the solvation free energy term ($\Delta G_{\text{nonpol, sol}} = -84.25$ kcal/mol). These two terms resulted in the total favorable non-polar interactions of -205.1 kcal/mol. The highly favorable non-polar part of the free energy might come from the hydrophobic interactions between MIG-6_s1 and EGFR kinase as well as the desolvation of the non-polar groups of MIG-6_s1 from water that aligned them with the binding interface [10]. Such a phenomenon could be seen in several protein-protein interactions including the interaction between MDM2 and p53, where van der Waals interaction was the major contributor to the inhibitor binding originated from the effects of solute-solvent attractive forces on hydrophobic and the dispersion [10,31].

On the other hand, the favorable columbic term in gas phase ($\Delta E_{\text{ele}} = -130.4$ kcal/mol) was completely cancelled by the unfavorable contribution of the polar part of the solvation free energy ($G_{\text{PB}} = 151.6$ kcal/mol), resulting in the total unfavorable electrostatic interactions of 21.2 kcal/mol. This compensation phenomenon was discussed in several studies of protein-protein interactions in solution, including the binding processes of IGF-II/IGF2R and hGH/GHR, giving the polar interactions ($\Delta E_{\text{ele}} + G_{\text{PB}}$) of 31.57 and 134.2 kcal/mol, respectively [10,32]. Moreover, the magnitude of $T\Delta S$ is in the range of +30 to -40 kcal/mol, being consistent with the previous results [32].

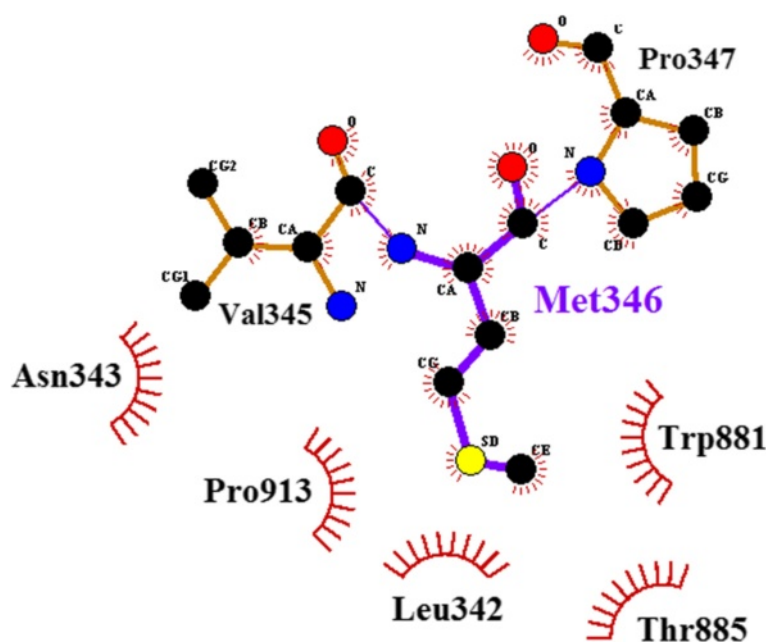
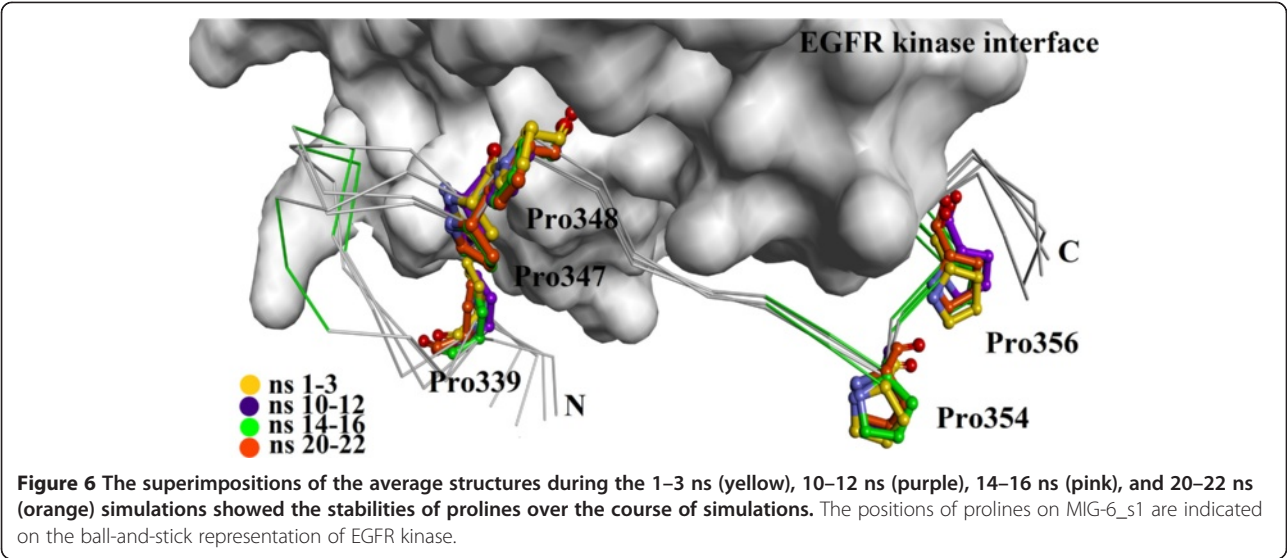


Figure 5 The interactions of Met346 with its nearby residues in the average structure (plotted by LIGPLOT). Met346 is shown in the ball-and-stick representation and its neighbours are shown in brown. Carbon, nitrogen and oxygen atom are shown in black, blue and red, respectively. Dark red “eyelashes” represent hydrophobic contacts between Met346 and its neighbours. Black eyelashes correspond to atoms involved in hydrophobic contacts.



Free energy decompositions of the interfacial residues

The technique of per-residue binding free energy decomposition can reveal the contributions of the key residues responsible for the protein-protein interactions at the interface. The total of 300 snapshots extracted from the last 3 ns of the MD trajectories were decomposed by the MM-PBSA method. The important binding residues of the EGFR kinase/MIG-6_s1 interface were extracted using the residue cut-off at $\Delta G_{\text{tot,PB}} \leq -1.0$ kcal/mol (favorable binding) (Figure 7). There were 13 important binding residues from MIG-6_s1 and 9 residues from EGFR kinase with binding free energies less than -1 kcal/mol (Figure 8). The 13 residues of MIG-6_s1 comprise Ser337, Leu338, Pro339, Tyr341, Met346, Pro348, Thr349, Gln350, Phe352, Lys357, Tyr358, Val359 and Ser361 whereas the 9 residues of EGFR kinase were Thr885, Glu904, Gly906,

Arg908, Pro910, Gln911, Pro913, Met928 and Ile929. Some of these residues, such as Pro910, Thr349, Gln350 and Tyr358, had the values of the binding free energies less than or equal to -4 kcal/mol. Five residues of MIG-6_s1, namely Leu342, Asn343, Ser351, Asp355, and Ser360 and seven residues of EGFR kinase, namely Glu907, Leu909, Pro912, Tyr920, Val924, Trp927 and Gly959, had their binding free energies approximately -1.0 kcal/mol. We further analyzed the binding free energy component of each residue based on the van der Waals energy and the sum of electrostatic contribution in the gas-phase and the polar part of the solvation energy (Figure 8). The results suggested that the main favorable contribution to the binding free energy was essentially the van der Waals interactions. Normally, the favorable electrostatic energies are opposed by the unfavorable polar

Table 1 The binding free energy of the EGFR kinase/MIG-6_s1 complex and its components calculate using the MM-PBSA method

Contribution	EGFR kinase/MIG-6_s1		EGFR kinase		MIG-6_s1		Delta	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Eele	-21099.37	5.29	-19392.37	4.9	-1576.62	1.48	-130.38	1.03
EvdW	-2536.90	1.24	-2296.92	1.2	-119.22	0.33	-120.75	0.31
Egas	-23636.27	5.35	-21689.29	4.96	-1695.84	1.59	-251.13	1.01
Gsol, np	2470.43	0.53	2273.28	0.48	281.40	0.21	-84.25	0.16
Gsol, PB	-3449.41	4.13	-3449.41	3.87	-351.98	1.08	151.59	0.93
Gsol	-978.98	4.03	-975.73	3.8	-70.58	0.94	67.34	0.93
ΔG_{cal}							-183.79	0.5
-T ΔS	-233583.65	1.09	-3282.65	3.36	-342.04	4.02	41.05	2.08
$\Delta G_{\text{cal}} - (-T\Delta S)$							-142.74	-

Mean energies are in kcal/mol, with the corresponding standard errors (σ).
Eele and Evdw are electrostatic and van der Waals contributions in gas phase.
GPB and Gnp are electrostatic and nonpolar contributions in solvation phase.
 ΔG_{cal} is total binding free energy.

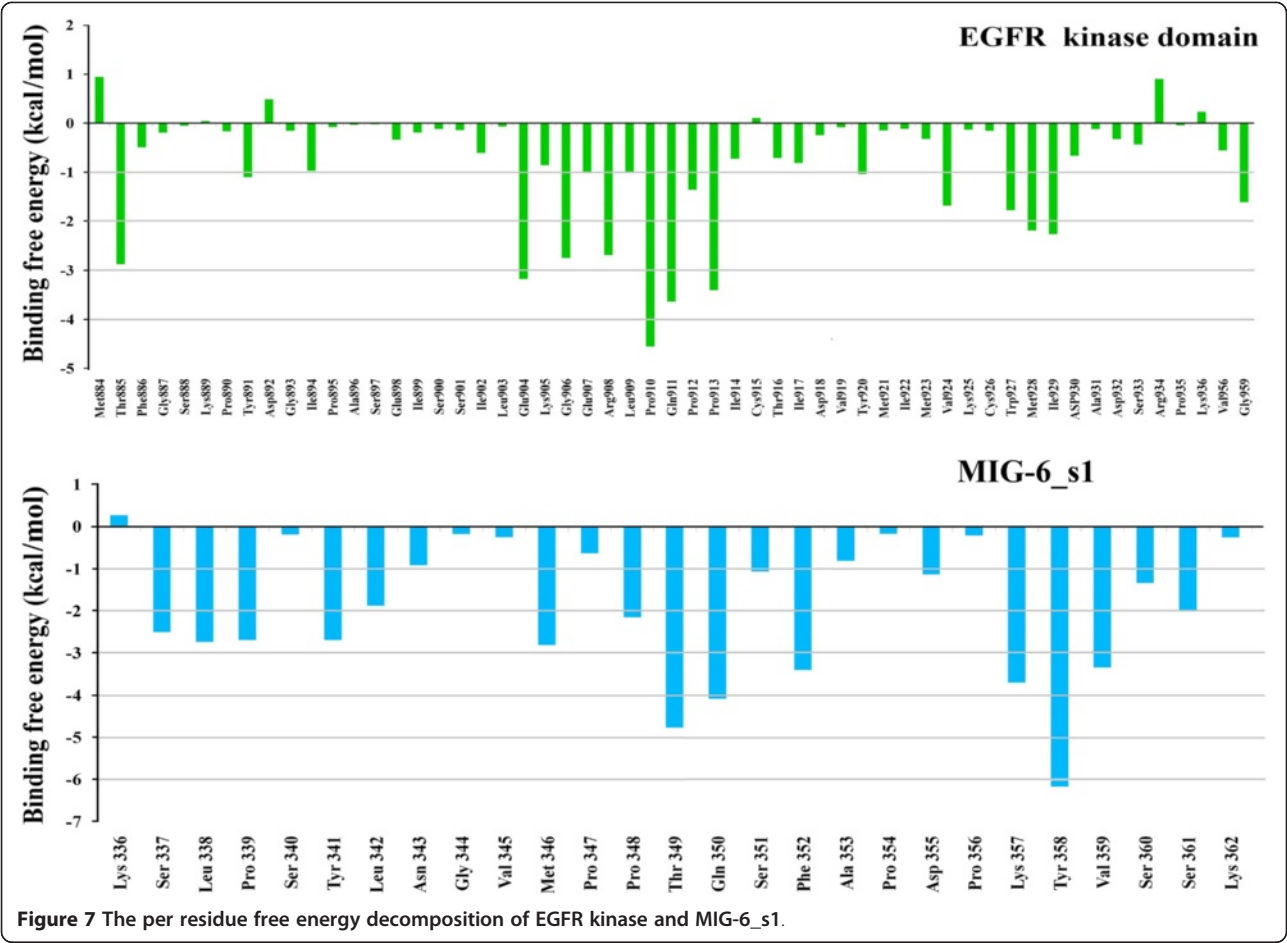


Figure 7 The per residue free energy decomposition of EGFR kinase and MIG-6_s1.

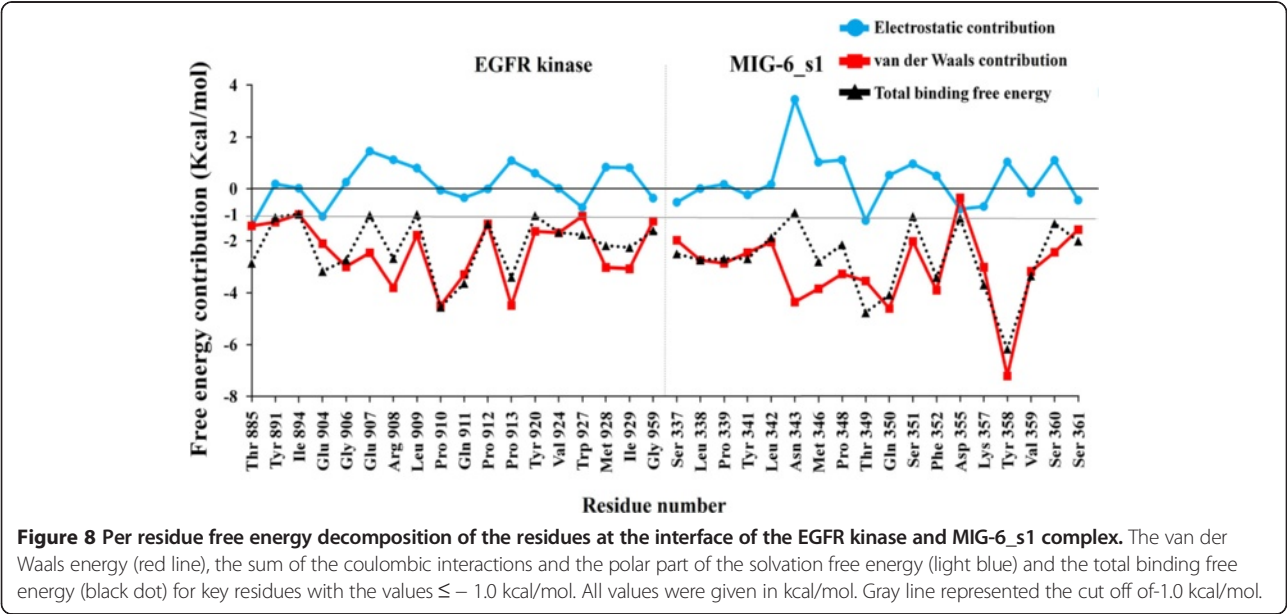


Figure 8 Per residue free energy decomposition of the residues at the interface of the EGFR kinase and MIG-6_s1 complex. The van der Waals energy (red line), the sum of the coulombic interactions and the polar part of the solvation free energy (light blue) and the total binding free energy (black dot) for key residues with the values ≤ -1.0 kcal/mol. All values were given in kcal/mol. Gray line represented the cut off of -1.0 kcal/mol.

parts of the solvation free energies (desolvation of the polar groups), resulting in unfavorable contributions. However, some residues had slightly favorable electrostatic contributions (the sum of the electrostatic energies and the polar parts of the solvation free energies) including Thr885, Glu904, Pro910, Gln911, Pro912 and Trp927 of EGFR kinase as well as Ser337, Leu338, Ser340, Tyr341, Val345, Thr349, Pro354, Asp355, Lys357, Val359 and Ser361 of MIG-6_s1.

Table 2 shows the information on hydrogen bonds between MIG-6_s1 peptide and EGFR kinase during the last 3 ns of the simulations. These are the hydrogen bonds formed between residues Ser337, Tyr341, Asn350, Phe352, Lys357, Tyr358 and Ser360 of MIG-6_s1 and the residues Glu904, Gly906, Arg908, Gln911 and Ile929 of EGFR kinase. Strong, medium and weak hydrogen-bond interactions were defined as having simulated hydrogen bond occupancy of > 75%, 50–75%, and < 50%, respectively. The hydrogen bond distance profiles of some of the strong hydrogen bond interactions during the 22 ns of the simulations are shown in Figure 9. Most of these hydrogen bonds were stable throughout the entire simulations (Figures 9A, B, C, D, E, F, and G), except for hydrogen bonds shown in Figures 9B, I and J. In particular, the hydrogen bond between Tyr341-OH and Gln911-O was briefly disrupted during 13.5–16 ns of the simulations (Figure 9B). The hydrogen bonds between Asn343-ND2-HD21 and Thr885-O (Figure 9I) and between Asn343-NH and Gly959-O (Figure 9J) were not stably formed until around 13 ns and 18 ns, respectively.

Table 2 Hydrogen bonds found in the last 3-ns of the simulation

Donor	Acceptor	Distance (Å) ± SD	Occupancy (%)
Arg908-NH	Gln350-O	2.916 (0.12)	99.87
Tyr341-OH	Gln911-O	2.838 (0.17)	99.27
Gln350-NH	Arg908-O	2.907 (0.13)	99.20
Ile929-NH	Lys357-O	2.918 (0.15)	98.53
Asn343-NH	Gly959-O	2.994 (0.16)	97.73
Ser360-NH	Glu904-O	2.841 (0.12)	96.83
Gln911-NE2-HE21	Ser337-O	2.954 (0.16)	96.10
Arg908-NH2-HH22	Tyr358-O	2.969 (0.16)	92.73
Asn343-ND2-HD21	Thr885-O	3.063 (0.18)	89.57
Phe352-NH	Gly906-O	3.077 (0.18)	87.43
Gln911-NH	Gln350-OE1	3.034 (0.20)	79.33
Thr349-OG1-HG1	Glu907-OE1	2.697 (0.13)	69.90
Ser361-NH	Glu904-O	3.112 (0.20)	69.37
Ile917-NH	Ser337-O	3.044 (0.17)	51.27
Thr349-OG1-HG1	Glu907-OE2	2.703 (0.18)	24.33

The distance cut-off was 3.5 Å and angle cut-off was 120°.

Computational alanine scanning technique

Computational alanine scanning is a powerful technique that not only monitors the key residues required for the interactions between the two protein partners at the interface but also computes the energetic contribution to the binding free energy of the individual side-chain based on each alanine mutation [33]. In order to identify the key residues from the 29 residues along the binding interface of the EGFR kinase/MIG-6_s1 complex, we substituted alanine in amino acids having the free energy contribution less than or equal to -1.0 kcal/mol in order to classify them as either “hot spot” or “warm spot” residues. Note that the higher negative value of $\Delta\Delta G$ of an alanine point substitution represents the better binding state of the residue. A residue is classified to be very important for binding to the protein partner (hot spot) [21,24,34], if its $\Delta\Delta G \leq -4$ kcal/mol. To a lesser extent, the “warm spot” will be classified if $-2 \text{ kcal/mol} \geq \Delta\Delta G > -4$ kcal/mol. In this analysis, proline and glycine residues were not included since their backbone conformations differ from alanine.

The results from the alanine scanning method (Figure 10A and B) and from the per residue free energy decomposition technique (Figure 8) are largely consistent. As determined by the free energy decomposition technique, most of the important residues were also classified as hot spot residues. We showed that there were six “hot spot” residues on EGFR kinase, namely Glu904, Glu907, Arg908, Gln911, Met928 and Ile929; and eight “hot spot” residues on MIG-6_s1, namely Leu338, Tyr341, Asn343, Met346, Thr349, Gln350, Phe352 and Tyr358. For the warm spot classification four residues, Thr885, Ile94, Tyr920 and Val924, were on EGFR kinase (Figure 10A) while Leu342, Val359, and Ser361 were on MIG-6_s1 (Figure 10B). However, due to the limitation of alanine scanning, Gly906, Pro910, Pro912, Pro913, Gly959 on EGFR kinase and Pro339, Pro348 on MIG-6_s1 could not be alanine scanned.

Note that by means of per residue free energy decomposition, eight residues (Tyr891, Leu909, Trp927, Ser337, Ser351, Asp355, Lys357 and Ser360) were identified as important residues and not by the alanine scanning approach. Such discrepancies may stem from the differences in the energy cut-offs used in these two methods. Moreover, the alanine scanning method verified only the contributions from the side-chains, while the free energy decomposition analysis considered both the contributions from the side-chains and the backbones. It may be possible that the contributions from the backbones of these residues may be more than those of the side-chains [18]. The common “key” residues identified by both alanine scanning and the free energy decomposition methods, such as Val924, Met346, Phe352 and Tyr358, are supported by the previous mutagenesis experiment as the alanine replacements of these residues abolished the peptide binding [6].

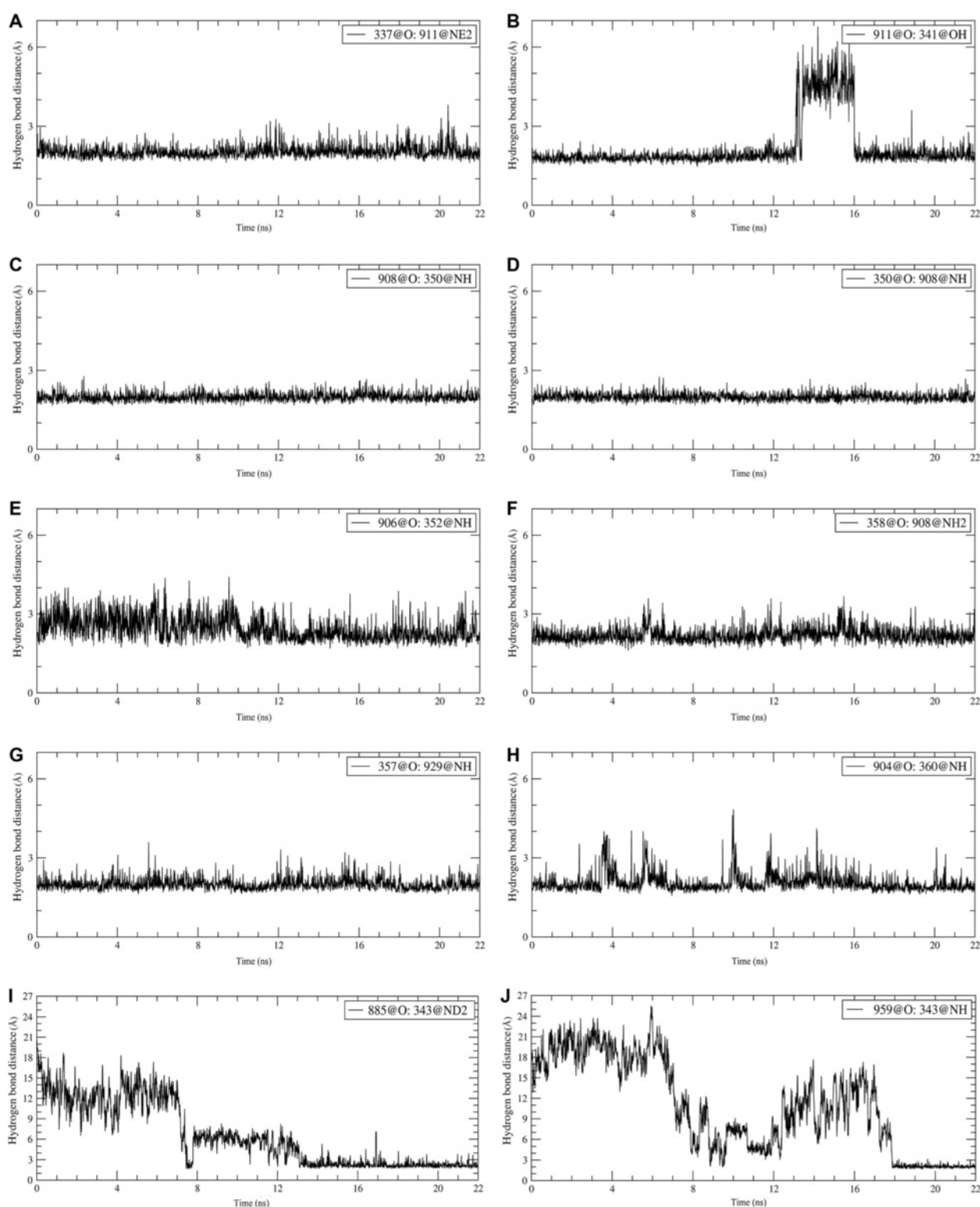
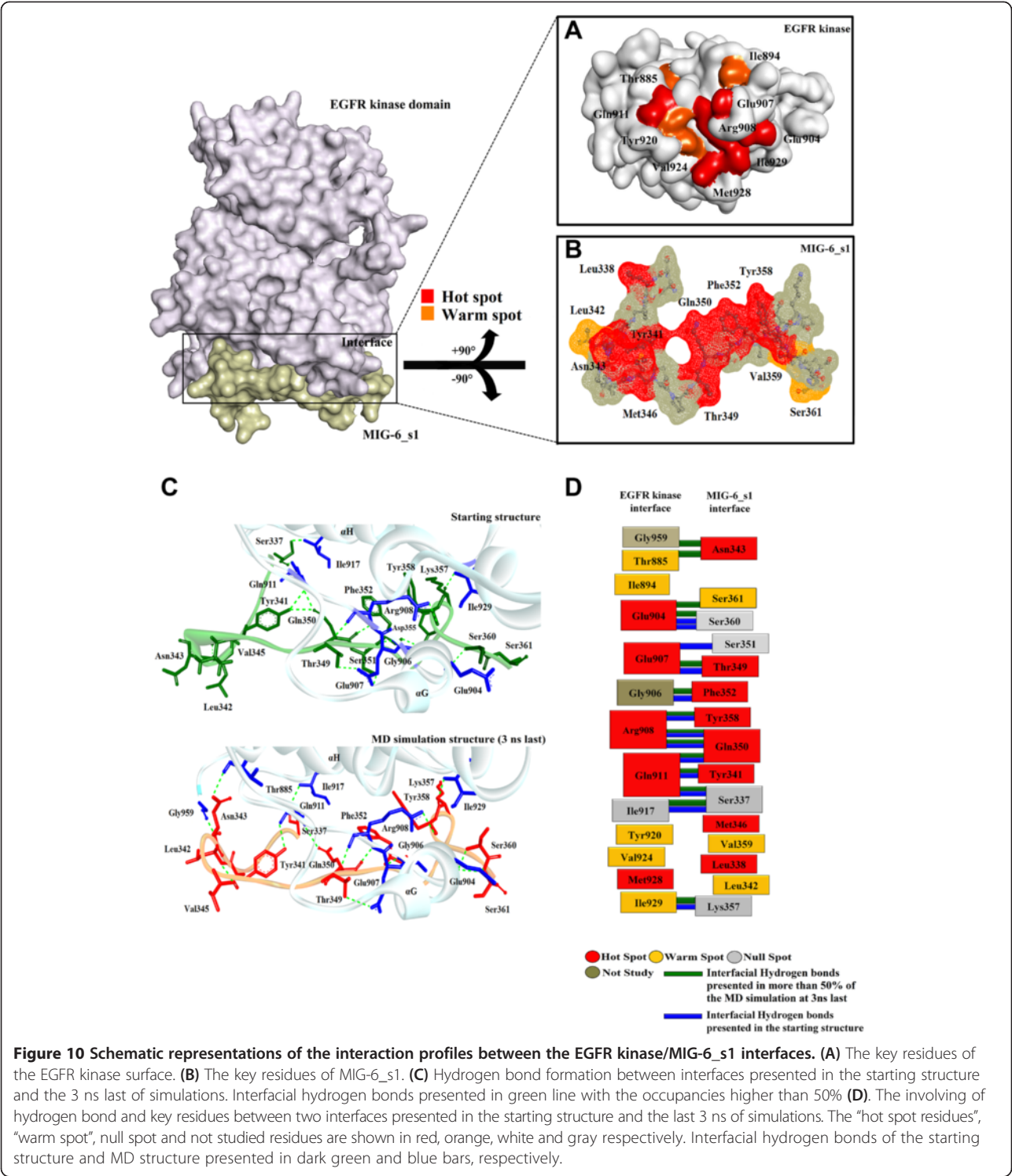


Figure 9 Hydrogen bond distance profiles. (A) Between Gln911-NE2-HE21 and Ser337-O. (B) Between Tyr341-OH and Gln911-O. (C) Between Gln350-NH and Arg908-O. (D) Between Arg908-NH and Gln350-O. (E) Between Phe352-NH and Gly906-O. (F) Between Arg908-NH2-HH22 and Tyr358-O. (G) Between Ile929-NH and Lys357-O. (H) Between Ser360-NH and Glu904-O. (I) Between Asn343-ND2-HD21 and Thr885-O. (J) Between Asn343-NH and Gly959-O.



The interactions of the key residues at the interface of the EGFR kinase/MIG-6_s1 complex

The common key-residue results from both the per residue binding free energy decomposition and computational alanine scanning technique predicted the most important residues at the binding interface that include

six residues of the EGFR kinase and eight residues of the MIG-6_s1 peptide segment (Figure 10A and B). These residues have their $\Delta\Delta G \leq -4$ kcal/mol after performing alanine substitution. Both different and consistent interaction profiles between the MD simulations and the starting structures were observed and presented using

hydrogen bonding network (Figure 10C). The resulting relationship does not only exhibit their interfacial residues but also some involvements of bonding in the important residues (Figure 10D). The MD simulations mostly reveal stable pattern of hydrogen bonds, with the exception for the bonding of Ser351 to Glu907. Interestingly, these hydrogen bonds were also involved in the important residues on the interface. This suggests that the main contributions to the binding interactions were from the hydrogen bonds.

Mutagenesis studies of MIG-6_s1 revealed the importance of Met346, Phe352 and Tyr358 for EGFR kinase binding [6]. With the free energy contribution of approximately -2.8 kcal/mol, Met346 is probably the most important residue of MIG-6_s1 that contributes to the hydrophobic interactions at the interface of the EGFR kinase/MIG-6_s1 complex. The side-chain of Met346 was buried in the hydrophobic pocket consisting of Leu343, Trp881, Thr885 (warm spot) and Pro913 of the EGFR kinase (Figures 4 and 5). The mutating Met346 to Ala346 costs approximately -8.1 kcal/mol of the magnitude of the binding free energy, supporting the importance of its side-chain that engages in the pocket of EGFR kinase. These results were consistent with the crystal structure showing Met346 buried in the hydrophobic pocket of the EGFR kinase, and the experimental results that the mutation of Met346 could abolish EGFR binding [6]. Phe352 and Tyr358 of MIG-6_s1 were also identified in this study as the essential residues for the interface binding with the energetic contributions of about -3.4 and -6.2 kcal/mol, respectively. Their backbones formed hydrogen bonds with hot spot residues, Gly906 and Arg908 (hot spot), respectively, and their side-chains were also buried in the hydrophobic cleft (Figure 11A, B and C). Moreover, the phenyl ring of Phe352 and the phenol ring of Tyr358 formed the parallel π - π interaction with the distance of 5.0 Å between two centroids (<12.0 Å) (Figure 11C) [29]. Therefore, the mutations of Phe352Ala and Tyr358Ala could result in the loss of van der Waals contributions from the hydrophobic interactions of their side-chains [6].

Electrostatic and hydrogen bond interactions were also crucial factors for the binding of MIG-6_s1 and EGFR kinase. Tyr341 of MIG-6_s1 was an important residue for the binding of MIG-6_s1 and EGFR kinase with per residue contribution of -1.88 kcal/mol. During the simulations, Tyr341 formed a hydrogen bond with Gln911 (hot spot) and was also buried in the hydrophobic pocket of Pro339, Pro347, Pro910 and Pro913 (Figure 11D). Moreover, the value of $\Delta\Delta G$ of the alanine substitution of Tyr341 is about -7.9 kcal/mol, which was influenced by the loss of van der Waals and electrostatic interactions. Another residue important for the binding of this complex is Thr349 of MIG-6_s1. Thr349 formed a hydrogen bond with

Glu907 (hot spot) of EGFR kinase (Figure 11E). The value of $\Delta\Delta G$ of the alanine substitution of Thr349 is about -7.4 kcal/mol, caused mainly by the loss of electrostatic contributions. Moreover, the backbone of Gln350 formed two hydrogen bonds with the backbone of Arg908 (hot spot), while its side-chain formed not only a hydrogen bond with Gln911 (hot spot) but also formed van der Waals interactions with Pro910 and Tyr920 (warm spot) (Figure 11F). The value of $\Delta\Delta G$ of the alanine substitution of Gln350 is about -4.3 kcal/mol, where van der Waals and electrostatic contributions were lost upon the side-chain mutation.

During the simulations, not only the new hydrogen bond between Ser361 and Glu904 was formed, but the hydrogen bond involving Asn343 at the loop region of MIG-6_s1 was also observed. The side-chain of MIG-6_s1 moved toward the EGFR kinase surface and formed two hydrogen bonds with Thr885 (warm spot) and Gly959 (Figure 11G). These hydrogen bonds started to form around 13 ns and 17 ns and became stable later on with the occupancies of 89.6% and 97.7% for Asn343-Thr885 and Asn343-Gly959, respectively (Figure 9I and J). The $\Delta\Delta G$ values of the alanine substitution of Asn343 was -4.9 kcal/mol, where the main contribution came from the loss of van der Waals interactions due to the increased cavity volume after alanine substitution (Table 3). The importance of this residue was also supported by the free energy decomposition per residue analysis of the 9- and 18-residue with the values of -0.86 and -1.11 kcal/mol, respectively. This Asn343 residue could play a very important role for a peptide inhibitor to bind to the EGFR C-lobe interface. However, more studies may be needed to further confirm the importance of this residue.

Furthermore, the two MIG-6_s1 prolines (Pro339 and Pro348) and those from EGFR kinase domain (Pro910 and Pro913) could also be essential for the interface binding. They not only gave the free energy contributions of about -2.8, -2.2, -4.55 and -3.40 kcal/mol, respectively (these values were larger than those of the other proline residues), but their free energy contributions were more likely to come from the feature of the side-chains that were -2.34, -2.15, -2.71 and -3.19 kcal/mol more than those of their backbones, respectively. Moreover, the proline rings could also form interaction with the side-chain of Tyr341 and Met346. However, the alanine scanning technique could not be done because the mutation of proline to alanine causes structural perturbations of the backbone [21-23].

We also performed alanine scanning on the EGFR kinase binding interface, on ten residues, namely, Thr885, Ile894, Glu904, Glu907, Arg908, Gln911, Tyr920, Val924, Met928 and Ile929. In particular, Thr885 is located within αF of kinase domain and the side-chain pointed to the

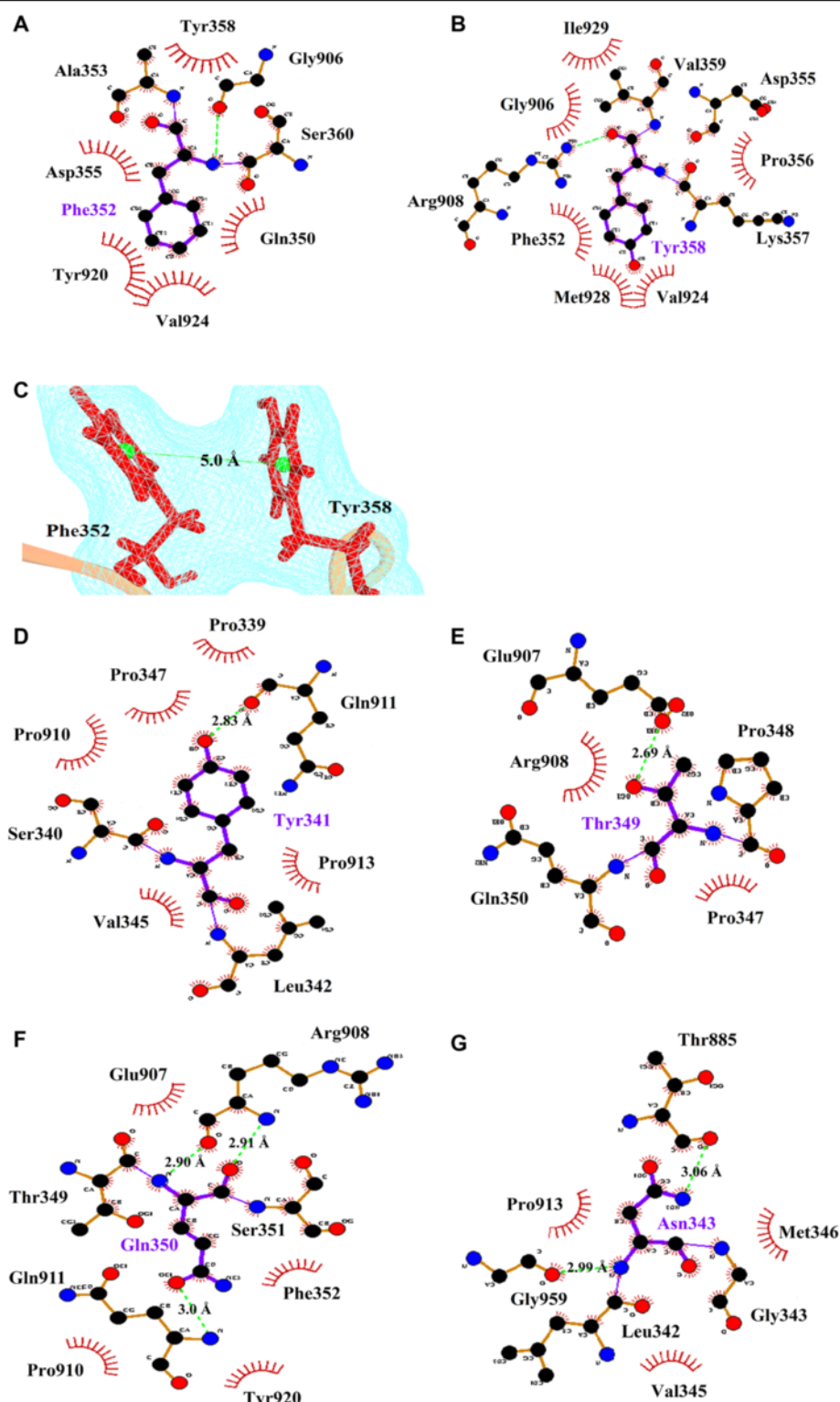


Figure 11 The interactions of the key residues of MIG-6_s1 with nearby residues. **(A)** The hydrophobic interactions and hydrogen bonds of Phe352 and nearby residues. **(B)** The hydrophobic interactions and hydrogen bonds of Tyr358 and nearby residues. **(C)** The π - π interactions between the aromatic rings of Phe352 and Tyr358. **(D)** The hydrogen bond and hydrophobic interactions of Tyr341. **(E)** The hydrogen bonds between polar residues of EGFR kinase Glu907 and Thr349 of MIG-6_s1. **(F)** The hydrogen bonds and hydrophobic interactions of Gln350. **(G)** The hydrogen bonds of Asn343 to Thr885 and Gly959. The important interactions were plotted by LIGPLOT program.

Table 3 The contribution components of binding free energies upon alanine substitutions

Contributions	Residue number									
	Thr885A	Tyr891A	Ile894A	Glu904A	Glu907A	Arg908A	Leu909A	Gln911A	Tyr920A	Val924A
$\Delta\Delta E_{vdw}$	-0.88	-1.22	-1.46	-0.96	-0.67	-4.12	-0.62	-3.12	-2.93	-1.35
$\Delta\Delta E_{electrostatic}$	-0.99	0.38	0.09	-51.15	-60.00	23.34	0.05	-7.21	0.07	0.08
$\Delta\Delta E_{gas}$	-1.85	-0.83	-1.37	-52.11	-60.66	19.22	-0.57	-10.33	-2.86	-1.27
$\Delta\Delta G_{sol, non\ pol}$	-0.14	-0.24	-1.20	-1.16	-1.73	-0.78	0.00	-1.30	-1.24	-0.93
$\Delta\Delta G_{sol, PB}$	-0.12	-0.18	-0.31	48.52	57.96	-34.08	-0.03	4.15	0.45	-0.55
$\Delta\Delta G_{solvation}$	-0.25	-0.42	-1.51	47.36	56.23	-34.86	-0.03	2.85	-0.79	-1.48
$\Delta\Delta G_{subtotal}$	-2.10	-1.25	-2.88	-4.75	-4.43	-15.65	-0.60	-7.48	-3.65	-2.75
	Trp927A	Met928A	Ile929A	Ser337A	Leu338A	Tyr341A	Leu342A	Asn343A	Met346A	Thr349A
$\Delta\Delta E_{vdw}$	-0.70	-3.44	-3.91	-0.43	-3.10	-3.46	-1.40	-4.65	-5.67	-1.10
$\Delta\Delta E_{electrostatic}$	-0.48	-2.63	-1.37	-0.69	0.11	-6.32	-0.05	-3.06	-1.12	-13.24
$\Delta\Delta E_{gas}$	-1.18	-6.07	-5.28	-1.12	-2.99	-9.78	-1.45	-7.71	-7.10	-14.34
$\Delta\Delta G_{sol, non\ pol}$	1.42	-2.22	-1.54	-0.03	-2.54	-1.20	-1.08	-2.19	-3.12	-1.82
$\Delta\Delta G_{sol, PB}$	0.00	2.99	0.13	1.20	0.73	3.04	0.12	4.98	2.10	8.81
$\Delta\Delta G_{solvation}$	1.42	0.77	-1.14	1.17	-1.81	1.84	-0.96	2.80	-1.02	6.99
$\Delta\Delta G_{subtotal}$	0.24	-5.30	-6.69	0.05	-4.80	-7.94	-2.41	-4.91	-8.11	-7.35
Contributions	Residue number									
	Gln350A	Ser351A	Phe352A	Asp355A	Lys357A	Tyr358A	Val359A	Ser360A	Ser361A	
$\Delta\Delta E_{vdw}$	-3.82	-0.77	-5.41	-0.18	-2.02	-8.02	-1.56	-1.36	0.15	
$\Delta\Delta E_{electrostatic}$	-6.61	-1.14	0.70	-5.91	1.38	0.14	3.74	-1.52	-8.37	
$\Delta\Delta E_{gas}$	-10.43	-1.91	-4.71	-6.10	-0.64	-7.88	2.18	-2.88	-8.22	
$\Delta\Delta G_{sol, non\ pol}$	-1.24	-0.07	-3.39	-0.00	-1.44	-4.57	0.41	-0.20	-0.35	
$\Delta\Delta G_{sol, PB}$	7.34	1.03	-0.46	5.13	0.16	2.50	-3.88	2.98	5.37	
$\Delta\Delta G_{solvation}$	6.10	0.96	-3.85	5.13	-1.28	-2.07	-4.29	2.77	5.02	
$\Delta\Delta G_{subtotal}$	-4.33	-0.95	-8.56	-0.97	-1.92	-9.95	-2.11	-0.11	-3.02	

$\Delta\Delta E_{eje}$ and $\Delta\Delta E_{vdw}$ are electrostatic and van der Waals contributions in gas phase.

$\Delta\Delta G_{sol,pb}$ and $\Delta\Delta G_{np}$ are electrostatic and nonpolar contributions in solvation phase.

interface. Thr885 can form hydrogen bonds with Asn343 (warm spot) as well as forming the hydrophobic pocket with side-chain of Met346. However, the loss of $\Delta\Delta G$ (-2.1 kcal/mol) was from the loss of van der Waals and electrostatic interactions but they were not large effects.

The negatively charged residues of EGFR kinase interface, Glu904 (αG) could form hydrogen bonds with the uncharged residues of MIG-6_s1, namely Ser360 and Ser361 (warm spot), while the side-chain of Glu907 (at the loop between αG and $-\alpha H$) formed hydrogen bond with Thr349 side-chain (hot spot). These Glu904 and Glu907 amino acids were classified as hot spot residues due to the mutation of two glutamic acids resulting with the $\Delta\Delta G$ costs of -4.75 and -4.43 kcal/mol, respectively. The loss of electrostatic contribution was the major reason of such results.

Likewise, the positively charged Arg908 located between αG and αH possibly made strong hydrogen bonds with the backbone of Gln350 (hot spot) and the side-chain of Tyr358 (hot spot). After alanine mutation, Arg908 was

identified as a hot spot residue because of its $\Delta\Delta G$ -15.65 kcal/mol that may have been influenced by the increasing cavity volume and the disrupting the conventional hydrogen bonds upon mutating the side-chain. Moreover, the “nitrogen-containing” side chains of Arg908 lining in parallel with the aromatic ring of Tyr358 also formed the cation- π interaction. Although this geometry layout is preferred in protein structures, the mutation of the side-chain could render large favorable electrostatic contribution in the polar solvation [35].

Gln911 (on the loop between αG - αH) was also indicated as a hot spot residue. It disrupted strong hydrogen bonds to Ser337 side-chain upon alanine substitution causing the binding free energy -7.48 kcal/mol, but did not disturb a hydrogen bond of the backbone-side chain between Gln911 and Tyr341. The alignment on αH of Tyr920, Val924 and Met928 showed that they are important to the dimer formation, which were called as “The core of the asymmetric EGFR kinase domain” [36].

These residues not only participate during the hydrophobic interactions that involve residues in the N-lobe of EGFR kinase receiver, but their mutations also abrogated the contribution of the asymmetric dimer [6,36].

These experimental results align with our *in silico* predictions, these residues were superimposed to those in different structures as shown in the RMSF plots (Figure 3) demonstrating the stability of their backbones. Moreover, we took the EGFR kinase activator interface when binding to MIG-6_s1 and compared with the one when binding to the EGFR kinase receiver. The important of these residues were observed on the above EGFR kinase activator interfaces. After alanine mutations, the loss of van der Waals interaction could originate from the disruption of hydrophobic interaction to their partner proteins. The binding energies of Tyr920, Val924 and Met928 values are -3.65, -2.75, and -5.30 kcal/mol (when binding to MIG-6_s1) and -3.46, -4.98, and -6.07 kcal/mol (when binding to receiver kinase), respectively. The sharing residues between two interfaces of the activator entail that MIG-6_s1 might bind the EGFR kinase activator in the manner similar to a binding of the receiver kinase, even though the function of MIG-6_s1 was opposed to the receiver kinase [6,37].

Conclusions

In this study, the 22 ns MD simulations of EGFR kinase/MIG-6_s1 complex was performed in order to provide more understanding about the binding of MIG-6_s1 peptide to the C-lobe of EGFR kinase that is an important target in cancer therapy. Using both per residue free energy decomposition and computational alanine scanning techniques, we found that the Asn343 residue situates on a flexible loop of MIG-6_s1. This particular loop was absent in the original X-ray crystal structure. Furthermore, the underlying loop potentially renders Asn343 a key residue to bind to EGFR kinase. This finding requires further experimental validation to better our understanding about Asn343 role in the protein-protein interaction.

In addition, we used the above energy calculation techniques to identify the hot spot residues (six residues on EGFR kinase and eight residues on MIG-6_s1) that play significant role in the binding between the two proteins. Particularly, the alanine scanning technique showed that the alanine mutations on these hot spot residues would worsen the EGFR kinase/MIG-6_s1 interaction. This technique also revealed the other important residues but to a lesser extent, called warm spot residues that might be worth to further investigate their binding roles. These key amino acids proposed in this work should play vital role for a peptide-based inhibitor to prevent the EGFR asymmetric dimer formation. This *in silico* study, hence, provides valuable information for designing new peptide-based cancer drugs.

Abbreviations

EGFR: Epidermal growth factor receptor; MIG-6_s1: Mitogen induce gene 6 segment1; MM-PBSA: Molecular mechanics/poisson-boltzmann surface area; RMSD: Root-mean-square-deviations.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NM carried out the molecular dynamic simulations, performed the analysis and drafted the manuscript. NS, SR, WM and SC participated in molecular dynamic simulations and provided technical advice. ST and KC conceived the initial concept of the study, and participated in its design and coordination of drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by National Science and Technology Development Agency (NSTDA) under the Thailand Graduate Institute of Science and Technology (TGIST) programme. The scholar student ID is TG-22-11-55-024 M and the grant number is TGIST 01-55-024. ST acknowledges Cluster and Program Management Office (CPMO), National Science and Technology Development Agency (NSTDA), grant no. P-10-10567 for partially supporting this work. KC is supported by Faculty of Science, Kasetsart University, and Thailand Research Fund (RSA5780066). SC is supported by Grants for Development of New Faculty Staff, Ratchadaphiseksomphot Endowment Fund, Chulalongkorn University. The authors would like to thank the National Center for Genetic Engineering and Biotechnology (BIOTEC) and the National Science and Technology Development Agency (NSTDA) for the use of high performance computer clusters.

Author details

¹Department of Biochemistry, Faculty of Science, Kasetsart University, 50 Ngam, Wong Wan Rd, Bangkok, Chatuchak 10900, Thailand. ²Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand. ³Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Pathum Wan 10330, Thailand. ⁴Genetic Engineering Interdisciplinary Program, Graduate School, Kasetsart University, 50 Ngam Wong Wan Rd, Bangkok, Chatuchak 10900, Thailand. ⁵Genome Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency (NSTDA), 113 Thailand Science Park, Phahonyothin Road, Khlong Nueng, Pathum Thani, Khlong Luang 12120, Thailand. ⁶Center for Advanced Studies in Tropical Natural Resources, National Research, University-Kasetsart University, Kasetsart University, Bangkok, Chatuchak 10900, Thailand.

Received: 8 September 2014 Accepted: 6 March 2015

Published online: 27 March 2015

References

- Hynes NE, MacDonald G. ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol*. 2009;21(2):177–84.
- Sebastian S, Settleman J, Reshkin SJ, Azzariti A, Bellizzi A, Paradiso A. The complexity of targeting EGFR signalling in cancer: from expression to turnover. *Biochim Biophys Acta*. 2006;1766(1):120–39.
- Wang DD, Zhou W, Yan H, Wong M, Lee V. Personalized prediction of EGFR mutation-induced drug resistance in lung cancer. *Sci Rep*. 2013;3:2855.
- Friedman R. Drug resistance missense mutations in cancer are subject to evolutionary constraints. *PLoS One*. 2013;8(12):e82059.
- Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A*. 2008;105(6):2070–5.
- Zhang X, Pickin KA, Bose R, Jura N, Cole PA, Kuriyan J. Inhibition of the EGFR receptor by binding of MIG6 to an activating kinase domain interface. *Nature*. 2007;450(7170):741–4.
- Zhang YW, Vande Woude GF. Mig-6, signal transduction, stress response and cancer. *Cell Cycle*. 2007;6(5):507–13.
- Zhang YW, Staal B, Su Y, Swiatek P, Zhao P, Cao B, et al. Evidence that MIG-6 is a tumor-suppressor gene. *Oncogene*. 2007;26(2):269–76.

9. Yang Y, Liu H, Yao X. Understanding the molecular basis of MK2-p38 α signaling complex assembly: insights into protein-protein interaction by molecular dynamics and free energy studies. *Mol BioSyst.* 2012;8(8):2106–18.
10. Guo J, Wang X, Sun H, Liu H, Yao X. The molecular basis of IGF-II/IGF2R recognition: a combined molecular dynamics simulation, free-energy calculation and computational alanine scanning study. *J Mol Model.* 2012;18(4):1421–30.
11. Case DA, Cheatham 3rd TE, Darden T, Gohlke H, Luo R, Merz Jr KM, et al. The Amber biomolecular simulation programs. *J Comput Chem.* 2005;26(16):1668–88.
12. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Physics.* 1983;79(2):926–35.
13. Darden T, York D, Pedersen L. Particle mesh Ewald: an n-log (N) method for Ewald sums in large systems. *J Chem Physics.* 1993;98(12):10089–92.
14. Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem.* 1992;13(8):952–62.
15. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res.* 2000;33(12):889–97.
16. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem.* 1999;20(2):217–30.
17. Zhong H, Carlson HA. Computational studies and peptidomimetic design for the human p53-MDM2 complex. *Proteins.* 2005;58(1):222–34.
18. Li T, Froeyen M, Herdewijn P. Computational alanine scanning and free energy decomposition for E. coli type I signal peptidase with lipopeptide inhibitor complex. *J Mol Graph Model.* 2008;26(5):813–23.
19. Bradshaw RT, Patel BH, Tate EW, Leatherbarrow RJ, Gould IR. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. *Protein Eng Des Sel.* 2011;24(1–2):197–207.
20. Saiz-Urra L, Cabrera MA, Froeyen M. Exploring the conformational changes of the ATP binding site of gyrase B from *Escherichia coli* complexed with different established inhibitors by using molecular dynamics simulation: protein-ligand interactions in the light of the alanine scanning and free energy decomposition methods. *J Mol Graph Model.* 2011;29(5):726–39.
21. Morrison KL, Weiss GA. Combinatorial alanine-scanning. *Curr Opin Chem Biol.* 2001;5(3):302–7.
22. Bullock BN, Jochim AL, Arora PS. Assessing helical protein interfaces for inhibitor design. *J Am Chem Soc.* 2011;133(36):14220–3.
23. Massova I, Kollman PA. Computational alanine scanning to probe protein – protein interactions: a novel approach to evaluate binding free energies. *J Am Chem Soc.* 1999;121(36):8133–43.
24. Sousa SF, Tamames B, Fernandes PA, Ramos MJ. Detailed atomistic analysis of the HIV-1 protease interface. *J Phys Chem B.* 2011;115(21):7045–57.
25. Espinoza-Fonseca LM, Trujillo-Ferrara JG. Conformational changes of the p53-binding cleft of MDM2 revealed by molecular dynamics simulations. *Biopolymers.* 2006;83(4):365–73.
26. Red Brewer M, Choi SH, Alvarado D, Moravcevic K, Pozzi A, Lemmon MA, et al. The juxtamembrane region of the EGF receptor functions as an activation domain. *Mol Cell.* 2009;34(6):641–51.
27. Bonvin AM, Brunger AT. Conformational variability of solution nuclear magnetic resonance structures. *J Mol Biol.* 1995;250(1):80–93.
28. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* 1995;8(2):127–34.
29. McGaughey GB, Gagne M, Rappe AK. π -Stacking interactions: alive and well in proteins. *J Biol Chem.* 1998;273(25):15458–63.
30. Stein A, Aloy P. Contextual specificity in peptide-mediated protein interactions. *PLoS One.* 2008;3(7):e2524.
31. Niu RJ, Zheng QC, Zhang JL, Zhang HX. Molecular dynamics simulations studies and free energy analysis on inhibitors of MDM2-p53 interaction. *J Mol Graph Model.* 2013;46:132–9.
32. Huo S, Massova I, Kollman PA. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J Comput Chem.* 2002;23(1):15–27.
33. Moreira IS, Fernandes PA, Ramos MJ. Hot spots—a review of the protein-protein interface determinant amino-acid residues. *Proteins.* 2007;68(4):803–12.
34. Pons J, Rajpal A, Kirsch JF. Energetic analysis of an antigen/antibody interface: alanine scanning mutagenesis and double mutant cycles on the HyHEL-10/lysozyme interaction. *Protein Sci.* 1999;8(5):958–68.
35. Crowley PB, Golovin A. Cation- π interactions in protein-protein interfaces. *Proteins.* 2005;59(2):231–9.
36. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell.* 2006;125(6):1137–49.
37. Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, et al. Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. *Cell.* 2009;137(7):1293–307.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



ies against tyrosine kinase domain of EGFR: A novel cancer drug

Tabtimmai^{a,*}, Praphasri Suphakun^a, Pichamon Kiatwuthinon^a, Wanpen
pa^b, and Kiattawee Choowongkomon^a

^aDepartment of Biochemistry, Faculty of science, Kasetsart University, Bangkok, Thailand
^bCenter for Research and Technology Development, Department of Parasitology, and
Center of Excellence on Therapeutic Proteins and Antibody Engineering, Faculty of
Siriraj Hospital, Bangkok, Thailand

Corresponding author: fsciktc@ku.ac.th

Antibody-based drug approach is one of the present day strategy for cancer therapy to use the immune system of patients to attack cancer cells. Moreover, an engineered antibody can also be used as a delivery system for translocalizing several lethal molecules such as drugs that enhances the activity of the drug. There are many molecules that can be used to modify antibody properties such as their effector functions, immunogenicity, and cell permeability. In this study, we have focused on the ability of a humanized heavy chain antibody fragment (VH/V_HH) from a nonimmune camel, *Camelus dromedaries*, against the tyrosine kinase domain of the epidermal growth factor receptor (EGFR). Firstly, EGFR-specific VH/V_HH clones were obtained from the humanized VH/V_HH phage display library. These clones were then expressed for obtaining EGFR specific VH/V_HH expressing clones. The cell permeability was introduced into each clones by adding 9 arginine residues (R9) at the C-terminus using a ligation independent cloning system. Three clones (V_HH35, VH35, and VH35R9) were produced in BL21 (rosetta) *E.coli* in a denatured condition by affinity chromatography (TALON). The concentration of the purified VH/V_HHs were determined by SDS-PAGE and subjected to a cell viability assay. The results show that the IC₅₀ of V_HH35, VH35, and VH35R9 were in the nanomolar range (9.61nM, 9.90nM, and 60nM, respectively) compared to the commercial drug (Erlotinib) (20.05μM). This indicates that VH/V_HH can be used for cancer treatment. Because of its activity was greater than the conventional antibody and cell-permeability, this antibody fragment can be developed as a therapeutic agent for cancer therapy.

Keywords: Epidermal growth factor receptor (EGFR), Nanobody, Cell-penetrating, Non-immune camel, lung cancer (NSCLC), VH/V_HH