THESIS APPROVAL

GRADUATE SCHOOL, KASETSART UNIVERSITY

Master of Science (Biochemistry)

DEGREE

Biochemistry	Biochemistry
FIELD	**DEPARTMENT**

TITLE: Molecular Dynamics Simulations of Mitogen-Induced Gene-6 Segment 1 (MIG-6_s1) Peptide to the Activated Epidermal Growth Factor Receptor Kinase Domain Targeting the Asymmertric Dimer Interface

NAME: Ms. Ninnutt Moonrin

THIS THESIS HAS BEEN ACCEPTED BY

THESIS ADVISOR

(Assistant Professor Kiattawee Choowongkomon, Ph.D.)

THESIS CO-ADVISOR

(Mr. Sissades Tongsima, Ph.D.)

DEPARTMENT HEAD

(Assistant Professor Amornrat Promboon, Ph.D.)

APPROVED BY THE GRADUATE SCHOOL ON _____

DEAN

(Associate Professor Gunjana Theeragool, D.Agr.)

THESIS

MOLECULAR DYNAMICS SIMULATIONS OF MITOGEN-INDUCED GENE-6 SEGMENT 1 (MIG-6_s1) PEPTIDE TO THE ACTIVATED EPIDERMAL GROWTH FACTOR RECEPTOR KINASE DOMAIN TARGETING THE ASYMMERTRIC DIMER INTERFACE

NINNUTT MOONRIN

A Thesis Submitted in Partial Fulfillment of
the Requirements for the Degree of
Master of Science (Biochemistry)
Graduate School, Kasetsart University
2014

Ninnutt Moonrin 2014: Molecular Dynamics Simulations of Mitogen-Induced Gene-6 Segment 1 (MIG-6_s1) Peptide to the Activated Epidermal Growth Factor Receptor Kinase Domain Targeting the Asymmertric Dimer Interface. Master of Science (Biochemistry), Major Field: Biochemistry, Department of Biochemistry. Thesis Advisor: Assistant Professor Kiattawee Choowongkomon, Ph.D. 201 pages.

The over-expression of EGFR is associated with human cancers and drug resistant of the mutated EGFR is also a problem thus the C-lobe of EGFR kinase might be a new target for EGFR kinase activity inhibition. MIG-6 is a key in controlling the EGFR over-expression. In order to design a potency MIG-6_s1, the molecular dynamics (MD) simulations and binding free energy were applied to examine the basis protein-peptide interactions between EGFR kinase/MIG-6_s1 on interface. The molecular mechanics Poisson-Boltzmann and Generalized-Born surface area (MM-PB/GBSA) methods were performed to predict binding free energy. Van der Waals and non-polar solvation were the major favorable driving forces for binding process. The crucial residues on interface were obtained by free energy decomposition and several key residues were also proven the important for binding by computational alanine scanning approach. We had used the basis information from our study of EGFR kinase/MIG-6_s1 interactions to design short peptide of MIG-6_s1 together with *in silico* site-direct mutagenesis. MM-PBSA calculation was able to successfully rank the binding affinities of all four models in the following order: T349R/S351R> T349R> wild-type> S351R. The double mutant T349R/S351R bound EGFR kinase was predicted the highest binding affinity. The increasing of electrostatic potential in all mutations was revealed in free energy shift, which might be influenced by charge-charge interaction with the capability of guanidinium side-chain of arginine to salt-bridge formations. To prove the electrostatic effect by charge-charge interaction, not only salt-bridge analysis could confirm this phenomenon through hydrogen bond interaction but the free energy components of the mutated residues were also analyzed. They were made sure that arginine substitutions could increase not only the electrostatic contribution but the capability of their side-chains could also increase van der Waals interaction. The synergistic effect was explained the effect of two mutated sites in double mutation in term of the coupling free energy. These results permitted the mutations of positions 349 and 351 in an optimized sequence of MIG-6_s1 to significantly enhance binding affinity. These approaches revealed in this study might be helpful in the design of new EGFR kinase inhibitors toward the C-teminal lobe of EGFR kinase interface.

_____ _____ ____ / ____ / ____

Student's signature Thesis Advisor's signature

ACKNOWLEDGEMENTS

The thesis would have not been completed without the support from many people who offer me the most valuable suggestions and encouragements throughout the whole time as a M.S student in Kasetsart University.

First, I would like to express my sincere gratitude to my advisor Assistant Professor Dr. Kiattawee Choowongkomon and co-advisor Dr. Sissades Tongsima, respectively as well as the external committee, Dr. Surasak Chunsrivirot for their valuable suggestions, comments and guidances for the completion of this thesis.

I would like to thank Dr. Napat Songtawee and Ms. Wanwimon mokmak who taught and advised me on the technique in Molecular dynamic simulations

My special appreciations are also expressed to Dr. Chomdao Sinthuvanis who advised and helped me.

This research was fullly supported by Thailand Graduate Institute of Science and Technology (TGIST: TG-๒๒-๑๑-๕๕-0๒๔M).

I would like to express my gratitude to the National Center for Genetic Engineering and Biotechnology (BIOTEC) for generously providing computational resource and also the staff at biostatistics and informatics lab.

Finally, I am deeply thankful to my parents, my aunt and my elder brother for their continuous encouragements, understanding and believing in me. Especially I would like to thank my late grandparents for your love and inspiration.

Ninnutt Moonrin

June 2014

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF ABBREVIATIONS

Å	=	Angstrom
ACE	=	Acetyl
ATP	=	Adenosine-triphosphate
AREG	=	Amphiregulin
BTC	=	Betacellulin
BF	=	Bernal-Fowler model
CT	=	C-terminal
CCD	=	Coiled-Coiled Domain
CCPs	=	Cell-penetrating peptides
DBD	=	DNA binding domain
DFG	=	Aspartate-Phenylalanine-Glycine
EGFR	=	Epidermal Growth Factor Receptor
EGF	=	Epidermal Growth Factor
Ele	=	Electrostatic
ErbB	=	Erythroblastosis oncogene B
EPG	=	Epigen
EREG	=	Epiregulin
EBR	=	EGFR binding region
EKTOT	=	Total kinetic energy
ETOT	=	Total energy
EPTOT	=	Total potential energy
GB	=	Generalized Born
HER	=	Human epidermal growth factor receptor
HB	=	Heparin-binding
ICD	=	Intracellular domain
JM	=	Juxtamembrane
JMAD	=	Juxtamembrane activation domain
LCPO	=	linear combinations of pairwise overlaps

LIST OF ABBREVIATIONS (Continued)

MD	=	Molecular Dynamics
MM	=	Molecular Mechanics
MIG	=	Mitogen-Induced Gene
MM/PBSA	=	Molecular Machanics/Poisson-Boltzmann Surface Area
MoAbs	=	Monoclonal Antibodies
NME	=	*N*-methylated amino
NPT	=	Isothermal-isobaric ensemble
NRG	=	Neuregulins
NSCLC	=	Non-small cell lung cancer
NVT	=	Canonical ensemble
OBC	=	Onufriev, Bashford and Case model
PB	=	Poisson-Boltzmann
PBC	=	Pericodic boundary condition
PDB	=	Protein Data Bank
PDK1	=	Phosphoinositide-dependent kinase-1
PI3K	=	Phosphoinositide 3-kinase
PIAS	=	Protein inhibitors of STATs
PTEN	=	Phosphatase and tensin homolog
RMSD	=	Root mean square deviation
RMSF	=	Root mean square fluctualtion
RTK	=	Receptor tyrosine kinase
SASA	=	Solvent Acessible Surface Area
SPC	=	Single point charge
SPC/E	=	Single point charge extend
STAT5	=	Signal Transducer and Activator of transcription
SOCS	=	Suppressor of cytokine signaling proteins
TKD	=	Tyrosine kinase domain
TGFα	=	Transforming growth factor
TIP3P	=	Transferable intermolecular potential 3 point

MOLECULAR DYNAMICS SIMULATIONS OF MITOGEN-INDUCED GENE-6 SEGMENT 1(MIG-6_s1) PEPTIDE TO THE ACTIVATED EPIDERMAL GROWTH FACTOR RECEPTOR KINASE DOMAIN TARGETING THE ASYMMERTRIC DIMER INTERFACE

INTRODUCTION

Cancer is the third leading cause of death worldwide about 14.1 million new cancer cases and 8.2 million cancer deaths estimated to have occurred in 2012 (Ferlay *et al*., 2013). World Heath Oranization (WHO) statistics estimate approximately 22.2 and 13.5 million for new cancer case and cancer-related death cases in 2030, respectively (Knaul and Chabner, 2013). The most commonly diagnosed cancers worldwide in 2012 are lung, breast and colorectal cancer that are estimate 13%, 11.9% and 9.7% of the total number of new cases, respectively. Among these, lung cancer is the most common causes of cancer-related deaths worldwide while Non-Small Cell Lung Cancer (NSCLC) is one of the most lung cancer types that cause to human lethal (Ferlay *et al*., 2013).

The Epidermal Growth Factor Receptor (EGFR) is expressed in the origin of epithelial, mesenchymal and neuronal tissue (Hynes *et al.*, 2001).The activation of EGFR can be resulted in a widly different ways of cellular responses depending on ligand availability and expression pattern (Kholodenko *et al.*, 1999). The over expression of EGFR lead to cell abnormality in functions and become tumor developments including to angiogenesis, metastasis and deadhesion of tumor cells especially in NSCLC tumors that found EGFR expression in range from 43 to 89% (Grunwald and Hidalgo, 2003; Scagliotti *et al.*, 2004).

EGFR is growth factor receptor in EGFR family involved with the biological processes including cell proliferation and differentiation that its activation is mediated by signal transduction networks (Yarden, 2001). The EGFR structure consists of three portions

are the ligand binding extracellular domain, a lipophilic transmembrane region and intracellular domain (ICD). ICD is composed of juxtamembrane region (JM), a signal-transducing tyrosine kinase domain (TKD) and caboxyterminal (CT) region (Flynn *et al.*, 2009; Hynes *et al.*, 2001).

However, the mechanism for activation of the EGFR kinase domain has been intensively studied. Before ligand binding, not only EGFR kinase domain is in an auto-inhibition as seen in Src/Cdk-like inactive conformation but the unintended activation of EGFR is also blocked by self-inhibitory constraints imposed on both the extracellular ligand binding domain and the intracellular catalytic domain. After ligand binding, the extracellular domains are driven dimerization throughout the juxtamembrane region and subsequent to forming of asymmetric dimers. Not only the carboxyl-terminal lobe (C-lobe) of one kinase (the activator) can abute the amino-terminal (N-) lobe of its partner (the receiver) and give rise EGFR kinase domain to the allosteric activation but also induces the auto-phosphorylation of tyrosine at C-terminal tail that is the initiation of downstream signaling. The important pathways of EGFR including to the Ras/Raf/MEK/ERK signaling, PI3K/Akt signaling and Jak/Stats signaling pathways, which can affect to cellular processes of a variety organisms (Yarden, 2001). EGFR signaling is tightly controlled by the role of negative regulatory circuits, which can drive ligand-bound EGFR complex into fast internalization. One such protein inhibitors, MIG-6 (also called Mitogen-Induced Gene-6) is negative feedback inhibitor of EGFR and transcriptionally induced by the over-expression of EGFR through a RAS/ERK pathway (Descot *et al.*, 2009). In addition, the over-expressed MIG-6 segment 1 (MIG-6_s1) completely blocks ligand-induced protein kinase by binding the distal surface portion of C-lobe in the part of kinase domain. The formation of asymmetric kinase dimer is inhibited before carrying of the internalized receptor into late endosomes and subsequently the lysosomal destruction compartment (Anastasi *et al.*, 2003; Frosi *et al.*, 2010). The hallmark feature of MIG-6 bound EGFR give rise to the suppression of EGFR catalytic activity and down-regulation of EGFR signaling. As the involving of the over-EGFR expression with the incidence of cancer and two small oral adenosine triphosphate (ATP)-compertitive EGFR-TK targeted inhibitors including erlotinib and gefitinib are clinically significant response with non-small-cell lung cancer

(NSCLC) patients (Masago *et al.*, 2010). Nevertheless, the activating mutations in EGFR kinase domain after treatment are critical roles in the activation of EGFR kinase activity and cancer development. The mutations such as the second mutation of T790M can be occurred in patients who have ever take the ATP-competitive kinase inhibitors leading to reducing the potency and dysregulation of small molecule inhibitors activity (Yun *et al.*, 2008). Therefore, the interactions feature of EGFR kinase and MIG-6_s1 was important for determining the complex binding process. It could be utilized to be guide for designing a new class of EGFR kinase inhibitors in the future with high degree affinity to new targeting EGFR kinase surface.

It is well known that the interaction of protein-protein complex is important role in both of biophysical processes and biological systems including signal transductions antibody-antigen interactions, hormone-receptor interactions and enzyme allostery (Horn *et al.*, 2009). Nevertheless, the deeply understanding of the binding of the two proteins at an atomic level in terms of energetic and dynamic modes is limited. The increasing number of X-ray crystallography obtained protein-protein complex, thermodynamic experiment and mutational studies have contribute to the understanding of protein-protein interactions but are often not sufficient to understand the broad range of biological activity. Although, it still lacks the detail information about the conformation changes during the interactions, free energy and driving forces for binding also the critical residues in the interactions interface (Haberl *et al.*, 2009).

The molecular dynamics (MD) methods have become an important computational tool for understanding the physical basis of the structure and function of biological macromolecules including to protein-protein, protein-lipid and protein-ligand complex (Horn *et al.*, 2009). These methods are utilized to estimate dynamic properties of complex systems and also for equilibrium state that cannot be calculated directly. Moreover, MD simulations are the interface between theory and experiment at the way of science. They have recently been applied for EGFR inhibition by MIG-6 bound kinase domain interface, which MIG-6 segment 1 (residues 337-361) lie on the distal surface of C-lobe of kinase domain adopts the Src/CDK–like inactive conformation formed by helices αG and αH and

loops connecting helices αF-αG, αG-αH and αH-αI of EGFR kinase. In the energetic of interactions, the polar interaction is mainly important, while the hydrophobic interaction is revealed in few residues from helix αH contribution to binding interface (Zhang *et al.*, 2007). However, the dynamics of atomic-level phenomena of MIG-6_s1 peptide/EGFR kinase complex is still not characterized.

Thus, the purpose of this study was to investigate the MIG-6 segment 1 (MIG-6_s1) peptide binding to the C-lobe of the activated EGFR kinase domain interface by using the MD techniques. This technique could help us to get a better understanding of the complex binding processes and the interactions. The information could be the basic knowledge for designing short peptides derived from MIG-6_s1 peptide. There were two strategies including the optimization sequence that covers key residues and computational site-direct mutagenesis. These strategies might provide for the potency inhibitor with high degree affinity to EGFR kinase (Figure1).

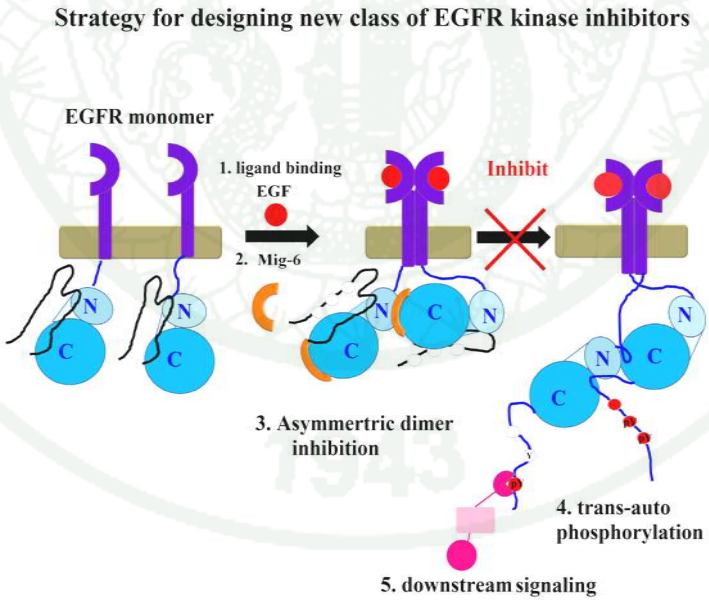


Figure 1 Model for inhibition of asymmertric EGFR dimer formation using MIG-6_s1 peptide based on protein-peptide interactions strategies.

Source: Modified from Zhang *et al.* (2006)

OBJECTIVES

The general purpose of this research is to fulfill the knowledge about the importance of MIG-6_s1 peptide blockage on the activated EGFR kinase interface in the molecular mechanism level. Taking an advantage of the binding interactions, the peptide derived from MIG-6_s1 peptide could be designed in order to carry out with high affinity for inhibition of the activated EGFR asymmertric dimer formation.

The objectives of this study were:

1. To investigate the overview of interactions between MIG-6_s1 peptide bound EGFR kinase monomer complex.

2. To design short peptide derived from MIG-6_s1 peptide for high affinity to the EGFR kinase domain.

3. To investigate the interactions in terms of the dynamics and stabilities of short peptides derived from MIG-6_s1 peptide with the EGFR kinase monomer.

LITERATURE REVIEW

1. Receptor tyrosine kinases (RTKs) family

Receptor tyrosine kinases (RTKs) are a family of transmembrane glycoproteins (Hubbard and Till, 2000) that act as receptors for cytokines, growth factors, hormones and other signaling molecules (Lemmon and Schlessinger, 1994). The receptors are activated by the binding of their ligands and they can transactivate the extracellular signal into the cytoplasm by phosphorylating tyrosine residues on their domains (autophosphorylation) and further downstream signaling proteins. RTKs are expressed in many cell types and involved in biological processes including to cell proliferation, differentiation, migration, angiogenesis or metabolic changes (Mlcochova *et al.*, 2013). The RTK family including platelet-derived growth factor receptor (PDGF), fibroblast growth factor (FGFR), vascular endothelial factor (VEGFR) and epidermal growth factor receptor (EGFR) with the various growth factors such as platelet-derived growth factor (PDGF), fibroblast growth factor (FGF), vascular endothelial factor (VEGF) and epidermal growth factor (EGF) are involved in cellular processes (Hubbard, 1999; Lemmon and Schlessinger, 1994). Epidermal growth factor receptor (EGFR) is the receptor tyrosine kinase cloned in 1984 and also plays important roles in tumorigenesis (Ullrich *et al.*, 1984). This protein is frequently overexpressed in non-small-cell lung, colorectal and in squamous-cell carcinomas of the head and neck (Hubbard, 1999). Moreover, the focus of this review is on the insights into EGFR studies. Before reviewing the mechanism of EGFR, the structural of EGFR was reviewed and its functions, mutation, inhibitions and treatments therapy were also presented in last part of EGFR introduction.

2. Architecture of Epidermal growth factor receptor (EGFR)

The EGF receptor family has four members including EGFR/HER1, ErbB2/HER2, ErbB3/HER3 and ErbB4/HER4 as seen in Figure 2. EGFR is the first family member though its human gene denoted to ErbB1 after the v-erbB oncogene of avian erythroblastosis virus (Roskoski, 2004).

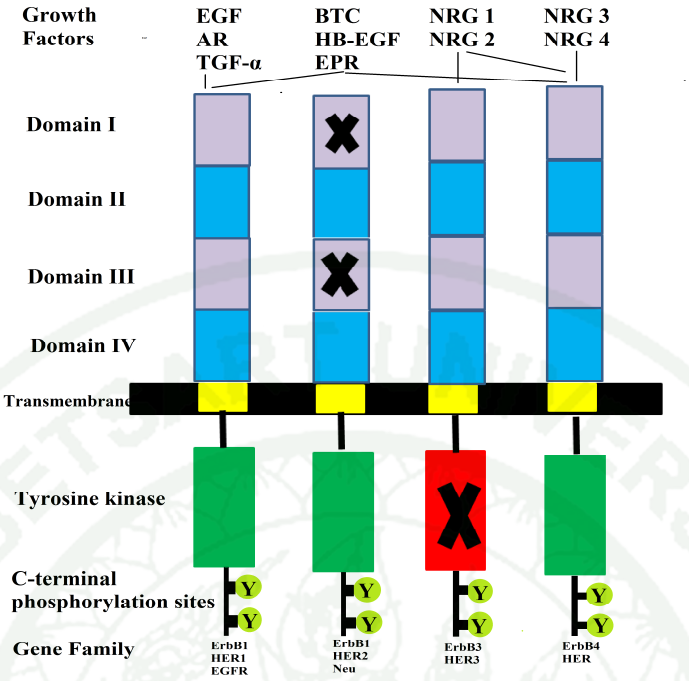


Figure 2 The four members of EGFR family and their ligands.

Source: Modified from Roskoski (2004)

Epidermal growth factor receptor (EGFR) is a type I single-pass transmembrane glycoprotein, which has a single isoform, from a single 28 exons gene located on chromosome 7p11-13 (Wang *et al.*, 2012). EGFR is a 170 kd glycoprotein (Merlino *et al.*, 1985) and initially synthesized from a 1,210 residues transmembrane precursor polypeptide and the 24 residues at the N-terminal signal sequence is cleaved then the cleaved sequence of 1186 residues protein sequence is translocated into the cell membrane (Ullrich *et al.*, 1984). The structure of EGFR is consisted of the ligand binding extracellular domain, a lipophilic transmembrane region (TM) and intracellular domain (ICD) that ICD is composed of juxtamembrane region (JM), a signal-transducing TK kinase domain (TKD) and a carboxyterminal region (CT) (Flynn *et al.*, 2009; Hynes *et al.*, 2001) as shown in Figure 3.

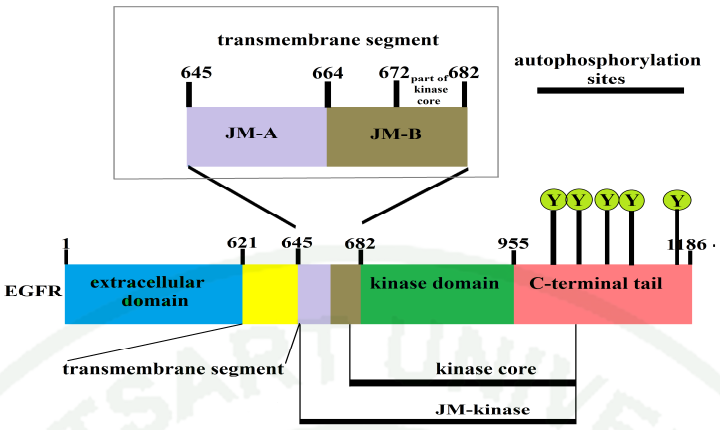


Figure 3 The structure of EGFR protein. It was consisted of the extracellular region (EC), transmembrane region (TM) and intracellular domain (ICD).

Source: Modified from Jura *et al.* (2009)

First of all, the EGFR extracellular region (residues 1-621) or ectodomain is consists of four subdomains that have functions in dimerization of the receptor. It is a crucial step required for tyrosine kinase activation. The subdomains of ectodomain are subdomain I (L1 where L: leucine-rich domain), subdomain II (CR1), subdomain III (L2) and subdomain IV (CR2) as shown in Figure 2 (Ogiso *et al.*, 2002). As can be seen in Figure 4, before ligand-bound EGFR, the unliganded ectodomain adopts a "tethered" inactive receptor state with low-affinity of EGFR. In presence of ligand, subdomains II and IV can intact each other and change to extended conformation with high-affinity of EGFR (Mattoon *et al.*, 2004). In this state, the ligand can induce sub-domain I to close subdomain III by swinging a 130° movement of subdomain I. This can lead to interacting between these subdomains. After swinging from the tethered to the extended conformation, the dimerization arm of subdomain II is driven exposure and contacted with the exposed domain II of the other partner monomer in order to mediate EGFR dimerization (Ogiso *et al.*, 2002). In addition, the highly containing of cysteine residues in the subdomain II and IV domains are the potential for forming disulfide bonds at intradomain that is important for maintenance the structure (Saxon and Lee, 1999).

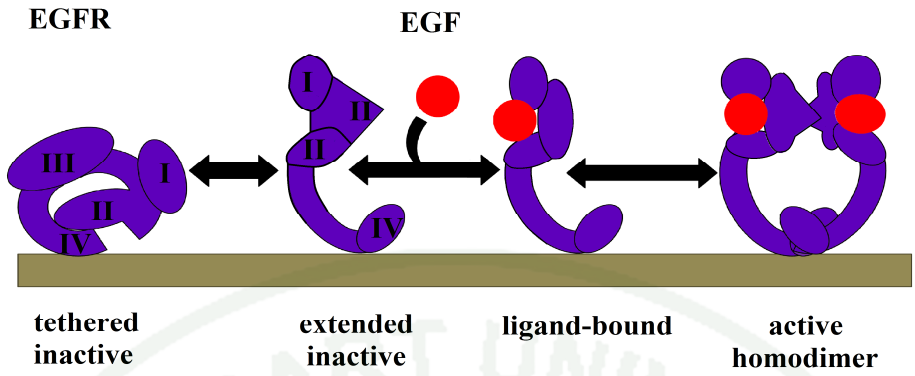


Figure 4 Schematic of EGFR ectodomain conformations after ligand binding

Source: Modified from Jura *et al.* (2009)

The second part of EGFR, a lipophilic transmembrane domain is consists of 23 amino acids as residues 622-644 that was predicted by visual analysis of EGFR sequence. This transmembrane is a single α-helical transmembrane that connects to juxtamembrane region and plays role in anchoring the receptor to the lipid bilayer of the cell (Saxon and Lee, 1999).

The third part of EGFR is intracellular domain as can be seen in Figure 5. It contains the juxtamembrane region (JM) of human EGFR (residues 645 to 682) that can connect between transmembrane region and tyrosine kinase domain. The JM region is composed of receptor-trafficking signal (Bao *et al.*, 2000), a putative calmodulin-binding site (residues 645-660 in JM region) (McLaughlin *et al.*, 2005) and a poly-basic region (Aifa *et al.*, 2005). The JM region can be divided into 2 major areas including juxtamembrane A and juxtamembrane B regions, which are called the JM-A and JM-B, respectively (Figure 6). The JM-A is located on the N-terminal half as residues 645-663 and formed short an alpha helix as observed in the crystal structure. It also connects to the C-terminal end of the transmembrane helice (Jura *et al.*, 2009) and forms an anti-parallel helical manner in the dimer form (juxtamembrane clasp) (Adak *et al.*, 2011). As recently study, Aislyn and colleagues in 2012 assayed peptide from the JM-A region for studying anti-cancer properties and its ability to modulate EGFR signaling. It was interesting that the

designed peptide residues 645-662 could bind EGFR at the JM region, inhibit EGFR dimerization and downstream signaling in Akt and Erk pathways. These pathways play role in the mediated cascade for signaling to regulate gene expression and prevent cell apoptosis. The results showed that this peptide derived from JM region could potentially serve as a therapeutic drug (Boran *et al.*, 2012; McCubrey *et al.*, 2006). Moreover, the JM-A region can be applied as a marker for determination ligand binding properties of the EGFR due to the losing all coorperativity in ligand binding after the deletion of JM-A region (Boran *et al.*, 2012). The other segments, juxtamembrane B located on the C-terminal half as residues 664-682 (Jura *et al.*, 2009). The important role of this segment is the forming a cradle by JM-B segment of receiver kinase around the C-lobe of activator kinase in asymmetric dimer. Red Brewer and colleagues in 2009 identified the critical role of JM-B that the C-terminal half of the JM region refers to juxtamembrane activation domain (JMAD) residues 664-682 can contribute to EGFR activation. It enhances the asymmetric dimer formation and also allosteric activation of the receiver kinase. However, the mutations in this region can give rise to the lossing of the EGFR kinase activity (Boran *et al.*, 2012; Red Brewer *et al.*, 2009). Jura and co-workers demonstrated the involving of JM-B segment in dimer interface refers to as juxtamembrane latch (residues 664-671) that is shielded by the C-terminal tails of the EGFR in the symmetric dimer structure to inactive kinase domain (Figure 7A). This potential mechanism prevents EGFR activation in the absence of its ligand condition (Jura *et al.*, 2009). Moreover, the residues motif in juxta-membrane latch as EPLTPS is identical in residues 323-327 of MIG-6 peptide (EGFR inhibitor) with sequence EPLSPS. This identical sequence of MIG-6 peptide can engage onto the C-lobe of the activator kinase domain as Figure 7B and 7C. The reminder residues of MIG-6 peptide are allowed to block asymmetric dimer formation by disruption the JM latch formation between the activator and receiver kinase domain (Jura *et al.*, 2009).

Due to the important roles of JM region including the stabilizing of EGFR active dimer and the mediating of EGFR signaling (Jura *et al.*, 2009). Macdonald-Obermann and Pike found the intracellular juxtamembrane domain A and B are associated with the allosteric regulation in ligand binding EGFR (Macdonald-Obermann and Pike, 2009). The ligand engages to the first site on EGFR dimer with high affinity in order to induce the

formation of the asymmertric dimer by allowing activation of one kinase. This interaction involves with the stabilization mediated by the formation of an anti-parallel helical dimer in JM-A and also the proximal JM-B residues, which can lock the JM-A helix against the C-lobe of the activator kinase. For second site on the dimer, the binding affinity of ligand is reduced comparing with the first site and required for disrupting these JM-A domain-mediated interaction in the initial asymmetric dimer. Kinase domain is allowed to the formation of reciprocal asymmetric dimer and activated other partner kinase domains. However, the deletion of JM can affect to EGFR dimerization, the abnormal of ligand binding and also the decreasing of the phosphorylation (Aifa *et al.*, 2005; Jura *et al.*, 2009; Thiel and Carpenter, 2007).

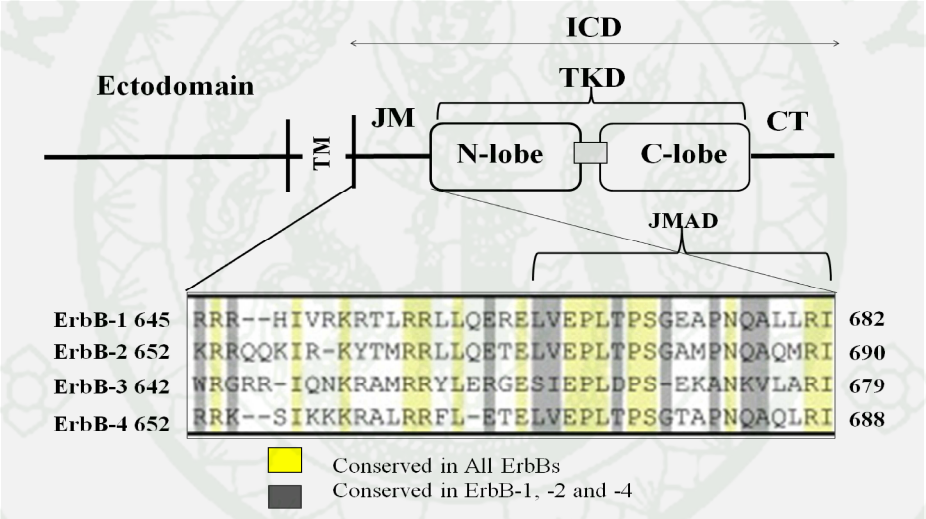


Figure 5 The juxtamembrane regions of the human EGFR members of EGFR family and their conserved residues in JM regions.

Source: Modified from Thiel and Carpenter (2007)

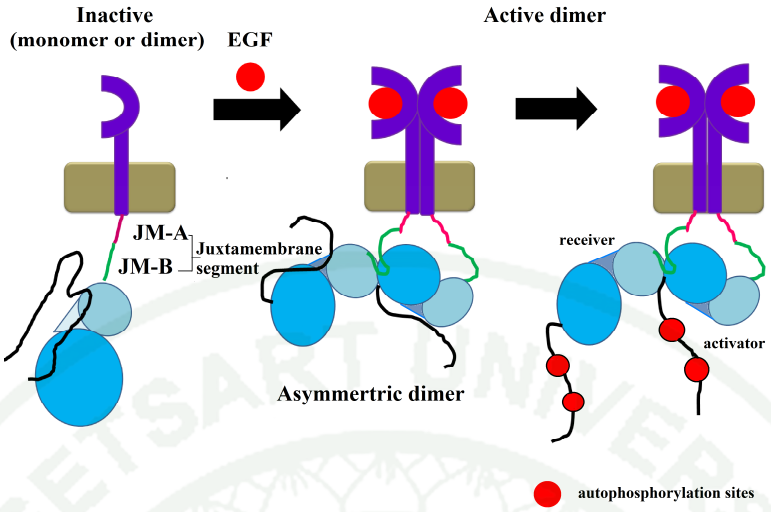


Figure 6 A schematic representation of the juxtamembrane A (pink) and B (green) in EGFR inactive monomer and active dimer.

Source: Modified from Jura *et al*. (2009)

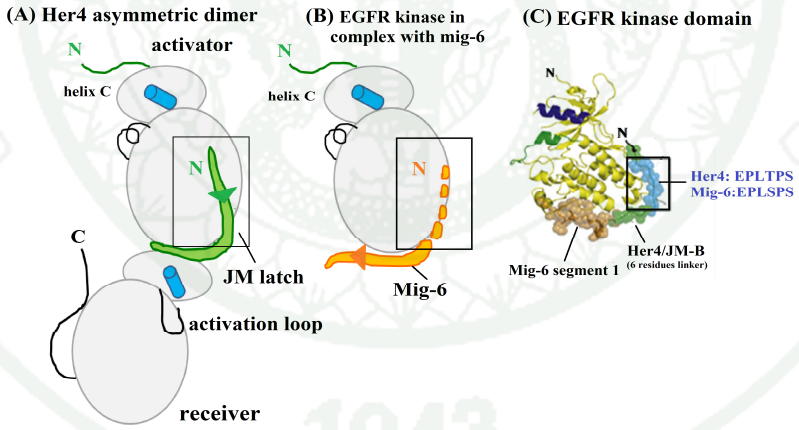


Figure 7 The asymmetric dimer formation of the EGFR Kinase Domain. (A) The JM latch engagement at the C-lobe of kinase in EGFR activation. (B) The blockage of negative feedback inhibitor MIG-6. (C) The conserved residues between JM latch and MIG-6 peptide.

Source: Modified from Jura *et al*. (2009)

After JM region, a signal-transducing TK kinase domain (TKD; residues 683-955) are formed by mainly β-sheet and α helix and adopts a bi-lobal structure that is composed of N-terminal lobe (N-lobe) and C-terminal lobe (C-lobe) as shown in Figure 8A (Macdonald-Obermann and Pike, 2009). The N-lobe contains the glycine-rich nucleotide phosphate-binding loop, while the C-lobe is composed of the DFG motif, the catalytic loop and the A-loop (Stamos *et al.*, 2002). Tyrosine kinase domain has the important structure required for controlling activation state including ATP-binding pocket and substrate binding site. In addition, the catalytic machinery within the ATP cleft, specific residues, loops and motifs must be in the proper positions for the phosphate transferring. However, the important residues are described. (i) The glycine-rich nucleotide phosphate-binding loop consists of glycine residues Gly695-Gly700 responsible for nucleotide binding and makes interactions with the β and γ-phosphates of ATP (Stamos *et al.*, 2002). (ii) Activation loop (Asp855-Val876) is not required phosphorylation for tyrosine kinase activation but there are many groups of residues, which interact either hydrophobically or with H-bonds leading to stabilizing the unphosphorylated activation loop in an active position. These H-bonds can be occurred at the early and central parts of the A-loop including part of Lys836-Leu838, Val810-Arg808 and also Tyr867-Arg812. The N-terminal region of the activation loop contained Asp855-Phe856-Gly857 (DFG) motif is also called the magnesium position loop because the side-chain of the aspartate can coordinate with magnesium ion. This coordination is required for ATP binding (Aertgeerts *et al.*, 2011). In inactive conformation, the flip-in (active state) of DFG motif is changed the conformation by exchanging position each other between Asp and Phe. The swinging out of Asp located on active site into hydrophobic environment can disturb the coordination between Asp and magnesium ion leading to the loss of ATP binding in the nucleotide binding cleft. The swinging out of Asp can make ion pair with Lys 721. Furthermore, the moving of Phe residue outwards from core the spine can lead to disruption of the regulatory spine that relates to the activation of kinase (Aertgeerts *et al.*, 2011). (iii) Helix αC in N-lobe is contained the conserve residues name Glu738. In active state, αC helix can swing in its conformation closely to ATP and permits Glu738 to form salt bridge with a conserved Lys721 in strand β3 coordinated with phosphates group of an ATP. In opposite to active state, the result form αC helix swinging out and pushing Glu738 outwards from active site

can lead to a short helix of DFG motif formation for stabilizing αC helix inactive conformation. The side-chains of two or three conserved hydrophobic residues are used for disturbing salt bridge (Glu738-Lys721) in active state by forming new salt bridge of Lys721-Asp 885 in DFG motif (Aertgeerts *et al.*, 2011). Besides of ATP binding cleft, the substrate peptide pocket is located in a part of the C-terminal region of activation loop, which is in open conformation provided for the substrate docking. The C-lobe of region containing sequence Leu955-Val956-Ile957 is also called LVI motif adjacent to the COOH terminal end, which it regulates the phosphorylation of substrate tyrosines in oligomerized EGFR family complexes. This motif is necessary for ligand independent dimerization of EGFR ICDs (Chantry, 1995). Nevertheless, the mutation can be occurred within this region leading to unable be as a substrate for phosphorylation (Aertgeerts *et al.*, 2011).

3. Mechanisms of activation and signaling pathways

3.1 Ligand binding

The ligands of EGFR family are recognized up to 13 types including EGF itself, heparin-binding (HB)-EGF, transforming growth factor (TGFα), amphiregulin (AREG), epiregulin (EREG), epigen (EPG), betacellulin (BTC) and neuregulins (NRG), which EGF and TGFα are play role in EGFR binding (Eccles, 2011). Human EGF is a 6045 Da protein with 53 amino acids and three intramolecular disulfide bonds (Cohen, 1962). Epidermal growth factor or EGF is ligand released for cell for modulation cell growth, proliferation and differentiation by binding to its partner receptor leading to dimer formation (Liu *et al.*, 2007). As seen in Figure 8, at the absence of ligand, the EGFR can form on cells as both monomers and ligand-independent dimers (pre-dimers). Some study reported that the cell surface can found EGFR in inactive dimer formation lower than monomer, thought the ligands display the binding with high affinity to EGFR dimer at cell surface. It was suggested that EGFR in pre-dimers formation can assist the active formation to diffuse along the plasma membrane, even at low EGF concentrations. However, because of pre-dimer is insufficient for activation, thus the ligand induced-inactive dimers and dimerization steps are required for activation (Liu *et al.*, 2007). As mention above, the

extracellular domain of EGFR is consisted of subdomain I through IV. Totally, the binding between ligand and receptor are formed a 2:2 complex formation on ectodomain EGFR at residues 1-621 on the cell surface for functional (Ogiso *et al*., 2002). In the outer part of EGFR portion is formed an EGF-binding domain and is comprised of four parts (I, II, III and IV). In monomeric form, a tethered conformation is found and stabilized by subdoman II and IV that also called dimerization arm of subdomain II and the tethering arm of subdomain IV (Ferguson *et al*., 2003). After ligand presentation, the tethered conformation is induced by opening two globular parts (I and III) for EGF binding, while two rod-shaped linkers (II and IV) contain 12 cysteine amino acids to help rigidified structure. After EGF binding to one monomer with low affinity, the second arm of other sites is induced conformational change for ligand binding. Then EGF can bind both sites of dimer leading to conformational change into active conformation of EGFR kinase domain and beginning signaling processes. To opposite with monomeric form for dimeric configuration, the EGF ligand can bind with high affinity. Nevertheless, EGF ligand can induce its receptor to form dimer that is the basic of positive cooperatively (Lemmon *et al*., 1997). Therefore, the role of extracellular domain is also balanced between inactive and active states of the kinase domain, which the inactive monomer and pre-dimer are presence markedly before ligand inducing and active dimer is outstanding after ligand simulation. Even though, the intracellular mechanisms of EGFR can achieve in the absence of ligand by Src kinase protein stimulation through G-protein-coupled receptor (Andreev *et al*., 2001).

3.2 EGFR activation

The activation of kinase domain are processed into two steps including receptor dimer formation in homo/hetero dimer and tyrosine phophorylation for serving as docking site of adaptor protein in the carboxyl terminal of EGFR by its own kinase domain. They are the beginning steps of EGFR activation (Yu *et al*., 2002). The activity of kinase is controlled by the conformation pattern within kinase domain including inactive and active dimeric conformations.

In active state**,** as Zhang and his colleagues presented that the active dimer can bring the C-terminal of the extracellular domain closely together then their transmembrane domains at N-terminus containing dimerization motif are induced to dimerize and contribute the juxtamembrane and kinase domain activation later (Zhang *et al*., 2007).

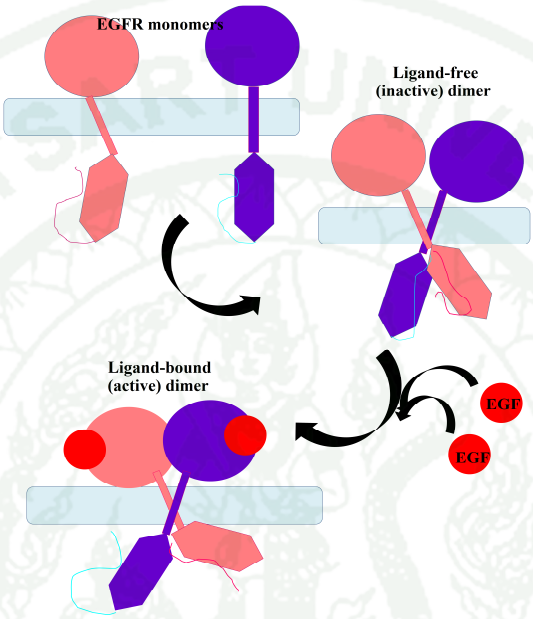


Figure 8 The model of EGF ligand binding with its receptor. The EGFR monomers can form pre-dimer in ligand independent or dependent way.

Source: Modified from Arkhipov *et al*., (2013)

In part of juxtamembrane region (JM) found that JM-B portion called as "juxtamembrane latch" part of monomer B can cradles the C-lobe of kinase monomer A to stabilize the asymmetric dimer (Theil and Carpenter, 2007; Zhang *et al*., 2007). Moreover, JM-A is required for dimerization and activation by two JM-A segments of dimer connecting with the C-terminal of transmembrane segment can form short helix and antiparallel manner leading to allow properly conformational change of monomer B for activation. As the pattern conformation associated with kinase activity, the asymmetric conformation is exactly active form of intracellular domain. It is formed between the C-

lobe of one kinase domain monomer (monomer A) to activate in part of N-lobe of the other (monomer B). The asymmetric interaction between them is resemblance between cyclin A and activated cyclin-dependent kinase 2 (Figure 9) (Zhang *et al*., 2006), while the structure of the C-lobe of the kinase domain is not related to that cyclins. However, this conformation in active state allow the activation loop extending from the cleft and permitting Glu residue in helix αC to form salt bridge with Lys residue that coordinates the α and β phosphate of ATP. This mechanism leads to phosphate transfer from ATP to peptide substrate in catalytic domain providing for tyrosine phosphorylation in C-terminal tail and allowing the downstream signaling (Figure 10) (Pellicena and Kuriyan, 2006).

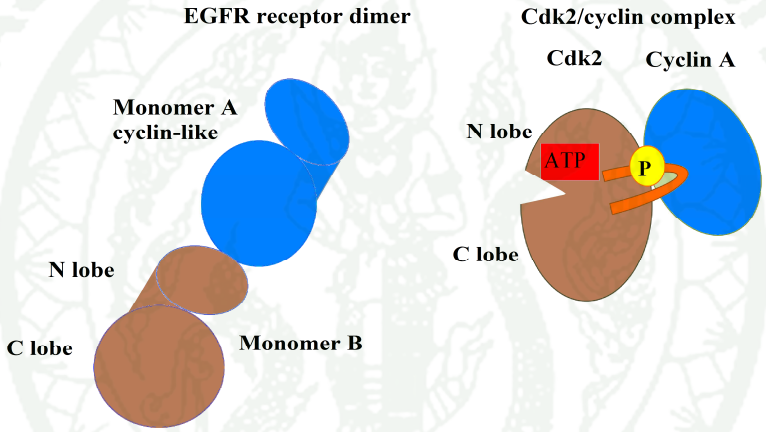


Figure 9 Comparison of EGF receptor activation by dimerization with the Cdk2-cyclin A complex.

Source: Modified from Pellicena and Kuriyan (2006)

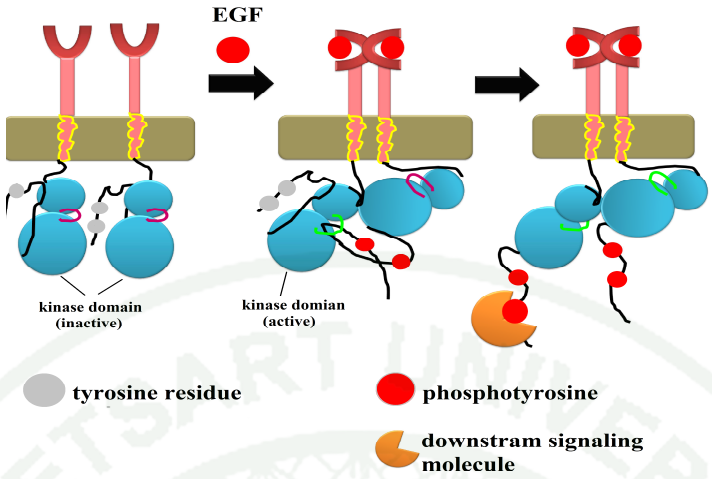


Figure 10 The model of the activation mechanism for the EGFR after ligands binding and inducing dimerization.

Source: Modified from Zhang *et al*. (2006)

In inactive state, at the cell surface can find both of monomer and pre-dimer isoforms (Figure 11). These are in inactive state resembles to Src and the serine/threonine cyclin dependent kinases (CDKs) formation, which helix Cα and activation loop are closely each other and the C-terminal tail is kept distance providing for preventing the suitable position of catalytically key elements in activation (Zhang *et al*., 2006). Moreover, Zhang and his co-workers reported that monomeric inactive conformation of EGFR kinase is stabilized by contacting between C-terminal tail and tyrosine kinase domain. When EGF induce native EGFR monomer, it take long for dimerization because the liganded-monomer need to move around the cell surface looking for its partner (Yu *et al*., 2002). In addition, Yu and colleague (2002) revealed that the size of EGFR dimer can be detected in sample from unstimulated Ba/F3 cells by SDS-PAGE and western blot analysis. It was corresponded to Zhang and his colleagues (2006), dimeric formation in the inactive conformation of the EGFR may similar to the symmetric dimer conformation found in crystal complex and not requires for ligand-induced activation (Figure 11). Symmetric dimer conformation is consisted of the two monomers independent ligand binding that position is back to back and head to tail between the N-lobe of its own kinase domain and

the C-lobe of one kinase partner (Figure 12) (Landau *et al.*, 2004). This conformation is auto-inhibition and not involved in activation process though ligand-binding can induce the predimerized EGFR to twist in portion of transmembrane domain and reorients into active conformation in part of intracellular domain (Lax *et al.*, 1991; Moriki *et al.*, 2001) by the juxtamembrane latch is suppressed by the C-terminal tail of monomer B to prevent kinase activation in the absent of ligand (Zhang *et al*., 2006). Furthermore, only the extracellular domain does not form dimer indicating that it is also important for preventing autoactivation (Lax *et al.*, 1991; Olayioye *et al.*, 2000).

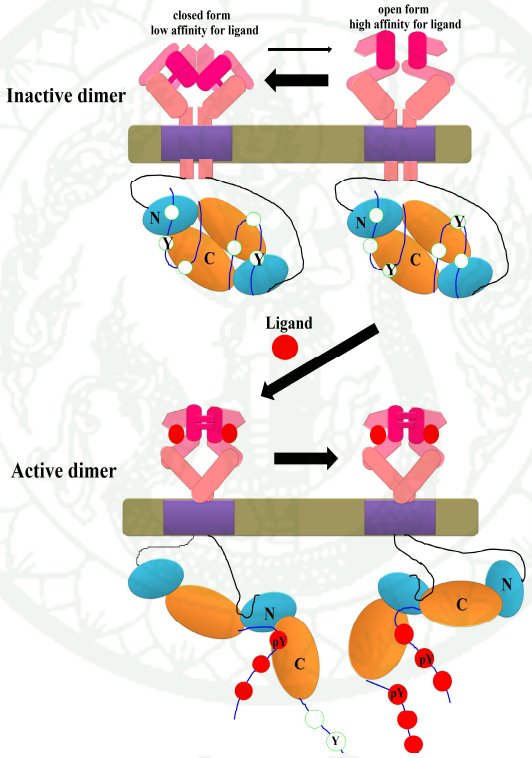


Figure 11 The pre-dimerized EGFR twist and rotation into active dimer in present of ligand

Source: Modified from Tao and Maruyama (2008)

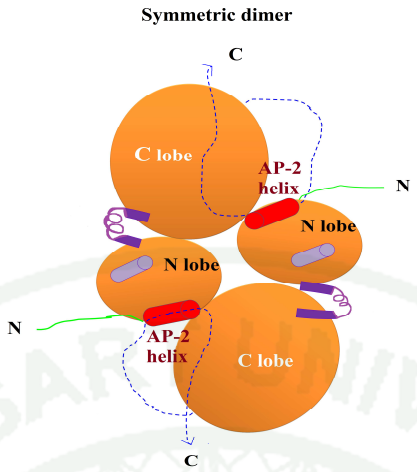


Figure 12 A symmetric inactive dimer of the EGFR Kinase Domain.

Source: Modified from Zhang *et al.* (2006)

Many receptor tyrosine kinases are regulated by the intracellular JM segment region and the C-terminal tail, which their C-terminal ends are brought closely together at cross-linked with the transmembrane segments. The JM-B section of juxtamembrane segment can form a clamp and cross between its the N-terminal lobe of the receiver kinase domain and the C-terminal lobe of the activator kinase. Ligand binding drives receptor to form homo- or hetero-dimerization leading to activation of tyrosine kinase and subsequent auto-transphosphorylation specific tyrosine residues located within C-terminal tail (Landau and Ben-Tal, 2008).

EGFR dimerization stimulates its intrinsic intracellular protein-tyrosine kinase activity leading to open of autophosphorylation sites. The autophosphorylation sites on specific residues (Tyr^{845}, Tyr^{891}, Tyr^{920}, Tyr^{1045}, Tyr^{1101} and Tyr^{1173}) on the C-terminal tail can be bound and phosphorylated with protein kinases and phosphatases that contain Src homology 2 (SH2) or phosphotyrosine binding (PTB) domain. These proteins including adaptor protein such as Src homology domain-containing adaptor protein (Shc), phospho-lipase C_Υ (PLC_Υ), Cbl, Src and and phosphatidylinositol-3-kinase (PI3K) (Well, 1999; Moriki *et al*., 2001). For example, up to date, Tyr^{845} is identified as non-autophosphory-

lation of EGFR catalytic domain that plays role in the activation loop stabilization, the active state enzyme maintenance and a binding residue for substrate proteins, which c-Src is related in phosphorylation of EGFR at Tyr^{845} (Figure 13) (Cooper and Howell, 1993; Hubbard *et al.*, 1994).

Intermolecular interaction is occurred in EGF-induced trans-auto phosphorylation of EGFR by one kinase acts as a substrate for another one in the dimer (Qian *et al.*, 1994). Trans-(auto) phosphorylation can control different functions of EGFR, including docking of kinases, ubiquitination and internalization. It can be devided the features into two roles are phosphorylation on activation loop to elevates enzyme activity and phosphorylation of JM region and C-terminal tail to produce various patterns for recruitment and phosphorylation of target substrates. Nevertheless, the phosphorylation of activation loop is not required for EGFR catalytic activition (Bonaccorsi *et al.*, 2006; Kumar *et al.*, 2008).

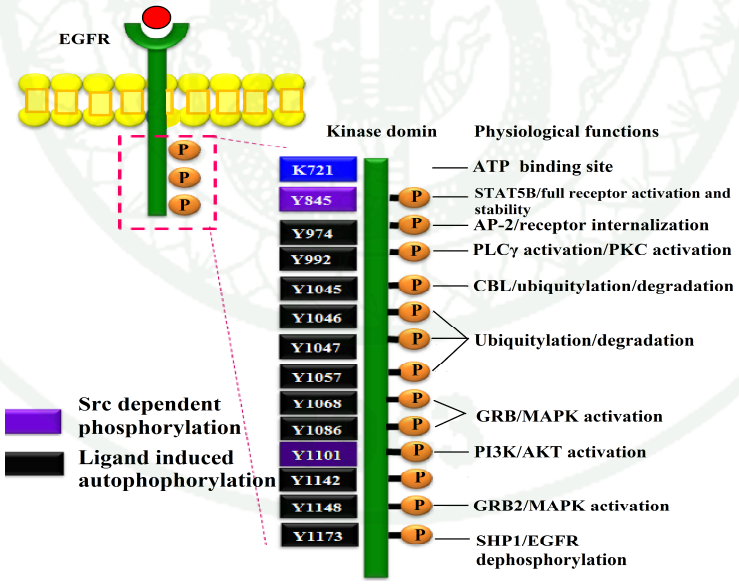


Figure 13 Recruited adaptor protein at specific phosphotyrosine residues on EGFR

Source: Modified from Wheeler *et al.* (2010)

3.3 Signaling pathways activated by EGFR

After tyrosine phosphorylation, some signaling pathway can present to initial recognition at the C-terminal phosphotyrosine by adaptor protein for signaling molecule partner (Hackel *et al.*, 1999). Then EGFR is induced migration by ligand binding and kinase activation into intercellular or intracellular signaling, which targets into nucleus of signalling for controlling biological functions such as cell proliferation, differentiation and angiogenesis (Frosi *et al.*, 2010; Mattoon *et al.*, 2004). Nevertheless, in non-signaling state, EGFR can produce low basal activity and increase substantially after ligand binding, receptor dimerization and phosphorylation step (Stamos *et al.*, 2002). Phosphotyrosine residues (pY) have been identified as specific interaction sites serving as docking sites for adaptor proteins in order to generate of various signaling cascades. The major cascades were Ras/Raf/MAPK kinase, (MEK)/extracellular-related kinase (ERK) and PI3K/AKt and other pathways such as PLC_{γ}/protein kinase C (PKC) (Yardan and Kelman, 1991; Yardan, 2001; Hirsh *et al*., 2003; Jorissen *et al*., 2003; Sebastian *et al*., 2006).

3.3.1 The Ras/Raf/MEK/ERK signaling

The Ras/Raf/MEK/ERK signaling pathway is studied in EGFR intracellular signaling, which the signaling can regulate many cellular processes including cell proliferation, differentiation, migration, apoptosis, angiogenesis and chromatin modeling, while these processes are also involved in cancer pathology (Dunn *et al.*, 2005). The Ras/Raf/MEK/ERK signaling pathway can be activated by growth factor receptor after dimerization and autophosphorylation on tyrosine residues. Next, the phophotyrosine residues recruits adaptor protein (Grb2) directly association with EGFR at tyrosine residues 1068 and 1086 or indirectly activated EGFR by Shc binding at tyrosine residues 1148 and 1173, then Grb2 recruits SOS guanine nucleotide exchange factor to the receptor complex for activation Ras proteins and localize to the cytosol (Batzer *et al*., 1994). The activated Ras protein can induce Raf family kinase activation. Then the activated Raf kinase can phosphorylate serine residues at 218 and 222 located on the activation loop in MEK1 and MEK2, which are further activation in ERK1 and ERK2 by phosphorylation (Figure 14).

Next, the phosphorylated ERKs can further phosphorylate cytoplasmic and nuclear target such as signal transducer and activator of transcription (STAT), c-Jun terminal kinase (JNK), c-Myc regulator gene (Yan and Templeton, 1994).

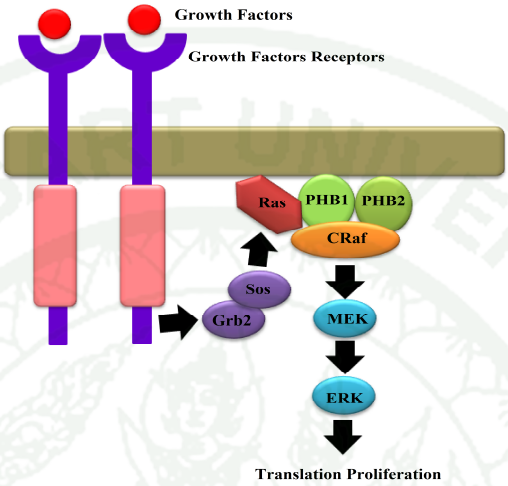


Figure 14 The mechanism and function of the Ras/Raf/MEK/ERK pathway.

Source: Modified from Polier *et al.* (2012)

3.3.2 PI3K/Akt signaling pathway

PI3K/Akt signaling pathway (Figure 15); phosphoinositide 3-kinase (PI3K) is a key signal transduction system that important in cellular functions such as cell survival, cell proliferation and also cell differentiation though it is also associated with oncogenes in human cancers (Zhao and Roberts, 2006). PI3K is a hetero-dimer consisted of regulatory subunit (p85) and catalytic subunit (p110). It can be divided into three classes according to their structural characteristics and substrate specificity including Class IA-IB, Class II and Class III. However, Class IB, II and III of PI3Ks are distincted in regulatory proteins ranging from 55 to 150 kD in size and genes, while in catalytic subunit are less well-known (Leevers *et al.*, 1999; Wymann and Pirola, 1998). The signaling mechanism, after ligand binding with the receptor tyrosine kinases, then p110 subunit of PI3K is allowed to transfer the γ-phosphate group from ATP into the D3 position of phosphor-

inositides-2-phosphate (PIP2). The phosphorylated PIP2 form or phosphoinositides-3,4,5-triphosphate (PIP3) can activate PDK1 and PKC. The recruited Akt of PIP3 at the membrane can result in the phosphorylation of Akt by mammalian targeting of rapamycin (mTOR)-rictor kinase complex and also 3-phosphoinositide-dependent kinase-1 (PDK1) (Martinez-Marti and Felip, 2011). Akt is protein kinase B classified into three isoforms; Akt1, Akt2 and Akt3. They are encoded by the genes PKBα, PKBβ and PKBγ, respectively (Dummler and Hemmings, 2007). Akt isoforms have a same structure such as N-terminal PH domain, a serine/threonine at center of catalytic domain and a small C-terminal regulatory domain (Morrow *et al.*, 2011). The phosphorylation of Akt generates the active form of Akt (p-Akt) enhancing cell survival and proliferation via the phosphorylation several proteins. For example, Bad and pro-apoptotic Bcl-2 family are suppressed for reducing apoptosis. GSK3β modulator is inhibited in multiple pathways and affected to cell proliferation. Moreover, Akt increases Mdm2-mediated ubiquitination and degradation of the tumor suppressor p53. Too much PI3K/p-Akt signaling can promote tumor growth. However, the removing of phosphate in PIP3 by the phosphatase and tensin homolog (PTEN) can affect in Akt suppression (Steelman *et al.*, 2004).

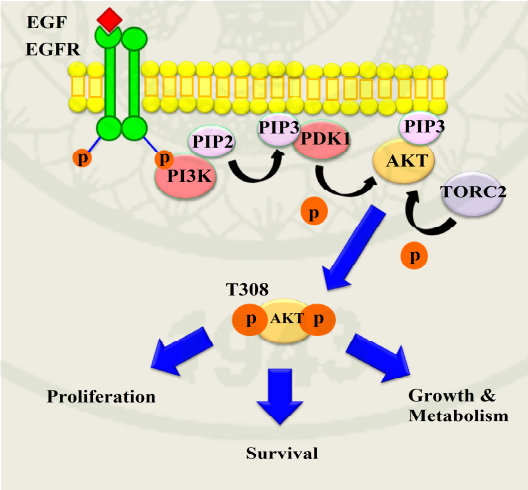


Figure 15 Exploiting the Pi3k/Akt pathway.

Source: Modified from Garcia-Echeverria and Sellers (2008)

3.3.3 The JAK/STAT signaling pathway

The JAK/STAT pathway is the main signaling cascade pathway played an important role in various cellular functions. Janus tyrosine kinases (JAKs) activation is responsible for cell proliferation, differentiation, cell migration and apoptosis. In mammals, there are four members in the JAK family including JAK1, JAK2, JAK 3 and Tyk2 (Rawlings *et al.,* 2004). The structure of JAK sequence exhibits seven janus homology (JH1-JH7). JH1 encode the kinase domain whereas JH2 is a pseudokinase domain needed for helping JH1 catalytic activity. Janus tyrosine kinases are contained Src homology-2 (SH2) domain at JH3-JH4 (O'Shea *et al.,* 2004). The amino-terminal JAK homology domains are JH5 to JH7 responsible for cytokine binding and regulating catalytic activity (Kisseleva *et al.*, 2002).

STATs (Signal Transducer and Activator of transcription) are cytoplasmic transcription factors recruited by an activated receptor complex. They are activated via JAK after recruitment into receptor. STATs family has seven isoforms including STAT 1 to 6 while STAT5a and STAT5b are also counted (Kisseleva *et al.,* 2002). For the structure of STATs includes the amino-terminal domain (NH2), the coiled-coiled domain (CCD), the DNA binding domain (DBD), the linker domain, the Src homology 2 (SH2) domain and tyrosine activation domain (TAD) (Kisseleva *et al.*, 2002).

In the JAK/STAT pathway (Figure 16), the downstream signaling of JAKs can occur after ligands binding to receptors and stimuliting the multi-merization of receptors. Then, two JAKs are closed approximately for trans-phosphorylation. After that, JAKs also phosphorylates the both of receptor and members of the cytoplasmic transcription factor STAT family containing SH2 domain. The phosphorylated STATs can dissociate from receptor to dimer formation and nucleus translocation. The binding specific regulatory section of dimer can lead to activating or suppressing target genes in transcription step. Nevertheless, the phosphotyrosine can be regulated by phosphatase enzymes for inhibiting the feature of receptor and also blocking transcription factor. The phosphatases were comprised the protein inhibitors of STATs (PIAS) and the suppressor of

cytokine signaling (SOCS) proteins activation (Greenhalgh and Hilton, 2001). Thus, the JAK/STAT pathway is a critical role in linking between cell surface receptors and nuclear transcriptional events leading to cell growth and can be controlled by negative regulator (Marrero, 2005).

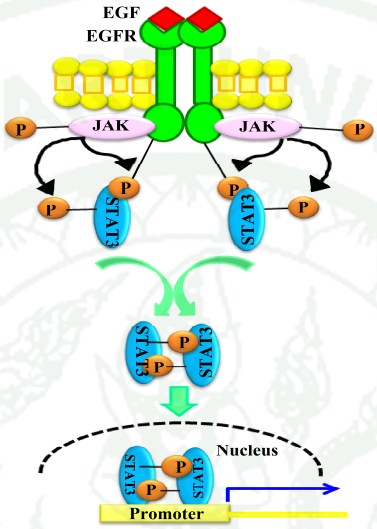


Figure 16 The JAK/STATs signaling cascade.

Source: Modified from Telser (2002)

4. Mutations in EGFR

4.1 Mutations in the extracellular domain

In the part of the extracellular domain, EGFR has three major types of the deletion mutations depending on the site and length of deletions. There are EGFR type vI, vII and vIII (Figure 17) (Yan and Templeton, 1994). EGFR type vI is total deletion of the extracellular domain and has only been observed in a xenograft derived from a malignant human glioma (Zhao and Roberts, 2006). EGFR type vII was found in gliomas. It contains a deletion of 83 amino acids at residues 521–603, which is part of the cysteine-rich domain IV crossed between the ligand binding domain III and the transmembrane domain V.

Moreover, the EGFR type vII is important for transducing EGF stimulation of cell proliferation and invasion *in vitro*, while its responds very similar to the wild-type of EGFR to growth factors (Leevers *et al.*, 1999). Of these mutant forms, EGFR vIII is the most common mutation in high- and low-grade lesion of gliomas (Wymann and Pirola, 1998), which lacks a large part of the extracellular portion domain I and II giving rise to constitutive dimerization and activation of the receptor. This mutation is detected in 5 of 32 lung carcinomas, 21 of 27 breast carcinomas, 4 of 6 paediatric gliomas, 6 of 7 medullo-blastomas and 24 of 32 ovarian carcinomas (Dummler and Hemmings, 2007; Martinez-Marti and Felip, 2011; Morrow *et al.*, 2011). In addition, Montagut and co-wokers revealed that the mutation from serine 429 to arginine can prevent cetuximab antibody inhibitor binding to EGFR ectodomain in colorectal cancer treatment (Montagut *et al.*, 2012; Zhang *et al.*, 2010).

4.2 Mutations in the juxtamembrane region (JM)

The juxtamembrane region (JM) domain plays role in promoting allosteric activation of the acceptor kinase domain for enhancing the formation of asymmetric dimer. However, many rare mutations in this domain have been identified in NSCLC patients such as V689M and L703F (Figure 17). These mutants make constitutively active and stabilizing acceptor/donor interactions. V689M mutant promotes cellular transformation *in vitro* and tumorigenic in a xenograft assay. Red Brewer and colleagues (2009) presented the results form alanine scanning mutagenesis that the deletion of residues in juxtamembrane region (JM) 645-662 or 645-676 can reduce the autophosphorylation in intracellular domain of cell, suggesting the JM region can contribute to tyrosine kinase domain activation (Red Brewer *et al.*, 2009; Zhang *et al.*, 2010).

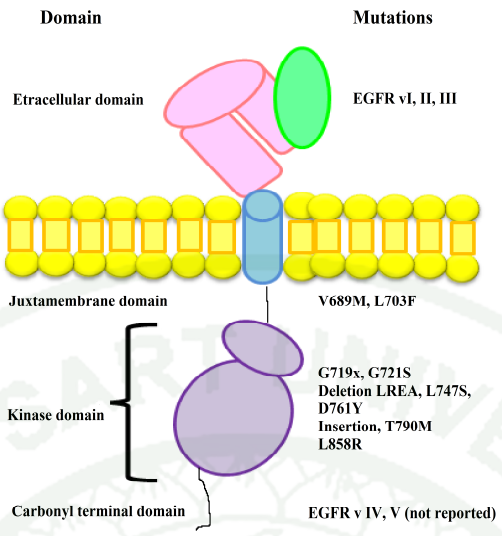


Figure 17 The positions of key EGFR mutant and variants.

Source: Modified from Zhang *et al.* (2010)

4.3 Mutations in the cytoplasmic domain

As mentioned the large portion of cytoplasmic EGFR domain, however, the mutations in the cytoplasmic domain have been investigated less than the extracellular domain (Voldborg *et al.*, 1997). Mutations in EGFR can find in exon 19 to 21 that cover most of tyrosine kinase domain of EGFR (Mir *et al.*, 2008). In tyrosine kinase domain, the mutations can be identified into four regions including an in-frame deletions in exon 19, substitutions in exon 21, the single-nucleotide substitutions (amino acid change) in exon 18 and an in-frame duplication/insertion in exon 20 (Wu *et al.*, 2008) (Figure 17). First mutant type is small in-frame deletions found around the conserved LREA motif of exon 19 (residues L747-A750, included the 24 residues signal sequence) (Mu *et al.*, 2005). Second type from single point mutation including L858R in exon 21, it is the most common mutation replaced by a large polar arginine. This point mutation is aligned in activation loop able to destabilize the inactive conformation into active EGFR-TK conformation (Kawahara *et al.*, 2011). Third, the position of G719 located on the glycine rich P-loop at exon 18 is found less frequency than L858R and can be mutated into serine, alanine or

cystein. The mutated G719S can lead to constitutively active state of EGFR (Yun *et al.*, 2007). Fourth type, the mutation in exon 20 can be in-frame duplication/insertion such as T790M, which amino acid alteration is acquired resistance to EGFR inhibitors. However, the mutations in tyrosine kinase domain are constitutively active and favor tumorigenicity (Balak *et al.*, 2006; Costa *et al.*, 2007). In part of the carboxyl terminal region in gliomas, two forms of deletion mutants have been reported as EGFR type vIV harbors an in-frame deletion and EGFR type vV has a carboxyl terminal truncation (Nicholas *et al.*, 2006). Deletions started at the same point but the size of the deletions varied from 254 bases to a premature termination of the transcript, resulting in the truncation of the mRNA. These mutants seem to be constitutively active (Balak *et al.*, 2006). However, the L858R in exon 21, G719S in exon18 of point mutations, as well as the exon 19 deletions and T790M in exon 20 of insertions are transformation of developing and progressing in human malignancies (Voldborg *et al.*, 1997).

5. EGFR inhibitions and inhibitors

Because EGFR activity is key cellular programs such as cell proliferation and cell required for normall cell growth, which the signaling of epidermal growth factor receptor (EGFR) must be regulated tightly by the endogenous proteins in signaling networks. Even though in cancer cell, these cellular responds are also needed in the over-expression state, while the protein inhibitors in the network can dysregulate cancer cell. Therefore, the synthesized inhibitors have been widely developed for the potency of suppressing of the abnormal signaling. However, both of the synthesized inhibitors and endogenous proteins inhibitors were detaied as below.

5.1 EGFR inhibitors in clinical utilities

Inhibitors of the EGFR have been improved to be a good anticancer agents and two therapeutic strategies have been developed are Monoclonal antibodies to target at the ligand or part of extracellular domain of the receptor, which molecule inhibitors can block

two EGFR receptors to prevent dimerization form and inhibit not only the phosphorylation but also the activation of downstream signaling pathways (Figure 18) (Okines *et al.*, 2011).

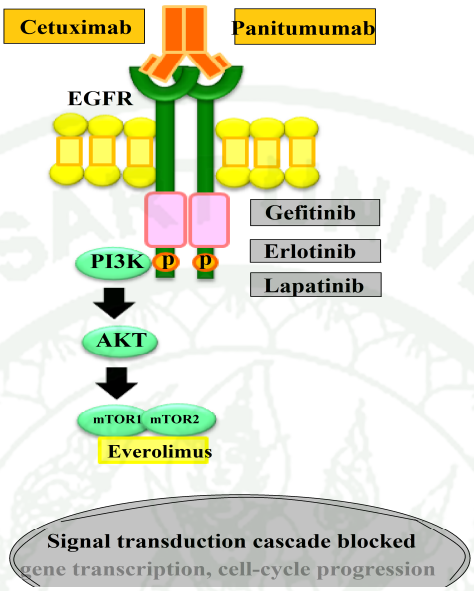


Figure 18 The overall working of monoclonal antibody inhibitors (Cetuximab) and Tyrosine Kinase Inhibitors (TKIs: Genfitinib and Erlotinib) of EGFR.

Source: Modified from Okines *et al.* (2011)

5.1.1 Epidermal growth factor receptor-targeted monoclonal antibodies or Monoclonal Antibodies (MoAbs). It can interact specifically in part of the extracellular EGFR domain and competitively with ligand for binding (Figure 19). Cetuximab is chimeric monoclonal immunoglobulin G1 antibody (Goldstein *et al.*, 1995; Okines *et al.*, 2011). It has been used for the treatment of metastatic colorectal cancer and head and neck cancer by combination with chemotherapy. It was approved by the FDA for advanced-stage in the treatment of KRAS wild-type colon cancer. Antigen binding fragment (Fab) part of cetuximab can bind with high affinity to ectodomain at domain III (L2) for directly blockage ligand induced EGFR dimerization. Morover, the domain I and III of extracellular domain cannot adopt conformation for engagement the ligand. Li and his colleagues (2005) described the X-ray structure of the antigen-binding (F_{ab}) fragment from cetuximab binding

with the soluble extracellular domain III of EGFR (sEGFR) that the side chain hydroxyl of F_{ab} at Y102 can construct hydrogen bonds to glutamine (Q384 and Q408). These hydrogen bond formations cannot observe in EGF-bound EGFR while only salt bridge formation detected between arginine (R41) of EGF with aspartate (D355). As high affinity of cetuximab binding to domain III than EGF ligand, the double mutant Q408M/H409E can disturb inhibitors binding higher than EGF ligand with a 150-fold reduction rate in cetuximab and only 3.5-fold in EGF.

5.1.2 Small-Molecule Tyrosine Kinase Inhibitors

As seen in Figure 19, tyrosine kinase inhibitors can act at the intracellular domain of the receptor that can be competitively with ATP for the intra-cellular catalytic site of EGFR. Tyrosine kinase inhibitors are able to inhibit tyrosine auto-phosphorylation leading to inhibiting EGFR signal transduction pathways. Two small-molecule inhibitors of EGFR/ErbB2 have been approved by the FDA including erlotinib (Tarceva) and gefitinib (Iressa) (Hirsch and Bunn, 2005; Yun *et al.*, 2007).

Erlotinib (trade name Tarceva) is oral ATP-competitive EGFR tyrosine kinase targeted inhibitors used for non-small cell lung cancer (NSCLC), pancreatic cancer and other cancers (Figure 19A). The selectively and reversibly of erlotinib can suppress the tyrosine kinase domain at ATP binding site of EGFR leading to reducing EGFR autophosphorylation in tumor cells, resulting in inhibition of EGFR-dependent cell proliferation and also blocking G1 phase in cell-cycle pro-gresssion. Nevertheless, the kinase domain can display the mutation in the majority of the tumors, which can response to erlotinib after 8-12 months of starting treatment. The second mutation of T790M is involved in inhibitors resistance at the ATP binding pocket of EGFR kinase, the increasing in ATP binding affinity of the L858R and reducing the inhibitory effect of erlotinib can be occurred similarly to the gefitinib resistance (Yun *et al.*, 2008).

Gefitinib (trade name Iressa) is first selective oral EGFR inhibitors used for lung and other cancers. Gefitinib is quinazoline-derivative molecule by competetively

to ATP binding site in tyrosine kinase domain (Figure 19B), resulting in blocking EGFR signal transduction, reducing of cell proliferation, inducing of cell cycle arrest, increasing of apoptosis and enhancing anti-angiogenetic effects (Sordella *et al.*, 2004). The EGFR gene point mutations including the position of G719S (Figure 19C) or L858R has been identified in the patients with non-small-cell lung cancer while L858R is found the most frequently point mutation. These patients can response to gefitinib inhibitors. However, the patient who is in an advanced non-small-cell lung cancer stage with EGFR-mutant and initially responds to gefitinib, the subsequently regressive can occure two years of starting treatment. It has been reported that the tumor biopsy specimen of patient can be detected the second point mutation in EGFR gene at position790 form threonine to methionine. The double mutations of G719S/T790M and L858R/T790M affect to a dramatic resistance to this inhibitor. The T790M mutant is not found in primary stage of non-small-cell lung cancer, Toyooka and his colleagues (2005) reported that a point mutation in T790M is aggressive with the gefitinib-treated patient (Kobayashi *et al.*, 2005; Sordella *et al.*, 2004; Toyooka *et al.*, 2005; Yoshikawa *et al.*, 2013). However in drug resistance phenomenon, the MD simulations were applied for interacting investigation by docking gefitinib into the crystal structure wild-type of EGFR (Figure 19B) and studying the effect of gefitinib in the EGFR T790M and L858R (Figure 19D) mutations. After the last 9 ns of MD simulation, gefitinib was moved closer to the hinge region in mutation L858R and escaped from the binding pocket in T790M mutant to similar with the observation in double mutant T790M+L858R. Gifitinib stopped to escape from the binding pocket after backward from M790 into the position of wild-type EGFR. From free energies computing of binding revealed that L858R mutation increased gefitinib binding affinity corresponding to experimental study, gefitinib and erotinib were about 10-fold more potent in the L858R than the wild-type EGFR (Liu *et al*., 2006). Thus, because of more effective in first line of gefitinib than second line, the potency in single mutant more wild-type and undetection of second mutation in primary NSCLC suggesting that gefitinib should be used for first-line treatment with the NSCLC and the sensitive EGFR mutations. If the patients receive in second or third-line treatment, they may miss to treatment with the high potency of gefitinib due to a rapidly progressive cancers (Argiris and Mittal, 2004; Lynch *et al.*, 2004; Maemondo *et al.*, 2010).

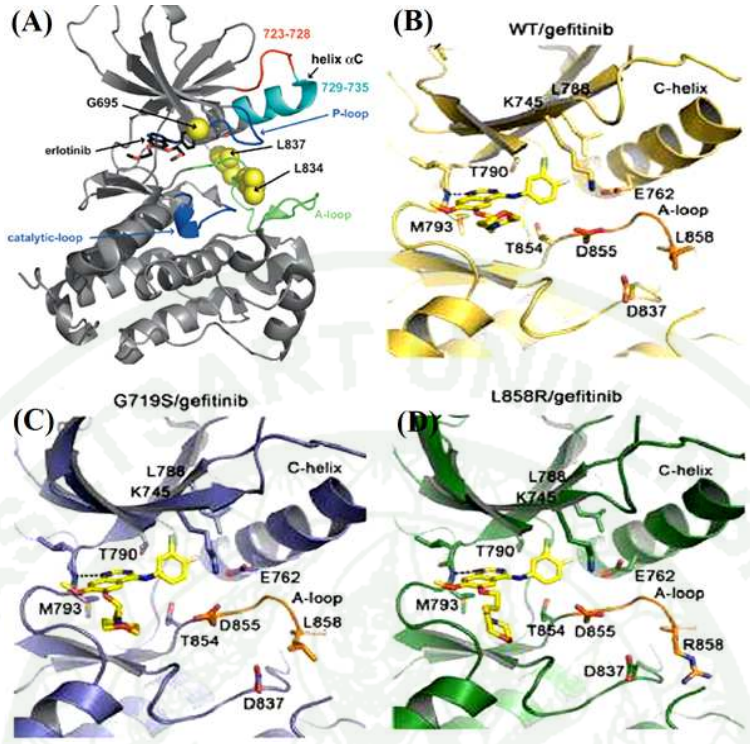


Figure 19 TKIs binding mode with wide-type and mutants EGFR kinase domain. (A) Erlotinib bound to active EGFR tyrosine kinase (PDB: 1M17). Erlotinib (black strick) locates in the ATP- binding site while G695, L837, and L834 are mutated in NSCLC (yellow). (B-D) Gefitinib bound to active EGFR tyrosine kinase wide-type and mutants. Gefitinib (yellow stick) locates in the ATP-binding site with L858R and G719S mutant represented in B and D, respectively.

Source: Modified from Choi *et al.* (2007); Yun *et al.* (2008)

5.2 Negative regulators/inhibitors of RTK signaling pathways.

EGFR can control cellular response and biological process by various signaling creations. The production of signaling has to appropriate strength and duration while the abnormal or excessive signaling involving in oncogenic threat (Sibilia *et al.*, 2007). EGFR is regulated by receptor itself and also endogenous protein inhibitors including positive and

negative feedback regulators. Negative feedback inhibitor such as LRIG1, SOCSs and MIG-6 can bind EGFR and suppress signaling in many mechanisms including catalytic inhibition and downregulation of receptor (Segatto *et al*., 2011).

5.2.1 Leucine- rich repeats and immunoglobulin- like domains 1 (*LRIG1*)

LRIG1 is newly discovered as negative feedback regulator of EGFR, which human LRIG family can be comprised into 3 members: LRIG1 (chromosomes 3p14.3), LRIG2 (chromosomes 1p13) and LRIG3 (chromosome 12q13). The structure of LRIG consisted of the extracellular part (a leucine-rich repeat (LRR) domain and three immune-globulin-like domains), a transmembrane region and a cytoplasmic tail (Figure 20A) (Nilsson *et al.*, 2001; Rubin *et al.*, 2005). *LRIG1* is defined as tumor suppressor gene. It can regulate either promoting or suppressing human cancer due to it is not down-regulated in all tumor types. *LRIG1* mRNA expression is induced by certain growth factors such as EGF and androgen. The translated LRIG1 can drive EGFR to degradation in lysosome by three steps (Figure 20B) (i) LRIG1 bind directly to EGFR inactive monomer in a ligand-independent fashion through its extracellular domain and immunoglobulin-like domains binding with extracellular region of EGFR. This mechanism directs the EGFR into lysosome. (ii) LRIG1 is released into extracellular space by ADAM (A Disintegrin And Metalloproteinase) cleavage and then it can act as a high-affinity competitive inhibitor to EGF ligand for EGFR binding. (iii) LRIG1-bound EGFR in active state, followed by the recruitment of E3 ubiquitin ligase called Cbl leading to the activation of EGFR ubiquitylation, internalisation and lysosomal degradation. However, Cbl is recruited in both LRIG1 independent and dependent fashions at the phosphorylated tyrosine residues (Segatto *et al.*, 2011).

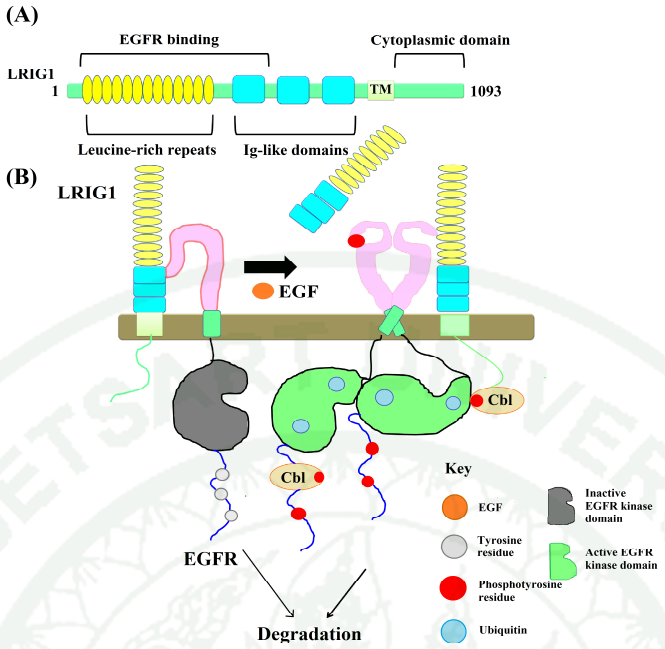


Figure 20 The structure and lysosomal degradation function of human LRIG1. (A) The structure consists of the extracellular part including a leucine-rich repeat (LRR) domain and three immunoglobulin-like domains, followed by a transmembrane region and a cytoplasmic tail. (B) LRIG1–bound EGFR undergoes to degradation in lysosome.

Source: Modified from Segatto *et al.* (2011)

5.2.2 Suppressor of cytokine signaling (SOCS)

SOCS proteins are family of intracellular protein identified as feedback inhibitors of cytokine signaling pathway including immune system. SOCS proteins stucture comprise of N terminal region kinase inhibitory region (KIR) for inhibiting kinase activity, the center Src homology-2 (SH2) domain for binding with specific phosphotyrosine residue and the last part SOCS box at C terminal region intact with ubiquitinating machinery (Figure 20A) (Akhtar and Benveniste, 2011). Moreover, some member of SOCS can feedback inhibitors to EGFR such as SOCS1 and SOC3 proteins that are negative feedback regulators related to cytokine signaling pathways through Jak-Stat signaling cascade. Many

ligands including interleukins, interferons and growth factor can simulate the signaling through Jak-Stat pathway. Among these stimuli, growth factor can activate EGFR resulting in the simulation of that pathway (Rawlings *et al.*, 2004). Furthermore, SOCS3, SOCS6, SOCS5 are inducible expressed in response to EGF, which these SOCS proteins can bind to the autophosphorylated EGFR by their Src homology 2 (SH2) domains and recruit E3 ubiquitin ligase complex by SOCS domains leading to the ubiquitinylation and degradation of EGFR (Figure 20B) (Yoshimura *et al.*, 2007).

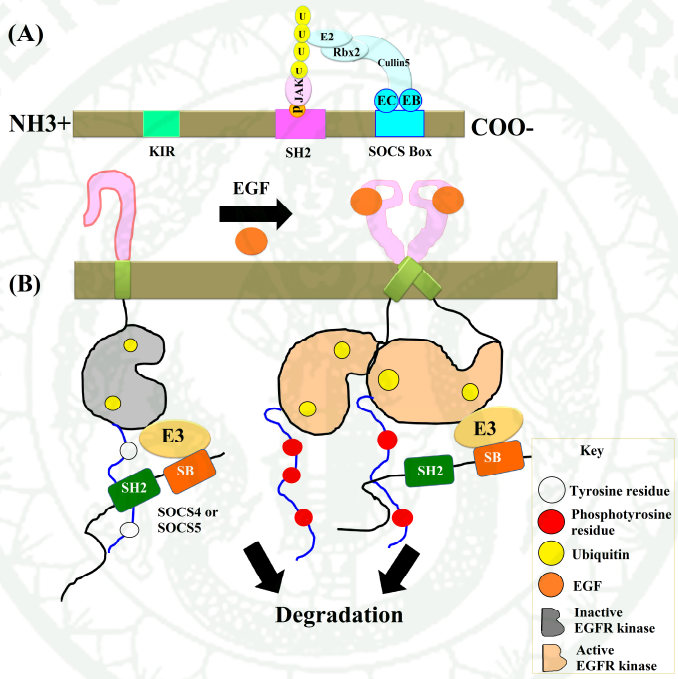


Figure 21 Structure and functions of suppressor of cytokine signaling (SOCS). (A) All SOCS proteins consist of kinase inhibitory region (KIR) to inhibit kinase activity, Src homology-2 (SH2) domain to bind with specific phosphotyrosine residue, and SOCS box intact with ubiquitinating machinery. (B) SOCS4 and 5 bind to EGFR in ligand- independent fashion to enhance EGFR tagging with ubiquitin, routing into MVBs, and fusing with lysosome.

Source: Modified from Akhtar and Benveniste (2011); Segatto *et al.* (2011)

5.3.3 The Mitogen-Inducible Gene-6 (*Mig-6;* known as receptor- associated late transducer or *RALT*)

MIG-6 is endogenous inhibitor protein encoded as gene-33, which is located on human chromosome 1p36 and the size of ~50 kDa. The isoforms of MIG-6 are identified into two isoforms including long and short isoforms. While long isoform is active conformation, inactive in short isoform is caused from alternative splicing process. In addition, the conformation of short isoform, amino acids of 67-142 between CRIB and SH3 binding regions are deleted and required for the stress-activated protein kinases (SAPKs) activation thus the isoform of MIG-6 may key for controlling the expression of *Mig-6* (Chrapkiewicz *et al.*, 1989; Makkinje *et al.*, 2000). MIG-6 protein in long isoform contains 459 amino acids, which there are 77 amino residues (amino acid 336 to 412) important for EGFR inhibition (Xu *et al.*, 2005; Zhang and Vande Woude, 2007). MIG-6 can be recruited to bind the EGFR kinase domain in a presence of ligand through Ras-MEK-ERK pathway (Figure 22B). The over-expression of MIG-6 can prevent asymmetric dimer leading to the inhibition of autophosphorylation, downstream signaling and then susbsequent to undergoing the lysosomal degradation (Zhang *et al*., 2007).

The structure of MIG-6 negative feedback inhibitor of EGFR contains the important protein-protein interaction domain/motifs (Figure 22A) such as NH_2-terminal Cdc42/Rac interaction and binding (CRIB) domain, a 14-3-3 protein binding motif, a Src-homology-3 (SH3) domain binding motif, PDZ binding domain and PEST sequences. Two PEST sequences involve in ubiquitination and proteosome degradations (Fiorini *et al.*, 2002). PDZ binding domain related to the assembly of ion channels, signaling proteins and cell junction (Fanning and Anderson, 1999; Shoelson, 1997). The COOH-terminal motif of MIG-6 is homologous to Ack-1 non-receptor identified as EGFR binding domain (EBR) (Xu *et al.*, 2005). The mechanism of MIG-6 is associated with the binding of ERB region to the C-lobe surface and the catalytic domain of the activated EGFR kinase domain (Figure 22C). The result of binding can block the activated EGFR kinase in asymmetric dimer activation and also lock the catalytic site into inactive configuration, in order to prevent the signaling production (Zhang *et al*., 2007). After MIG-6-bound the activated

EGFR, the EGFR molecule is downregulated and undergone into the lysosomal for degradation, which EGFR has own endoytosis pathway and endocytic machinery for sorting into mutivesicular bodies (MVBs) (Sorkin and Goh, 2008). Ralt endocytic domain (RED) of MIG-6 can bind to AP-2 protein adaptor for sorting MIG-6/EGFR complex into the vesicle of clathrin-coated pits. Moreover, intersectins are SH3-domain-containing accessory protein participated with AP-2 adaptor for sorting and involved in the mutation of clathrin-coated pits (Segatto *et al.*, 2011). In late endosomes, EGFR is carried by SNARE protein synataxin 8 trafficking from early to late endosome in the absent of ubiquitylation (Ying *et al.*, 2010) .

Notably, MIG-6 protein contains 77 amino responsible for EGFR inhibition. This can be calssified into two segments, segment1 (336-364) and segment 2 (365-412) (Wang *et al.*, 2013). Segment 1 can extend its segment across the C-lobe of kinase domain to prevent asymmetric dimer formation with the N-lobe of partner and inhibit kinase activity of dimer, while the basal activity of monomer in solution cannot be reduced. It is appreciated that, residues of EGFR bound with MIG-6 segment 1 on interface are conserved, which can be implied that MIG-6 may bind with other EGFR members (Zhang *et al.*, 2007). Other segments, the C-terminal region of segment1 (amino acid residues 365-412) is denoted as segment 2, which is uncertain structure. The interaction of segment 2 is closely to binding site of peptide kinase inhibitor and its sequence similarity to region in ACK1 (Wang *et al.*, 2013). Thus, segment 2 is responsible for the inhibiting of the activated EGFR kinase domain and resulting in kinase activity inhibition (Zhang *et al.*, 2007). After MIG-6 binding, EGFR is adopted the conformation from CDK/cyclin-like active structure to the CDK/Src-like inactive structure (Shen *et al.*, 2007). However, the over-expression of EGFR mutations in L858R and exon19 deletion (EGFR746-750) can resistant to MIG-6 inhibition as observed in lung cancer cell line, suggesting that the loss of MIG-6 promotes tumor formation (Bose and Zhang, 2009; Wang *et al.*, 2013). All the feedback inhibitors described might be useful as cancer biomarkers to predict the efficacy of targeting drugs because the EGFR family members involve in cancer manner. Moreover, small molecules or peptides can enhance the potent of inhibitors would possibly to be act

novel drugs. Although, the physiological and pathological in these inhibitors have to be validated and elucidated in advance personalized medicine.

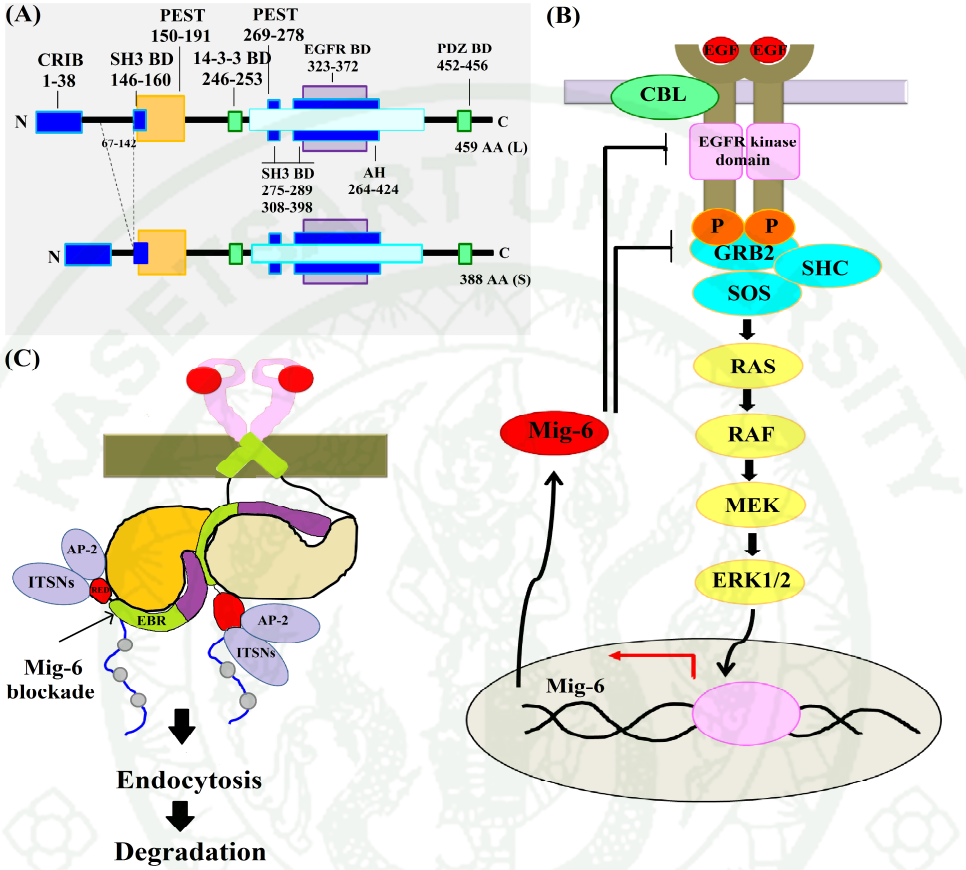


Figure 22 The overview of Mitogen-Inducible Gene-6 (MIG-6). (A) The structure of MIG-6 in long and short isoforms. (B) The MIG-6 regulates EGFR signaling by negative feedback loop. (C) The mechanisms of EGFR inhibition by MIG-6 segment 1 (green) and segment 2 (purple) participated with accessory adaptor proteins bound RED of MIG-6 for sorting promotion.

Source: Modified from Xu *et al.* (2005); Segatto *et al.* (2011); Zhang and Vande Woude (2013)

6. Intermolecular interactions in biomolecular system

The intermolecular interactions are crucial for cellular activities such as cell development, communication, operation and regulation (Dong *et al*., 2008b). Intermolecular interactions including van der Waals, hydrogen bonding, ionic attractions and electrostatics contributions are a non-covalent interaction that presents the driving forces in interactions between biomolecules (Telser, 2002). These interactions plays role in the adhesion, immune defense, complexation, signal transduction and regulation systems (Schreiber and Fleishman, 2013). For investigation these interactions, computational methodology was applied for monitoring the phenomenon of the interactions and also free energy of intermolecular interactions (Foloppe and Hubbard, 2006). According to previous studies, it was used for extending into the detailed non-covalent interactions between protein-protein, protein-peptide, protein-ligand and protein-DNA on the atomic scale (Wan *et al.*, 2007; Yan and Wang, 2013). This information gives rise to better understanding how their functions that can be taken this advantage in order to develop or design new inhibitors.

6.1 Electrostatic interactions

Electrostatic interactions are also called charge-charge interactions essential for determining the thermodynamics of binding and also the overall stability of protein-protein complexes (Perez-Jimenez *et al.*, 2004). These interactions can be either the attractive or repulsive forces depended on the signs of two charge molecules, *Qi* and *Qj*. The force is attractive and can move the molecules forward each other if *Qi* and *Qj* have opposite sign (+,-) or (-, +). In contrast, the force is repulsive and can push the molecules away each other if the charges have the same sign (-,-) or (+, +). The strength of forces influenced on the difference of two charges (*Qi* and *Qj*), their distance (*rij*) and dielectric constant (ε) of the medium. If two charges are opposite signs and resulted in attractive each other, the energy is decreased and implied that the interaction is favorable. Whenever the distance (*rij*) is increased, the interaction is less favorable with 1/rij. The strength of electrostatic interaction is decreased when two charge molecules are changed from vacuum into higher

dielectric constant (ε) of medium (solvent). Therefore, the electrostatic interaction has long-range interaction potential. Electrostatic free energy can be described by Coulomb's law (eq 1). Coulomb's law is a model for explained the strength of forces between two charged particles, which r_{ij} is the distance between charge Q_i and Q_j and ε_p is the dielectric coefficient of the solute (Dong *et al*., 2008a). Furthermore, the total electrostatic free energy (eq 2) can be decomposed into the favorable charge-charge interaction (denoted as Coulomb's law; E_{coul}) in a homogeneous medium (vacuum) with solute dielectric constant and the energy penalty associated with moving charge from vacuum to its position in the solution (homogeneous to inhomogeneous) (Sheinerman and Honig, 2002).

$$G_{coul} = \sum_{ij} \frac{1}{4\pi\varepsilon 0\varepsilon p} . \frac{QiQj}{rij} \quad (1)$$

$$G_{pol,ele} = E_{coul} + \Delta G_{pol,\ solv} \quad (2)$$

In addition, the interactions between two amino acids with oppositely charged side-chains are also called "Salt Bridge". It is one type of hydrogen-bonded ion pais interactions between two groups of oppositly charged residues that close each other by electrostatic attraction in order to contribute the protein structure. Salt bridge is consisted both of electrostatic interaction and special form of hydrogen bond, which the oppositely charged residues can be considered into positive charges from Lys, Arg and His with negative charges from Asp or Glu (Kumar and Nussinov, 2002). A fully charged of ion pair residues lead to higher the bonding energy than that of a hydrogen bond, which only one has fully charged or both resides have uncharged side-chains (Petsko and Ringe, 2004). The energies of surface and buried salt-bridge are found to contribute 1–5 kcal.mol^{-1} (Horovitz, 1996). As seen in Figure 27, the salt bridge can form between the side-chain of Arg57 and point mutation of HIV-1 protease at Glu35 (Meiselbach *et al.*, 2007). Because the side-chains of these residues are pH-dependent thus the free energy contributions of salt bridges are also respected to pH. The total electrostatic free energy of a salt bridge can be divided into three contributions including coulombic (charge-charge) interactions, the interactions of charge with permanent dipoles and the desolvation of charge. The strength

of the coulombic interaction is depended on pH, the geometry of salt-bridge and charge-charge interacting distance. Coulombic interaction between the charges is energetically favorable opposed by unfavorable desolvation of charge. Salt bridge is classified by the location and number of interacting charges. Some salt bridge can be exposed on protein surface encountering with an aqueous environment, while some bridge is buried in protein interior (Figure 23) (Bosshard *et al.*, 2004).

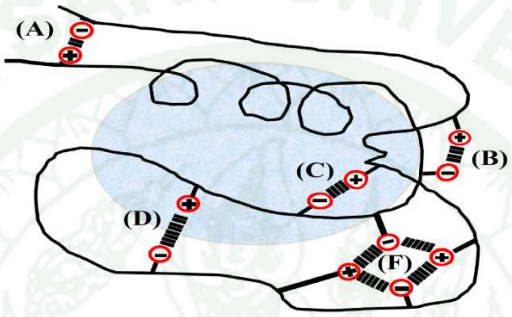


Figure 23 Classification of salt bridges in a protein (a heavy line). A and B represent the exposed salt bridges on protein surface in a high dielectric environment or their completely buried in the protein interior in a low dielectric environment (C), or half-buried (D). Moreover, salt bridges are formed by only two opposite charge (A-D), but several charges may create network in a complex salt bridge (F).

Source: Modified from Bosshard *et al.* (2004)

6.3 Van der Waals interactions

Van der Waals forces are in a sense electrostatic nature that important for hetero-protein and protein-ligand interactions. When any two atoms are closed each other, they can create a weak attractive and weak repulsive force between atoms, molecule or surface. Van der Waals forces can occur in both polar and nonpolar molecules with the interaction energy about 1 kcal.mol^{-1}. Van der Waals are consisted of dipole-dipole interactions (Keesom forces), dipole-induced dipole interaction (Debye force) and induced dipole-induced dipole forces (London dispersion forces) as shown in Figure 24A, 24B and

24C, repectively. Dipole-dipole interactions are electrostatic origin from permanent dipoles in two neutral polar molecules that associated with electronegative atoms. The forces of interactions are proportional to $1=r^{3}$**.** For Debye force, it is dipole-induced dipole interaction, which the permanent dipole come form the polar molecules that can induce the dipole moment in second atom (non polar molecule lacking permanent dipole). The force of interaction is proportional to $1=r^{5}$. For London dispersion forces, it is the induced dipole-induced dipole forces created by a temporary change in dipole moment, which results from a fluctuation of electrons cloud in one side of one atom or molecule. The temporary dipole molecule can distort a neighboring atom and induce that atom having partial charge by the electrostatic forces. The side of electron clouds has partial negative charge (δ^{-}) and another side no electron density has partial positive charge (δ^{+}) in moment. This phenomenon is called an instantaneous dipole. The opposite partial charges between one atom and a neighboring atom cause attractive each other for short time. London dispersion forces are extremely short-range in action (depending on $1=r^{6}$).

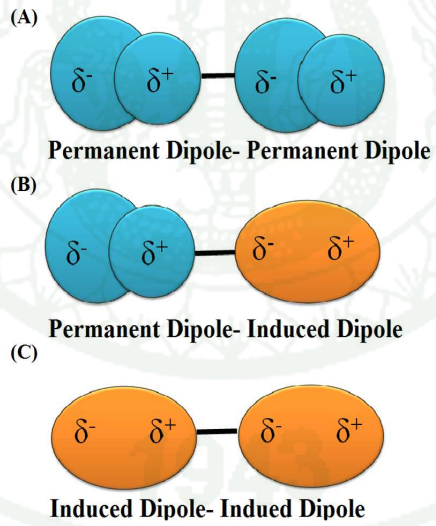


Figure 24 Interactions between molecules–temporary and permanent dipoles. Permanent Dipole-Permanent Dipole forces (A). Permanent Dipole-Induced Dipole (B). Dipole-Induced Dipole (C).

Source: Modified from Leite *et al.* (2012)

6.5 Hydrogen bonding

Hydrogen bond is special case of dipole forces and the attractive interaction between two electronegative atoms, donor and acceptor through the intermediate of hydrogen atom, which a particularly strong dipole-dipole attractions (Figure 25). The strength of the hydrogen bonds in proteins are weak about 1-2 kcal.mol^{-1} as seen in N-H..O (1.9 kcal.mol^{-1}) for contributing to protein-protein associations (Telser, 2002). However, the strength of hydrogen bond can reach to 40 kcal.mol^{-1}, is depended on the types of donor-acceptor pair atom, angle and length (Hubbard and Kamran Haider, 2001; Saggu *et al.*, 2012). The folding of hydrogen bond is stronger than a van der Waals, but weaker than covalent and ionic interactions. The definition of bonding is the forming of the interaction between the partners D^{δ}-$H^{\delta+}$ and $A^{\delta-}$, which $D^{\delta-}$ and $A^{\delta-}$ represents atoms of higher electro-negativity atoms than hydrogen atom such as C, N, P, O, S, F, Cl, Br or I. Inaddition, a highly electronegativity is bound with hydrogen atom (H) called hydrogen bond donor ($D^{\delta-}$) and hydrogen atom can create the hydrogen bond with others highly electronegativity atom called hydrogen bond acceptor ($A^{\delta-}$) (Grabowski, 2006). However, hydrogen bonds between O-H..O, N-H..O, O-H..N and also N-H..N may be stronger than C-H..O, C-H..N, O-H..π, NH..π andC-H..π. In part of the geometrical criteria, a detailed analysis of hydrogen bond is shown in Figure 25, which *d* distance between donor and acceptor; θ angle between acceptor-donor-acceptor; *r* distance between hydrogen and acceptor; γ angle between hydrogen-acceptor-acceptor antecedent (Baker and Hubbard, 1984; Xu *et al*., 1997; Torshin *et al*., 2002; Sarkhel and Desiraju, 2004; Bertolasi *et al*., 2011). Because hydrogen bonds are long-range interactions, thus donor group ($D^{\delta-}$-$H^{\delta+}$) can form more than one acceptor ($A^{\delta-}$) at the same time. Hydrogen atom can form with two acceptors called “Bifurcated donor”, while two hydrogen atoms can form with one donor called “Bifurcated acceptor” (Figure 26) (Steiner, 2002; Tiwari and Panigrahi, 2007).

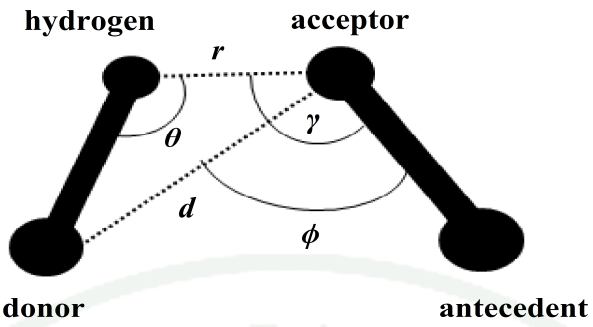


Figure 25 The geometry of hydrogen bonding analysis.

Source: Modified from Xu *et al*. (1997)

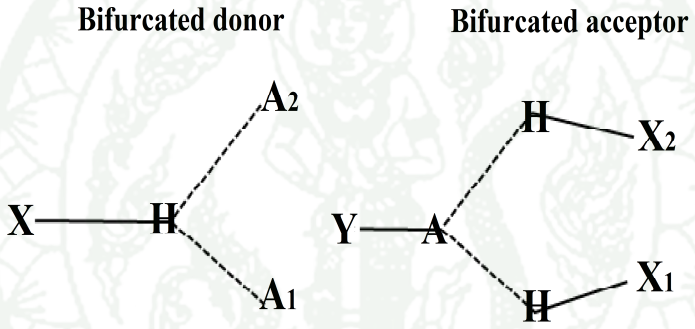


Figure 26 The bifurcated hydrogen bond types. .

Source: Modified from Steiner (2002)

6.6 Hydrophobic interactions

Hydrophobic interactions (literally, “water-hating”) are one of important driving force in protein structure stabilizing that the side-chain of non-polar amino acids are made up of long chain hydrocarbon atoms (Matthews, 2001). These non-polar side-chains of amino acids are “oily” characteristic that prefer to be segregated in contact with each other and dislike interacting with water molecules as opposed to hydrophilic character of the charged side-chain (Figure 27). The hydrophobic effect represents the effect of hydrophobic substances or non-polar atoms dissolving in water molecules, which the effect

is resulted from the preclusion of highly dynamic hydrogen bonds network between molecules of liquid water (David, 2005). The hydrogen bonding network of water molecules reduce the entropy of system, making energetically unfavorable (Yokota *et al.*, 2003). Thus, the disruption of hydrogen bond formation between water molecules is the energetically favorable to maintain protein folding and stability (Kratochvil *et al.*, 1972; Matthews, 2001) .

Hydrophobic cores and hydrophobic interactions in proteins are not only important for protein folding but also essential for protein-protein and protein-ligand binding. In protein-protein binding, the interfaces between proteins are high hydrophobic than the solvated surface area and can contribute to complex formation. (Jones and Thornton, 1996; Matthews, 2001).

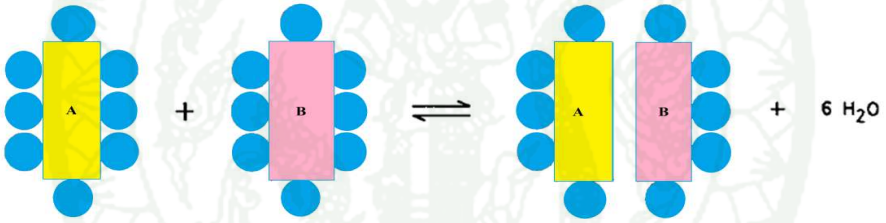


Figure 27 The formation of hydrophobic interaction between molecule A and B leading to the disrupting and losing of hydrogen bonds in water molecules network. Blue circle represents water molecule.

Source: Modified from Kratochvil *et al.* (1972)

7. Modeling of intra- and intermolecular forces

Hence, intermolecular interactions including van der Waals, hydrogen bonds and salt bridge are electrostatic in origin. Computational simulations have been developed in order to model biological systems of interest and study the interactions occurred in biomolecular systems.

7.1 Modeling molecular solvation

It is well-known that, the mechanisms of life are depended on water at the macroscopic and molecular levels, which plays important role in a variety of cellular functions, being the solvent of most biological molecules, helping an enzymatic catalysis, and a building block of macromolecules. Water is the environment effect on the conformational stability and protein folding through the interactions between the solvent-protein and solvent-solvent, the classifying main characteristics of the real physical system (a protein in solution) with inexpensive computer are challenged in explaination how the intermolecular interactions. Thus, the various effects from intermolecular interactions may be replaced by suitable averages. Nevertheless, water can be modeled by two main approaches ways as explicit and implicit solvent models (Figure 28) (Drost-Hansen, 1969; Levy and Onuchic, 2006).

7.1.1 Explicit solvation model

Firstly, explicit solvent is biomolecular solvation applied to explain the interaction between ions, solvent and solute atoms in terms of non-bonding interactions (van der Waals; E_{vdW} and coulombic interaction; E_{ele}) and bonding interaction (angles, bonds, torsion), which are calculated by molecular mechanics force fields to approximate quantum mechanical energies of these terms (Ponder and Case, 2003). The explicit solvent model treats a solvent by using thousands of water molecules and salt ions (Figure 28) (Bizzarri and Cannistraro, 2002). Because solvent molecules can be directly involved in a solute that give rise physically to more correct, thus the molecular details of each solvent molecule are obtained.

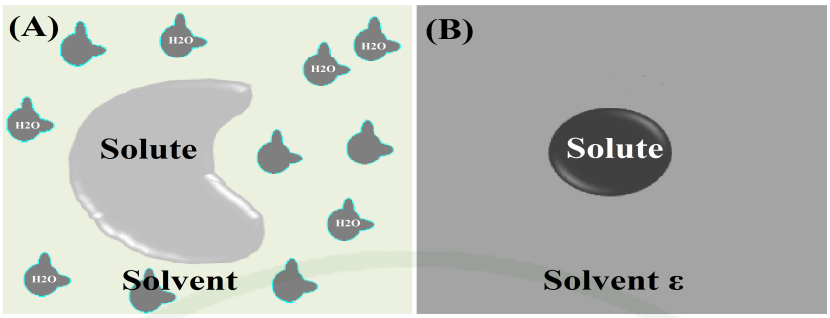


Figure 28 A schemic presentation of two solvent models. (A) The explicit solvent models, the solute is surrounded with water molecule. (B) The implicit solvent models, the solute is surrounded with the average effect of real water and water is entered only in term of dielectric permittivity (ε).

Source: Modified from Dong *et al.* (2008a); Lima *et al.* (2009)

There are a large variety of explicit water models including: (i) simple models (SPC, SPC/E, BF, TIP3P or ST2) that water molecules are rigid and rely on non-bonded interactions. Electrostatic (E_{ele}) and van der Waals (E_{vdW}) interactions are modeled by Coulomb and Lennard-Jones potentials, respectively. (ii) Flexible models involve in the internal (bond stretching, angle bending) conformational change of water molecules and (iii) Polarisable water models are considered in the liquid phase under an ambient conditions involving in the contribution of the electrostatic field of ions and permanent charged groups (Levy and Gallicchio, 1998). According to the computational consuming of the last two types, thus three rigid models metioned in first type are most frequently applied; SPC, TIP3P and TIP4P (Mahoney and Jorgensen, 2000). Neuraminidase-inhibitor complex is one of the explicit solvation samples. Before entering into minimization and MD simulations, the enzyme-inhibitor complex was performed in a TIP3P water model of the explicit solvation. The particle-mesh Ewald (PME) method was used for long-range electrostatic interactions. In van der Waals calculation, Lennard-Jones interactions were used with cutoff of 8 Å. The SHAKE algorithm was applied to constrain all covalent bonds involving hydrogen (Bonnet and Bryce, 2004). Here again, the explicit solvation model can provide the most details of the interactions between the solute and solvent molecule

resulting in more accuracy, but it uses high computer consumption in calculations (Tanizaki and Feig, 2005; Tanner *et al.*, 2011). To solve computational cost from the explicit model, the calculation of the intermolecular interactions is replaced by the alternative way with suitable averages parameter as described below (Lima *et al.*, 2009).

7.1.2 Implicit solvation model

In contrast to the computationally-expensive of the explicit model, an implicit solvent method is lower computational-comsuming and become popular, although it is lower-accuracy (Dong *et al.*, 2008b). It represents a solvent as continuous medium having the average properties of its coordinates and momenta of the real solvent on the solute (Chen *et al.*, 2008). Water is entered in calculations by its bulk dielectric constant (ε), density and surface tension (Lima *et al.*, 2009). The solvation energy can be described by the interaction between solute-solvent, which the free energy of transferring the solute from a vacuum to the solvent environment (Figure 29).

The energy calculation of the implicit model is contributed the same as the explicit model but the solvation contribution (G_{solv}) is an extra terms of the force fields added up in the implicit model. In term of the solvation contribution, two basic models applied in the implicit solvent; continuum electrostatic models (PB/GB model) and approaches based on solvent accessible surface area (SASA) (Lima *et al.*, 2009).

The solvation energies are the combination of non-polar and polar solvation terms (Dong *et a*l., 2008a). (i). Non-polar solvation is calculated using method association with short-range repulsive interactions to make up the solvent cavity where the solute is located. The non-polar solvation is depends on solvent-accessible surface area (SASA); *A* through a surface tension; *γ*, the equation presented as below (Wagoner and Baker, 2006).

$$G_{\text{non-polar}} = \gamma.\text{A} \qquad (1)$$

Nevertheless, because short range interaction can large affect in non-polar solvation, the equation as above can be alternatively divided into a repulsive cavity term (ΔG_{cavity}) and an attractive solute-solvent dispersion interaction term (ΔG_{vdW}).

$$\Delta G_{non\text{-}polar} = \Delta G_{cavity} + \Delta E_{vdW} \quad (2)$$

Where ΔG_{cavity} is referred to the initial free energy of cavity creation in water molecules and ΔE_{vdW} is sign the free energy for establishing the solute-solvent vdW dispersion interactions. These terms of non-polar solvation have opposite signs (Levy *et al.*, 2003).

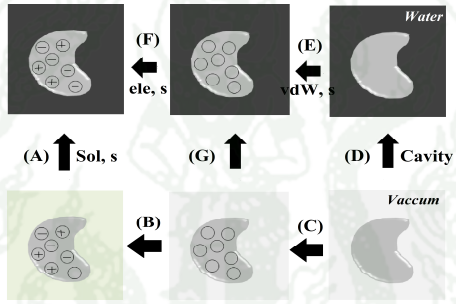


Figure 29 Thermodynamic cycle presenting the bimolecular solvation process. The overview steps of total solvation energy: charging the solute in vacuum (B) and solvent (F), the attractivesolute-solvent dispersion interaction in vacuum (C) and solvent (E) and the cavity creation involved in repulsive solute-solvent interactions (D). The solvation energy (A) involved in polar solvation energy (B, F) and nonpolar solvation (G). The nonpolar solvation energy (G) associated with van der Waal interaction between solute-solvent (C, E) and the cavity formation (D). The polar solvation energy originates from the difference of electrostatic free energy between in a vacuum, which is the homogenous of dielectric coefficient (B) and the in solvent, which is dielectric coefficient is in the inhomogenous medium (F).

Source: Modified from Chen *et al.* (2008); Dong *et al.*, (2008b)

(ii) In part of polar solvation, it is related to a difference in free energy of charge in vacuum and solvent conditions for monitoring the electrostatic free energy (Figure30). The implicit solvent models treat a solvent as a continuous dielectric medium that used for calculating the electrostatic contribution of the energetic effects of solvation. The electrostatic behavior is often identified by the properties of the system including the charges, radii and dielectric constant.

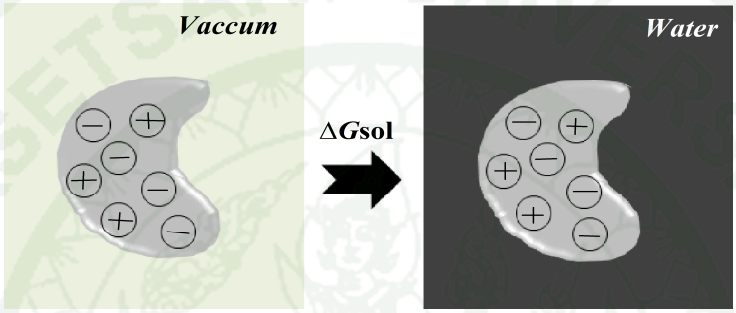


Figure 30 A schemic presentation of polar solvation free energy calculation (ΔG_{solv}). It obtained form transferring the charged biomolecule from vacuum (homogeneous medium) to solvent (inhomogeneous medium).

Source: Modified from Dong *et al*., (2008a)

The polar interaction in implicit is often used Poisson equation and computed by Poisson-Boltzmann (PB) and Generalized Born (GB) models. The solvent is presumed a continuum high-dielectric-constant medium, which responds to a low-dielectric constant of a solute (Wagoner and Baker, 2006). The polar solvation equation is presented as below (Dong *et al*., 2008a).

$$\Delta G_{\text{polar sol}} = G_{\text{ele (2)}} - G_{\text{ele (1)}} \quad (3)$$

The polar solvation free energy ($\Delta G_{\text{polar solv}}$) originates from the transferring the charged solute from homogenous medium to inhomogeneous medium (Figure 31), where the electrostatic free energy $G_{\text{ele (2)}}$ and $G_{\text{ele (1)}}$ are come from a

homogeneous medium and inhomogeneous medium, respectively (Figure 31). While the solute dielectric coefficient (ε_p) is equal to the dielectric coefficient constant (ε_w) of the system in homogenous medium (A), the solute has dielectric coefficient (ε_p) smaller than in the bulk solvent of inhomogeneous medium (B), that a dielectric constant (ε_w) increase from a value between 1 and 2 for the interior solute and 80 for the surrounding solvent (Tanizaki and Feig, 2005)

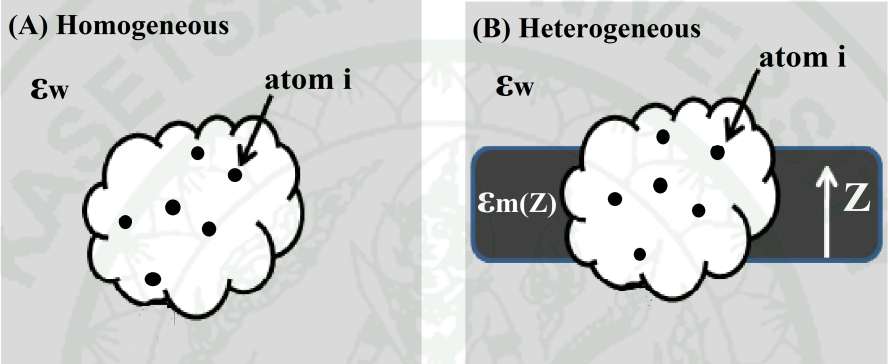


Figure 31 A schematic illustrating two different approaches to approximate the distributions of dielectric in the solute and solvent system. A homogenous standard implicit solvent model (A) has the dielectric coefficient constant of a protein and solvent (ε_p and ε_w), respectively. An inhomogenous solvent environment (B), the solute has dielectric coefficient smaller than in the bulk solvent. ε_m is dielectric constant of a membrane.

Source: Modified from Tanizaki and Feig (2005)

8. Applications

In previous part, we guide in some basic concept of intermolecular interaction and using the computational tool to simulate interactions in systems. In this part, we pointed out the application of these basically.

8.1 Binding free energy calculation

Molecular Machanics/Poisson-Boltzmann Surface Area (MM/PBSA) is a hybrid method that combine molecular mechanics (MM), continuum solvent (PB or GB model) and non-polar contribution to solvation (SA) for analyzing the free energy of binding and the free energies of different sample conformations also (Bonnet and Bryce, 2004; Gohlke and Case, 2004). MM/PBSA method takes conformations of the solute from a MD trajectory achieved in explicit solvent using a periodic box with water and counterions. The binding free energy is computed as eq 4 and illutrated in Figure 32 (Stoica *et al.*, 2008; Xue *et al.*, 2012).

$$\Delta G_{binding} = G^{complex} - (G^{receptor} + G^{ligand}) \quad (4)$$

$$\Delta G = E_{MM} + G_{sol} - T\Delta S \quad (5)$$

$$E_{MM} = E_{internal} + E_{vdW} + E_{ele} \quad (6)$$

$$E_{internal} = E_{angles} + E_{bond} - E_{torsin} \quad (7)$$

$$G_{solv} = G_{PB/GB} + G_{non\text{-}pol} \quad (8)$$

$\Delta G_{binding}$ is the difference of binding free energy in water among complex, receptor and ligand, which free energy (G) of each species can be calculated from eq 5. ΔE_{MM} is molecular mechanism interaction energy between the protein and the ligand in vacuum. ΔG_{sol} refers to the solvation free energy of each species and $-T\Delta S$ is the change of the conformational entropy upon binding that approximated by normal mode analysis of the extracted MD snapshots (Gouda *et al.*, 2003).

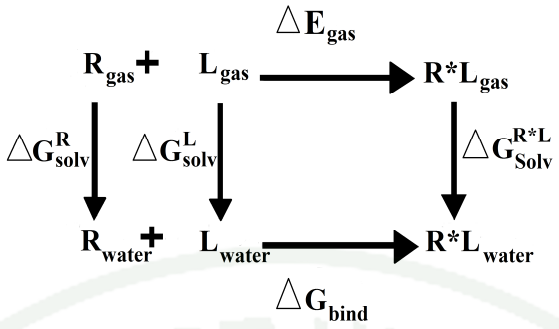


Figure 32 The binding free energy cycle**.** R, L and R*L refers to receptor, ligand and the complex, respectively

Source: Modified from Huang *et al.* (2006)

Molecular Mechanism interaction energy (E_{MM}) is originated from the sum of bonded and non-bonded term (eq 6). Non-bonded term is the sum of van der Waals and electrostatic interaction, whereas bonded term stands for the internal energy terms (including bond stretching, angle bending and torsional angle) using eq 6 and eq7. Because the cancelled internal energy terms of bonded part can significantly reduce the noise in the computed binding free energies. Thus, this term is cancelled in the single-trajectory approximation (Gohlke and Case, 2004; Hou *et al*., 2011).

The solvation energy in eq 8 (Guo *et al*., 2012), ΔG_{solv} is originated from the sum of the polar and non-polar contribution, which non-polar solvation part is calculated by the solvent accessible surface area (SASA). The electrostatic solvation or polar part can be calculated using the Poisson-Boltzmann (PB) equation based on PB or GB approximations (Thomas, 2003). According to parameters used in the solvation free energy calculation depending on PB or GB models, these parameters are contained in Amber program version 12 (Case *et al.*, 2005). In GBSA, the polar solvation free energy was defined by the modified GB model of Onufriev and colleague (GB^{OBC1}) (Onufriev *et al.*, 2004). In PBSA, the polar solvation free energy was employed by the PB solver in the pbsa module of AMBER12. The radii optimized by Tan and Luo with relating to the reaction field energies using in the TIP3P explicit solvents (Tan *et al.*, 2006). Furthermore, the SA

in PBSA scheme of Amber program version 12 has been developed a new non-polar solvation term that contains the sum of the cavity (repulsive) and dispersion (attractive) term. In eqs 9 and 10, we used the solvent-accessible-surface area (SASA) that correlate with the cavity creation and van der waals interactons to the solute while the attractive dispersion term (G_{disp}) was set as zero (Tan *et al.*, 2007). In term of G_{cavity}, The cavity_surften is set to 0.0378 kcal $mol^{-1}.A^{-3}$ for PB model and SASA is the solvent-accessible surface area of molecule determined by LCPO algolithm with the cavity_offset is set to -0.5692 kcal mol^{-1} (Tan *et al.*, 2007). These parameters respect to the TIP3P explicit water model (Tan *et al.*, 2008).

In PB model

$$G(\text{np}) = G_{disp} + G_{cavity} \quad (9)$$

$$G(\text{np}) = G_{disp} + (\text{cavity\_surften} * \text{SASA}) + \text{cavity\_offset} \quad (10)$$

In part of MM/GBSA (simple term), G(np) is non-polar solvation free energy removed all charge, obtained to be proportional to the Solvent-Accessible Surface Area (SASA) of the molecule with a proportionality constant derived from experimental solvation energies of small nonpolar molecules (Onufriev *et al*., 2004). The parameters in eq 11; "γ" is a surface tension parameter set to 0.0072 (Still *et al.*, 1990). SASA is the solvent-accessible surface area of molecule determined by LCPO algorithm (Weiser *et al.*, 1999) and "*b*" is a parameterized value set to 0.0 (Still *et al.*, 1990).

In GB model

$$G(\text{np}) = \gamma\text{SASA} + b \quad (11)$$

Finally, the entropy ($-T\Delta S$) is divided into translational, rotational and vibrational contributions, which estimated from a normal-mode analysis (Brooks *et al.*, 1995). Normal mode analysis (nmode) is one of the major simulation tool applied to

observe the large-scale, shape-changing motions in biological molecules, where large system with many snapshots may be needed for more stable and accurate prediction (Hayward and de Groot, 2008). However, because nmode calculations are high computationally expensive for large systems and the calculation of the relative free energy between relating systems are often assumed that the entropy of solute will be the same for each system being compared. Thus, nmode is often except for many cases of computational simulations (Brooks *et al.*, 1995; Massova and Kollman, 1999).

8.2 Energy decomposition per-residue

To understand the detail of protein-peptide interaction at interface, the free binding energy can be decomposed into the contribution from each residue. The interaction of each residue can be estimated using MM/PBSA approach based on the same snapshots as used in the binding free-energy calculation. The residue is splitted the individual interaction into backbone and side-chain that the separate contributions of them are calculated from the relevant atoms. In addition, the contributions of non-mutatable fuctional group such as backbone atom can be calculated at the atomic level (Li *et al.*, 2008). This tool can help better insight into understanding the importance of that residue to its partner.

8.3 Computational alanine scanning analysis

The computational alanine scanning method is applied for virtual screening the hot spots residues based on MM/PBSA calculation. The feature of side-chain functions at specific locations and the contribution in energetic of individual side-chain can be detected by alanine substitution leading to the change of binding free energy (Moreira *et al.*, 2007b)

Herein, the substitution a side-chain of the mutated residue with a hydrogen atom of alanine residue is monitored the difference in binding free energy between wild-type and mutated residues in eq 12. It should correspond to the binding affinity of that residue (Bradshaw *et al.*, 2011). Hot spot residues can be identified by alanine scanning in

post-processing protocol. Alanine and glycine, which smaller or equal sizes to alanine are cancelled the same as proline residues due to the difference of backbone conformation in comparison to alanine (Ribeiro *et al.*, 2012). Moreover, the binding free energy of wild-type and mutant is the different binding free energy of the complex, receptor and ligand (eq 13) (Moreira *et al.*, 2008).

$$\Delta\Delta G_{\text{binding}} = \Delta G_{\text{wild-type}} - \Delta G_{\text{mutant}} \quad (12)$$

$$\Delta G_{\text{binding-molecule}} = G_{\text{complex}} - (G_{\text{receptor}} - G_{\text{ligand}}) \quad (13)$$

The computational alanine scanning carries out into two ways including full molecular dynamics protocol and post-processing protocol. Full MD protocol that separate the initial structure, one is wild-type and another is mutant structure. Both of them are run parallel. Post-processing protocol is one of two ways in alanine scanning protocols. Only the wild-type structure is run and post-processed to generate the topology and coordinate files for both wild-type and mutant complexes, which the mutation of the desired residue was truncated at the side-chain (C_γ) before using in this step (Bradshaw *et al.*, 2011). In addition, the result from post-process alanine-scanning protocol developed by Massova and Kollman (1999) presented that the consistence with the experiment within accuracy 1.0 kcal.mol^{-1} by roughly (Massova and Kollman, 1999).

After computing, the result can be classified into 3 orders of magnitude including the hot-spots, warm spots and null spots. To have a strong impact for binding where alanine replacements are caused a significant negatively in the binding free energy higher than -4 kcal.mol^{-1}, also called hot-spot. Warm sports are the binding free energy differences between -2 to -4 kcal.mol^{-1} though it gets enough for statistical study (Moreira *et al.*, 2007a; Perez *et al.*, 2011). The binding free energy is lower than -2 kcal.mol^{-1} referred to null spots that are not the important residues for protein-protein binding (Perez *et al.*, 2011). The negative and positive signs mean highly unfavorable substitutions and preference for the mutated position by alanine, respectively (Massova and Kollman, 1999; Zhong and Carlson, 2005).

9. Molecular Dynamics (MD) simulations of biomolecules

Molecular dynamics (MD) simulations are very powerful toolbox based on computer simulation for understanding the physical basis of structure and function of biological macromolecules such as nucleic acids and proteins. Atoms and molecules in system are allowed to interact in order to reveal the dynamic characteristics of the system depending on evolution time by solving Newton's Second law of motion to generate trajectories (Karplus *et al.*, 1987; Lindahl, 2008). MD trajectories provide data to explain how the atomic positions, velocities, accelerations and single-point energies at that time (Adcock and McCammon, 2006) .

9.1 Classical Molecular Dynamics

In molecular dynamics, the motion of every atom in system is monitored by solving of Newton's equation (eq 14 and eq 15) as depending on time (Adcock and McCammon, 2006).

$$F_i = m_i a_i = m_i \frac{d^2 r}{dt^2}, \quad i = 1, 2, .., N, \quad where \quad a = \frac{d^2 r}{dt^2} \tag{14}$$

The net force, F_i exerted on atom *i*, m_i is the mass of atom *i*, a_i is the acceleration of atom *i*, *N* is the number of that atom in the system and *r* is Cartesian coordinates of that *i* atom. However, the acceleration can be given as the derivative of the potential energy with corresponded to the position, r_i (eq 14).

$$a = -\frac{1}{m}\left(\frac{dE}{dr}\right) \tag{15}$$

The force, F_i is the negative gradient of potential energy V (r_1, r_2,…,r_N) with respect to the position of atom *i* (eq 16).

$$F = -\nabla_I V_r \tag{16}$$

Where V_r is the potential energy of the system, however, a classical molecular mechanical mechanical model (MM, also known as “force field”) used potential energy functions (V_r), which consists of bonded (bond, angle, angle) and non-boned interactions including van der Waals (Lennard-Jones) and electrostatic (Coulombic) for treating the atomic motion of biological molecules (Adcock and McCammon, 2006).

From eq 14 and eq 16, two equations can be combined to obtain new equation as below (eq 17).

$$-(dV/dr_i) = m_i\,(d^2r/dt^2) \qquad (17)$$

In order to obtain the trajectory, the initial position of atoms origining from X-ray crtstallography and NMR spectroscopy are needed together with an initial distribution of velocities and the acceleration (Figure 33) (Nowak, 2012).

10. Application of Molecular dynamics in study of the EGFR kinase domain structures

10.1 A summarized introduction of the molecular dynamics simulations

Obsevervation of crystallographically investigated the complex of protein-protein, parts of thermodynamic experiments and mutational studies have supported to better understanding in protein-protein interactions especially the study of binding mode interaction between interface of hetero complex. The molecular dynamics (MD) methods have become an important computational tool for understanding the physical basis of the structure and function of biological macro-molecules including to protein-protein, protein-lipid and protein-ligand complex.

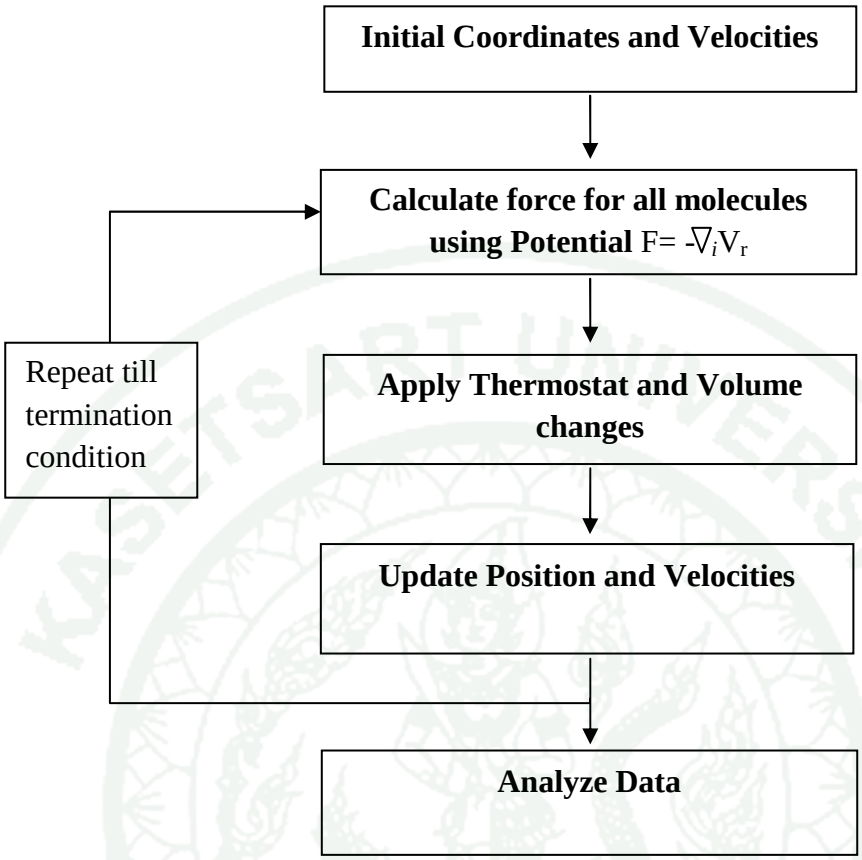


Figure 33 A flow chart illustrating of the standard Molecular dynamics (MD) procedure. Each of run needs the input parameters of initial coordinates and randomly initial velocities till reach to desire temperature.

Source Modified from Nowak (2012)

10.2 Molecular dynamics studies of the EGFR kinase with its drug inhibitor

The molecular dynamic simulations have been successful in the proving for EGFR interactions with its drug inhibitors and also insight a dynamic assessment of the effect from EGFR mutations (Papakyriakou *et al.*, 2009; Songtawee *et al.*, 2013). Erlotinib (tradename Tarceva) is one small molecule inhibitors used for non-small cell lung cancer (NSCLC), pancreatic cancer and other cancers treatments targeting at ATP binding site of EGFR kinase domain (Yun *et al.*, 2008). Park and co-workers studies the binding of small

inhibitor called erlotinib to EGFR kinase domain by performing in crystallographic study and computational methodology. They reported that erlotinib can bind in both inactive and active conformation of EGFR kinase domain after prediction by computational approach. The information was proven with testing in the crystal by soaking EGFR kinase domain in active and inactive forms with erlotinib. The erlotinib inhibitor can bind to wild-type EGFR kinase, which it can always adopts the crystal structure in active conformation. However, erlotinib can also bind to inactive conformation of the mutated kinase domain (V940R mutant) after co-crystallization. The results indicated that erlotinib can bind to both active and inactive EGFR kinase formations and the results was related to the computational methodology (Park *et al.*, 2012). Moreover, the MD simulations can gain the novel information, which the crystal structure cannot present. The novel EGFR information including intermolecular interactions can help us to better understanding the features in the structure basis of EGFR activation (Papakyriakou *et al*., 2009). In addition, the conformation from crystallographically revealed only snapshot of the actual pathway, while nonactive conformation can be stabilized by crystal contacts. An inaccuracy of functional interpretations may be occurred leading to presenting of artifact. Thus, MD is technique suitable for effectively determining the conformational fluctuation of protein molecule (Karplus *et al.*, 1987; Karplus and Kuriyan, 2005). Papakyriakou and colleagues (2009) investigated in the conformational change in EGFR activation in order to solve the point that how the mutant residues can activate EGFR (Papakyriakou *et al.*, 2009). The targeted MD simulations were performed to compare between the simulations of monomer and activated dimer. The results revealed that computational molecular dynamics simulations approach can predict the novel intermolecular interactions by forming salt-bridge between the mutated L834R with neighboring acidic residues leading to destabilizing of hydrophobic group the inactive conformation.

11. Current studies of new cancer therapeutic strategy: Blocking the interaction of protein-protein binding

Inhibition of protein-protein interactions targeting protein interfaces are an attractive drug discovery approach that a widely studied in current years. In previous

studies, there are several investigations reporting the importance of interactions between protein-protein in the atomic level using computational methodology.

11.1 Insulin-like growth factor-II and insulin-like growth factor-II receptor, IGF-II or Insulin-like growth factor-II is important regulator of cellular process including cell growth, survival and differentiation and also needed in many cancers. IGF2R or insulin-like growth factor-II receptor is a key as being tumor suppression, plays role in the maintenance of IGF-II levels in the circulation and in target tissues. MD simulations were applied to study the IGF-II/IGF2R complex in order to describe the detailed interaction, which greater than the obtained information from crystal structure. The given MD trajectories were performed using MM/PBSA to calculate the binding free energy and understanding the driving force of protein binding. Taken together with per residue binding free energy, it used to identify key residues for interaction and found that Glu12, Phe19 and Leu53 on the IGF-II surface and residues in IGF2R including Tyr1542, Ser1543, Phe1567, Thr1570, Arg1571, Ile1572, Pro1599, Ser1600 and Lys1631 were important to the binding process, yielding >2 kcal mol^{-1} (Figure 34). The information of IGF-II/IGF2R complex can guide us to better understanding the binding process that can be used in designing molecular probe to control tumor development (Guo *et al.*, 2012).

11.2 The study of MDM2-p53 interaction (Figure 35A), p53 is well-known as tumor suppressor that has essential role in biological processes including the cell cycle, DNA repair, apoptosis, angiogenesis and cellular metabolism (Vousden and Lu, 2002). The lost of p53 is not only inactivated by tumor but can be occurred by the over-expression of MDM2 protein, which MDM2 act as negative feedback inhibitor of p53 (Picksley and Lane, 1993; Vousden and Lu, 2002). Inhibitors of the p53-MDM2 interaction might be attractive new anticancer, design of non-peptide small-molecule inhibitors might block the interaction of MDM2-p53 (Shangary and Wang, 2008). Kussie and co-workers reported that human MDM2-complexed with short p53 peptides have performed atomic detailed of the interaction. This investigation found the important hydrophobic pocket of MDM2 involving with three key residues of p53 is Phe19, Trp23 and Leu26 shown in Figure 35B (Kussie *et al.*, 1996).

This interaction of hydrophobic pocket and key residues of MDM2 and p53, respectively can be guided to design non-peptide small molecule inhibitors based on the computational structure-based screening. Three key hydrophobic binding residues in p53 (Phe19, Trp23, Leu26) were modeled to search for small-molecule inhibitor from database of compound using the GOLD program and ranking score by Chemscore and X-score (Eldridge *et al.*, 1997). WK23 is structure-based design from p53 that is a member in the imidazole-indole family, an inhibitor based on aromatic groups. The Phe19' pocket is filled by the phenyl ring in a similar pattern but not as deeply into pocket as phe19 of p53, not affect the high affinity. The bottom of the Trp23 pocket is inserted by the indole group, which it contain the chlorine atom. The indole nitrogen atoms can form hydrogen bonds to Leu54 that the pattern is similar to Trp23 of p53. Leu26 pocket is submerged by the para-chlorobenzyl group (Figure 36) (Popowicz *et al.*, 2011).

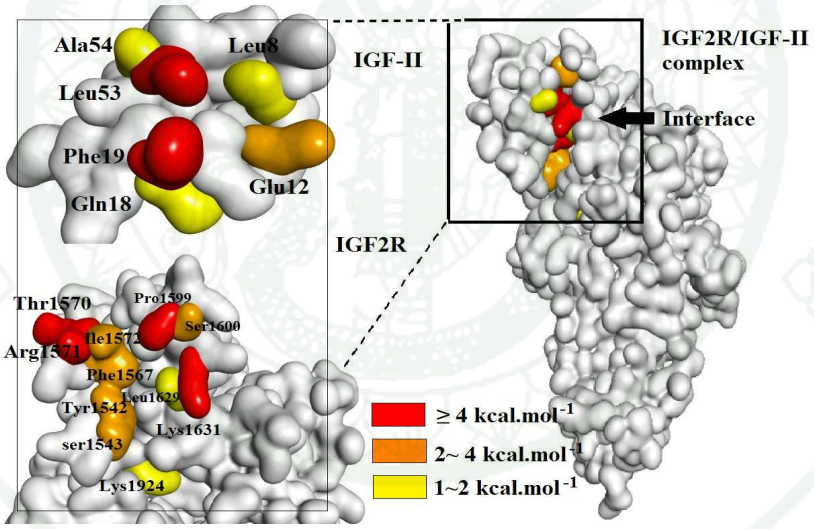


Figure 34 Surface representations of key residues between interface of IGF2R/IGF-II complex. Each color presented the level of free energy contribution.

Source Modified from Brown *et al.* (2008)

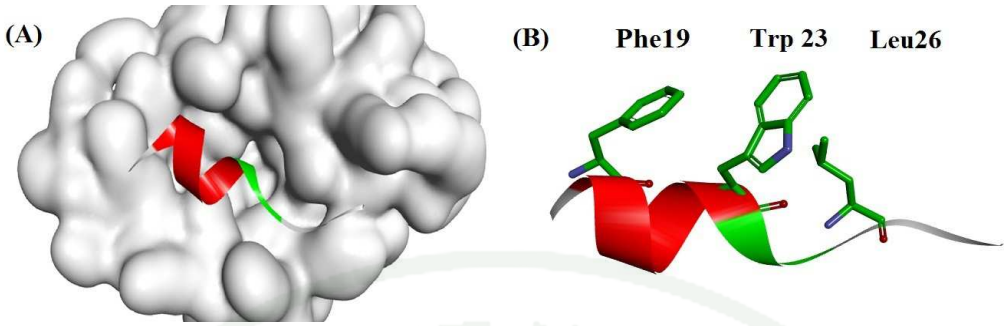


Figure 35 The crystal structure of human MDM2/p53 complex. The binding of p53 in complex (A) and the important side-chains for binding (B) were shown.

Source: Modified from Kussie *et al.* (1996)

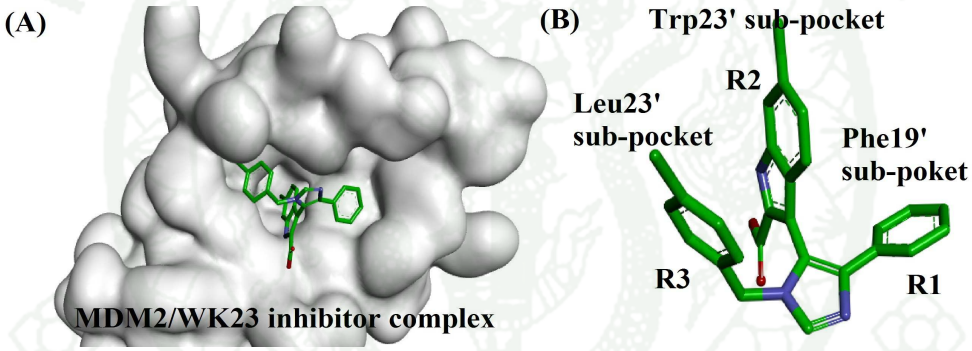


Figure 36 The crystal structure of human MDM2/WK23 complex. The binding of WK23 in complex (A) and the mimicked side-chain of WK23 (B) were shown.

Source: Modified from Shangary and Wang (2009)

MATERIALS AND METHODS

Computational methods

1. Initial structures preparation

From the Protein Data Bank (PDB), we received the starting geometry structure of the complexes for MD simulations. The MIG-6 bound EGFR complex was based on the crystal structure (PDB code: 2RFE) whereas the free MIG-6_s1 peptide (chain E) together with the free EGFR kinase (residues 678-959, chain B, act as activator) were also extracted from this complex (Zhang *et al*., 2007). MIG-6_s1 peptide sequences models used in this study and their properties were revealed in Table 1. The properties including isolectrostatic point (pI) and molecular weight (M.W) were performed by Expasy online tool (Wilkins *et al*., 1999).

As a-27 mer wild-type was lied on the structure of MIG-6 segement 1, then it was reduced the number of residues into a 20-mer wild-type for studying the effect of the double and single mutant models. All crystallographic atoms of bound ligands, water molecules and other heteroatoms were removed and after that the protein structures were modeled for missing and disordered residues by using the Swiss-PdbViewer v4.0.1. The protein structures of non-overlapping disordered regions were applied as multiple templates on the MODELLER 9v8 to model the EGFR kinase monomers. The PROCHECK v3.5.4 determined the stereochemical quality for all complexes in term of the ramachandran plot (Appendix Figure A1). Then, the complexes were constructed using the complex between Mig6-bound EGFR (PDB: 2RFE). In addition, the mutated sequences were trunicated the side chain by Discovery studio v3.5 (for education only). In order to keep the uncharged ends before performing a molecular dynamics simulations, an acetyl (ACE) and a methylated amino (NME) groups were used to cap at the N- and C-terminal end of each amino acid chain of every sequence, respectively (Nymeyer and Garcia, 2003; Zuo *et al.*, 2012). All missing hydrogen atoms of protein were added using the tleap program in

AMBER 12 software suite (Case *et al.*, 2005) and the AMBER ff03 force field (Duan *et al.*, 2003) were also used for describing the investigated protein. The systems were solvated by using atomistic TIP3P water (Jorgensen *et al.*, 1983) in octahedron truncated box with a distance of at least 7 Å between the wall of the box and the closest atom of the solute. To neutralize the charge, the counterions of Na^+ or Cl^- were added to maintain the electroneutrality of all the systems (Zuo *et al.*, 2012).

In this study, we processed into three parts depending on the objectives and the summary flowchart of the investrigation was present as Figure 37.

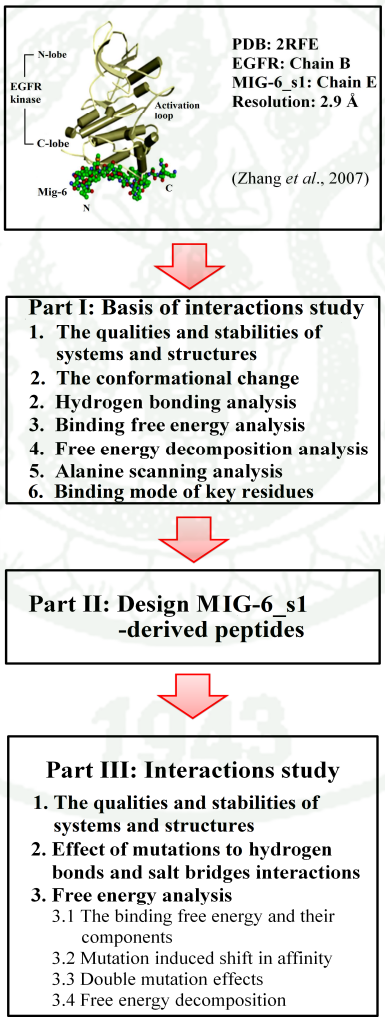


Figure 37 Mapping workflow on this study.

Table 1 Sequences of MIG-6_s1 peptide wild-types and mutants used in this experiment.

Nomenclature of Peptides.	Peptide sequence systems	Total residues	pI	M.W.
A 27-mer wild-type	ACE- ^{336}KSLPSYLNGVMPPT$_{349}$QS$_{351}$FAPDPKYVSSK^{362} -NME	27	9.4	2939.3
A 20-mer wild-type	ACE- ^{341}YLNGVMPPT$_{349}$QS$_{351}$FAPDPKYVS^{360} -NME	20	5.83	2211.5
T349R/S351R	ACE- ^{341}YLNGVMPPR$_{349}$QR$_{351}$FAPDPKYVS^{360} -NME	20	9.7	2335.7
T349R	ACE- ^{341}YLNGVMPPR$_{349}$QS$_{351}$FAPDPKYVS^{360} -NME	20	8.5	2266.6
S351R	ACE- ^{341}YLNGVMPPT$_{349}$QR$_{351}$FAPDPKYVS^{360} -NME	20	8.5	2280.6

pI is isoelectric point and M.W. is the molecular weight (g.mol^{-1})

2. Molecular dynamics simulations

Molecular Dynamics (MD) simulations were performed by using the sander module in AMBER 12. To remove the bad contacts in the crystal structure, each system was energy minimized in three steps before using the structure in further step. Positional restraints were applied to the whole system in the first and second steps with a force constant of 10 kcal.mol^{-1}Å^{-2} and 2 kcal.mol^{-1}Å^{-2}, respectively. Then, the whole system was allowed all atoms to move freely without restraint. The energy minimization in each step was carried out using 5000 cycles of minimization (MAXCYC) with the first 2000 cycles of steepest descent (NCYC), the remainders were conjugate gradient minimization. After minimization stage, the system was gradually heated from 0 to target temperature of 300 K over 100 ps in the NVT ensemble with the force constant 2.0 kcal.mol^{-1}Å^{-2} on all residues atoms. This followed by constant temperature equilibration at 300 K for another 500 ps under 1 atm pressure in NPT ensemble simulation. Finally, the production runs were carried out at 1 atm, 300 K, 2 fs integration time step and 10 Å nonbonded cutoff. During all the MD simulations, the particle mesh Ewald (PME) method (Darden *et al.*, 1993) and the SHAKE algorithm (Ryckaert *et al.*, 1977) were applied to approximate long-range electrostatics interactions and constrain the bond lengths involving bonds to hydrogen atoms, respectively. The size of systems used for the simulations and total 300 snapshots extracted from the last 3 ns of each simulation for further analysis were shown in Table 2.

Table 2 Summary of MD simulations performed in present study.

Nomenclature of peptides.	Total residues	Simulation length
A 27-mer wild-type	27	22 ns
A 20-mer wild-type	20	16 ns
T349R/S351R	20	16 ns
T349R	20	16 ns
S351R	20	16 ns

3. Post-processing analysis

2.1 Molecular dynamics quality analysis of the protein structures

The stereochemical quality of the ensemble of structures was checked by the Ramanchandran plot within PROCHECK program before using in further study (Laskowski *et al.*, 1993). This plot is between the Φ main-chain torsion angle versus the Ψ main-chain torsion angle for every amino acid residues in complex (Glycines and prolines were not included from calculation). The measure of the ramachandran plot is in term of percentage of residues in each region. The regions could be classified into 4 areas including the most favored regions, the additional allowed regions, the generously allowed regions and disallowed regions (Laskowski, 2009). If the residues were in the disallowed regions higher than 20% of all residues, it could be indicated that the ensemble structure had serious problems (Kleywegt and Jones, 1996). In addition, if the majority were in the most favored regions, the structure was acceptable. After that the structure was used further study. Before determining the dynamic stability of EGFR kinase/MIG-6_s1 peptide complexes, the quality of the MD simulations was required for assessment. The total (ETOT), kinetic (EKTOT) and potential energies (EPTOT) were observed along the MD simulations to ensure that the proper equilibration had been achieved. Furthermore, the Root-Mean-Square-Deviation (RMSD) of backbone atoms in system was calculated with respective to the initial minimized-structure over simulations.

2.2 Hydrogen bonding analysis

The average structure collected in equilibrium was superimposed with the initial-minimized structure for conformational change observation and then we checked the hydrogen bond formation or disruption by using the ptraj module of AMBER 12 suite. The criteria considerations for hydrogen bonding were: (1) acceptor–donor heavy atoms distances of less than 3.5 Å and (2) acceptor-H-donor cut-off bond angle of 120°. These parameters are generous for accepting weak hydrogen bonds formation, which the backbone oxygen could be related to more than one backbone hydrogen bonds. The results

were reported in the percentage of hydrogen bond occupancy term that the presenting of the hydrogen bond relatived to the total time length of the selected trajectory. Moreover, for checking the stability of hydrogen bond, the distance fluctuations between donor and acceptor H of the important atom pairs could be observed the distance dynamically along simulations (Xu *et al*., 1997).

2.3 Salt bridges analysis (computed only in part III)

The average structure were monitored salt bridge formation according to the distance between two oppositely charged pairs of difference amino acid, which the donor atoms from N^{ζ} of Lys, $N^{\delta1}$ and $N^{\varepsilon2}$ of His, N^{ζ}, $N^{\eta1}$ and $N^{\eta2}$ of Arg and the amide N of the N-terminus and the acceptor atoms from $O^{\varepsilon1}$ and $O^{\varepsilon2}$ of Glu, $O^{\delta1}$ and $O^{\delta2}$ of Asp and the two carboxy oxygen atoms of the C-terminus were observed. The distance is $\leq$ 4.0 Å between of these, was counted as a salt bridge (Barlow and Thornton, 1983; Kumar and Nussinov, 2002). However, if the geometry was in hydrogen bonding criteria, a salt bridge was also counted as a hydrogen bond (Xu *et al*., 1997).

2.4 The hydrophobic interaction analysis (computed only in part I)

LIGPLOT program was used to monitor the hydrophobic interaction, which hydrophobic contacts were presented by an arc with spokes radiating towards the atoms they contact. The contacted atoms were shown with spokes radiating back (Figure 38) (Wallace *et al.*, 1995). The hydrophobic contact threshold was set at default in range of 2.9-3.9 Å.

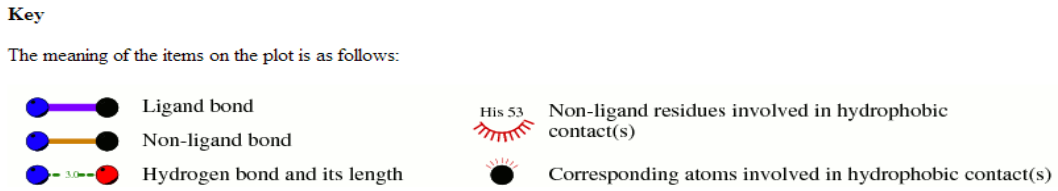


Figure 38 The meaning of the interactions signs on LIGPLOT program.

Source: Modified from Laskowski and Swindells (2011)

2.5 Binding free energy by using MM/PBSA calculation

The MD trajectories were used for calculating protein-protein binding free-energy by using MM/PBSA approach. The 300 snapshots were extracted for calculating binding free energy. Each molecular species (Complex, EGFR kinase and MIG-6_s1) were calculated the free energy from each snapshot and the binding free energy were estimated as $\Delta G_{binding} = G_{(complex)} - (G_{receptor} + G_{ligand})$.

The free energy (G) for each molecular species can be computed using MM-PBSA method as $G = E_{(gas)} - G_{(sol)} - T\Delta S$. Because E_{gas} is gas-phase free energy, which consisted of the internal energy (E_{int}) including the bond (E_{bond}), angle (E_{angle}) and torsion ($E_{torsion}$) energies. While $E_{(vdw)}$ and $E_{(ele)}$ are the van der waals and coulomb energies, respectively, also in the internal energy. E_{gas} was computed using ff03 molecular mechanics force field. The solvation free energy (G_{sol}) term is divided into polar and nonpolar contribution. $G_{(PB)}$ is the polar solvation calculated by solving the Poisson-Boltmann (PB) model. Dielectric constants of 1.0 and 80.0 were used for solute and solvent, respectively. The nonpolar part ($G_{nonpol, solv}$) is the sum of repulsive cavity term (ΔG_{cavity}) and an attractive solute-solvent dispersion interaction term (ΔG_{vdW}). And the last terms, T and S are the temperature and the total solute entropy, respectively. These terms could be estimated by classical thermodynamic, using normal-mode analysis (NMA). However, the entropy contribution was not calculated in Part III due to the complexity for calculating entropy accurately from normal mode analysis and the high computational

requirement for a large protein-protein complex (Hou *et al.*, 2009). In addition, the calculation of binding free energy based on single-trajectory simulation employs simulations of the single state (the complex) to generate conformations for all three states of complex including complex, protein and ligand (Hou *et al*., 2011). This single-trajectory simulation cancels internal energy term for reducing the computational requirement and is generally remained to sufficient for accuracy (Hayes and Archontis, 2012).

2.6 Double mutation effect analysis (computed only in part III)

In order to insight the effect of arigine substituations to protein stability and binding affinity, the effect of double mutant was analyzed. The Gibbs free energy (ΔG) of each species was compared to wild-type as shown in equation 18 (Doura and Fleming, 2004). Then, the difference of free energies of double mutant and two single mutants were computed by equation 19 (Jang *et al.*, 2004).

$$\Delta G=\Delta G_{MT} - \Delta G_{WT} \quad (18)$$

$$\Delta G_i=\Delta G_{1+2} - (\Delta G_1+\Delta G_2) \quad (19)$$

The Gibbs free energy change or the coupling energy (ΔG_i) is obtained by the difference between double mutant (ΔG_{1+2}) and the sum of mutated residues in two single mutations ($\Delta G_1+\Delta G_2$). Generally, the interpretion results can be classified into 5 classes including additive, partial additive, synergistic, antagonistic and absent in double mutant effects (Mildvan, 2004). Additive effect can occurr when the sum of two single mutants is equal to the free energy of the double mutation leading to the zero value in the coupling energy ($\Delta G_i=0$). The synergistic effect is ($\Delta G_{1+2} > \Delta G_1+\Delta G_2$) and ($\Delta G_i<0$), the result from the sum of two single mutants ($\Delta G_1+\Delta G_2$) is weaker than the free energy of double mutant (ΔG_{1+2}) (Jang *et al.*, 2004). The negative coupling energy is implied that the degree of destabilization by the double mutantion is smaller than the sum of those residues in two single mutations.

2.7 Per-residue interaction decomposition analysis

To gain detailed insight into the residue contribution, the important residue of MIG-6_s1 responsible for binding EGFR surface was calculated using decomposition per-residues process in the *mm_pbsa* module in Amber 12. This investigation based on the same snapshots as those used in the binding free-energy calculation. The binding interactions of each peptide residue to protein were computed by equation as below.

$$\Delta G_{\text{(protein-residue)}} = \Delta E_{\text{(vdw)}} + \Delta E_{\text{(ele)}} + \Delta G_{\text{(sol, PB)}} \quad (20)$$

The following MM/PBSA method used to perform per-residue decomposition based on PB model, while PB non-polar solvation energy was currently not decomposable (Berhanu and Masunov, 2012; Yang *et al.*, 2014).

2.8 Alanine scanning methodology (computed only in part I)

In order to validate the contributions of free energy components on per residues, in this study, alanine scanning was carried out by post-processing protocol to compute the energetic contributions of the individual side-chains to the protein binding (Moreira *et al.*, 2008). The difference of binding free energy between the wild-type and mutant complexes was computed using MM/PBSA method. The difference of them should correspond to the binding affinity of that residue, computed as eq 21

$$\Delta\Delta G_{\text{binding}} = \Delta G_{\text{binding, wild-type}} - \Delta G_{\text{binding, mutant}} \quad (21)$$

As post-processing protocol is one of two ways in alanine scanning calculation while only wild-type structure is run in MD simulations to collect trajectories (Saiz-Urra *et al.*, 2011). After running, the 300 snapshots were used to generate the topology and coordinate files both of wild-type and mutant. These snapshots were sampled form the equilibrium MD trajectories. Before free energy calculation step, the mutant structure was generated base on the structures of the collected snapshots by truncating the mutated

residue at the side-chain (C_γ) and replacing with a hydrogen atom of alanine (Massova and Kollman, 1999). Glycine and proline were not mutated because the substitution might cause a conformational change in the protein backbone (Bullock *et al.*, 2011). After calculation, the different of binding free energy from wild-type and mutant could be indicated in positive and negative values, which were preferable and unfavorable for the alanine substitution, respectively. However, the important residues had been clarified cut-off as those sites had a significant decrease in the binding free energy of at least > -2.0 kcal.mol^{-1}. Being strong interaction in binding should decrease the free energy higher than 4 kcal.mol^{-1} (Morrison and Weiss, 2001; Sousa *et al.*, 2011)

Computational Resources

Computer cluster from Biochemistry department of Science, Kasetsart University and High Performance Computer (HPC) cluster of National center for genetic engineering and Biotechnology Center (BIOTEC), Pathum Thani, THAILAND.

RESULTS AND DISCUSSIONS

As previously reported, EGFR kinase can be inhibited in the kinase activity by endogenous protein called MIG-6. The segment 1 (337-361) of MIG-6 (MIG-6_s1) can cradle the C terminal lobe of EGFR in order to prevent asymmetric dimer formation before entering into the auto-phosphorylation step (Zhang *et al*., 2007). Thus, besides of the most widely-studied in extracellular and ATP-binding site of the targeted EGFR, we interested in new targets at the C terminal lobe part of EGFR kinase surface for studying the inhibiting interactions of MIG-6 (Kuriyan *et al.*, 2012).

For results and discussions, we followed the objectives of this study at first part involving in basis of EGFR kinase/MIG-6_s1 peptide interaction based on MD simulations. In second part, we applied the understanding of these relationships to design short peptide inhibitors using MIG-6_s1 peptide model. Last part, we investigated the interactions of short peptides comparing to MIG-6_s1 reference structure.

In order to design the potency of MIG-6_s1 peptide as EGFR kinase inhibitors, the understanding interactions between them at atomic level were the first important for investi-gation. The results and discussions were described together in part I.

Part I: Preliminary analysis of MD simulations of EGFR kinase/MIG-6_s1 interface

1. Molecular dynamics quality of the protein structures

Before using the structure in next step, the stereochemical quality of protein structures was monitored with PROCHECK program to check the quality of structure (Laskowski *et al.*, 1993). The ordinary ramanchandran plot was first explained in term of steric clashes by Ramanchandran and co-workers (Ramachandran *et al.*, 1963). The Ramanchandran plot is a way to monitor backbone torsion angles, psi (ψ) against phi (φ) of each amino acid in protein structure. There are four regions contribution in plot including red region for the most favorable that corresponding to the possible conformation without steric hindrance between atoms, yellow regions are the additional allowed regions indicated the allowing of atom to come a little closer together whereas the regions of cream are also called name “the generously allowed regions” presenting the closer contact between atoms than those in yellow regions. However, the steric hindrance between atoms is shown in the white regions, which are the disallowed regions except glycine residue. The EGFR kinase/MIG-6_s1 peptide ensemble structure was acceptable geometry. The 78.4% of all residues were in the most favored regions, 20.1 % in the additional allowed regions, 0.7% in the generously allowed regions and a few in the disallowed regions 0.7 % (Appendix Figure A1). As the most of residues were located on the red regions, the modeled structure can be used in next step.

After that the physical properties of the simulations were addressed. The total energy (ETOT) is obtained by the sum of potential (EPTOT) plus kinetic (EKTOT) energies. As presented in Figure 39, the total energy, potential energy and kinetic energy were investigated for all simulations starting from the heating step with 100 picoseconds (ps), the equilibrating of the system with 500 ps until to the completely production runs of 22 nanoseconds (ns). The results showed that three energies of the system turn into equilibrium within 600 ps (0.6 ns) before entering into the production run step as shown in small Figures 39 and the systems could constant through 22-ns of MD simulation. These

results indicated that the systems had been reached equilibrium and could rely to estimate the binding free energy

Next property was the assessment of the dynamic stability of EGFR kinase/MIG-6_s1 peptide complex. The convergence of the root-mean-square deviation (RMSD) based on the backbone (Cα, N and O) atoms for each trajectory was plotted relative to the initial-minimized structure. The conformational flexibility is indicated by the fluctuation of RMSD without significantly with the variances of 1 Å or more shown a large change in the protein folding (Kannan and Zacharias, 2014).

In this study, the MD trajectories of EGFR kinase and MIG-6_s1 peptide in bound states and unbound state were performed in explicit water for a 22-ns production trajectory at 300 K. The backbone root-mean-square deviation (RMSD) along a 22-ns MD simulation was shown in Figure 40. The relative fluctuations in the RMSD of all systems were within reasonable fluctuation except the unbound MIG-6_s1 peptide.

In comparison among EGFR kinase, MIG-6_s1 peptide and complex systems in bound state, protein EGFR kinase (the red line) and EGFR kinase/MIG-6_s1 complex (the black line) reached to the plateau about after 2 ns and was stable over the time scale of the simulations, but MIG-6_s1 peptide in complex system (the green line) increased in the RMSD observed between 7 and 12-ns of MD simulation and trended to be stable the rest of the simulation phase. Form the overview of the average RMSD from a total 22-ns production run, all RMSD of each system except MIG-6_s1 peptide trended to reach equilibrium state with the convergent around 2.5–3.5 Å. Nevetheless, MIG-6_s1 had larger RMSD fluctuation and higher the average RMSD value (~4.2 Å) than that of complex (~2.5 Å) and EGFR kinase (~2.9 Å), which were quite less fluctuations respected with the initial-minimized structure.

Taken together, to compare with unbound state, MD trajectories of EGFR kinase and MIG-6_s1 were performed and the levels of conformational drift were monitored. We prepared the individually crystallized EGFR kinase and MIG-6_s1 peptide (PDB code:

2RFE). The unbound simulations systems were compared with the one in bound system, the result showed that EGFR kinase in both states (bound and unbound) were similar pattern of RMSD fluctuations and the RMSD average values (~3.0 Å for unbound and ~2.9 Å for bound EGFR kinase). In addition, the unbound state of MIG-6_s1 peptide showed significantly higher the RMSD fluctuation and RMSD average value than the one in bound state (~ 6.4 Å for unbound and ~4.2 Å for bound MIG-6_s1). The results could be suggested that EGFR kinase was quite stable not only in the unbound state but the introducing of EGFR kinase could help MIG-6_s1 peptide structure and the overall complex more stability.

As the free moving of short peptide was often observed in solution (Espinoza-Fonseca and Trujillo-Ferrara, 2006), it might be explained by the observation of Zhong and Carlson, which the great deal of the flexibility in free p53 could not stabilize in solution. However, it could be stabilized by the hydrophobic binding cleft of MDM2 protein (Zhong and Carlson, 2005). It was interesting that MIG-6_s1 peptide might present to follow this trend and the binding of EGFR kinase in bound state could assist MIG-6_s1 peptide more stable in structure and maintain peptide to the complex formation (Massova and Kollman, 1999). The differences in RMSD were monitored in the unbound and bound states supporting the fact that unbound systems both of EGFR kinase and MIG-6_s1 peptide were more dynamic change and less stable compared to bound state, similar results of MDM2-p53 and other studies (Espinoza-Fonseca and Trujillo-Ferrara, 2006; Yang *et al.*, 2012).

The results from the physical properties of the simulations and the stabilities of the simulated structure observation could be concluded that the qualities of the MD simulations were reliable and sufficient for further investigations. Based on the observation, we extracted 300 snapshots from the last 3 ns of MD simulation to further study in the intermolecular interactions (hydrogen bonds and hydrophobic interactions), binding free energy components, decomposition of effective energies on key residues and also alanine scanning analysis.

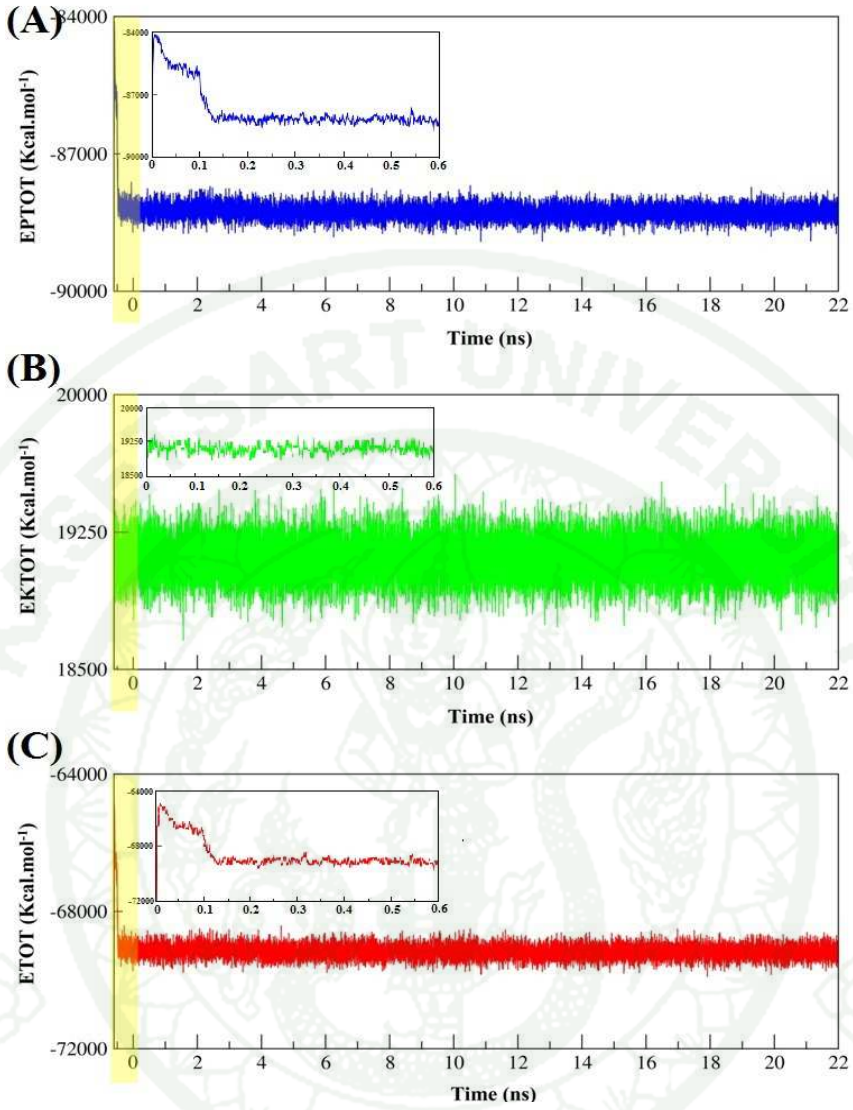


Figure 39 The energies value starting from heating steps, equilibration the system until production runs for the complex formation. The potential energy; EPTOT (A), the kinetic energy; EKTOT (B) and the total energy; ETOT (C) were plotted. The yellow areas represented the energies in heating-equilibrating step, which the scale was extended and showed in small figures with a total of 600 ps or 0.6 ns.

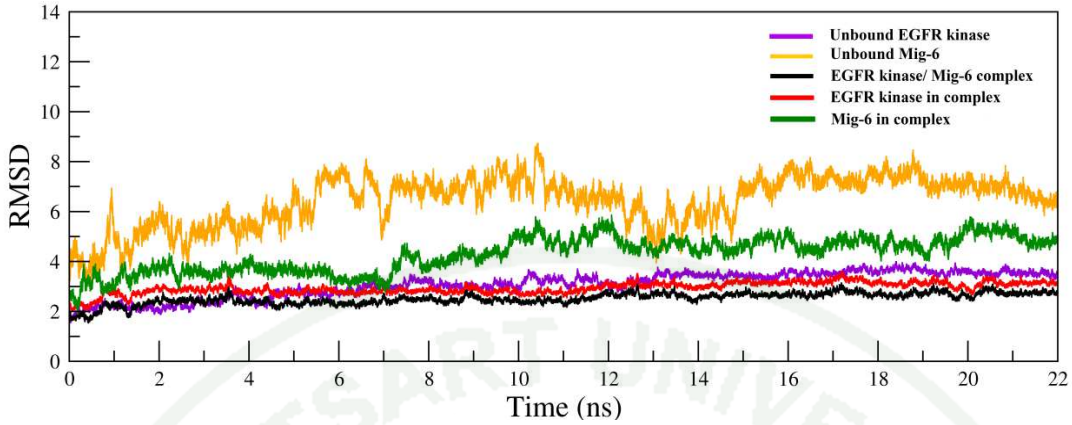


Figure 40 Root-mean-square displacement (rmsd) analysis for the EGFR kinase/ MIG-6_s1 peptide backbone respected to the initial minimized structure in bound and unbound state of a 22-ns MD simulation.

2. Conformational change of MIG-6_s1 peptide

MIG-6_s1 peptide of residues 336-362 can intact at the distal surface of the C-lobe kinase domain for preventing EGFR kinase to act as a cyclin-like activator for other kinase (Figure 41) (Zhang *et al*., 2007). In this part, the phenomena studies of the average structure were observed by the evolution time. To gain the molecular levels of the difference between the initial-minimized structure and MD-stimulated structure, the last 3 ns of production run (referred to as equilibrium simulation) was averaged at the interval of 10 ps and superimposed with the initial-minimized structure. The superimposition between the 300 snapshot averaged structure and the initial-minimized structure of EGFR kinase/ MIG-6_s1 complex was shown in Figure 42. It revealed that EGFR kinase structure (orange) in complex took the similar orientation as seen in the initial-minimized structure (gray) corresponding to RMSD results. The average structure of MIG-6_s1 peptide (orange) had a remarkable motion different from the initial-minimized structure (green) at loop region. Nevertheless, this particular case, the conformational flexibility of MIG-6_s1 peptide did not large effect to the global dynamics of the complex as seen in RMSD plotting of complex.

In the conformational flexibility of loop region, it presented the interesting that the residue Asn343 of MIG-6_s1 peptide was departed from the initial-minimized structure by shifting the side-chain up closely to EGFR surface (Figure 42 and 43). Howerver, at loop region was normally flexible together with the crystal packing effect. Therefore, the conformational distortion of the flexible region could be occured leading to unobservation in the crystal structure and also did not be reported in the experimental study (Megy *et al*., 2005; Shahsavar *et al*., 2012). This novel observation could be proven and explained the feature interactions in further analysis for example the intermolecular interactions and decomposition binding free energy on per-residue basis.

Moreover, as proline rich in peptide structure might affect to the rigidity of structure, five prolines of MIG-6_s1 were found in *cis*-configuration located on turn and the dispersion of five prolines could be also stabilized by aromatic stacking along MD simulations (Figure 44). Four of them Pro347, Pro348, Pro354 and Pro356 excepted Pro339 were remained to bending form the interface (Zhang *et al*., 2007). From this observation, it could be suggested that proline residues on turn might participate in the peptide-stabilized structure. To similar manner of proline in SH3 domain, the positions of two proline were relatively fixed for constraining the structure flexibility (Stein and Aloy, 2008). As hydrophobic interactions were a leading forces in protein-protein interactions, in this study was revealed the hydrophobic pocket of EGFR kinase interface that consisted of Trp881, Pro913 and Thr885. This pocket was plugged by the side-chain of Met346 which was stable on time evolutions (Figure 45).

From these results, most of these phenomena were consistency with the experiment observation of Zhang and co-workers in the crystallography study (Zhang *et al*., 2007). To concluded that the superimposition of the average structure had the major deviation from the initial-minimized structure at loop region.

The rest of MIG-6_s1 residues were accepted with slightly shift of the orientations and positions. The degree of conformational change in this observation revealed the interesting points for further analysis in more detail.

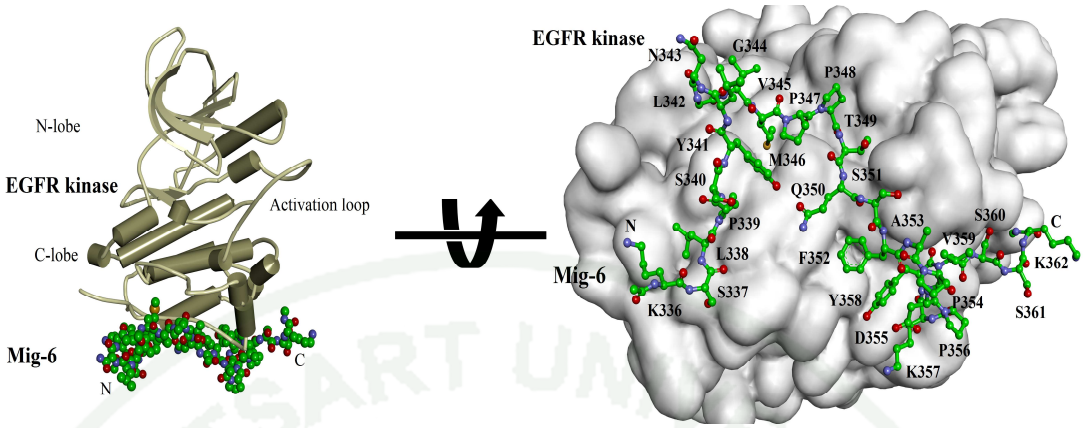


Figure 41 Cartoon representation of the crystal structure of the EGFR kinase domain /MIG-6_s1 peptide (Lys336-Lys362). The positions of the interested MIG-6_s1 residues were shown on the C lobe of EGFR kinase interface.

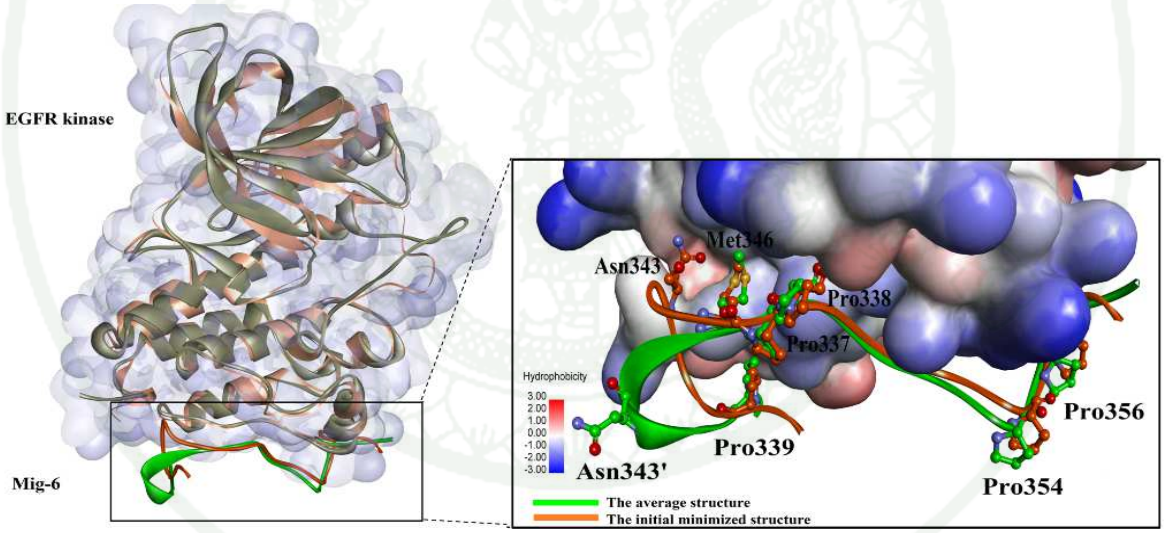


Figure 42 The superimposition of the average structure over the 300 snapshots using the respective initial minimized structure. The residues of MIG-6_s1 distributed on the C lobe of EGFR kinase interface.

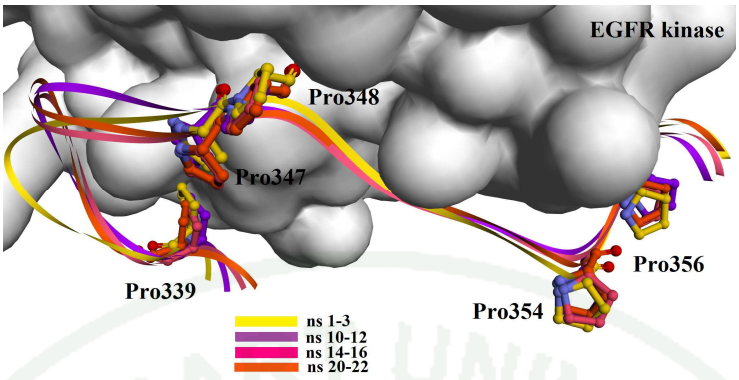


Figure 43 The superimposition of five proline residues on the average structure of MIG-6_s1 peptide on time evolutions.

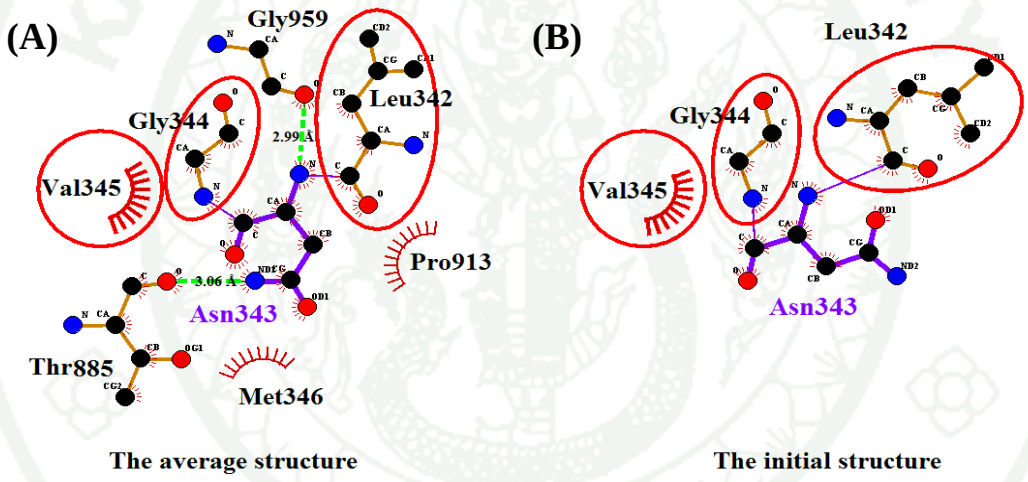


Figure 44 Comparing the interaction of Asn343 and its relevant residues between the average structure (A) and the initial structure (B) plotted by LIGPLOT. The average structure shown Asn343 as ribbon with purple color and neighbors residues were brown. Carbon, nitrogen and oxygen atom were exhibited as black, blue and red balls. Green dash line presented hydrogen bonds and dark red “eyelashes” were hydrophobic contacts between residues to Asn343. Black eyelashes corresponded to atoms involved in hydrophobic contact.

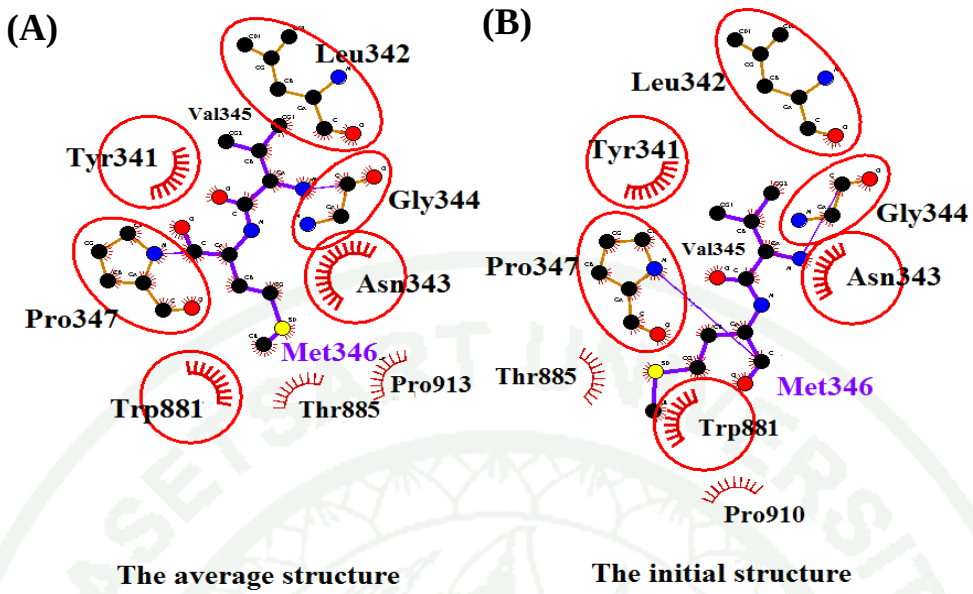


Figure 45 Comparing the interaction of Met346 and its relevant residues between the average structure (A) and the initial structure (B) plotted by LIGPLOT. The average structure shown Met346 as ribbon with purple color and neighbors residues were brown. Carbon, nitrogen and oxygen atom were exhibited as black, blue and red balls. Green dash line presented hydrogen bonds and dark red “eyelashes” were hydrophobic contacts between residues to Met346. Black eyelashes corresponded to atoms involved in hydrophobic contact

3. Hydrogen bonding analyses

Hydrogen bond (HB) is the important one of various intermolecular driving forces in protein-protein interactions (Baker and Hubbard, 1984). Therefore, in our analysis, the phenomena of hydrogen bonds were observed and shown in Figure 46. The criterions consideration for hydrogen bonding were: (1) acceptor-donor heavy atom distances of less than 3.5 Å and (2) acceptor H-donor cut-off bond angle of 120° based on Amber12 suite calculation (Case *et al.*, 2005). These parameters could allow the construction of weak hydrogen bonds, which the backbone oxygen may be involved in more than one backbone hydrogen bond. However, the hydrogen bonding of the backbone (-CO-) group with the partner amino (-NH-) donor group was stronger and remained the most outstanding (Laurence and Berthelot, 2000). The existence of how often and how long hydrogen bond formation were counted in term of "% occupancies" (Khandelia and Kaznessis, 2007) and listed in Table 3 and 4.

In this study, 300 snapshots of the last 3-ns of MD simulation were averaged and observed. Taken together with the distance fluctuations profiles of the important HB along a 22 ns of productions run was also observed and presented. They were proven that some of the hydrogen bond distance presented the stabilizing of distance within the cut-off through the simulations. During the breaking of distances in some simulation time, however, they could remain to reform in the rest time. In addition, the totally of 13 intermolecular hydrogen bonds were found in the initial-minimized structure as shown in Table 3 and Figure 46A. Although, a total of 15 hydrogen bonds were found in the time-averaged MD structure with the occupancies higher than 20% and presented in Table 4 and Figure 46B. To compare with the initial-minimized structure, the results revealed that four original hydrogen bonds obtained from the initial-minimized structure were lost in the time-averaged MD structure, whereas six new HBs were created in the time-averaged MD structure and two of six new HBs involved in the conformational change. These formations could be revealed and described as below.

Before MD simulations, the initial-minimized structure found the hydroxyl oxygen OG of Ser337 in MIG-6_s1 peptide accepted hydrogen bond from Ile917 the nitrogen backbone of EGFR kinase (Figure 46A). However, as the superimposition results seen in the average structure (Figure 46B), the rotation 90° of the backbone and side-chain of Ser337 (excepted for the backbone C=O) could result in HB construction of its oxygen backbone with Ile917 N-H of EGFR kinase (occupancy of 51%, distance ~3.0 Å), whereas the bonding of Ser337 the hydroxyl oxygen OG was disturbed. Moreover, the carboxyl O of Ser337 could participate with Tyr341 in both the initial-minimized and MD-simulated structures for stabilizing peptide binding by accepting and maintaining a strongly HB from the NE2 atom of Gln911 in EGFR kinase with the occupancy of 96% and distance ~3.0 Å. As seen in Figure 47A, the hydrogen bond distance profile of Ser337 with Gln911 could stabilize throughout the MD stimulation and the distance fluctuation showed the dynamic lesser than 3.5 Å which could be partly resulted in the stability of the interface binding.

On the loop connecting of helices αG-αH in EGFR kinase could detect several HB formations with MIG-6_s1 peptide. In initial-minimized structure, the side-chain (OH) in Tyr341 and the atom OE1 in Gln350 of MIG-6_s1 peptide accepted the H atom from the nitrogen atom in the backbone of Gln911. While in the average structure, the side-chain (OH) in Tyr341 was switched the role from acceptor to be donor and could donate H-atom to acceptor the carboxyl O of Gln911 (occupancy of 99.27 %, distance ~2.8 Å) and Gln350 of MIG-6_s1 also remained to accept a HB form Gln911 with the occupancy value of 79.33%, distance ~3.0 Å. The strong HBs of them were supported by distance profile (Figure 47B), the distance fluctuation of Tyr341 was kept for hydrogen bonding during all the MD processes within the cut-off, while was disturbed only during the 13-16 ns. The results indicated that Try341 could play role in participation of the stability complexation, in spite of losing (OH) in acceptor role seen in the initial-minimized structure.

Table 3 List of intermolecular hydrogen bonding between EGFR kinase domain and MIG-6_s1 peptide interface of the initial-minimized structure. The criterions were at a cut-off distance of 3.5 Å between donor and acceptor atoms and a DH-A cut-off bond angle of 120°. The occupancies of 100% were shown.

Donor	Acceptor	Distance (Å)
Arg908-NH2	Tyr358-O	2.844
Ile929-NH	Lys357-O	2.965
Gln911-NH	Gln350-OE1	2.979
Arg908-NH	Gln350-O	3.014
Gln911-NH	Tyr341-OH	3.276
Gln911-NE2-HE21	Ser337-O	2.926
Ile917-NH	Ser337-OG	3.314
Gln350-NH	Arg908-O	2.844
Thr349-OG1-HG1	Glu907-OE2	2.651
Ser351-OG-HG	Glu907-OE1	2.591
Phe352-NH	Gly906-O	2.927
Ser360-NH	Glu904-O	2.966
Ser360-OG-HG	Glu904-OE2	2.544

Table 4 List of intermolecular hydrogen bonding between EGFR kinase domain and MIG-6_s1 peptide interface of the average structure. The criterions were at a cut-off distance of 3.5 Å between hydrogen bond donor and acceptor atoms and a DH-A cut-off bond angle of 120°. The occupancies higher than 20% were shown.

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

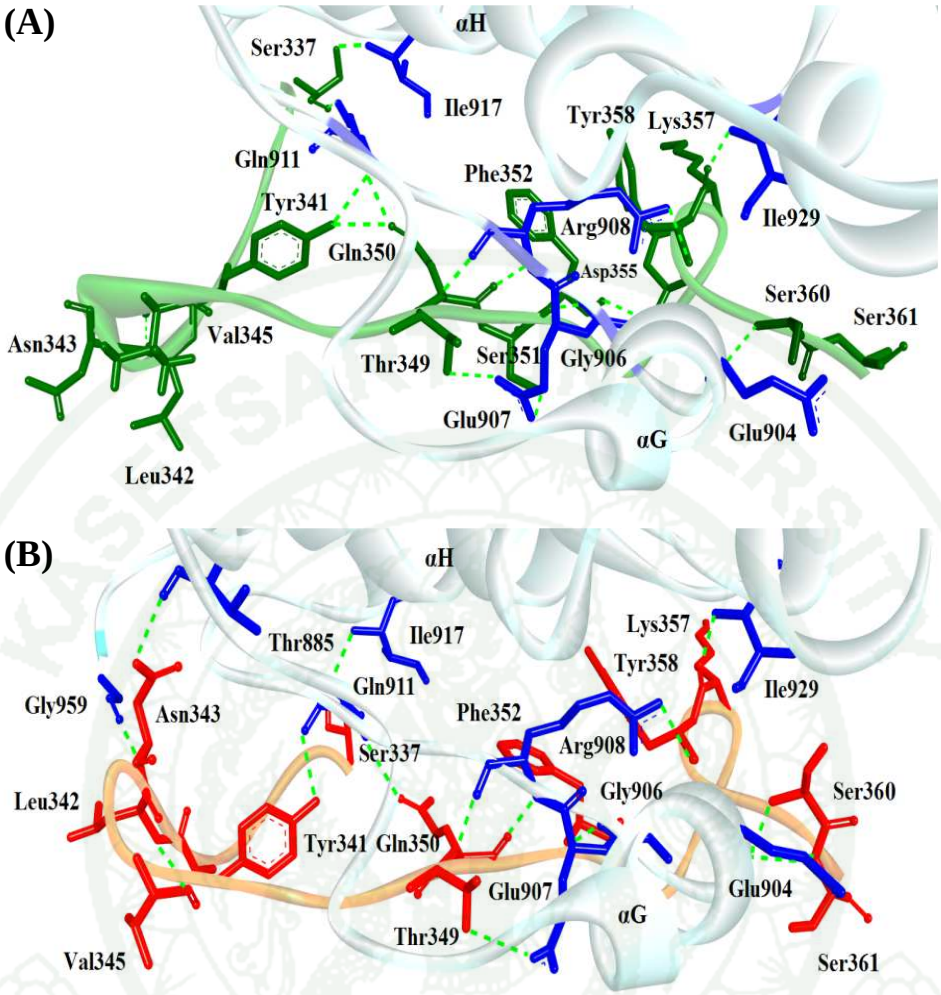


Figure 46 The hydrogen bonds network of EGFR kinase/MIG-6_s1 peptide complex. (A) The initial minimized structure presented the residues involving in hydrogen bonds of MIG-6_s1 peptide (green) and EGFR kinase (blue). (B) The average structure presented the residues involving in hydrogen bonds of MIG-6_s1 peptide (orange) and EGFR kinase (blue). Hydrogen bond was presented as green dot.

Next hydrogen bond analysis, the oxygen atom in the side-chain of Thr349 (OG1) and Ser351 (OG) in MIG-6_s1 were detected in the initial-minimized structure providing intermolecular HB for the oxygen side-chain of Glu907 in EGFR kinase. However, at the last 3-ns of production run, only Thr349 (OG1) could form both the oxygen side-chain of Glu907 (the occupancies of 24% and 69.90%, distance ~2.7 Å in both), whereas the bonding between Ser351and Glu907 was disturbed.

In initial-minimized structure detection, the residues Gln350, Phe352, Tyr358 and Ser360 of MIG-6_s1 peptide could be aligned parallel to each other for forming HB. The involving Arg908 of EGFR kinase could form with two bond of Gln350 and one bond of Tyr358. Moreover, Gly906 and Glu904 of EGFR kinase could form with Phe352 and Ser360, respectively. Based on the selected simulation data, we found all residues as aforementioned that could remain HBs establishment. Gln350 could form HB between the backbone with Arg908 (-N..O-) of EGFR kinase, over 99 % occupied and distance ~2.9 Å. Furthermore, NH2 atom of Arg908 donated HB to Tyr358 of MIG-6_s1 peptide with the occupancy of 92.73% and kept the distance along 22 ns of the MD simulation within ~3.0 Å (Figure 47F). The carboxyl O of Gly906 and Glu904 could form HBs with the nitrogen backbone of Phe352 and Ser360 (occupancies of 87.43%, distance ~3.0 Å and 96.83%, distance ~2.8 Å), respectively. As seen in Figure 47C, 47D and 47E, the distance profiles of these HBs were stable with less fluctuation within the cut-off through a 22-ns of the MD simulation, opposed by Ser360 in Figure 47G that little fluctuation over 3.5 Å in some trajectory. From the profiles, these residues might be crucial for EGFR kinase/MIG-6_s1 peptide complex stability.

In αH helices of EGFR kinase, the backbone NH of Ile929 donated H atom to the carboxyl O of Lys357 in a 3-ns of MD process (occupancy of 98.53%, distance ~2.9 Å), which was also conserved in the initial-minimized structure. The distance profile of hydrogen bond was stable within the hydrogen bond cut-off during long MD simulations, indicating that these residues might also participate in the important feature for binding stability.

Moreover, during MD-stimulated structure, Ser361 of MIG-6_s1 moved the side-chain facing on the interface and pushed the carboxyl backbone out, resulting in the stabilizing Glu904 residue in αG helices by H-bonds construction (occupancy of 69.37 %, distance ~3.1Å) although hydrogen bonding of side-chain in Ser360 was distrubed. According to Table 3 and 4, besides of the construction of intermolecular hydrogen bonds, the intramolecular H-bonding within MIG-6_s1 peptide could also find two the conserved HBs of Leu342-Val345 (occupancy of 94%) and Phe352-Asp355 (occupancy of 48%), data not shown.

Interestingly, as we found the major conformational change of MIG-6_s1 by the ability of Asn343 in shifting the side-chain up closely to EGFR kinase surface, this residue could form two hydrogen bonds. New hydrogen bonds of the Asn343 were revealed that its side-chain (ND2) could construct with Thr885 the carboxyl O of EGFR kinase located on αF helices (occupancy of 89.57%, distance ~3.0 Å). Another one with the C=O of Gly959 (occupancy of 97.73%, distance ~3.0 Å). The difference distances between them to Asn343 were continued to decrease until the HB could be formed and stabilized with less fluctuations, when the processes reached at the 13- and 17-ns for Thr885 and Gly959, respectively. They could stabilize the protein-peptide complex at the average structure from MD simulations. However, the conformational shift phenomena were missed in the initial-minimized structure.

Based on the aforementioned all HBs detections, it could be concluded that during the time-averaged MD structure, most of residues involved in hydrogen bonds could be maintained. Here again, because of the conformational change detection in the average trajectories of MIG-6_s1, the stabilities and dynamics of H-bonding along the 22 ns of simulation should be observed. The results presented the disturbing and bonding contact between interfacial EGFR kinase and MIG-6_s1 along MD simulations based on the criteria. Despite of the generous cut-off, the most strongly and hydrogen bonds stability remains dominant in the backbone C=O group with the N-H donor group.

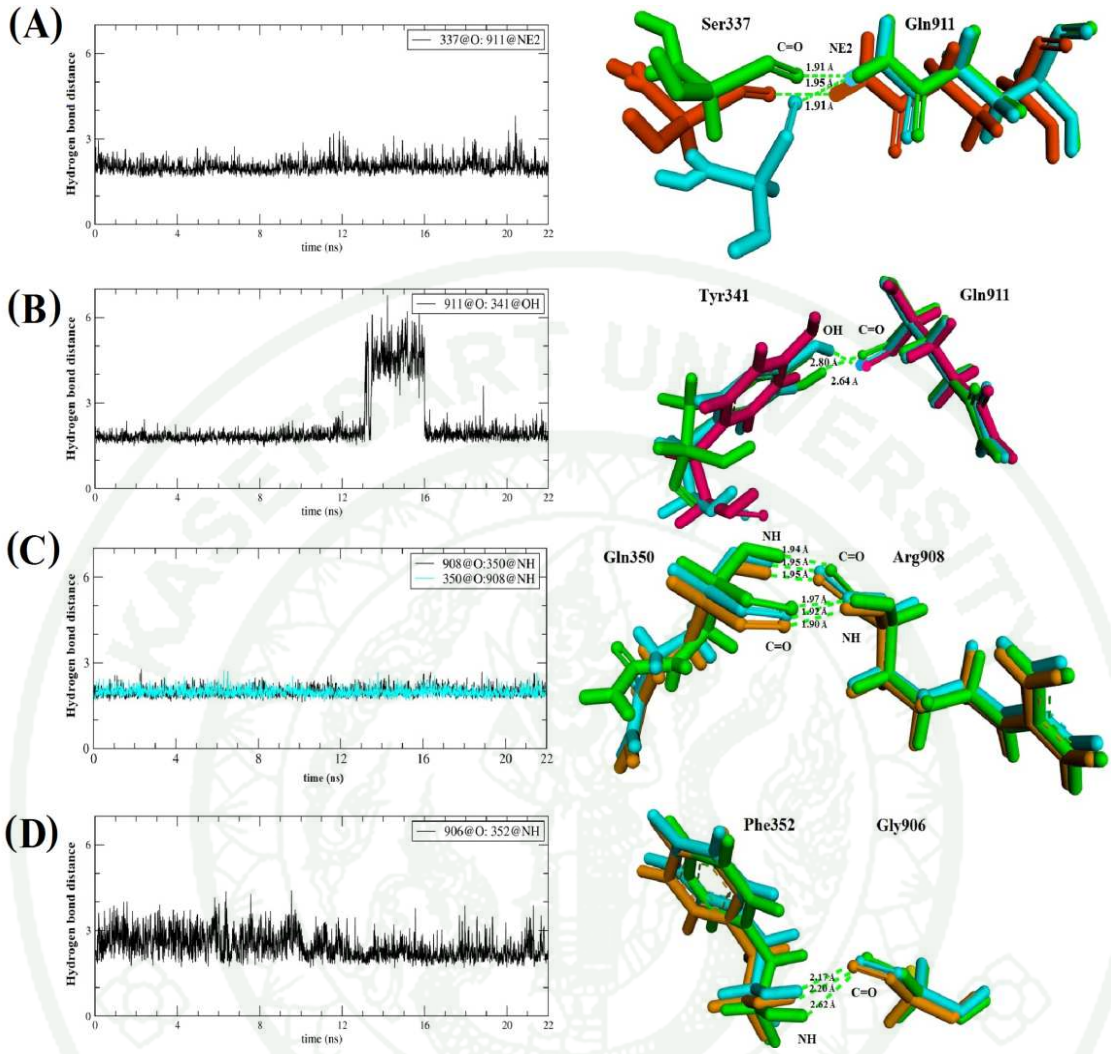


Figure 47 The distance variations (left) during a 22-ns MD simulation and the superimpose results (right) of hydrogen bonding between pair-residues during MD simulations. Each pair was presented as (A) Ser337:Gln911, (B) Tyr341:Gln911, (C) Gln350:Arg908, (D) Phe352:Gly906, (E) Lys357:Gly929, (F) Tyr358:Arg908 (G) Ser360:Glu904, (H) Asn343:Thr885 and (I) Asn343:Gly959.
The color in right side represented the average trajectories of each period simulation time with green (the first 3 ns), orange (during 10-12 ns), pink (during 14-16 ns) and light blue (the last 3 ns).

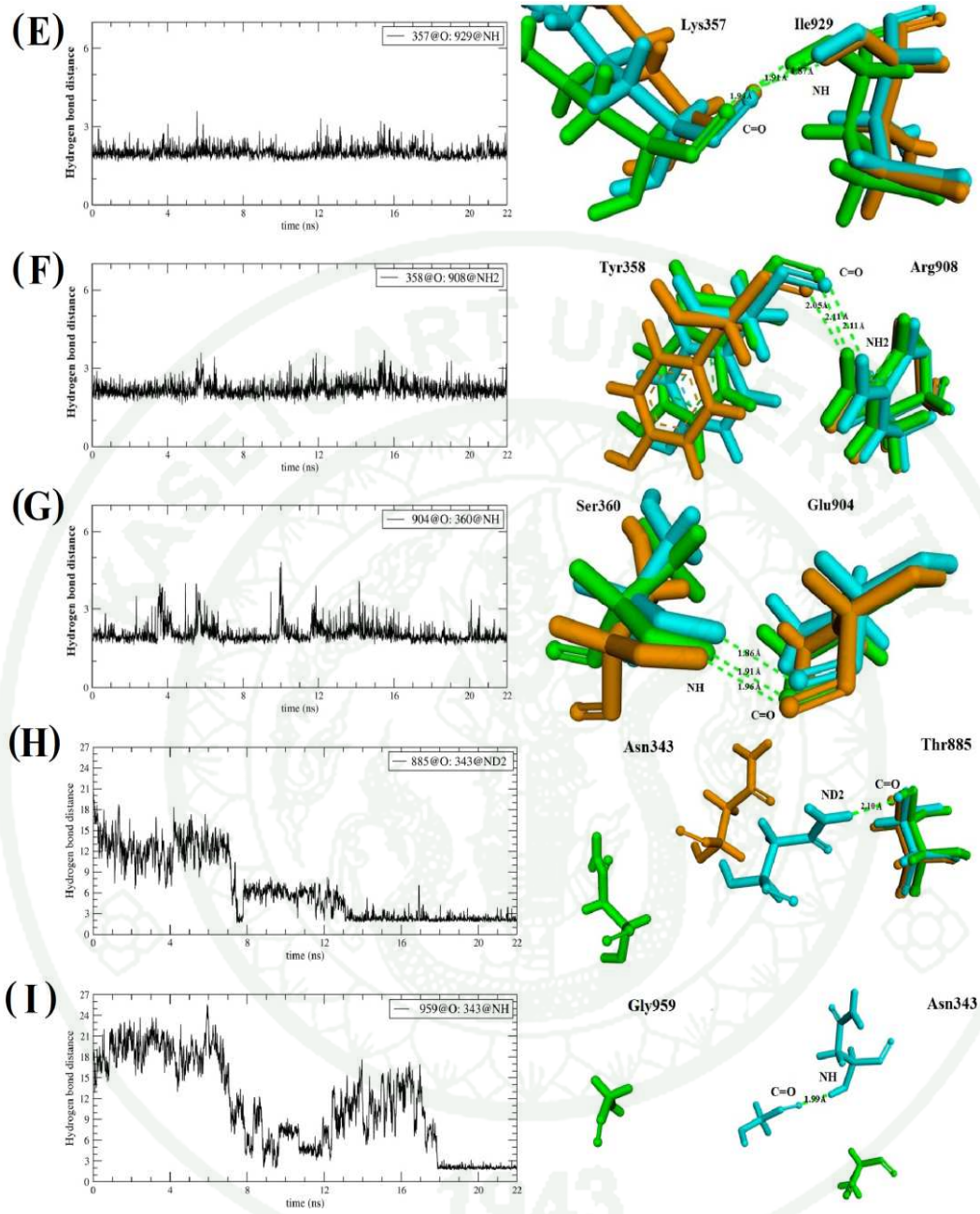


Figure 47 (Continued)

4. The binding free energy calculations

To exposure the binding mode interactions between protein EGFR kinase and MIG-6_s1 peptide, the binding free energy was calculated by MM-PBSA approach based on PB and GB modules in Amber 12 (Case *et al.*, 2005; Salomon-Ferrer *et al.*, 2013). The 300 snapshots sampling from the last 3-ns of the equilibrated simulation were used. The single-trajectory protocol was applied in the MD simulations because it was faster than separate-trajectory protocol (Dong *et al.*, 2008b). Here, we compared the performance of two methods on prediction to binding free energy.

4.1 PBSA variant analysis

As first try, we predicted the binding free energy of complex by using PB model. The results of MM-PBSA approach were presented in Table 5. The total binding free energy was predicted to -142.7 kcal.mol $^{-1}$, including nmode term. The contributions of the molecular mechanics part (ΔE_{MM}) and the solvation energy (ΔG_{sol}) was computed to be -251.13 kcal.mol^{-1} and 144.69 kcal.mol^{-1}, respectively, while the internal energy term (ΔE_{int}) were cancelled for single trajectory calculation (Yang *et al.*, 2012). As revealed in Table 5, van der Waals was the major favorable component to binding (ΔE_{vdw} = -120.75 kcal.mol^{-1}). Together with, the non-polar solvation ($\Delta G_{nonpol,\ sol}$ = -84.25 kcal.mol^{-1}) was also favorable the binding, which the highly favorable non-polar part of solvation free energy probably caused from the hydrophobic effects. The association of the electrostatic contribution ($\Delta G_{ele,\ PB} = \Delta E_{ele} + \Delta G_{PB}$), coulombic term in gas phase (ΔE_{ele} = -130.38 kcal.mol^{-1}) was completely compensated by unfavorable contributions of polar part (ΔG_{PB} =151.59 kcal.mol^{-1}) in the solvation free energy, resulting in totally unfavorable contribution (21.21 kcal.mol^{-1}). However, in protein-protein interactions, columbic interaction is always favorable to binding. For part of entropy, the estimation of entropic component using NMODE method was time-consuming, so we used only 3 of 300 snapshots to compute the average. The entropy results of complex were -41 kcal.mol^{-1}.

4.2 GBSA variant analysis

To reveal the binding mode interactions between EGFR kinase and MIG-6_s1 peptide, the data was generated with 300 snapshots sampling from the last 3-ns of single run using the MM-PBSA approach based on GB module (using the modified OBC-I GB Model) (Onufriev *et al.*, 2004).

The results of calculation were presented in Table 6. The total binding free energy was predicted to -65.39 kcal.mol^{-1}, including nmode value. The contributions of the molecular mechanics part (ΔE_{MM}) and the solvation energy (ΔG_{sol}) were computed to be -251.13 kcal.mol^{-1} and 144.69 kcal.mol^{-1}, respectively, while the internal energy term (ΔE_{int}) was cancelled for single trajectory calculation (Yang *et al.*, 2012). As shown in Table 6, the major favorable components to binding was van der Waals (ΔE_{vdw} = -120.75 kcal.mol^{-1}) term in gas-phase. The non-polar solvation (ΔG_{surf} = -15.51 kcal.mol^{-1}) was also slightly favorable the binding, which was probably caused by the hydrophobic effects. The association of the electrostatic contributions ($\Delta G_{ele,\ PB} = \Delta E_{ele} + \Delta G_{PB}$) presented that the favorable coulombic force in gas phase (ΔE_{ele} = -130.38 kcal.mol^{-1}) was over-compensated by unfavorable contributions of polar solvation (ΔG_{GB} = 160.20 kcal.mol^{-1}), resulting in totally unfavorable contribution (29.92 kcal.mol^{-1}). Our binding free energy value based on GB model in Amber program corresponded to previous study including the binding free energy of MP1-p14 complex was -69.20 kcal.mol^{-1} (Cui *et al.*, 2008) and also the binding of EGFR with lapatinib was -68.12 kcal.mol^{-1} (Ahmed *et al.*, 2013).

4.3 MM-PBSA approach comparative between PB and GB calculation.

The binding free energy was obtained from the sum of gas-phase energies (van der Waals and coulombic energy), solvation free energies (non-polar and polar part obtained from PB and GB models in Amber 12) and entropic terms. The results showed that the binding free energy of MM-PBSA based on PB model was -142.74 kcal.mol^{-1}, higher than -65.39 kcal.mol^{-1} of GB model.

Because the using of the same force field in non-bonded interaction energies of gas phase, thus van der Waals and coulombic electrostatic interactions were similar in two models of the free energy to the binding. Columbic and van der Waals interactions in both model were remained the major favorable forces of MIG-6_s1 peptide binding to EGFR kinase, which can be seen in the study of protein-protein interactions in solution (Chen *et al.*, 2013; Gouda *et al.*, 2003). Additionally, the magnitude of TΔS was the range of +30 to -40 kcal.mol^{-1} that consistency with previous results and indicating a qualitatively reasonable free energy of binding (Huo *et al.*, 2002).

Nevertheless, two methods were different each other in the binding free energy of -87.35 kcal.mol^{-1}. The difference between them was originated from the contribution of the solvation free energy, which was larger in the MM-GBSA than MM-PBSA scheme ($G_{solv,\ GB}$ = 144.69 kcal.mol^{-1} and $G_{solv,\ PB}$ = 67.34 kcal.mol^{-1}) and resulted in higher favorable of binding free energy in PB model (ΔG_{bindng} = -142.74 kcal.mol^{-1}) than GB (ΔG_{bindng} = -65.39 kcal.mol^{-1}) model. However, the difference between PB and GB parameters in the solvation energy could be explained that during the calculation of the solvation energy, the MM-PBSA approach based on PB scheme was more sensitive than the GB scheme.

Athough, as theoretically foundation, PB model is rigorous framework than GB (Feig *et al.*, 2004; Hou *et al.*, 2011), which it does not mean that PB model could give accuracy predictions than GB model but was depended on parameters in applying (Hou *et al.*, 2011). Nevertheless, GB model could predict the binding free energy relative to the binding free energy of other protein-ligand interactions including EGFR bound lapatinib and MP1-p14 complex as mentioned above. Thus, in this study could be suggested that GB-based variance of MM-PBSA procedure might give the binding free energy better reasonable in the binding free energy calculation than PB model. However, in relative binding free energy, MM-PBSA based PB model remains sufficient and popular in the binding affinity of quite different compounds and systems (Merz Jr *et al.*, 2010; Michel *et al.*, 2010; Wan and Coveney, 2011). As previous study of Xu and co-wokers on 2013 reporting that PB model with the Tan's parameters predicts better the capability of ranking

than MM-PBSA based GB model using OBC1 parameter (Xu *et al.*, 2013). Because the accuracy of ranking ligands affinities can still satisfy in the relative binding free energy omitting conformational entropy term in MM-PBSA approach (Hou and Yu, 2007). Thus, to save the time, the entropy calculation could be negotiated in part III.

Table 5 The binding free energy calculations using MM-PBSA calculation.

Contribution	EGFR kinase/MIG-6_s1	EGFR kinase	MIG-6_s1	Delta
E_{ele}	-21099.37	-19392.37	-1576.62	-130.38
E_{vdw}	-2536.90	-2296.92	-119.22	-120.75
E_{gas}	-23636.27	-21689.29	-1695.84	-251.13
G_{PB}	-3449.41	-3449.41	-351.98	151.59
$G_{sol\text{-}np}$	2470.43	2273.28	281.40	-84.25
$G_{sol,\ PB}$	-978.98	-975.73	-70.58	67.34
$\Delta G_{binding}$ [a]				-183.79
Std. Dev[c]				8.61
ΔTS	3583.65	3282.65	342.04	-41.05
$\Delta G_{binding}$ [b] (Nmode)				-142.74
Std. Dev[c]				9.34

All values were given in kcal mol^{-1}

[a] The predictions of binding energy did not include entropy effect

[b] The predictions of binding energy included entropy effect

[c] Std. Dev: Standard Deviation

Table 6 The binding free energy using MM-GBSA calculation

Contribution	EGFR kinase/MIG-6_s1	EGFR kinase	MIG-6_s1	Delta
E_{ele}	-21099.37	-19392.37	-1576.62	-130.38
E_{vdw}	-2536.90	-2296.92	-119.22	-120.75
E_{gas}	-23636.27	-21689.29	-1695.84	-251.13
G_{GB}	-3822.75	-3449.41	-351.98	160.20
G_{surf}	110.34	104.75	21.10	-15.51
$G_{sol, GB}$	-3712.41	-3462.33	-394.77	144.69
$\Delta G_{binding}$[a]				-106.44
Std. Dev[c]				5.41
ΔTS	3583.65	3282.65	342.04	-41.05
$\Delta G_{binding}$[b] (Nmode)				-65.39
Std. Dev[c]				6.05

All values were given in kcal mol^{-1}

[a] The predictions of binding energy did not include entropy effect

[b] The predictions of binding energy included entropy effect

[c] Std. Dev: Standard Deviation

5. Free energy decomposition of MIG-6_s1 residues on EGFR kinase interface

The analysis of per–residue binding energy decomposition played role for understanding the contribution of key residues at the interaction interface responsible for the protein-protein interactions (Tue-ngeun *et al.*, 2013; Wan and Coveney, 2011; Wang *et al.*, 2014). In present study, the 300 snapshots extracted from the last 3 ns of MD trajectory were decomposed by MM-PBSA method. The per-residue basis is contributed into van der Waals energy, the sum of coulombic interaction and polar solvation free energy and the contribution of non-polar solvation. The purpose of this part was identify key amino acids in protein-peptide complex, which we defined as the one making the contribution of binding free energy greater than -1 kcal.mol^{-1} (Bharatham *et al.*, 2011; Guo *et al.*, 2012)

In overview, the residues of MIG-6_s1 peptide and EGFR kinase domain were distributed at discontinuous contact regions based on the energy contribution shown in Figure 48, which the major favorable energy contributions were predominantly originated from Thr885, Thr891, Ile894, Glu904, Gly906, Glu907, Arg908, Leu909, Pro910, Gln911, Pro912, Pro913, Tyr920, Val924, Trp927, Met928, Ile929 and Gly959 of EGFR kinase domain and Ser337, Leu338, Pro339, Tyr341, Leu342, Asn343, Met346, Pro348, Thr349, Gln350, Ser351, Phe352, Asp355, Lys357, Tyr358, Val359, Ser360 and Ser361 residues of MIG-6_s1 peptide, each yielding higher favorable than -1.0 kcal mol^{-1}. Furthermore, the levels of free energy contributions of key residues were clarified, the key residues with the free energy value of > -4 kcal. mol^{-1} were Pro910 of EGFR kinase and Thr349, Gln350, Tyr358 of MIG-6_s1, while most of EGFR kinase and MIG-6_s1 residues revealed the interaction energy in range of -2 to -4 kcal.mol^{-1}, which were found in Thr885, Glu904, Gly906, Arg908, Gln911, Pro913, Met928, Ile929 and Gly959 of EGFR kinase and MIG-6_s1 including Ser337, Leu338, Pro339, Tyr341, Met346, Pro348, Phe352, Lys357, Val359 and Ser361.

As seen in Figure 49, the each key residue yielding > -1.0 kcal.mol^{-1} was further identified the binding free energy components. It was well-known that, the binding free energy contributions were comprised of van der Waals and electrostatic interactions in gas-

phase plus and polar solvation energy. Our results suggest that most of the key residues of EGFR kinase domain and MIG-6_s1 made van der Waal energy contribution for supposing the total binding free energy of each residue (Figure 48). In addition, the electrostatic part in gas phase was totally compensated with the polar solvation part leading to unfavorable contribution for binding in the most residues. however, the polar interaction ($\Delta G_{ele,\ PB}$) could be contributed favorable in several residues including Thr885, Glu904, Pro910, Gln911, Pro912, Trp927, and Gly959 of EGFR and seven residues of MIG-6_s1; Ser337, Tyr341, Thr349, Asp355, Lys357, Val359 and Ser361. This favored polar interaction was mostly originated from hydrogen bond formation proved by the existence of hydrogen bond during simulations.

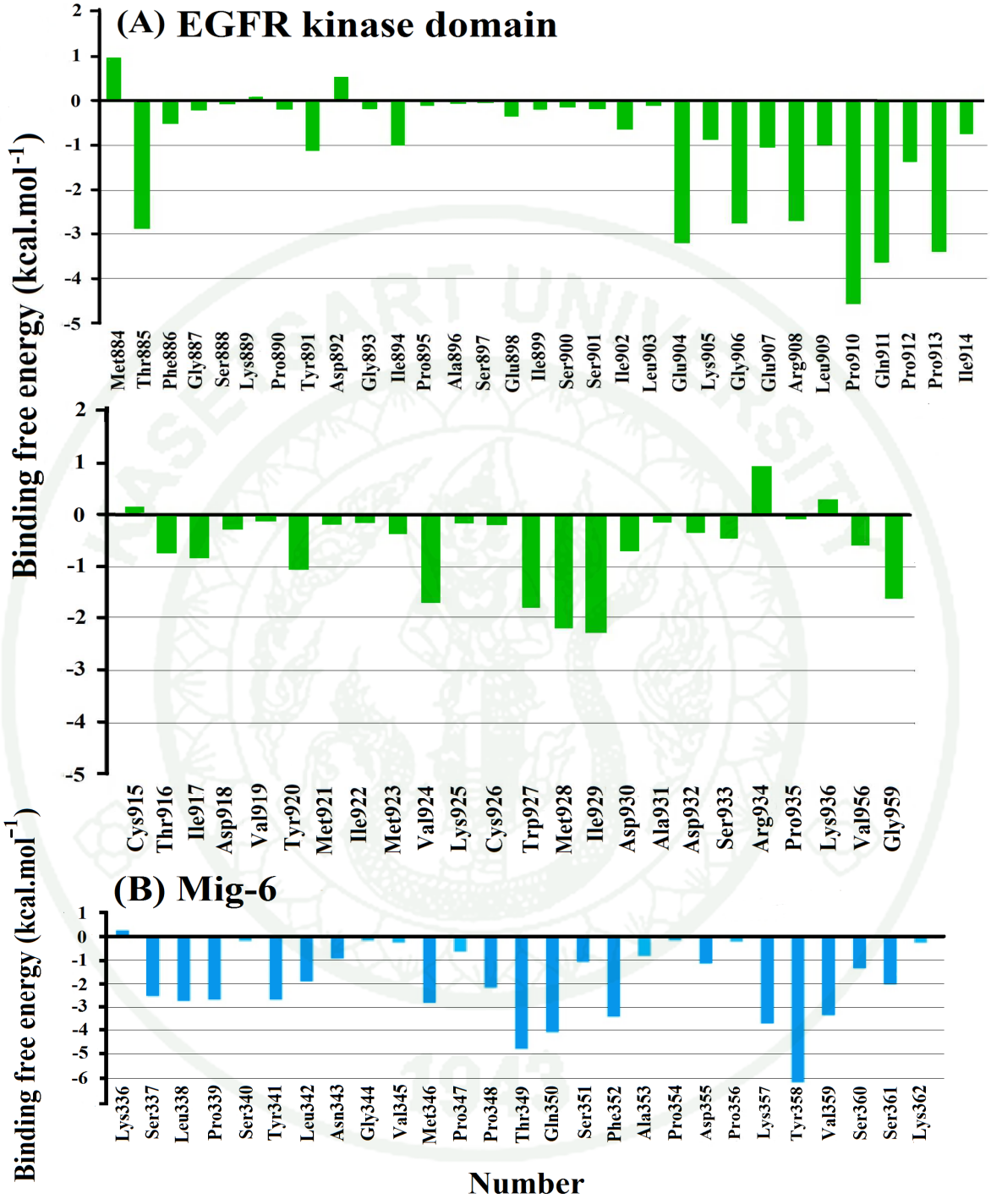


Figure 48 Decomposition binding free energy of EGFR kinase/MIG-6_s1 interfacial complex. Per-residue basis of EGFR kinase (A) and MIG-6_s1 peptide (B).

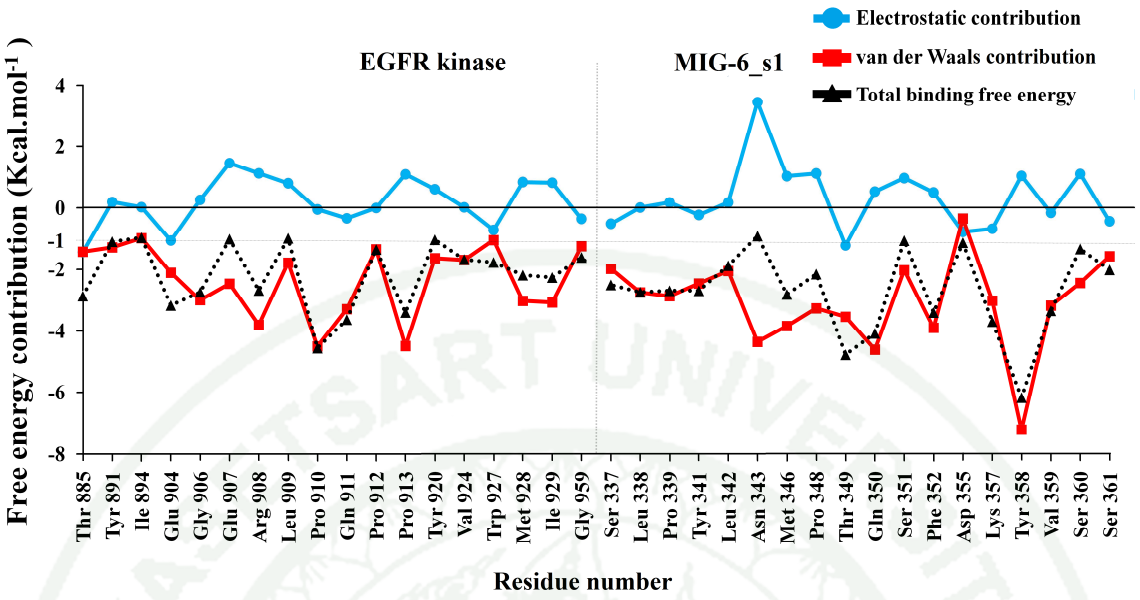


Figure 49 Decomposition free energies components for key residues on interfacial EGFR kinase/MIG-6_s1 complex. The components were van der Waals energy (red), sum of coulombic interaction plus polar solvation free energy (light blue) and the total binding free energy (black dot) for key residues giving energy higher than -1.0 kcal.mol^{-1} of EGFR kinase and MIG-6_s1 peptide. Gray line under zero bar represented the threshold scale of -1 kcal.mol^{-1}.

6. Computational alanine scanning mutagenesis

Computational alanine scanning methodology is powerful method not only monitoring the key residues between two components that responsible for making up interface on interface but also analysis the energetic contributions of the individual side chain upon alanine mutation in term of binding free energy. In previous studied reported that computational alanine scanning method and free energy decomposition protocol could be used together for relief their weakness, for example the possibility of the systems perturbations during virtual mutation and less accuracy in calculation of binding free energy decomposition. So as to validate the essential residues identified from free-energy decomposition analysis, the eighteen and eighteen residues of MIG-6_s1 peptide and EGFR kinase, respectively. Those residues were obtained from Figure 49 with the greater impact higher than -1 kcal.mol^{-1} to further verify the hot spot residues by computational alanine methodology based on MM-PBSA calculations (Guo *et al.*, 2012). Proline and glycine residues were not included due to the difference of their backbone conformations from alanine (Massova and Kollman, 1999).

According to the binding free energy difference between wild-type and mutant is computed as the equation $\Delta\Delta G = \Delta G_{\text{wild-type}} - \Delta G_{\text{mutant}}$, the value of positive and negative referred to favorable and unfavorable substitutions, respectively (Massova and Kollman, 1999). A key residue was defined a strong effect as hotspot if the resulting computed between wild-type and mutant was at least -4 kcal.mol^{-1} and warm spot was a range of -2 to -4 kcal.mol^{-1}, which also significant affects on binding but weaker than hotspot (Moreira *et al.*, 2007a).

By comparing of Figure 49 and Table 7, the results from decomposition per residue basis and computational alanine scanning revealed consistent each other, which most of important residues were also defined as key residues in computational alanine scanning. It could be indicated in the reliability of our results. As seen in Table 7, the summarization of computational alanine scanning results was based on the selected residues with the binding free energy higher than -1 kcal.mol^{-1} from decomposition on per-residue basis. Most of the

selected residues were unfavorable for the binding after alanine substitution. However, W927A and S337A of EGFR kinase and MIG-6_s1, respectively, provided a better for binding than wild-type. Furthermore, Table 7 showed the decreasing of the interaction energy greater than -2 kcal.mol^{-1} or namely “Hot spot” and “Warm spot” presenting as bold and underline. These important were defined as the essential residues for the protein-protein or protein-peptide binding. Six residues of EGFR kinase including E904A, E907A, R908A, E911A, M928A, I929A and eight residues of MIG-6_s1 peptide including L338A, Y341A, N343A, M346A, T349A, Q350A, F352A, Y358A made large unfavorable of > -4 kcal.mol^{-1} called hot spot. Warm spots residues, which also affects to binding but the important role was weaker than hot spot residues consisted of T885A, I894A, Y920A, V924A of EGFR kinase and L342A, V359A, S361A of MIG-6_s1.

As seen Figure 50 illustrated the interface of EGFR kinase/MIG-6_s1 complex (A) and the key residues were presented on the surface of EGFR kinase (B) and MIG-6_s1 (C). The results of computational alanine scanning were correlated with binding free-energy decomposition results and some of these was proven and agreed with the experimental study of Zhang and co-workers. They reported that Val 924 located on the C-lobe of EGFR kinase domain and Met346, Phe352 and Try358 of MIG-6_s1 peptide were participated in the interactions (Zhang *et al*., 2007). The results of experimental Ala replacement of Val924, Met346, Phe352 and Try358 resulted in the abolishing of the complex formation. It can be indicated that our analysis was proven and reliable, while the remainder important residues from Figure 50 might be the important residues the same as the experimental study. However, the residues including Tyr891, Gly906, Leu909, Trp927, Pro910, Pro912, Pro913 and Gly959 of EGFR kinase and Ser337, Pro339, Pro348, Ser351, Asp355, Lys357, Ser360 of MIG-6_s1 peptide were recognized as crucial residues from decomposition per-residue analysis only. In this point could be suggested that the contribution of them might cause from the backbone atoms more than the side-chain, while alanine substitutions method probe only the importance of side-chains of residues.

Table 7 Computational alanine scanning results of EGFR kinase/MIG-6 _s1 complex.

Contributions[a]	**The muated residue number**									
	T885A	Y891A	I894A	E904A	E907A	R908A	L909A	Q911A	Y920A	V924A
$\Delta\Delta E_{vdW}$	-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
$\Delta\Delta G_{non\ pol,\ PB}$	-0.14	-0.24	-1.20	-1.16	-1.73	-0.78	0.00	-1.30	-1.24	-0.93
$\Delta\Delta G_{PB}$	-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
$\Delta\Delta G_{subtotal,\ PB}$	**-2.10**	-1.25	**-2.88**	**-4.75**	**-4.43**	**-15.65**	-0.60	**-7.48**	**-3.65**	**-2.75**
	W927A	M928A	I929A	S337A	L338A	Y341A	L342A	N343A	M346A	T349A
$\Delta\Delta E_{vdW}$	-0.70	-3.44	-3.91	-0.43	-3.10	-3.46	-1.40	-4.65	-5.97	-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
$\Delta\Delta G_{non\ pol,\ PB}$	1.42	-2.22	-1.54	-0.03	-2.54	-1.20	-1.08	-2.19	-3.12	-1.82
$\Delta\Delta G_{PB}$	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
$\Delta\Delta G_{subtotal,\ PB}$	0.24	**-5.30**	**-6.69**	0.05	**-4.80**	**-7.94**	**-2.41**	**-4.91**	**-8.11**	**-7.35**

Table 7 (Continued)

Contributions[a]	**The mutated residue number**								
	Q350A	S351A	F352A	D355A	K357A	Y358A	V359A	S360A	S361A
$\Delta\Delta E_{vdW}$	-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
$\Delta\Delta G_{non\ pol,\ PB}$	-1.24	-0.07	-3.39	-0.00	-1.44	-4.57	0.41	-0.20	-0.35
$\Delta\Delta G_{PB}$	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
$\Delta\Delta G_{subtotal,\ PB}$	**-4.33**	-0.95	**-8.56**	-0.97	-1.92	**-9.95**	**-2.11**	-0.11	**-3.02**

[a]The alanine scanning of EGFR residues (Thr885-Ile929) and MIG-6_s1 residues (Ser337-Ser361)

$\Delta G_{subtotal,\ PB} > -2$ kcal.mol^{-1} values were displayed in bold and underlined.

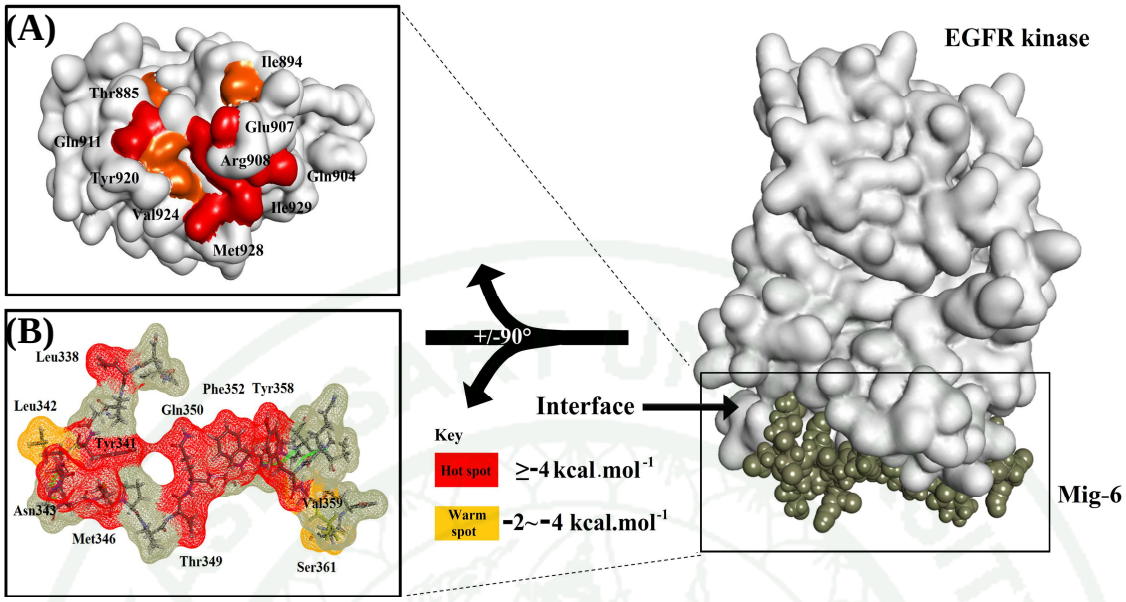


Figure 50 Surface represention of key residues on EGFR kinase/MIG-6_s1 interface. Key residues located on the interface between EGFR kinase (A) and MIG-6_s1 (B). The color of each residue referred to the levels of binding contributions. Higher than -4 kcal.mol^{-1} is defined as hot spot (red), which shows the most important residue. The free energy of -2 to -4 kcal.mol^{-1} is defined as warm spot (yellow) also being important residue but weaker than hot spot.

7. Binding mode of MIG-6_s1 key residues on interface

In order to insight the effect of key residues binding to the interface of complex, the binding mode of key residues were predicted the relationship between their structure. We used the residues-based on hydrogen bond, hydrophobic interaction, the free energy decomposition and computational alanine scanning results for explaining the phenomenon.

From the average structure information, as five proline residues on turn of MIG-6_s1 could stabilize on time evolutions and be stabilized by aromatic ring, which this phenomenon might related to the stability of MIG-6_s1 conformation. In addition, we found the most important residues on interfaces including six residues of EGFR kinase and eight residues of MIG-6_s1 peptide that their side-chains were responsible for binding with the energy contribution stronger than -4 kcal.mol^{-1}. Most of these residues were also involved in the main binding attractions from hydrogen bonding and/or hydrophobic interactions.

As the conformational change of loop region as monitored, it is worth paying attention to Asn343, which the side-chain could shift closely to EGFR kinase surface and formed two hydrogen bonds with Thr885 and Gly959. This capability could lead to powerful in polar interaction. Although, the distrubtion of Asn343 side-chain could be caused in the dropping binding free energy into -4.9 kcal.mol^{-1} after alanine substitution. Most of dropping was originated from the loss of van der Waals interaction due to the increasing of cavity volume after alanine substitution.

The hydrophobic interaction is important force in protein-protein interactions. Met346 produced the interaction free energy of -2.8 kcal.mol^{-1}. The important of Met346 was originated from the capability of side-chain forward into hydrophobic cleft that consisted of Trp881, Thr885 and Pro913 residues of EGFR kinase (Figure 51). The distrubtion of Met346 side-chain could lead to a large dropping in binding free energy (-8.1 kcal.mol^{-1}) upon alanine replacement, which was originated from the loss of van der Waals contribution in hydrophobic interaction. It was suggested that the side-chain of Met346 residues was crucial for binding to hydrophobic pocket of EGFR kinase. Corresponding to

the crystal structural reported from previous study, Met346 was oriented toward the hydrophobic pocket formed by residues of EGFR kinase and the mutation of M346A could abolish EGFR binding (Zhang *et al*., 2007).

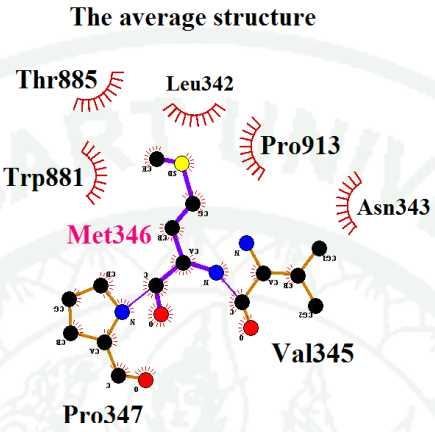


Figure 51 Hydrophobic interaction of Met346. The pointing outward the hydrophobic cleft of EGFR kinase were plotted by LIGPLOT program. The average structures shown as ribbon with purple color and its relevant residues were brown. Carbon, nitrogen and oxygen atom were exhibited as black, blue and red balls. Green dash line presented hydrogen bonds and dark red “eyelashes” were hydrophobic contacts.

It is well-known that electrostatic contributions are important forces involved in complex formation. Hydrogen bond is electrostatic origin, thus most of important residues were observed hydrogen bonds. First of all, Tyr341 was critical for EGFR kinase binding with per residue contribution of -1.88 kcal.mol^{-1}. The important of Tyr341 was contributed from the side-chain, which not only hydrogen bond could be constructed with Gln911 in MD processes but the hydrophobic pocket with Pro910, Pro913 of EGFR kinase and Pro339 and Pro347 of MIG-6_s1 (Figure 52A). Therefore, the distrubtion of Tyr341 side-chain upon alanine replacement could drop the binding free energy with the contribution of -7.9 kcal.mol^{-1}.

Next key residue was Thr349 polar residue, defined as hot spot after alanine substitution. The important of Thr349 was the contribution of side-chain to from hydrogen

bond with Glu907 polar residue on interface (Figure 52B). The distrubtion of it side-chain could large drop the binding free energy into -7.35 kcal.mol^{-1}, which originated from the loss of electrostatic interaction the most.

The residue Gln350 could contribute to binding free energy of -4.08 kcal.mol^{-1}. These contribtution was not only originated fron the capability of backbone to act in both donor and receiver H atom to form HBs with Arg908 but the side-chain of Gln350 could.also form HB with residue Gln911 similar to Tyr341. Moreover, the side-chain of Gln350 was oriented toward the hydrophobic environment created by Phe352, Pro910 and Try920 (Figure 52C). Therefore, the alanine replacement resulted to dropping the free energy into -4.3 kcal.mol^{-1}, which caused by the distrubtion of side-chain feature.

In addition, two residues Phe352 and Tyr358 reported as the important residues with interaction of -3.4 and -6.2 kcal.mol^{-1}, respectively. The importants of these residues were not only originated from the abilities of their backbones to form hydrogen bonds but the side-chains could also plug into hydrophobic cloud (Figure 53). Moreover, the phenyl of Phe352 and the phenol of Tyr358 side-chains could create the paralleled π-π interaction each other with the distance between centroid-centroid of aromatic rings less than 12.0 Å in proteins (Figure 54). The importants of Phe352 and Tyr358 residues were also reported in experimental alanine mutagenesis study that the mutations of Phe352A and Try358A could disrupt the binding (Zhang *et al*., 2007). Nevertheless, the distrubtion of their side-chains could be large dropping in binding free energy contribution into -8.96 and -9.95 kcal.mol^{-1}, respectively, which originated from the loss of van der Waals interaction the most.

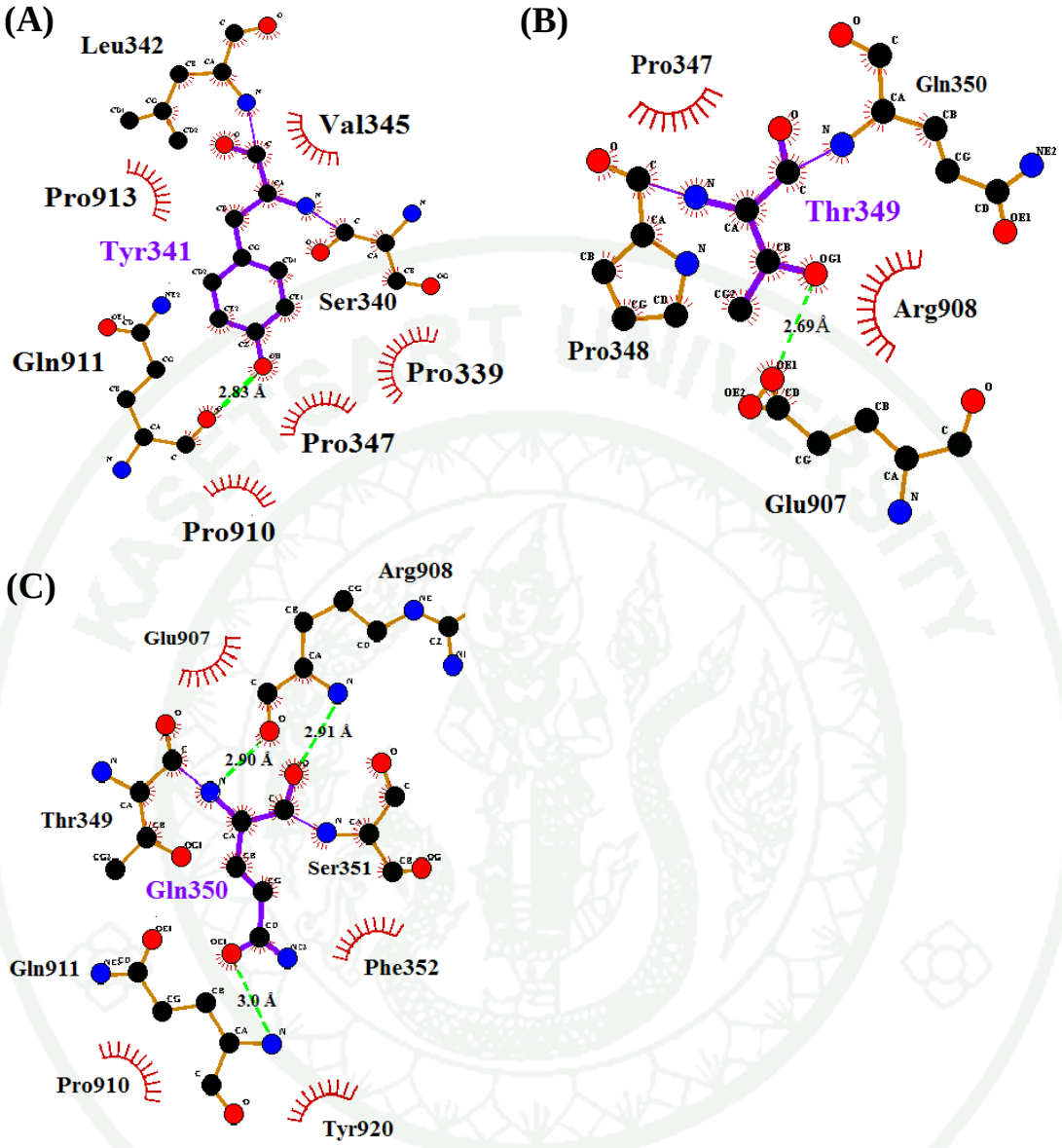


Figure 52 The interactions of key residues of MIG-6_s1. The hydrogen bonding and hydrophobic interaction of Tyr341 (A). The hydrogen bonding between EGFR kinase Glu907 polar residues to Thr349 polar residues of MIG-6_s1 (B). The hydrogen bonding and hydrophobic interaction of Gln350 (C). These were plotted by LIGPLOT program.

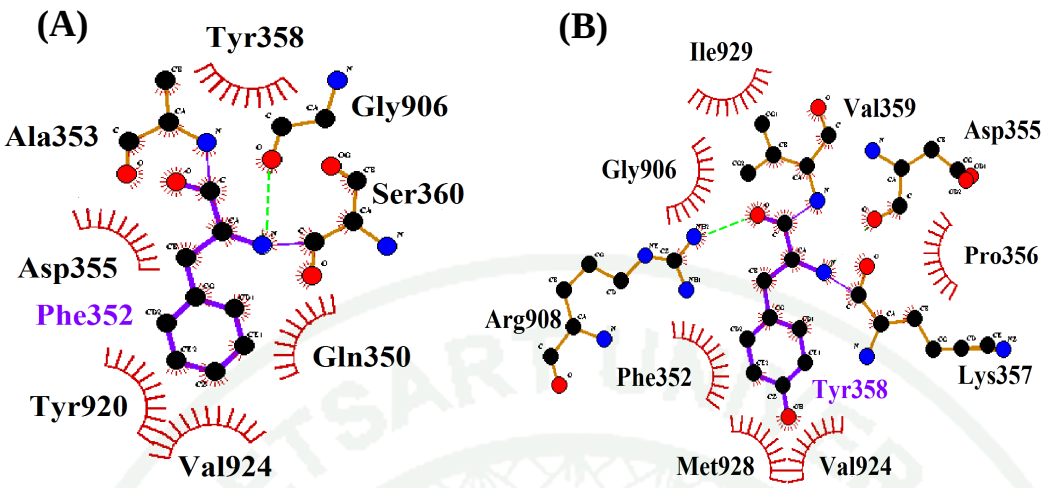


Figure 53 Hydrophobic interaction of Phe352 and Tyr358. The pointing outward the hydrophobic cleft of EGFR kinase by Phe352 (A) and Tyr358 (B) were plotted by LIGPLOT program. The average structures shown as ribbon with purple color and its relevant residues were brown. Carbon, nitrogen and oxygen atom were exhibited as black, blue and red balls. Green dash line presented hydrogen bonds and dark red “eyelashes” were hydrophobic contacts between residues to Phe352 and Tyr358.

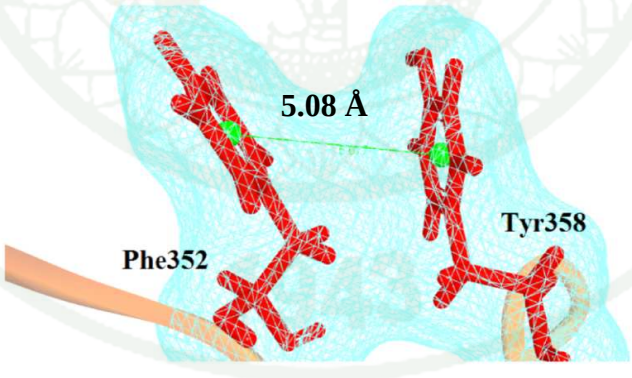


Figure 54 The paralleled π-π interaction between the phenyl of Phe352 and the phenol of Tyr358. The distance between centroid-centroid of two aromatic rings was 5.08 Å.

8. Comparison of the EGFR dimer interfaces and MIG-6_s1 binding interface

According to the helping of juxtamembrane (JM) segment of receiver can to stabilize asymmetric dimer. The spanning of residues 664-682 in JM-B portion can cradle the C terminal lobe of activator in asymmertric TKD dimer and residues of 672-682 were crucial for EGFR activation as previous studies (Zhang *et al*., 2006). As seen in Figure 55, EGFR kinase of receiver could bind to the C-lobe of EGFR activator in the same manner to MIG-6_s1 peptide by laying their structures across to the interface of kinase activator structure.

From the results of decomposition binding free energy per-residue presented that kinase activator made 12 predominant residues favor to binding with receiver. The 12 residues gave binding free energy higher than -1 kcal.mol^{-1} including Lys905, Gly906, Thr916, Ile917, Tyr920, Met921, Val924, Trp927, Met928, Arg949, Gln952 and Arg953 (Figure 56A). Detailed view of the interface between activator and receiver, Tyr920, Val924, Trp927 and Met928 interfacial residues on the C-lobe of activator was shared not only important for binding to receiver but also crucial for MIG-6_s1 peptide (Figure 56B). Moreover, Tyr920, Val924 and Met928 residues were considered as the core of asymmetric EGFR dimer (Figure 57). The mutations of these residues also abrogated the contribution of asymmertric dimer interface (Zhang *et al.*, 2006). It was distrinct that MIG-6_s1 could bind EGFR kinase domain similar manner to EGFR kinase receiver resulting in blocking the asymmertric dimer formation (Zhang *et al*., 2007). Meanwhile, the kinase activator can be found many hydrogen bonds to MIG-6_s1, some of the hydrogen bond was also found between activator and receiver dimer interface. Gly906 and Arg908 residues of interfacial kinase activator could form hydrogen bonds to the receiver the same as observed in MIG-6_s1 (Figure 58). In addition, Arg949 and Arg953 were crucial for activator binding to receiver based on per-residue basis. The mutations of Arg949 and Arg953 were reported by Jura and co-workers that they can affect to EGFR auto-phosphorylation (Jura *et al.*, 2009)**.** However, the activated MIG-6_s1 peptide plays role in inhibiting EGFR kinase into inactive form, opposed to the role of receiver in supporting asymmetric dimer formation (Jura *et al.*, 2009).

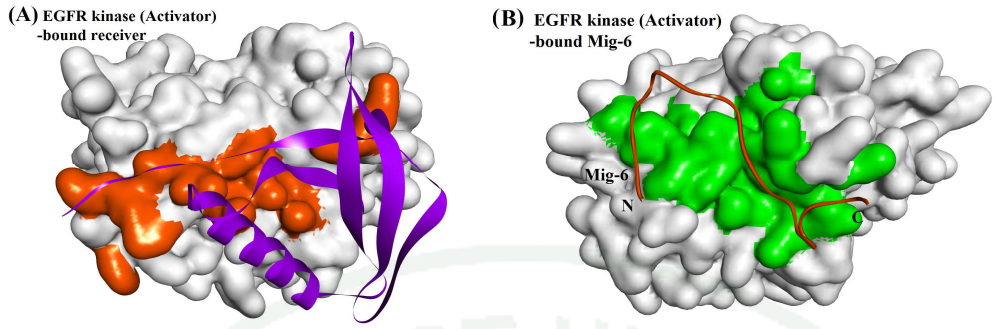


Figure 55 Surface represention of the binding of EGFR kinase (activator) to EGFR kinase (receiver) (A) and MIG-6_s1 peptide (B).

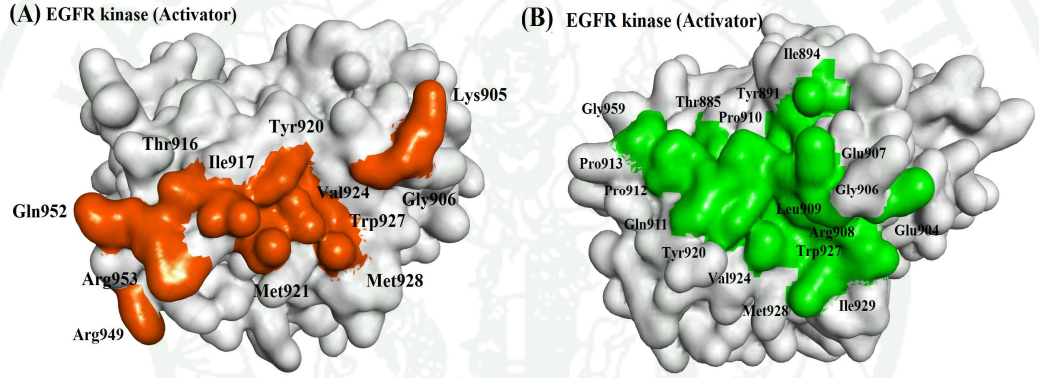


Figure 56 Surface represention of the important residues on interfacial EGFR kinase (activator) bound EGFR kinase (receiver) (A) and MIG-6_s1 peptide (B).

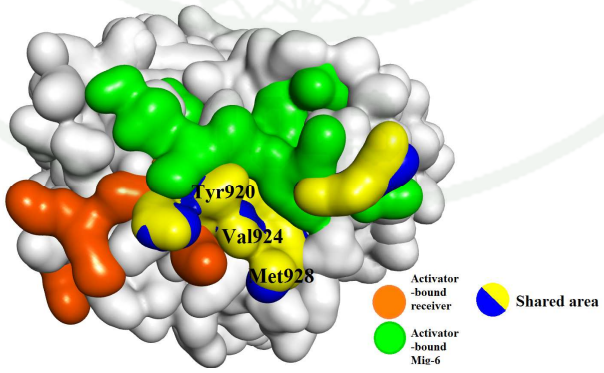


Figure 57 The core of asymmetric EGFR dimer on EGFR kinase activator. The shared area between receiver and MIG-6_s1 on EGFR kinase activator interface was also shown.

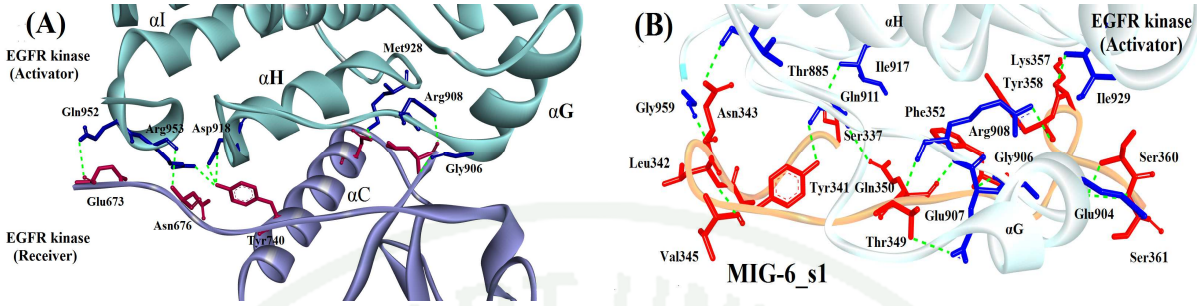


Figure 58 The hydrogen bond constructions on EGFR kinase activator bound with EGFR kinase receiver (A) and MIG-6_s1 (B) at the average structures.

In summary of part I, as the designing optimal kinase-inhibitory compounds needs an understanding of EGFR kinase structure binding partners. Therefore, in first part, a 22-ns MD simulation of EGFR/MIG-6_s1 complex was performed with the aim carrying out for the better understanding of MIG-6_s1 peptide binding to EGFR kinase domain. The conformational changes and interaction between both proteins interface were monitored. The results form superimposition indicated that Asn343 of MIG-6_s1 could from two hydrogen bonds with Thr885 and Gly959 of EGFR that absent in the initial-minimized structure and influenced on the conformational change in loop region at the averaged structure of 20-22 ns molecular dynamics simulations. In part of energetic properties performed by combination with MMPBSA calculation, Van der Waals contribution was the main driving force for the binding. The favorable electrostatic contribution in gas phase was compensated with unfavorable polar solvation leading to unfavorable total electrostatic interaction as seen in phenomenon of other protein-peptide interactions.

The hot spot residues were identified by per-residue contribution and proven by computational alanine scanning. The results agreed with each other that six residues of EGFR kinase including Glu904, Glu907, Arg908, Gln911, Met928, Ile929 and eight residues of MIG-6_s1 including Leu338, Try341, Asn343, Met346, Thr349, Gln350, Phe352 and Try358 decreased the binding affinity upon alanine substitutions. It could be implied that these residues played role in the binding affinity. Most of these residues could be also monitored the hydrogen bond formations, which could explain the influential of the hydrogen bonds to the stability of complex formation. However, Tyr891, Gly 906, Leu909, Pro910, Pro912, Pro913, Trp927 and Gly 959 of EGFR kinase and Ser337, Pro339, Pro348, Ser351, Asp355, Lys357 and Ser360 of MIG-6_s1, which some of these was limited by alanine scanning protocol and some of these was recognized as crucial residues from decomposition per-residue analysis only. Because the alanine substiutions method probed only the important of side-chains of residues and the contribution of them might cause from the backbone more than the side-chain atom. From this, it could be implied that these residues might help protein structure to the stability more than the binding affinity. In addition, Tyr920, Val924, Trp927 and Met928 of EGFR kinase surface could be shared by MIG-6_s1 and EGFR kinase receiver. It was predicted that MIG-6_s1 could bind EGFR

kinase domain similar to EGFR kinase receiver leading to blocking the asymmertric dimer formation.

The first part of this study could guide us for better understanding of binding interactions between EGFR kinase and MIG-6_s1 peptide on their interfaces. It might be considered as the first step in order to design EGFR kinase inhibitors by MIG-6_s1 peptide modeling to new targeting interfacial EGFR kinase with high affinity.

Part II: Design strategies for short peptides based on MIG-6_s1 peptide structure

This part toke the advantage of the knowledge from MD simulations of the EGFR kinase/MIG-6_s1 complex to design short peptide based on MIG-6_s1 modeling for highly affinity to EGFR kinase (Figure 18). We had a previously reported that a 30-residue fluorescein-labelled MIG-6_s1 peptide (residues 334-363, spanning on the segment1) could bind to EGFR kinase domain with micromolar affinity and the dissociation constant was 13.0 ± 1.3 μM. Therefore, MIG-6_s1 (337-361) could inhibit EGFR kinase activity by blocking asymmetric dimer formation (Zhang *et al*., 2007).

In this study, MIG-6_s1 was designed into short peptides and predicted the interactions feature to asymmertric dimer on C-lobe of EGFR kinase domain based on computational tool. Concepts to design short peptides derived from MIG-6_s1 structure were described as below.

1. The optimization of MIG-6_s1 sequence

Based on the contact residues strategy (Rudgers *et al.*, 2001), we selected and considered the number of residues based on these results.

1.1 The result of decomposition binding per residue gave the discontinuous region of binding free energies into 3 partitions (Figure 48B). Thus, it was be possible to separate the residues of MIG-6_s1 up to these partitions with containing 9-mer including Lys336-Gly344, Val345-Ala353, Pro354-Lys362 and 18-mer including Lys336-Ala353, Val345-Lys362. The total binding affinities of five segments were observed and revealed as -56.21, -82.34, -57.23, -145.06, -119.74 and -142.49 kcal.mol^{-1}, respectively.

1.2 After MD simulations of these segments, the rmsd fluctuations of these were plotted and revealed the high average of rmsd values and fluctuations (data not shown).

1.3 In a 27-mer wild-type: Leu338, Try341, Asn343, Met346, Thr349, Gln350, Phe352, Tyr358 and Val359 were detected as important residues after alanine substitution approach. Met346, Phe352 and Tyr358 were first reported the important for EGFR kinase binding in experiment studied by Zhang and co-workers (Zhang *et al*., 2007).

1.4 As the structural instability of short peptide could be geneally observed in solution, we extend the range of residues in order to increase the stability and remain to give high binding free energy. A 20-mer contained Tyr341-Lys360 was a suitable candidate for further investigation, which it could be explained as below. These residues in a 27-mer wild-type covered the most of hydrogen bonds forming with the important residues of EGFR kinase (Figure 46B) that defined as keystone residues (Figure 48A). After MD simulations, a 20-mer wild-type structure was stable than orther segments and gave binding free energy at high level of -142.49 kcal.mol^{-1}.

Based on the results as mentioned, we considered and compared in the structure and function of each segment. However, because the shorter MIG-6_s1 peptide sequences (9- and 18-mer) had the structural instability in solution and the sequences were not covered some of the important residues. Therefore, a 20-mer of Tyr341-Lys360 was a suitable candidate for further investigation.

2. *In silico* site directed mutagenesis

In order to enhance the binding affinity of MIG-6_s1 to EGFR kinase and after the obtaining of a 20-mer segment, the mutagenesis was further used to study the effect of the optimized sequence to the structure and energy information.

Before selecting the mutated positions, the MD simulations of a 20-mer was preliminary investigated. A 16-ns MD simulation time was run and scanned for the important interacting formations in order to support the information from a 27-mer wild-type before considering the positions of mutation.

We considered and mutated the interested positions and residues as below (Betts and Russell, 2003).

2.1 In 20-mer wild-type model, we found the interesting that Thr349 of MIG-6_s1 could form hydrogen bond to the negative charged residues of EGFR kinase (Glu907) on interface and this bond pair was also conserved in a 27-mer native structure. This residue was considered to arginine substitution in order to construct salt-bridge with Glu907 and increase the electrostatic potential from bridge formation.

2.2 In the initial-minimized structures of a 27-mer and 20-mer wild-type models, Gln350 had a conserved hydrogen bond with positively charged amino acid (Arg908) of EGFR kinase. Based on the formation of charge-charge interaction, Gln350 was mutated by glutamic acid residues and expected that this residue could form salt-bridge with Arg908 leading to increasing the electrostatic potential.

2.3 In the initial-minimized structure of a 27-mer and a 20-mer wild-type model, Ser351 could construct hydrogen bond to Glu907. Together with this residue in a 27-mer native structure gave the favorable for alanine substitution, thus Ser351 was candidate to alter the side-chain with arginine. The same as T349R, the mutation of serine at position of 351 to arginine was expected to increase the electrostatic potential by charge-charge formation.

2.4 Lys357 had hydrogen bond in both of a 27-mer and a 20-mer wild-types models. Thus, we considered to mutate the positively charged lysine with arginine residue. Because the capability of arginine side-chain (guanidinium group) could form the multible hydrogen bonds that this interaction could support favorable contribution. In addition, arginine was preferred to substitute for lysine. So, this mutation was expected the changing to longer side-chain of arginine could not only maintain hydrogen bond with Ile929 but it also created hydrogen bonds with the neighboring residues.

To summarize, as we known that the electrostatic interaction was the major force in the stability of protein on surface (van den Burg and Eijsink, 2002). Thus, we considered to increase the electrostatic force by mutating the uncharged polar amino acid of MIG-6_s1 (Thr349, Gln350 and Ser351) and positively charged amino acid (Lys357) replacing with the charged polar residue in terms of single and double mutations for making salt-bridge through hydrogen bond interaction and increasing the possibility of hydrogen bonds formation with other residues (Strub *et al.*, 2004). Because the geometric effect of positively charged in arginine could make a larger number of electrostatic interactions (salt-bridges and hydrogen bonds) and generate stable ionic interaction than other positively charge residues (Sokalingam *et al*., 2012). Therefore, we considered arginine to replace wild-type residues at Thr349 Gln350 and Ser351 for creating salt-bridge interaction and also wild-type residue at Lys357 for increasing the number of hydrogen bonds. These mutated positions were considered in terms of four single (T349R, Q350E, S351R and K357R) and six double mutations (T349R/Q350E, T349R/S351R, T349R/K357R, Q350E/ S351R, Q350R/K357R and S351R/K357R). However, the result showed that the pair of double mutant T349R/S351R gave the most binding affinity (-163.8 kcal.mol^{-1}) than a 20-mer wild-type (-142.49 kcal.mol^{-1}) and other mutations (data not shown). Therefore, the double mutation of T349R/S351R together with their single mutations were further studied in the detail of interactions and the effects of mutagenesis on binding affinities comparing to wild-type in Part III. The initial-minimized structure of a 20-mer wild-type (Figure 59A) was compared to the selected mutations (Figure 59B, 59C and 59D).

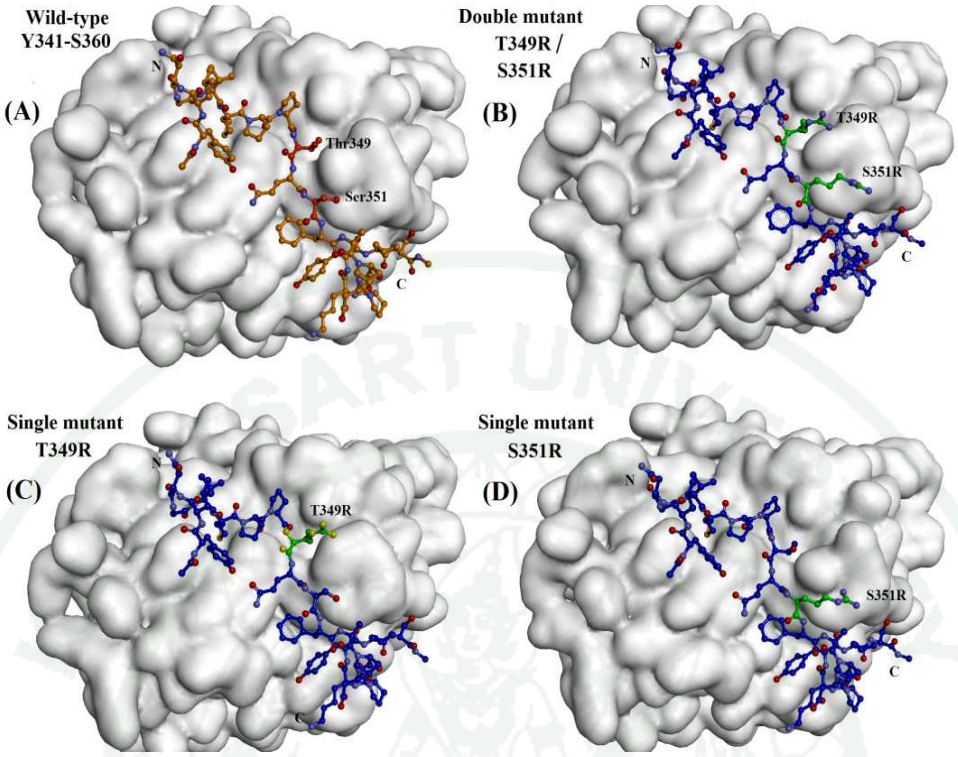


Figure 59 The initial structures of EGFR kinase binding with a 20-mer wild-type and mutant variants. A 20-mer wild-type (A), double mutation of T349R/S351R (B) and single mutation of T349 (C) and S351 (D) were shown. The positions of the mutations were presented in green.

Part III: MD simulations analysis of EGFR kinase and short peptides derived from MIG-6_s1 peptide structure

After analysis of part I and II, in part III, we performed the simulations of wild-type and mutant models. There were four 20-mer short peptide sequences derived from a 27-mer of MIG-6_s1 staring structure, the sequences were shown as below.

A 20-mer wild-type: ACE-^{341}YLNGVMPPTQSFAPDPKYVS^{360}-NME

Double mutant T349R/S351R: ACE-^{341}YLNGVMPP**R**Q**R**FAPDPKYVS^{360}-NME

Single mutant T349R: ACE-^{341}YLNGVMPP**R**QSFAPDPKYVS^{360}-NME

Single mutant S351R: ACE-^{341}YLNGVMPPTQ**R**FAPDPKYVS^{360}-NME

1. Molecular dynamics qualities of the protein structures

Because we changed the type of amino acids, the stereochemical quality of protein structures should be checked by the ramachandran plot before using them in next step (Ramachandran *et al.*, 1963). These plots were outputted by PROCHECK program. (Laskowski *et al.*, 1993). The ramachandran plot statistics were shown as Table 8 and Appendix Figure A2. Because the major residues were located on the most favored regions and only few was located on the disallowed regions, thus all structures were interpreted as good models that were acceptable and reliable for further analysis.

To further validate the equilibrium, the total energy (ETOT), potential energy (EPTOT) and kinetic energy (EKTOT) fluctuations of four 20-mer of wild-type and mutant complexes were monitored along the MD simulations. The results presented those three eneries of a 20-mer wild-type and mutant systems turned into equilibrium before entering into the production run and stabilized through simulations. These results showed that the systems were reliable (Figure 60).

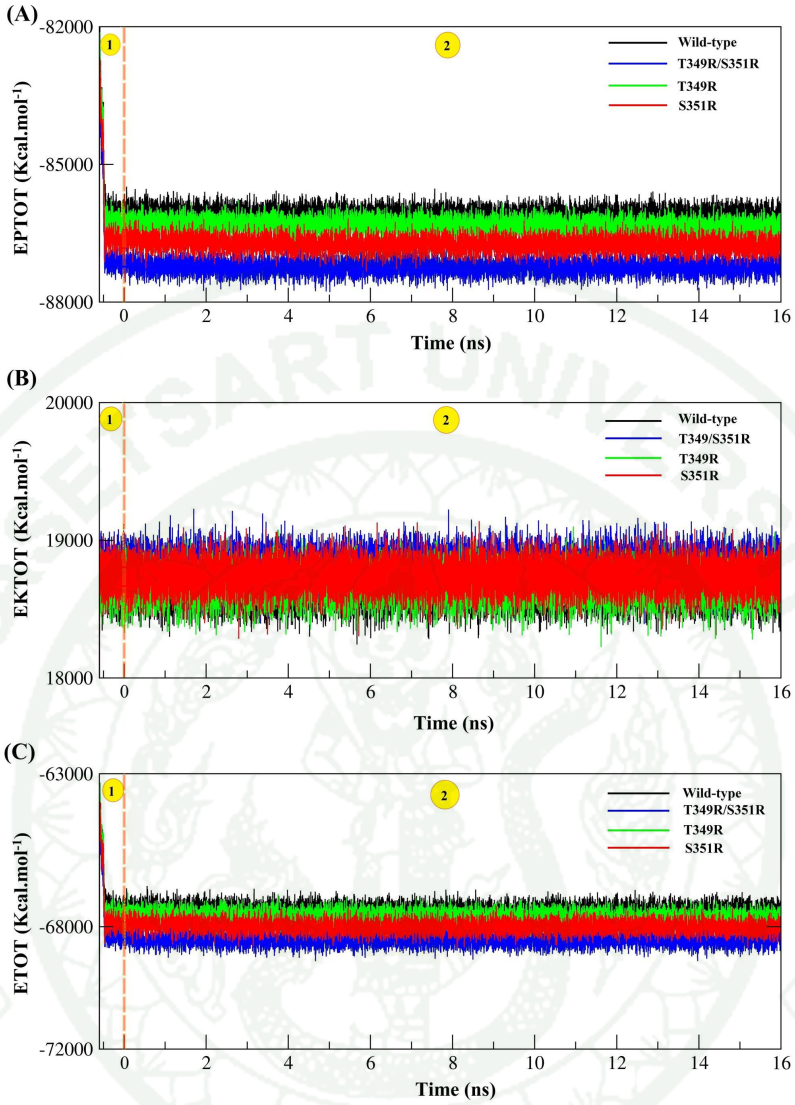


Figure 60 The energies plotting of four systems. The potential energy; EPTOT (A), the kinetic energy; EKTOT (B) and the total energy; ETOT (C) were plotted starting from heating to equilibration steps through the production runs (no.1 and no.2) for EGFR kinase/a 20-mer wild-type (black), EGFR kinase/ double mutant T349R/S351R (blue), EGFR kinase/single mutant T349R (green) and EGFR kinase/single mutant S351R (red) complexes.

Table 8 Ramachandran plot statistics for 20-mer of wild-type and three mutants complexes. The percentage of residues in the most favored, additional allowed, generously allowed and disallowed regions were presented below.

Peptides	The most favored regions	The additional allowed regions	The generously allowed regions	The disallowed regions
Wild-type	79.8%	19.0%	0.8%	0.4%
T349R/S351R	79.8%	19.0%	0.8%	0.4%
T349R	79.8%	19.0%	0.8%	0.4%
S351R	79.8%	19.0%	0.8%	0.4%

Another one important parameters was the root-mean-square deviation (RMSD). The stabilities of MD simulations were judged by the convergence of RMSD plot based on the backbone (Cα, N and O) atoms depending on time. The initial coordinates of the structures were defined after energy minimization. The average RMSD value of 20-mer wild-type and three mutants systems were shown in Table 9 and Figure 61. From RMSD plot showed that the overall complex structures in both of a 20-mer wild-type and their mutants were stable along MD simulations with the average RMSD value less than 3 Å. A 20-mer wild-type complex showed the most stable with the average value of 2.02 Å, as expected of a stabilized native protein conformation. Although, the average RMSD values of three 20-mer mutant complexes were higher than a 20-mer wild-type suggested that the overall structures might be affected by single and double mutations. Among three mutant systems, we interested in seeing that the S351R mutant structure in complex showed the most stable (2.82 Å), whereas double mutant strucrture showed the least stable in the average RMSD value (4.6 Å). However, the complexes of these two systems were stablilized by EGFR kinase, resulting in high stablility within 3 Å (2.62 and 2.55 for single S351R and double mutant structures, respectively). It could be implied that after introducing EGFR kinase and short peptide model in system, there was not effect to the global conformational change in the complex relatived to the initial minimized structures and seem not affecting to the complex stability.

In addition, largest of the average RMSD reaching 4.6 Å in double mutant might be affected by two arginines substitutions, due to a long flexible side-chain with positively-charged end. Consistency in the reporting of Vaughan and co-workers, the mutation was resulted in the structure shift that could be seen in double and triple mutations (Sinha and Nussinov, 2001; Vaughan *et al.*, 2002).

Next, the detailed analysis of Root-Mean-Square-Fluctuation (RMSF) versus a 20-mer residue for wild-type and three mutants were monitored with the calculation of the 300 snapshots from the last 3 ns of simulation time (Figure 62). RMSF respected to the average structure by residue or atom, thus RMSF could analyze the region that affect to the overall structure as seen in RMSD fluctuation. The RMSF value of protein could be monitored that protein is highly flexible will show large RMSF value, while is low RMSF value in the constrained protein. In RMSF plots in our study showed that the structure models of mutants had individual dynamic behaviors respected to their average structure and seem to deviate from the dynamic structure of a 20-mer wild-type. All three mutant structures were stable than a 20-mer wild-type with the averaged RMSF values of 4.22, 3.83, 2.89 and 2.57 Å for a 20-mer wild-type, double T349R/S351R, single T349R and single S351R mutant models, respectively, while single mutation of S351R model showed the most stable. In addition, five residues (Tyr341-Val345) in mutant models (excepted for in the mutant S351R models) located on flexible loop region of N-terminal segment showed the high fluctuation at least 4 Å, whereas the mutant S351R model remained to constrain the loop flexibility within 1 Å. In agreement with Sinha and Nussinov (2001) reported that the particular area including loop region could be presented the largest deviations due to the crystal-packing effect (Sinha and Nussinov, 2001).

However, two single mutant structures of T349R and S351R were contained one arginine substitution. This arginine residue not only supported the neighbored residues to more constraint in the average structure but also made its backbone more stable than a 20-mer wild-type. In double mutant structure, two arginine residues showed higher fluctuation than arginine of two single mutant structures that reflexed the impact of arginine substituations. It could be suggested that the flexibility of one side-chain arginine might

affect its neighbored residues and also other mutation sites of arginine guanidinium side-chain.

Table 9 The average RMSD value of each system along MD simulations.

Peptides models	Complex (Å)	EGFR kinase in complex (Å)	peptide in complex (Å)
Wild-type	2.02Å	2.06Å	2.23Å
T349R/S351R	2.55Å	2.94Å	4.60Å
T349R	2.38Å	2.81Å	3.11Å
S351R	2.62Å	3.11Å	2.82Å

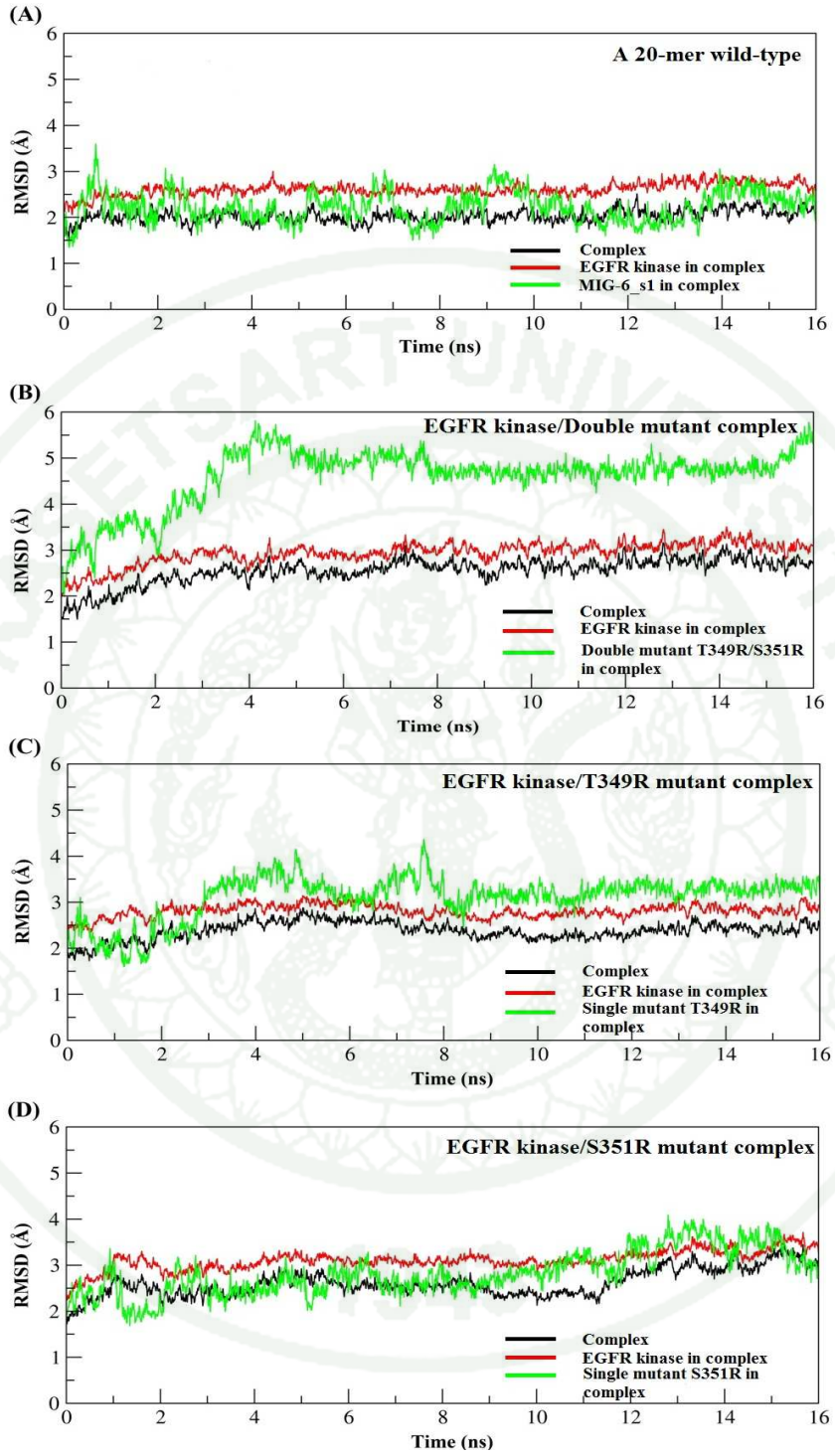


Figure 61 RMSD of the Cα backbone atoms for a 20-mer wild-type and mutants. A 20- mer wild-type (A) T349R/S351R (B), T349R (C) and S351R (D) mutant complexes were shown.

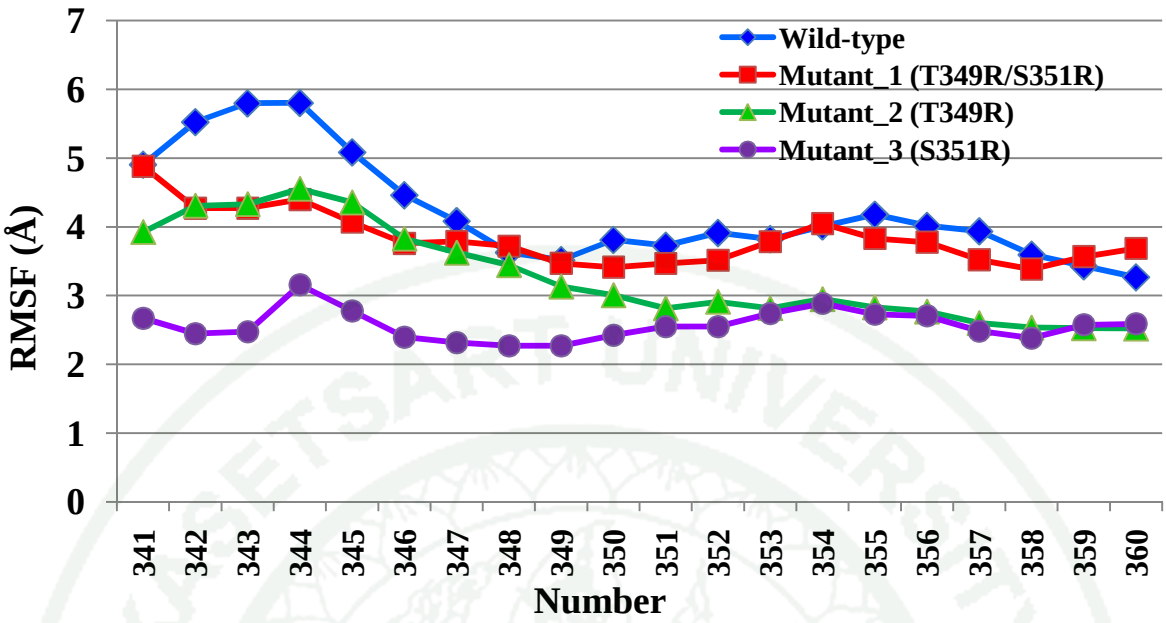


Figure 62 RMSF of Cα carbon atoms for a 20-mer wild-type and mutants. A 20-mer wild-type (blue), T349R/S351R (red), T349R (green) and S351R (purple) short peptides models were exhibited**.**

2. Effect of single and double mutations T349R, S351R and T349R/S351R in hydrogen bonds and salt bridges interactions

Because hydrogen bond and salt-bridge are important in protein stability, thus hydrogen bonding and salt-bridge interactions were monitored during MD simulations. As one of two strategies for design peptide revealed in part II, the interested residues were mutated from both uncharged polar residues (Thr349 and Ser351) to the charged polar residue of arginine. Based on preliminary information of two wild-type structures, we expected the substitutions of them by arginine leading to hydrogen bond and also salt-bridge constructions that participated to increase the stability of protein structure and affected to the binding affinity also. Nevertheless, salt-bridge formation could affect to destabilize or stabilize the complex and decrease or increase to binding free energy (Hendsch and Tidor., 1994).

The results were summarized in Table 10, suggested that there were 7, 5 and 5 pairs of residues from T349R/S351R, T349R and S351R models, respectively, which had the conserved hydrogen bonds (shown in underline) based on the wild-type observation. As expected, we found salt-bridge formations between the mutated residues to their partner residues on the dimer interfaces (Figure 63). Detailed results could be explained as below.

2.1 The optimized-MIG-6_s1 wild-type

The sequence of MIG-6_s1 was reduced from a 27-mer to a 20-mer (Tyr341-Ser360). A 20-mer wild-type was monitored the bonding interactions, which this information was joined with the obtained information from a 27-mer before applying in site-direct mutagenesis study. Hydrogen bond analysis revealed that eight pairs of residues could form hydrogen bonds during the 3-ns last of simulation, which some a given pair might contain more than one hydrogen bond. Because of the stability of MD simulated structure of a 20-mer wild-type as shown in RMSD and the superimposition results (Appendix Figure B1), the MD simulated structure could remain to conserve the hydrogen bond as monitored in a 27-mer and a 20-mer wild-type model. Moreover, the MD

stimulated structure did not form salt-bridge between their interfaces as seen in parent structures. Nevertheless, a native residue S351 in the average structure was disrupted hydrogen bond formation while Thr349 was remained to form the hydrogen bond with Glu907.

2.2 Double mutation of T349R/S351R complex

From the results of double mutation, nine pairs of residues constructed HBs. Not only seven of bonds were also found in a 20-mer wild-type, but both of Lys905 and Glu911 in EGFR kinase could also create new bonds with S351R. However, because of the conformational shift in N-teminal segment, the bonding between Gln911 and Tyr341 was disturbed. However, T349R residue could also form salt-bridge to Glu907 within hydrogen bond cut-off 3.5 Å, as expected (Figure 63A). Briefly, T349R NH2 atom formed bonding to atoms OE1/OE2 of Glu907 by occupancies of 71 and 55% and T349R NE atom with two atoms of Glu907 by occupancies of 58 and 48%, respectively.

While the bonding of a native S351 residue was lost in a 20-mer wild-type, it was surprising that the mutated S351R residue in double mutant could create new bonding in hydrogen bonded salt-bridge with Glu907 at OE2, the occupancies of 20%.

Furthermore, double T349R/S351R mutant complex showed the bonding number of the highest residue pairs though it was not significantly different from other mutant variants. While double mutant peptide model showed the high flexibility in RMSD and RMSF values, the effect from many hydrogen bonds and salt-bridges might help free double mutant structure to more stabilize after introducing of EGFR kinase. These actions could affect to a low average RMSD value of complex (~2.5 Å).

2.3 Single mutation of T349R

From the results of single mutant T349R, seven pairs of residues had hydrogen bonds formation (Table 10). Five of them were also found in a 20-mer wild-type. Two hydrogen bonds between Gly906-S351 and Gln952-Tyr341 were new made up, while the disruptions of hydrogen bonding occurred in three residues pairs of Gly906-Phe352 and Gln911-Tyr341. Tyr341 and T349R residues could switch their side-chains to construct hydrogen bonds with other residues in the MD-simulated structure as mentioned above. In contrast to Tyr341 and T349R residues, Phe352 could not switch the side-chain to form hydrogen bond due to the swinging-in behavior of the aromatic ring in Phe352, whereas aromatic rings of other mutants were quite restricted in the same manner (Figure 63 and Appendix Figure B1). Therefore, it was not surprised if the side-chain of Phe 352 could not form hydrogen bond at all. Moreover, as expected, single mutant T349R structure was found a salt-bridge interactions between NE atom to Glu907 OE1/OE2 with the distance ~3.8 Å by the occupancy of 23 and 19 %, respectively (Figure 63B).

2.4 Single mutation of S351R

From the results of single mutant S351R structure, eight pairs of residues were constructed hydrogen bonds as shown in Table 10. Five of them were also found in a 20-mer wild-type. While three new hydrogen bonds were constructed between Lys905-S351R, Glu907-S351R and Gly959-Asn343, three pairs of residues including Gln911-Tyr341, Gln911-Gln350 and Ile917-Lys357 were destroyed during MD simulations.

As theories, aromatic residues can often precede the *cis*-form of proline in turn for helping stabilization to prolin conformer by CH-π interaction (Juy *et al.*, 1983; Wu and Raleigh, 1998), so the aromatic ring of Tyr341 might be possibly affected from the steric influence of *cis* proline 347 structure (Appendix Figure B1). This phenomenon might lead to the disruption of hydrogen bonding between Tyr341-Gln911. As the bonding of a native S351 residue was lost in a 20-mer wild-type during the 3 ns last of simulation. However, after replacement of a native S351 by arginine, the mutated S351 could create new bonding

of hydrogen bonded salt-bridges as expected (Figure 63C). The NE and NH2 atoms of S351R could form bonding with Glu907 (OE2), the occupancy of 20% and 21%, respectively. In addition, based on the overlay of wild-type and three mutant structures in Figure 64, we found the structural shift in Gln350, Phe352, Tyr358 and Lys357 of S351R mutant structure, which were swunng-out from wild-type and other mutants as shown in purple atoms. Howerver, Phe352 and Tyr358 could remain to conserve the hydrogen bond construction, while Gln350 and Lys357 were disturbed hydrogen bonding with Gln911 and Ile917, respectively. The S351R mutant structure could also present the conformational shift the same as double mutant at N-terminal region, but the C terminal segment of S351R structure could create hydrogen bond between Gly959 and Asn343 for stabilizing segment (Figure 65). This hydrogen bond pair was also seen in the conformational change at loop region of a 27-mer wild-type. Nevertheless, It was absent in 20-mer of wild-type and other mutants.

Here again, besides of intermolecular bonding observation, this mutant model could found intrahydrogen bond between Leu342-Val345 located on the loop region. This pair of residue could also found in a 27-mer of wild-type and single mutant T349R structure, while it was absented in a 20-mer wild-type structure and also double mutant structure. To compare with RMSF of a 20-mer in loop region, this pair might participate in helping loop region to more rigid and stable than a 20-mer wild-type and double mutant as presented in RMSF plot.

From all hydrogen bond and salt-bridge analysis showed in this part, we found the disruptions and new bridge constructions of hydrogen bonds in residues pairs. It might cause from the conformational change resulted from the mutagenesis of the interested residues, when compared to the observation in a 20-mer wild-type. In addition, this event could be found in non-helical portion of protein, which the hydrogen bond is usually lost and reformed several times, in spite of having high occupancy. Thus, this event could be explained by the short lifetimes of bonding property, which the fluctuation of bonding might depend on the time period chosen for observation of binding and unbinding events. Herein, we succeed in the study of the effect of mutations to hydrogen bond and salt-bridge

interactions, which the mutations of polar uncharged residues to polar charged residue not only remained to maintain the hydrogen bonding but also able to create the salt-bridge (strong hydrogen bond) in the same time. Next, the binding free energy was interpreted as we expected that the single or double mutations should affect to the binding affinity.

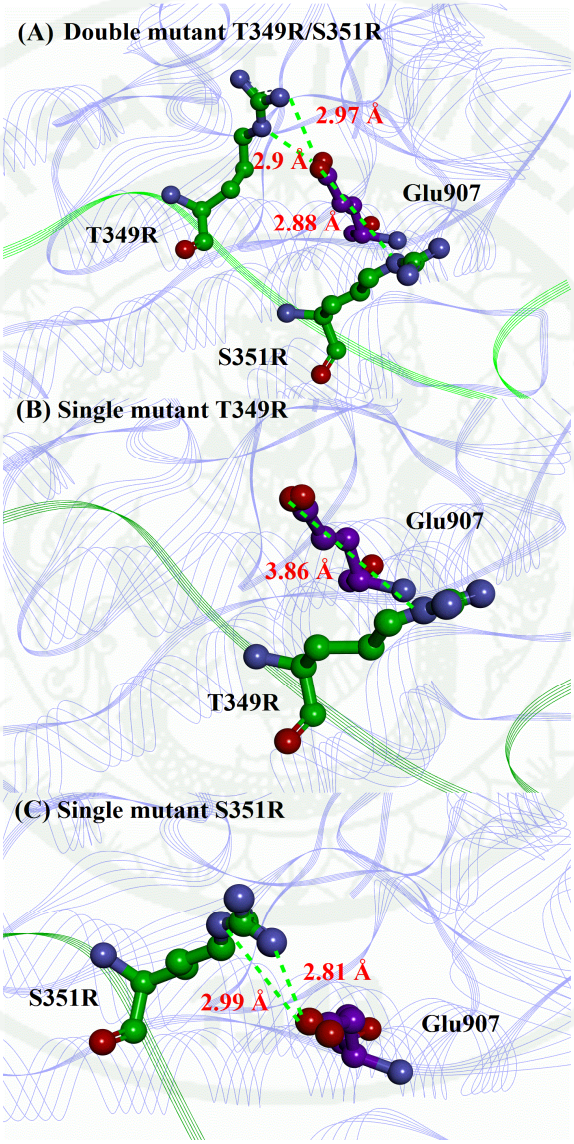


Figure 63 The schematic presentation of salt bridges formation within 4 Å of double T349R/S351R , two single; T349R and S351R mutations. The pairs of oppositely charged residues (salt-bridges) in T349R/S351R (A), T349R (B) and S351R (C) mutation structures were presented by green.

Table 10 List of hydrogen bonds and salt bridges across the interface of EGFR kinase with a 20- mer wild-type and three mutants. The distance cut-off of hydrogen bond and salt-bridge are 3.5 and 4.0 Å, respectively. The occupancies higher than 20% were shown. The same pair of residues was counted as one in each simulation**.**

Complex models	Hydrogen bonds		Salt-bridges	
	EGFR kinase residue	**MIG-6_s1 residue**	**EGFR kinase residue**	**MIG-6_s1 residue**
A 20-mer wild-type MIG-6_s1	Glu904	Ser360	-	-
	Gly906	Phe352		
	Glu907	Thr349		
	Arg908	Gln350		
		Tyr358		
	Gln911	Tyr341		
		Gln350		
	Ile917	Lys357		
Double mutation (T349R/ S351R)	Glu904	Ser360	Glu907[a]	T349R[a]
	Lys905	S351R	Glu907[a]	S351R[a]
	Gly906	Phe352		
	Glu907	T349R		
		S351R		
	Arg908	Gln350		
		Tyr358		
	Gln911	Gln350		
	Ile917	Lys357		

Table 10 (Continued)

Complex models	Hydrogen bonds		Salt-bridges	
	EGFR kinase residue	**MIG-6_s1 residue**	**EGFR kinase residue**	**MIG-6_s1 residue**
Single mutation (T349R)	Glu904	Ser360	Glu907[b]	T349R[b]
	Gly906	Ser351		
	Arg908	Gln350		
		Tyr358		
	Gln911	Gln350		
	Ile917	Lys357		
	Gln952	Tyr341		
Single mutation (S351R)	Glu904	Ser360	Glu907[a]	S351R[a]
	Lys905	S351R		
	Gly906	Phe352		
	Glu907	Thr349		
		S351R		
	Arg908	Gln350		
		Tyr358		
	Gly959	Asn343		

[a] The pair salt-bridge was also counted as hydrogen bonding pair within 3.5Å.

[b] The pair of residues was counted as the non-hydrogen bonded salt-bridge

Underline indicated the conserved hydrogen bond observed in wild-type.

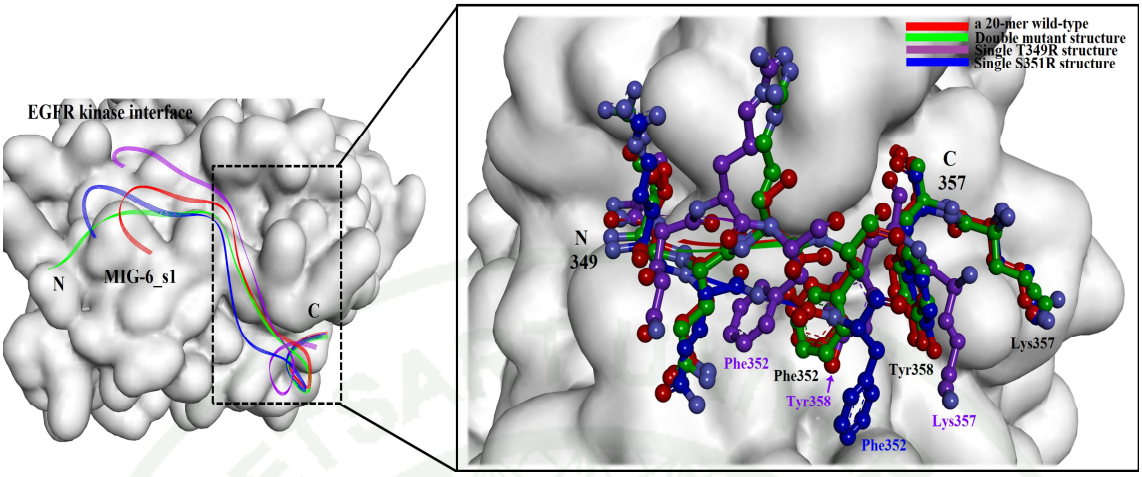


Figure 64 The superimposition of MIG-6_s1 structure among a 20-mer wild-type and mutant structures. The interested residues of MIG-6_s1 were shown and distributed on the C lobe of EGFR kinase interface.

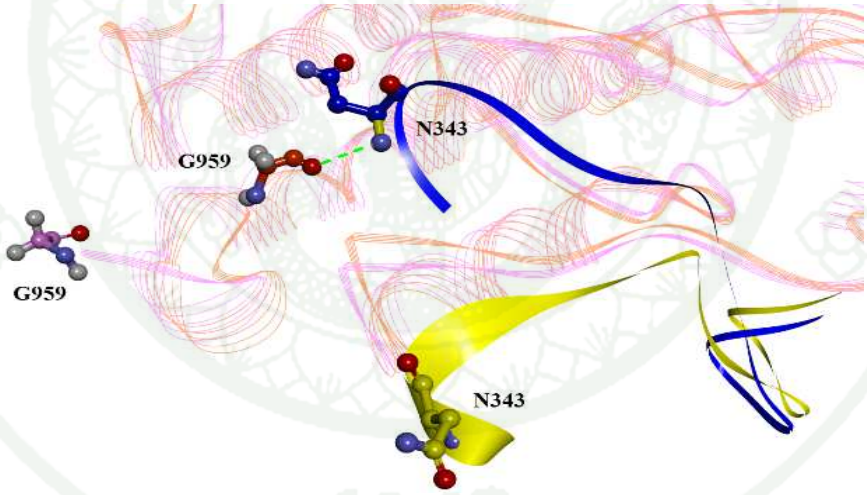


Figure 65 The hydrogen bond formation observed only in single mutation of S351R structure. The MD-stimutated structure (blue ribbon) was superimposed with the initial-minimized structure (yellow ribbon). The capability of Asn343 forming bond to Gly959 was shown (green dot) and absented in the initial minimized structure.

3. Free energy analysis

3.1 The binding free energy and their components

To insight the contribution spectrum of binding free energy for EGFR kinase with a 20-mer wild-type, double mutant T349R/S351R, two single mutants T349R and S351R structures, then the binding free energy of each complex was computed using the MM-PBSA method based on PB model. The binding free energies and their components of all complexes were summarized in Table 11 and Figure 66.

As presented in the Table 11 and Figure 66, due to the double mutant T349R/ S351R was clarly shown the highest in binding free energy among other the mutant types and a 20-mer wild-type. However, we used the two-tailed Z-scores for observating the statistically significant (p-value < 0.01) of difference in binding free energy (Appendix C). The results showed that single mutant T349R structure was significantly stronger than wild-type (z-score = -11.05) and S351R (z-score = -16.69), while S351R mutant structure showed the weakest binding affinity to EGFR kinase with a significant difference of 8.07 comparing with wild-type. The negative and positive z-scores indicated stronger and weaker binding affinity than that of a compared type (Yamasaki *et al.*, 2009). Therefore, the binding free energies of all mutants comparing to a 20-mer wild-type could be ranked in the order of the double mutant > T349R > a 20-mer wild-type > S351R complexes, which the values were -163.80 > -149.62 > -142.49 > -136.45 kcal.mol^{-1}, respectively, without considering the entropy effect.

The free energy components of double and single mutants complexes were evaluated the interaction mechanisms. In Table 11 and Figure 66, for all complexes, electrostatic interactions (E_{ele}) and van der Waals (E_{vdW}) interactions in the gas phase gave the major favorable contributions for binding. While non-polar solvation free energies (ΔG_{np}) were also slightly favorable contributions, polar solvation free energies (ΔG_{pb}) were unfavorable contributions to the binding affinity. As seen in previous protein-peptide interactions, the favorable electrostatic interactions (E_{ele}) in the gas phase were over-

compensated by the unfavorable desolvation of polar group (ΔG_{pb}) leading to unfavorable total electrostatic interaction ($\Delta G_{ele,pol}$) (Guo *et al.*, 2012; Xue *et al.*, 2012).

From the results, the weakest of binding free energy of S351R mutant structure could be discussed with RMSD and RMSF results, S351R mutant structure in complex revealed the most stable, while its complex structure showed the least stable and weakest in binding free energy. This phenomenon might be observed and explained by the structural shift in some residues of S351R mutant structure differing from those of other mutants, the steric cash in Tyr341 and the higher total electrostatic interaction than other mutants. These factors might be cause of the lowest binding affinity.

In contrast to S351R mutation, double mutation gave the highest free energy in all species, which not only the overall electrostatic interaction was more favorable than others but the sum of coulombic (E_{ele}) and van der Waals (E_{vdW}) interactions could also completely compensate the solvation free energy (ΔG_{solv}) in the most. Although, the RMSD and RMSF results showed the least stability of double mutant structure in complex (4.6Å), which the complex structure could stabilize than S351R complex after introducing of EGFR kinase. Therfore, not only the quality of structure but also the reducing of unfavorable total electrostatic interaction could support to the highest binding affinity.

In case of double and S351R mutant structures could be observed in to three points. First point, the competive properties between the stabilities and functions could be occurred in the protein mutagenesis studies (Bloom *et al.*, 2004). Second point, an improving protein stability by the changing of the amino acid could also be possible result in lose of the function (Shoichet *et al.*, 1995). Third point, the structure inspection alone was not sufficient to design peptide, though the polar solvation was also important to monitor (Kar *et al.*, 2013). However, the influence of binding free energy components in overall structure and also each residue could be revealed in futher analysis.

3.2 Mutation induced shift in affinity

As seen in Table 11, the total binding free energies of wild-type, double T349R/S351R, two single T349R and S351R mutations were predicted as -142.49, -163.80, -149.62 and -136.45 kcal.mol^{-1}, respectively. To compare with wild-type complex, the difference of standard binding free energy for each mutant complex was defines as $\Delta G^{'}=\Delta G_{\text{mutant}} - \Delta G_{\text{wild-type}}$ (Istomin *et al.*, 2008). From equation calculation, the double mutant T349R/S351R and two single mutants T349R and S351R had the alteration in binding free energy by -21.31, -7.13 and 6.04 kcal.mol^{-1}, respectively (Table 12). Negative sign indicated a stabilization of interaction, while positive value was decrease a stabilizing interaction.

In details, as the favorable diving forces were coulombic, van der Waals and non-polar solvation free energy, the coulombic interaction (E_{ele}) was the most favorable shifted the range from -17.66 to -92.24 kcal.mol^{-1}. The large shift contribution of the columbic interaction could present in all mutants, it was partially originated from charge-charge interactions in term of salt-bridge formations (Table 10). In similar favorable fashion, van der Waals (E_{vdW}) revealed a slightly shift of the free energy and varies from -10.94 to -18.85 kcal.mol^{-1}, which double mutant side-chain might be allowed to remove the inner vdW cavity volume due to a proper orientation of the mutant residues (Jiang *et al.*, 2001).

Besides of the aforementioned favorable contributions, the favorable non-polar solvation free energies were a slightly enhanced in yielding the contributions of -6.79 to -15.89 kcal.mol^{-1}. Nevertheless, these favorable contributions could be compensated by large unfavorable polar solvation in the ranging from 41.43 to 105.4 kcal.mol^{-1}. Moreover, because the polar part in solvation free energy was predominant by unfavorable contribution, demonstating to unfavorable total electrostatic contributions in range of 13.16 to 23.77 13.16 kcal.mol^{-1}. This trend could be seen in many studies (Bradshaw *et al.*, 2011; Guo *et al.*, 2012; Kar *et al.*, 2013).

Among three mutation complexes, the double mutant model showed the largest shift of all free energy components in both gas phase and solvation state. The largest shift in the coulombic attraction and the repulsive desolvation of double mutant model increased more 2.5-fold and 2-fold than T349R mutation and those of 5-fold and 2.5 fold than S351R mutation, respectively. Because electrostatic interactions were generally involved in the contact between charged or polar groups on protein-protein interfaces for many biological functions (Thomas, 2003). Therefore, the additional of two positively charged residues of arginine on interface might affect to significant increasing of the positive electrostatic potential at the interface higher than two single mutant models leading to highly electrostatic interactions in gas phase and solvation phase (Strub *et al.*, 2004). In the same word, the increasing of coulombic attraction in double mutation was proportional higher than the repulsive desolvation and larger than other mutants, influencing to relieve the unfavorable overall electrostatic contribution in the most of case. It could be suggested that the total electrostatic contribution to binding free energy was more favorable if molecules have opposite charges (Schapira *et al.*, 1999).

3.3 Double mutation effects

As presented in Table 12, because the effect of mutations on a 20-mer wild-type-induced the affinity shift especially was clearly seen in double mutation that it could support the most of the binding affinity in our simulations. Therefore, it was interesting to investigate the quantitative effect of double mutation. The effect of double mutation was calculated in terms of the coupling free energy (ΔG_i) by classical double mutant cycle analysis (Horovitz, 1987; Jang *et al.*, 2004; Otzen, 2005). Based on the Gibb free energy of each species obtained in quation 18 ($\Delta G = \Delta G_{MT} - \Delta G_{WT}$), then the coupling energy (ΔG_i) was defined by equation 19; $\Delta G_i = \Delta G_{1+2} - (\Delta G_1 + \Delta G_2)$. Where the ΔG_{1+2} value was the free energy of double mutant, each of ΔG_1 and ΔG_2 were the free energy for mutant 1 (more damage) and mutant 2 (less damage), respectively, these values were compared to wild-type before putting into each term (Vaughan *et al.*, 2002). ΔG_i could be positive ($\Delta G_i > 0$) or negative ($\Delta G_i < 0$) values depending on the interaction between the mutant side-chain could destabilize or stabilize the measured functional property (Doura and Fleming, 2004).

If the effects of the mutations were independent each other also called additive effect, the free energy for the double mutation was equal to the sum of two single mutations ($\Delta G_{1+2} = \Delta G_1+\Delta G_2$; $\Delta G_i=0$) while the synergistic effect in coupled interaction was the change free energy of the double mutation higher than the sum of two single mutants ($\Delta G_{1+2} > \Delta G_1+\Delta G_2$). As shown in Table 13, the ΔG_{1+2} value of double mutant T349R/S351R (-21.13 kcal.mol^{-1}) were larger than the sum of those of single mutant (T349R+S351R = -1.09 kcal.mol^{-1}) or two single mutants alone, the effect revealed the coupling free energy being negative value (-20.04 kcal.mol^{-1}). The negative sign suggested that two residues interacted to increase the binding affinity, in other words, the degree of stabilization by double mutation T349R/S351R was stronger than the sum of those by two single mutations ($\Delta G_1+\Delta G_2$) (Jang *et al.*, 2004).

As be often observed in the binding effect of charged residues on other interfaces, the effect of arginine in double mutation was defined as the synergistic effect, which the energetic of the interactions were dependent on each other. Although the mutation at one residue could perturb interactions to the other mutated positions, the effect might be resulted in increase the binding affinity for this study (Mildvan, 2004). Because of the columbic interaction and van der Waals predominantly in double mutant, thus the change free energy in two single mutations could not support to the additive effect.

3.4 Free energy decomposition of individual residue

To confirm the capability of mutated residues and insight the basis of the potency from the mutations, the key residues were obtained from the average structure by per-residue energetic decomposition analysis. The results were presented in Figure 67 that quite similar in the overall interaction spectra of five structures including a 27-mer wild-type (observed only 20 residues), a 20-mer optimized wild-type, T349R/S351R double mutant, T349R and S351R two single mutants. Following the individual attractive contribution results found in a 27-mer wild-type residues (Part I), the contribution of individual residues of a 20-mer wild-type and three mutant complexes could also find the ten predominant residues including Tyr341, Met346, Pro348, T/R349, Gln350, S/R351,

Phe352, Tyr358, Val359 and Ser360 with the contribution higher than -1.0 kcal.mol^{-1}. It was noted that the crucial residues were remained to conserve as key residue in a 27-mer wild-type though the optimized sequence and the mutated sequences. This conserve event was the important property of protein-protein recognition (Zhong and Carlson, 2005). In addition, it was consistent with Zhang and co-workers as fist reported the important of Met346, Phe352 and Tyr358 after alanine substitutions (Zhang *et al*., 2007). This was extremely useful to imply that a 20-mer optimized sequence in part II did not destroy the feature interactions of crucial residues. For all MIG-6_s1 structures, the most significant contributions were originated from Tyr358 to the binding free energy with the range between -3.67 to -6.6 kcal.mol^{-1}, whereas the residue Thr349 showed the most significant contributions in the single mutant S351R structure (-4.2 kcal.mol^{-1}). For these Tyr358 and Thr349 residues, most of the contributions came from van der Waals interactions. In addition, we could see that Pro347, Ala353, Pro354 and Pro356 were not change in binding free energies values of all complexes, while the individual free energies were not large (ΔG > -1.0 kcal.mol^{-1}). This could be suggested that these proline and alanine residues might involve in the structure flexibility more than the binding contribution. In Table 14, the given free energy of each residue in higher than -1 kcal.mol^{-1} was further clarified into energies compositions, which an emphasis on the native residues in wild-type and the mutation of native resides in the mutated sequence. They were described as below.

In double mutation of T349R/S351R structure, as we expected, two arginine residues could affect to the individual contributions by the increasing of coulombic interaction higher than the wild-type and two single mutants alone (excepted for S351R residue in double and single mutations). The favorable electrostatic interactions were originated from charge-charge interaction in term of salt-bridges. Even though, salt-bridge could cause also the increasing of the unfavorable electrostatic in solvation, which it could over-compensate the coulombic interaction and give rise to a little unfavored contribution. However, two arginines in the positions of 349 and 351 in double mutation could also add up the van der Waals interaction higher than other models (-6.10 and -3.35 kcal.mol^{-1}, respectively), which was the major driving force participated more than half of total binding free energy of each residue. The higher van der Waals interaction in double

mutation could cause from reducing the vdW cavity volume done by the arginine side-chains. Furthermore, the major contribution feature of van der Waals was be observed in a general phenomenon of protein-protein interactions. Moreover, as the contribution of T349R and S351R residues were -6.0 and -2.4 kcal.mol^{-1}, respectively, the side-chain were promoted more than half of free energy of them (-4.74 and -1.34 kcal.mol^{-1}).

For a T349R single mutant model, the free energy contribution of T349R residue gave the value of -4.9 kcal.mol^{-1}. The electrostatic contributions in the gas phase (columbic interaction) and in solvation phase (polar solvation) were significant increase to -25.6 and 24.7 kcal.mol^{-1}, respectively from the wild-type (-8.2 and 7.1 kcal.mol^{-1}). The large of the electrostatic interactions could be predicted by new formation of salt-bridge between T349R-Glu907. Nevertheless, after compensation with polar solvation, total electroststic interactions were not large favored and differed from a 20-mer wild-type. It was agreed with other studies reported that interfacial salt-bridge can favorable to contribute the stability of protein and the binding free energy by the interaction of columbic attraction, though these effects are often balanced by the unfavorable desolvation repulsion (Kundrotas and Alexov, 2006; Wong *et al.*, 2013). While total electrostatic interactions could be favored little free energy contribution, the most favorable contribution of T349R mutant residue could be originated from the van der Waals interaction (-4.38 kcal.mol^{-1}) and side-chain played role the most (-3.45 kcal.mol^{-1}).

A part of a mutated S351 in a single mutant S351R model, the individual contribution of the mutated residue was -3.14 kcal.mol^{-1}. The major supporting force was originated from van der Waals interaction (-4.03 kcal.mol^{-1}), which revealed the most favorable than that residue in double mutant and two wild-type models. However, van der Waals force was partially disturbed by the unfavored total electrostatic interactions (0.9 kcal.mol^{-1}) leading to decreasing the free energy contribution into -3.14 kcal.mol^{-1}, while the arginine side-chain was larger contribution than its backbone. Other interested residues in the spectra of S351R mutant structure, we could see the structural shift of Tyr341, Gln350, Phe352, Lys357 and Tyr358 differed from other structures (Figure 64). These shifts could result not only in the loss of binding affinities of their residues (Figure 67), but

might affect to lowest of overall binding affinity. The swinging in of Tyr341 position closing to proline 347 could be caused in the structural clash leading to giving a weakest binding affinity (-1.32 kcal.mol^{-1}) than other structures but quite similar value to Tyr341 of double mutation (-1.42 kcal.mol^{-1}). In addition, the swinging out of Gln350 position in S351R mutant structure could cause a weaker in binding (-3.04 kcal.mol^{-1}) than others. The same as Lys357 residue in S351R mutant structure, the structural shift could lead to the large unfavorable binding affinity (4.9 kcal.mol^{-1}) that opposed to the favorable of other structures. This could be explained by the disruption of hydrogen bond between Gln350-Gln911 and Lys357-Ile917. Nevertheless, Gln350 could construct one HB with the conserved Arg908 and remain to give favorable contribution. Unfortunately, Lys357 could not construct HB at all thus the large unfavorable of the electrostatic contribution in the gas phase (4.72 kcal.mol^{-1}) could be occurred. However, this unfavorable contribution could compensate the favorable polar solvation (-0.79 kcal.mol^{-1}) and van der Waals interactions (-3.06 kcal.mol^{-1}). Therefore, the free energy contribution of Lys357 gave rise to the unfavorable for binding (0.88 kcal.mol^{-1}). Likewise, the residue Tyr358 of S351R mutation had a trend to weaker favorable than other structures (-2.07 kcal.mol^{-1}). It might originate from the closing of structure into hydrophilic environment than that in other structure, leading to the decreasing of van der Waals contribution (-5.23 kcal.mol^{-1}) (Appendix Figure B1). However, because Tyr358 could remain to construct the conserved hydrogen bond to Arg908 as other structures, thus the total electrostatic interaction ($\Delta E_{ele}+\Delta G_{pb}$) was remained similar to others. In addition, this native Tyr358 in S351R mutant structure could be maintained as keystone residue (-3.67 kcal.mol^{-1}) though the structure was shifted as described.

Based on all spectrum of decomposition per residue basis, we could observe the mutation effect of arginine substitution. As seen in two single mutant models, when a native Thr349 was replaced by arginine in single T349R mutation, it could conserve the binding free energy to similar with a 20-mer wild-type but it seems to affect the native S351 by decreasing the contribution into -0.68 kcal.mol^{-1}. In the other word, the increasing of binding free energy contribution for Ser351R was revealed in single S351R mutation (-3.14 kcal.mol^{-1}) but the individual contribution of native Thr349 was dropped more than

wild-type (-4.22 kcal.mol^{-1}). This effect could also be seen in double mutation. The mutated Thr349 gave a highest contribution (-6.01 kcal.mol^{-1}) than other structures, even though its mutation might affect to Ser351R by dropping the free energy contribution larger than Ser351R in single mutation (-2.04 and -3.14 kcal.mol^{-1} for S351R in double and one in single mutations, respectively). However, the contribution of S351R remained to higher than native. These observations could be suggested that the effect of mutation in one residue could be resulted in slightly increasing of its free energy, though it also be limited the contribution of participating residue in the same position. In this studied, the phenomenon could be explained by the geometric effect of arginine side-chain involving in the electrostatic interactions (salt-bridge and hydrogen bond), which the negatively charged groups, the long and bulky group of side-chain might interfere other residues and be possible to low affinity of binding (Dornieden, 2012). However, the effect of two arginine substitutions in double mutation at positions 349 and 351 were balanced each other leading to supporting the free energy of them than wild-type and also contributing the overall binding affinity. This interaction was consistented with the synergistic effect as reported in double mutation effect analysis.

Table 11 Binding free energy components for the EGFR kinase with a 20-mer wild-type and three mutant complexes using MM-PBSA method.

Component[a,b]	**Wild-type**	**T349R/S351R**	**T349R**	**S351R**
ΔE_{ele}	-99.74	-191.98	-136.26	-117.40
ΔE_{vdW}	-90.45	-109.03	-104.27	-101.39
ΔE_{gas}	-190.19	-301.01	-240.53	-218.79
ΔG_{np}	-63.65	-79.54	-72.41	-70.44
ΔG_{pb}	111.35	216.75	163.32	152.78
ΔG_{sol}	47.70	137.21	90.91	82.34
$\Delta G_{ele,pol}$	11.61	24.77	27.06	35.38
ΔG_{total}	-142.49	-163.8	-149.62	-136.45
Std. Dev[c]	7.32	10.12	8.37	10.84

[a] All values were given in kcal.mol^{-1}

[b] Components: ΔE_{ele}: Electrostatic energy in the gas phase;

ΔE_{vdW} : van der Waals in the gas phase; ΔE_{gas}: The sum of $\Delta E_{ele} + \Delta E_{vdW}$;

ΔG_{np}: Non-polar solvation energy; ΔG_{pb}: Polar solvation energy;

ΔG_{sol}: The sum of $\Delta G_{np} + \Delta G_{pb}$;

$\Delta G_{ele,pol}$: Total electrostatic interaction ($\Delta E_{ele} + \Delta G_{pb}$);

ΔG_{total}: Total binding free energy ($\Delta E_{gas} + \Delta G_{sol}$)

[c] Std. Dev: Standard Deviation of total binding free energy

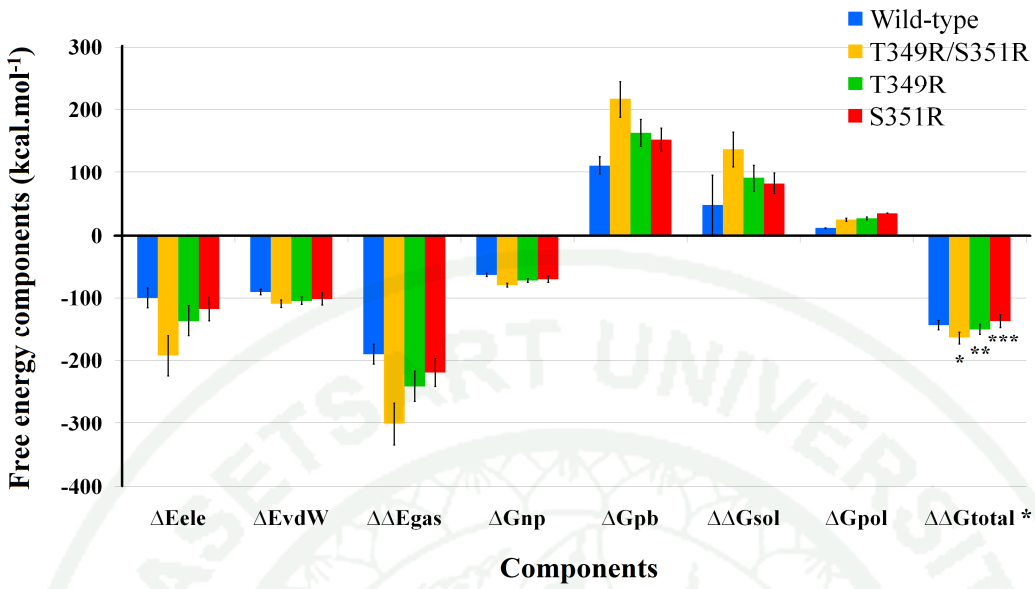


Figure 66 Energy components (kcal.mol^{-1}) for EGFR kinase with a 20-mer wild-type and mutant complexes. Components: ΔE_{ele}: Electrostatic energy in the gas phase ; ΔE_{vdW} : van der Waals in the gas phase; ΔE_{gas} : The sum of ΔE_{ele} + ΔE_{vdW} ; ΔG_{np}: Non-polar solvation energy; ΔG_{pb} : Polar solvation energy ; ΔG_{sol} : The sum of ΔG_{np} + ΔG_{pb}; ΔG_{pol} : Total electrostatic interactions ($\Delta G_{pol} = \Delta E_{ele} + \Delta G_{pb}$); ΔG_{total} : Total binding free energy ($\Delta E_{gas} + \Delta G_{sol}$). The statistically significant difference (z-scores $p < 0.01$) in binding free energy were presented as *, ** and *** underneath the error bars.

Table 12 Mutation-induced shifts in the binding free energy components of mutant variants respected to a 20-mer wild-type complex.'

Component[a,b]	**Wild-type - Wild-type**	**T349R/S351R - Wild-type**	**T349R - Wild-type**	**S351R - Wild-type**
ΔE_{ele}	0	-92.24	-36.52	-17.66
ΔE_{vdW}	0	-18.58	-13.82	-10.94
ΔE_{gas}	0	-110.82	-50.34	-28.6
ΔG_{np}	0	-15.89	-8.76	-6.79
ΔG_{pb}	0	105.4	51.97	41.43
ΔG_{sol}	0	89.51	43.21	34.64
$\Delta G_{ele,pol}$	0	13.16	15.45	23.77
ΔG_{total}	0	-21.13	-7.13	6.04

[a] All values were given in kcal.mol^{-1} and each species calculated as equation by ΔG_{mutant}-$\Delta G_{wild\text{-}type}$ with ΔG_{mutant} and $\Delta G_{wild\text{-}type}$ obtained in Table 11. Negative and positive values indicate that the mutant peptide binds protein stronger and weaker than the wild-type, respectively.

[b] The following notation was used: ΔE_{ele}: Electrostatic energy in the gas phase; ΔE_{vdW} : van der Waals in the gas phase; ΔE_{gas}: The sum of ΔE_{ele} + ΔE_{vdW}; ΔG_{np}: Non-polar solvation energy; ΔG_{pb}: Polar solvation energy; ΔG_{sol}: The sum of ΔG_{np} + ΔG_{pb}; $\Delta G_{ele,pol}$: Total electrostatic interaction (ΔE_{ele} + ΔG_{pb}); ΔG_{total}: Total binding free energy (ΔE_{gas} + ΔG_{sol})

Table 13 The coupling free energy respected to a 20-mer wild-type free energy.

Mutants	**Free energy shift[a] (kcal.mol^{-1})**
ΔG_1: T349R - Wild-type	-7.13
ΔG_2: S351R - Wild-type	6.04
ΔG_{1+2}: T349R/S351R - Wild-type	-21.13
Coupling free energy $\Delta G_i = \Delta G_{1+2} - (\Delta G_1 + \Delta G_2)$	-20.04

[a] All values were given in kcal.mol^{-1}.

Difference of binding free energy of mutants respect to a 20-mer wild-type was calculated by ΔG_{mutant}-$\Delta G_{wild\text{-}type}$ and the coupling free energy calculated by $\Delta G_i = \Delta G_{1+2} - (\Delta G_1 + \Delta G_2)$; positive ($\Delta G_i > 0$) or negative ($\Delta G_i < 0$) values depending on the interaction between the mutant side-chain can destabilize or stabilize the measured functional property, respectively.

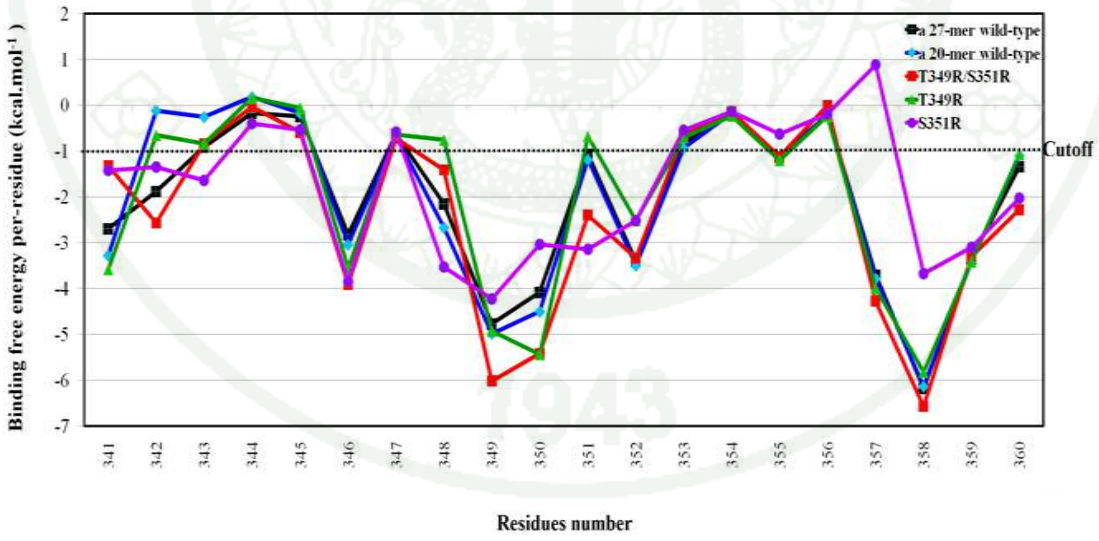


Figure 67 Decompostion of $\Delta G_{binding}$ into individual residue. The residues with the total binding free energy higher than -1 kcal.mol^{-1} were presented under the cut-off line.

Table 14 Distribution of the binding free energies components (ΔG) from individual residue.

Residue[a,b]	ΔE_{ele}	ΔE_{vdW}	ΔG_{pol}	ΔG_{tot}	ΔG_{sc}	ΔG_{bb}
A 27-mer wild-type						
Y341	-2.90	-2.47	2.67	-2.70	-2.93	0.21
M346	0.14	-3.84	0.90	-2.81	-2.15	-0.34
P348	-3.22	-3.27	4.34	-2.15	-1.34	-0.87
T349	-8.72	-3.54	7.50	-4.77	-3.18	-1.44
Q350	-7.56	-4.61	8.09	-4.08	-1.53	-2.72
S351	-2.40	-2.03	3.37	-1.07	-0.37	-0.58
F352	-2.31	-3.91	2.81	-3.41	-2.59	-0.60
D357	-5.12	-3.02	4.44	-3.70	-1.78	-3.13
Y358	-2.97	-7.21	4.01	-6.17	-3.81	-2.29
V359	-1.34	-3.18	1.18	-3.34	-1.30	-2.00
S360	-4.47	-2.44	5.58	-1.34	0.22	-1.59
A 20-mer wild-type						
Y341	-3.22	-1.91	1.84	-3.26	-3.15	-0.12
M346	0.14	-3.84	0.65	-3.01	-2.73	-0.31
P348	-3.68	-3.66	4.67	-2.73	-1.52	-1.15
T349	-8.24	-3.85	7.10	-4.90	-3.44	-1.55
Q350	-8.34	-4.65	8.49	-4.51	-1.84	-3.39
S351	-2.63	-2.09	3.54	-1.20	-0.46	-0.73
F352	-2.27	-3.95	2.72	-3.45	-2.73	-0.77
D357	-7.35	-2.94	6.50	-3.75	-0.91	-2.87
Y358	-3.02	-7.36	4.25	-6.17	-3.99	-2.14
V359	-1.45	-3.24	1.28	-3.36	-1.42	-1.98
S360	-5.31	-2.41	6.61	-1.10	0.58	-1.78

Table 14 (Continued)

Residue[a,b]	ΔE_{ele}	ΔE_{vdW}	ΔG_{pol}	ΔG_{tot}	ΔG_{sc}	ΔG_{bb}
Double mutant T349R/S351R						
Y341	0.05	-1.49	0.13	-1.32	-0.83	-0.48
M346	-0.92	-4.59	1.60	-3.91	-3.30	-0.60
P348	-2.26	-1.91	2.76	-1.41	-0.53	-0.87
T349R	-33.7	-6.10	33.8	-6.01	-4.74	-1.26
Q350	-8.52	-4.61	7.72	-5.41	-2.07	-3.34
S351R	-17.7	-3.35	18.7	-2.40	-1.34	-1.05
F352	-1.37	-4.06	2.11	-3.33	-2.56	-0.77
D357	-12.5	-2.91	11.1	-4.27	-1.21	-3.06
Y358	-2.82	-7.48	3.74	-6.56	-3.90	-2.66
V359	-1.32	-3.17	1.20	-3.30	-1.36	-1.94
S360	-10.90	-1.62	10.30	-2.28	-0.78	-1.50
Single mutant T349R						
Y341	0.05	4.14	0.51	-3.58	-2.41	-1.16
M346	-1.03	-4.09	1.57	-3.54	-3.03	-0.51
P348	-1.39	-1.43	2.06	-0.76	-0.32	-0.43
T349R	-25.6	-4.08	24.7	-4.92	-3.45	-1.47
Q350	-8.81	-5.37	8.76	-5.43	-2.89	-2.53
S351	-3.97	-2.34	5.63	-0.68	-0.68	0.00
F352	-0.16	-3.75	1.40	-2.51	-2.55	0.04
D357	-9.32	-3.05	8.40	-3.98	-1.04	-2.93
Y358	-3.29	-6.41	3.88	-5.81	-3.61	-2.21
V359	-1.49	-3.25	1.33	-3.41	-1.25	-2.16
S360	-4.24	-2.61	5.78	-1.07	0.64	-1.71

Table 14 (Continued)

Residue[a,b]	ΔE_{ele}	ΔE_{vdW}	ΔG_{pol}	ΔG_{tot}	ΔG_{sc}	ΔG_{bb}
Single mutant S351R						
Y341	-0.23	-2.01	0.82	-1.42	-1.34	-0.09
M346	-0.41	-4.98	1.56	-3.84	-3.62	-0.21
P348	-3.10	-4.68	4.26	-3.53	-2.00	-1.52
T349	-7.24	-4.27	7.28	-4.22	-2.78	-1.44
Q350	-6.84	-3.53	7.33	-3.04	-0.60	-2.43
S351R	-18.70	-4.03	19.60	<u>-3.14</u>	-1.95	-1.19
F352	-0.56	-3.49	1.53	-2.52	-2.10	-0.42
D357	4.72	-3.06	-0.79	0.88	0.60	0.28
Y358	-2.39	-5.28	3.99	-3.67	-2.20	-1.47
V359	-1.41	-2.90	1.21	-3.10	-1.42	-1.68
S360	-10.50	-1.55	9.98	-2.03	-0.70	-1.32

[a] All values were given in kcal.mol^{-1}.

[b] The contributions from the electrostatic interaction (ΔE_{ele}), van der Waals interaction (ΔE_{ele}) as well as the polar solvation energy (ΔG_{pol}) and the total binding free energy of a given residue (ΔG_{tot}) were presented.

ΔG_{sc} and ΔG_{bb} referred the side chain and backbone contribution, respectively.

PB non-polar solvation energy was currently not decomposable**.**

The residues with the total binding free energy higher than -1 kcal.mol^{-1} were exhibited in table.

The total binding free energy of the mutated positions was underlined.

In summary of part II and part III analysis, we applied the basis knowledge from part I to design the peptides in order to obtain short peptides with high affinity to EGFR kinase. We obtained a 20-mer (Tyr341-Ser360) of MIG-6_s1 peptide as the most suitable candidate from the optimization step and sucessed in the mutation of threonine and serine residues at position 349 and 351, respectively to arginine in terms of single and double mutations. Because of the properties of arginine substitutions in bond formations, arginine in the mutated positions not only could conserve hydrogen bonds as monitored in native residues of wild-type, but also form salt-bridges to negatively charged amino acids of EGFR kianse as expected.

The effects of the mutations were performed using MM-PBSA tool to evaluate the binding free energy. The results of MM-PBSA showed the double mutant T349R/S351R structure could give the highest binding affinity to EGFR kinase. The strongest binding free energy could be not only contributed by the increasing of van der Waals interactions from the losing of vdW cavity volume, but the reducing of total unfavorable electrostatic interactions could also support van der Waals interactions to binding affinity as seen in Table 12. Moreover, we could prove that the mutated residues could give the energy contribution higher than native residues expecially for favorable columbic and van der Waals interactions. In other word, unfavorable polar solvation could also increase to balance coulombic interaction leading to unfavorable overall electrostatic interaction. This phenomenon made van der Waals to become the most favorable contribution, which the most of total free energy contributions originated from the capability of the mutated side-chains the most. In particular, due to increasing of the binding free energy of double mutant higher than others, double mutation effect in term of coupling free energies were revealed in Table 13. The effect of mutation was synergistic effect interaction, which the effect of interaction in two sites could enhance the binding affinity.

However, our findings highlight the crucial of considering the positively charged amino acid substitution in peptide design strategy for enhancing of binding affinity in terms of electrostatic and van der Waals interactions.

CONCLUSION

As peptide-based inhibitors has become an established approach for drug discovery and pharmaceutical investigation. In this study, the protein-peptide interaction (PPI) was basis used for investrigation of the important interactions between interfacial EGFR kinase/MIG-6_s1 peptide. To develope EGFR kinase inhibitors and reduce the limitations of drugs such as having side effects from non-specific and ineffective of these in patients who have second mutation (Davies *et al.*, 2000; Doebele *et al.*, 2010; Rosti *et al.*, 2010). The C terminal lobe part of EGFR kinase surface was new targets for studying in the interactions of MIG-6_s1 peptide (Kuriyan *et al.*, 2012).

In order to design potent MIG6_s1 peptide as EGFR kinase inhibitors, the preliminary investigations were conducted by molecular dynamics (MD) simulations to observe structural properties and the energetic properties including the MM-PBSA calculation based PB and GB models, binding free energy per residue and computational alanine scanning. The interactions of EGFR kinase/MIG-6_s1 peptide were performed a 22-ns molecular dynamic simulation. The results of conformational change and the interactions found on their interface were revealed. While the MD simulated structure of MIG-6_s1 peptide could conserve the most of hydrogen bonds, the flexible loop region was shifted up and constructed new hydrogen bonds. The slightly shift in loop region of MIG-6_s1 might cause from the crystal packing of flexible region leading to unobservation in the initial-minimized structure. The binding of EGFR kinase and MIG-6_s1 peptide was the most driven by van der Waals contribution, while non-polar solvation contribution seem to more sensitive in PB model during calculation and subsequent to the overestimate of binding free energy than GB model. The most important residues on the interfacial EGFR kianse/MIG-6_s1 peptide including six residues of EGFR kinase (Glu904, Glu907, Arg908, Gln911, Met928 and Ile929) and eight residues of MIG-6_s1 (Leu338, Try341, Asn343, Met346, Thr349, Gln350, Phe352 and Tyr358) were considered after identification and vertification by a per-residue decomposition and computational alanine scanning studies in MM-PBSA tool, respectively. Some of them including Met346, Phe352 and Tyr358 were reported as the crucial residues in experimental mutagenic study. In

addition, most of the most important resideus were also involved in hydrogen bond formations, which might participate in protein binding.

In second and third parts, the design process was carried out by taking benefit from the understanding of interactions. A 16-ns MD simulation time was performed by molecular dynamics simulations and free energies were estimated by MM-PBSA method. Two strategies were introduced. First, the MIG-6_s1 peptide was optimized to increase the potency inhibitor by removing the less important residues based on our previous studies. Second, a 20-mer optimal sequence was mutated from polar uncharged residues of Thr349 and Ser351 by replacing with positively charged residue of arginine in terms of single and double mutations for monitoring the effect of mutants to binding free energy. As expected, due to salt-bridge formations of two arginine residues at the mutated positions, the capability of guanidinium group of arginine in double mutant could enhance the binding affinity by increasing of electrostatic potential and filling of the vdW cavity volume. In particular, the complementary surfaces of two arginines in double mutant could increase more favorable total electrostatic interactions leading to supporting van der Waals for binding process. The effect of two sites in double mutant T349R/S351R was interpreted to be the results of the synergistic effect, the pertubtion of them could enhance the binding affinity. Moreover, the capable of a 20-mer MIG-6_s1 wild-type and mutant types in key residues conservation were consistented as the protein-protein recognition feature.

From this study, we had successful benmarked a computational approach for the investigation of basis interaction between interfacial EGFR kinase/MIG-6_s1 peptide complex and *in silico* design of peptidic inhibitors targeting EGFR kinase interface.

Of note, peptide drugs give the benefit of high specificity, selectivity interact-tion and low toxicity, although, the drugable peptides have been challenged in part of low stability under physiological condition (Albericio and Kruger, 2012; Craik *et al.*, 2013). The modifications of a peptide drug (eg. unnatural amino acid substituations) or design peptide drug carrier (ex. liposome and CCPs) should be attended in order to increase the stability by protecting peptide drug from proteiolytic enzyme (Olmez and Akbulut, 2012).

LITERATURE CITED

Adak, S., K.S. Yang, J. Macdonald-Obermann and L.J. Pike. 2011. The membrane-proximal intracellular domain of the epidermal growth factor receptor underlies negative cooperativity in ligand binding. **J. Biol. Chem.** 286(52): 45146-45155.

Adcock, S.A. and J.A. McCammon. 2006. Molecular dynamics: survey of methods for simulating the activity of proteins. **Chem. Rev.** 106(5): 1589-1615.

Aertgeerts, K., R. Skene, J. Yano, B.C. Sang, H. Zou, G. Snell, A. Jennings, K. Iwamoto, N. Habuka, A. Hirokawa, T. Ishikawa, T. Tanaka, H. Miki, Y. Ohta and S. Sogabe. 2011. Structural analysis of the mechanism of inhibition and allosteric activation of the kinase domain of HER2 protein. **J. Biol. Chem.** 286(21): 18756-18765.

Ahmed, M., M.M. Sadek, K.A. Abouzid and F. Wang. 2013. In silico design: Extended molecular dynamic simulations of a new series of dually acting inhibitors against EGFR and HER2. **J. Mol. Graph. Model.** 44(0): 220-231.

Aifa, S., J. Aydin, G. Nordvall, I. Lundstrom, S.P. Svensson and O. Hermanson. 2005. A basic peptide within the juxtamembrane region is required for EGF receptor dimerization. **Exp. Cell Res.** 302(1): 108-114.

Akhtar, L.N. and E.N. Benveniste. 2011. Viral exploitation of host SOCS protein functions. **J. Virol.** 85(5): 1912-1921.

Albericio, F. and H.G. Kruger. 2012. Therapeutic peptides. **Future Medicinal Chemistry** 4(12): 1527-1531.

Anastasi, S., L. Fiorentino, M. Fiorini, R. Fraioli, G. Sala, L. Castellani, S. Alema, M. Alimandi and O. Segatto. 2003. Feedback inhibition by RALT controls signal output by the ErbB network. **Oncogene** 22(27): 4221-4234.

Argiris, A. and N. Mittal. 2004. Gefitinib as first-line, compassionate use therapy in patients with advanced non-small-cell lung cancer. **Lung Cancer** 43(3): 317-322.

Baker, E.N. and R.E. Hubbard. 1984. Hydrogen bonding in globular proteins. **Prog. Biophys. Mol. Biol.** 44(2): 97-179.

Balak, M.N., Y. Gong, G.J. Riely, R. Somwar, A.R. Li, M.F. Zakowski, A. Chiang, G. Yang, O. Ouerfelli, M.G. Kris, M. Ladanyi, V.A. Miller and W. Pao. 2006. Novel D761Y and common secondary T790M mutations in epidermal growth factor receptor-mutant lung adenocarcinomas with acquired resistance to kinase inhibitors. **Clin. Cancer Res.** 12(21): 6494-6501.

Bao, J., I. Alroy, H. Waterman, E.D. Schejter, C. Brodie, J. Gruenberg and Y. Yarden. 2000. Threonine phosphorylation diverts internalized epidermal growth factor receptors from a degradative pathway to the recycling endosome. **J. Biol. Chem.** 275(34): 26178-26186.

Barlow, D.J. and J.M. Thornton. 1983. Ion-pairs in proteins. **J. Mol. Biol.** 168(4): 867-885.

Berhanu, W.M. and A.E. Masunov. 2012. Alternative packing modes leading to amyloid polymorphism in five fragments studied with molecular dynamics. **Biopolymers** 98(2): 131-144.

Betts, M.J. and R.B. Russell. 2003. Amino acid properties and consequences of substitutions. **Bioinformatics for geneticists** 317: 289.

Bharatham, N., S.W. Chi and H.S. Yoon. 2011. Molecular basis of Bcl-X(L)-p53 interaction: insights from molecular dynamics simulations. **PLoS One** 6(10): e26014.

Bizzarri, A.R. and S. Cannistraro. 2002. Molecular Dynamics of Water at the Protein−Solvent Interface. **J. Phys. Chem. B.** 106(26): 6617-6633.

Bloom, J.D., C.O. Wilke, F.H. Arnold and C. Adami. 2004. Stability and the evolvability of function in a model protein. **Biophys. J.** 86(5): 2758-2764.

Bonaccorsi, L., M. Muratori, S. Marchiani, G. Forti and E. Baldi. 2006. The androgen receptor and prostate cancer invasion. **Mol. Cell. Endocrinol.** 246(1-2): 157-162.

Bonnet, P. and R.A. Bryce. 2004. Molecular dynamics and free energy analysis of neuraminidase-ligand interactions. **Protein Sci.** 13(4): 946-957.

Boran, A.D., J. Seco, V. Jayaraman, G. Jayaraman, S. Zhao, S. Reddy, Y. Chen and R. Iyengar. 2012. A potential peptide therapeutic derived from the juxtamembrane domain of the epidermal growth factor receptor. **PLoS One** 7(11): e49702.

Bose, R. and X. Zhang. 2009. The ErbB kinase domain: structural perspectives into kinase activation and inhibition. **Exp. Cell Res.** 315(4): 649-658.

Bosshard, H.R., D.N. Marti and I. Jelesarov. 2004. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. **J. Mol. Recognit.** 17(1): 1-16.

Bradshaw, R.T., B.H. Patel, E.W. Tate, R.J. Leatherbarrow and I.R. Gould. 2011. Comparing experimental and computational alanine scanning techniques for probing a prototypical protein-protein interaction. **Protein. Eng. Des. Sel.** 24(1-2): 197-207.

Brooks, B.R., D. Janežič and M. Karplus. 1995. Harmonic analysis of large systems. I. Methodology. **J. Comput. Chem.** 16(12): 1522-1542.

Brown, J., C. Delaine, O.J. Zaccheo, C. Siebold, R.J. Gilbert, G. van Boxel, A. Denley, J.C. Wallace, A.B. Hassan, B.E. Forbes and E.Y. Jones. 2008. Structure and functional analysis of the IGF-II/IGF2R interaction. **EMBO J.** 27(1): 265-276.

Bullock, B.N., A.L. Jochim and P.S. Arora. 2011. Assessing helical protein interfaces for inhibitor design. **J. Am. Chem. Soc.** 133(36): 14220-14223.

Case, D.A., T.E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R.J. Woods. 2005. The Amber biomolecular simulation programs. **J. Comput. Chem.** 26(16): 1668-1688.

Chantry, A. 1995. The kinase domain and membrane localization determine intracellular interactions between epidermal growth factor receptors. **J. Biol. Chem.** 270(7): 3068-3073.

Chen, H., C.F. Xu, J. Ma, A.V. Eliseenkova, W. Li, P.M. Pollock, N. Pitteloud, W.T. Miller, T.A. Neubert and M. Mohammadi. 2008. A crystallographic snapshot of tyrosine trans-phosphorylation in action. **Proc. Natl. Acad. Sci. U S A.** 105(50): 19660-19665.

Chen, J.J., S. Han, Y. Cao and J.Z. Chen. 2013. The agonist binding mechanism of human CB2 receptor studied by molecular dynamics simulation, free energy calculation and 3D-QSAR studies. **YaoXue XueBao** 48(9): 1436-1449.

Choi, S.H., J.M. Mendrola and M.A. Lemmon. 2007. EGF-independent activation of cell-surface EGF receptors harboring mutations found in gefitinib-sensitive lung cancer. **Oncogene** 26(11): 1567-1576.

Chrapkiewicz, N.B., C.M. Davis, D.T. Chu, C.M. Caldwell and D.K. Granner. 1989. Rat gene 33: analysis of its structure, messenger RNA and basal promoter activity. **Nucleic Acids Res.** 17(16): 6651-6667.

Cohen, S. 1962. Isolation of a mouse submaxillary gland protein accelerating incisor eruption and eyelid opening in the new-born animal. **J. Biol. Chem.** 237: 1555-1562.

Cooper, J.A. and B. Howell. 1993. The when and how of Src regulation. **Cell** 73(6): 1051-1054.

Costa, D.B., B. Halmos, A. Kumar, S.T. Schumer, M.S. Huberman, T.J. Boggon, D.G. Tenen and S. Kobayashi. 2007. BIM mediates EGFR tyrosine kinase inhibitor-induced apoptosis in lung cancers with oncogenic EGFR mutations. **PLoS Med.** 4(10): 1669-1679; discussion 1680.

Craik, D.J., D.P. Fairlie, S. Liras and D. Price. 2013. The future of peptide-based drugs. **Chem. Biol. Drug Des.** 81(1): 136-147.

Cui, Q., T. Sulea, J.D. Schrag, C. Munger, M.N. Hung, M. Naim, M. Cygler and E.O. Purisima. 2008. Molecular dynamics-solvated interaction energy studies of protein-protein interactions: the MP1-p14 scaffolding complex. **J. Mol. Biol.** 379(4): 787-802.

Darden, T., D. York and L. Pedersen. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. **J. Chem. Physics.** 98(12): 10089-10092.

David, C. 2005. Interfaces and the driving force of hydrophobic assembly. **Nature** 437(7059): 640-647.

Davies, S.P., H. Reddy, M. Caivano and P. Cohen. 2000. Specificity and mechanism of action of some commonly used protein kinase inhibitors. **Biochem. J.** 351(Pt 1): 95-105.

Descot, A., R. Hoffmann, D. Shaposhnikov, M. Reschke, A. Ullrich and G. Posern. 2009. Negative regulation of the EGFR-MAPK cascade by actin-MAL-mediated Mig6/Errfi-1 induction. **Mol. Cell.** 35(3): 291-304.

Doebele, R.C., A.B. Oton, N. Peled, D.R. Camidge and P.A. Bunn, Jr. 2010. New strategies to overcome limitations of reversible EGFR tyrosine kinase inhibitor therapy in non-small cell lung cancer. **Lung Cancer** 69(1): 1-12.

Dong, F., B. Olsen and N.A. Baker. 2008a. Computational methods for biomolecular electrostatics. **Methods Cell Biol.** 84: 843-870.

Dong, F., J.A. Wagoner and N.A. Baker. 2008b. Assessing the performance of implicit solvation models at a nucleic acid surface. **Phys. Chem. Chem. Phys.** 10(32): 4889-4902.

Dornieden, S., A. Müller-Schiffmann, H. Sticht, N. Jiang, Y. Cinar, M. Wördehoff, C. Korth, S.A. Funke and D. Willbold. 2013. Characterization of a Single-Chain Variable Fragment Recognizing a Linear Epitope of Aβ: A Biotechnical Tool for Studies on Alzheimer's Disease? **PloS one** 8(3): e59820.

Doura, A.K. and K.G. Fleming. 2004. Complex Interactions at the Helix–Helix Interface Stabilize the Glycophorin A Transmembrane Dimer. **J. Mol. Biol.** 343(5): 1487-1497.

Drost-Hansen, W. 1969. The Structure and Properties of Water. *In* D. Eisenberg and W. Kauzmann., Eds. **Science**. Science Oxford University Press, New York.

Duan, Y., C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. **J. Comput. Chem.** 24(16): 1999-2012.

Dummler, B. and B.A. Hemmings. 2007. Physiological roles of PKB/Akt isoforms in development and disease. **Biochem. Soc. Trans.** 35(Pt 2): 231-235.

Dunn, K.L., P.S. Espino, B. Drobic, S. He and J.R. Davie. 2005. The Ras-MAPK signal transduction pathway, cancer and chromatin remodeling. **Biochem. Cell Biol.** 83(1): 1-14.

Eccles, S.A. 2011. The epidermal growth factor receptor/Erb-B/HER family in normal and malignant breast biology. **Int. J. Dev. Biol.** 55(7-9): 685-696.

Eldridge, M.D., C.W. Murray, T.R. Auton, G.V. Paolini and R.P. Mee. 1997. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. **J. Comput. Aided Mol. Des.** 11(5): 425-445.

Espinoza-Fonseca, L.M. and J.G. Trujillo-Ferrara. 2006. Conformational changes of the p53-binding cleft of MDM2 revealed by molecular dynamics simulations. **Biopolymers** 83(4): 365-373.

Fanning, A.S. and J.M. Anderson. 1999. Protein modules as organizers of membrane structure. **Curr. Opin. Cell Biol.** 11(4): 432-439.

Feig, M., A. Onufriev, M.S. Lee, W. Im, D.A. Case and C.L. Brooks. 2004. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. **J. Comput. Chem.** 25(2): 265-284.

Fiorini, M., C. Ballaro, G. Sala, G. Falcone, S. Alema and O. Segatto. 2002. Expression of RALT, a feedback inhibitor of ErbB receptors, is subjected to an integrated transcriptional and post-translational control. **Oncogene** 21(42): 6530-6539.

Flynn, J.F., C. Wong and J.M. Wu. 2009. Anti-EGFR Therapy: Mechanism and Advances in Clinical Efficacy in Breast Cancer. **J. Oncol.** 2009: 526963.

Foloppe, N. and R. Hubbard. 2006. Towards predictive ligand design with free-energy based computational methods? **Curr. Med. Chem.** 13(29): 3583-3608.

Frosi, Y., S. Anastasi, C. Ballaro, G. Varsano, L. Castellani, E. Maspero, S. Polo, S. Alema and O. Segatto. 2010. A two-tiered mechanism of EGFR inhibition by RALT/MIG6 via kinase suppression and receptor degradation. **J. Cell Biol.** 189(3): 557-571.

Garcia-Echeverria, C. and W.R. Sellers. 2008. Drug discovery approaches targeting the PI3K/Akt pathway in cancer. **Oncogene** 27(41): 5511-5526.

Gohlke, H. and D.A. Case. 2004. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. **J. Comput. Chem.** 25(2): 238-250.

Goldstein, N.I., M. Prewett, K. Zuklys, P. Rockwell and J. Mendelsohn. 1995. Biological efficacy of a chimeric antibody to the epidermal growth factor receptor in a human tumor xenograft model. **Clin. Cancer Res.** 1(11): 1311-1318.

Gouda, H., I.D. Kuntz, D.A. Case and P.A. Kollman. 2003. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. **Biopolymers** 68(1): 16-34.

Grabowski, S.J. 2006. **Hydrogen Bonding - New Insights: New Insights.** Springer Netherlands, Netherland.

Greenhalgh, C.J. and D.J. Hilton. 2001. Negative regulation of cytokine signaling. **J. Leukoc. Biol.** 70(3): 348-356.

Grunwald, V. and M. Hidalgo. 2003. Developing inhibitors of the epidermal growth factor receptor for cancer treatment. **J. Natl. Cancer Inst.** 95(12): 851-867.

Guo, J., X. Wang, H. Sun, H. Liu and X. Yao. 2012. The molecular basis of IGF-II/IGF2R recognition: a combined molecular dynamics simulation, free-energy calculation and computational alanine scanning study. **J. Mol. Model.** 18(4): 1421-1430.

Haberl, F., O. Othersen, U. Seidel, H. Lanig and T. Clark. 2009. **Investigating Protein-Protein and Protein-Ligand Interactions by Molecular Dynamics Simulations**. Springer Berlin Heidelberg, Germany.

Hackel, P.O., E. Zwick, N. Prenzel and A. Ullrich. 1999. Epidermal growth factor receptors: critical mediators of multiple receptor pathways. **Curr. Opin. Cell Biol.** 11(2): 184-189.

Hayes, J.M. and G. Archontis. 2012. MM-GB (PB) SA calculations of protein-ligand binding free energies. **Molecular dynamics–studies of synthetic and biological macromolecules.** In Tech, Kroatien.

Hayward, S. and B.L. de Groot. 2008. Normal modes and essential dynamics. **Methods Mol. Biol.** 443: 89-106.

Hendsch, Z.S. and B. Tidor. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. **Protein Science** 3(2): 211-226.

Hirsch, F.R. and P.A. Bunn. 2005. Epidermal Growth Factor Receptor Inhibitors in Lung Cancer: Smaller or Larger Molecules, Selected or Unselected Populations? **J. Clin. Oncol.** 23(36): 9044-9047.

Horn, J.R., T.R. Sosnick and A.A. Kossiakoff. 2009. Principal determinants leading to transition state formation of a protein-protein complex, orientation trumps side-chain interactions. **Proc. Natl. Acad. Sci. U S A.** 106(8): 2559-2564.

Horovitz, A. 1987. Non-additivity in protein-protein interactions. **J. Mol. Biol.** 196(3): 733-735.

_____. 1996. Double-mutant cycles: a powerful tool for analyzing protein structure and function. **Fold Des.** 1(6): R121-126.

Hou, T., J. Wang, Y. Li and W. Wang. 2011. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. **J. Chem. Inf. Model.** 51(1): 69-82.

Hou, T., Z. Xu, W. Zhang, W.A. McLaughlin, D.A. Case, Y. Xu and W. Wang. 2009. Characterization of domain-peptide interaction interface: a generic structure-based model to decipher the binding specificity of SH3 domains. **Mol. Cell Proteomics.** 8(4): 639-649.

Hou, T. and R. Yu. 2007. Molecular dynamics and free energy studies on the wild-type and double mutant HIV-1 protease complexed with amprenavir and two amprenavir-related inhibitors: mechanism for binding and drug resistance. **J. Med. Chem.** 50(6): 1177-1188.

Huang, N., C. Kalyanaraman, K. Bernacki and M.P. Jacobson. 2006. Molecular mechanics methods for predicting protein-ligand binding. **Phys. Chem. Chem. Phys.** 8(44): 5166-5177.

Hubbard, R.E. and M. Kamran Haider. 2001. **Hydrogen Bonds in Proteins: Role and Strength**. John Wiley & Sons, Inc, UK

Hubbard, S.R. 1999. Structural analysis of receptor tyrosine kinases. **Prog. Biophys. Mol. Biol.** 71(3-4): 343-358.

Hubbard, S.R. and J.H. Till. 2000. Protein tyrosine kinase structure and function. **Annu. Rev. Biochem.** 69: 373-398.

Hubbard, S.R., L. Wei, L. Ellis and W.A. Hendrickson. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. **Nature** 372(6508): 746-754.

Huo, S., I. Massova and P.A. Kollman. 2002. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. **J. Comput. Chem.** 23(1): 15-27.

Hynes, N.E., K. Horsch, M.A. Olayioye and A. Badache. 2001. The ErbB receptor tyrosine family as signal integrators. **Endocr. Relat. Cancer.** 8(3): 151-159.

Istomin, A.Y., M.M. Gromiha, O.K. Vorov, D.J. Jacobs and D.R. Livesay. 2008. New insight into long-range nonadditivity within protein double-mutant cycles. **Proteins** 70(3): 915-924.

Jang, D.S., H.J. Cha, S.S. Cha, B.H. Hong, N.C. Ha, J.Y. Lee, B.H. Oh, H.S. Lee and K.Y. Choi. 2004. Structural double-mutant cycle analysis of a hydrogen bond network in ketosteroid isomerase from Pseudomonas putida biotype B. **Biochem. J.** 382(Pt 3): 967-973.

Jiang, X., J. Kowalski and J.W. Kelly. 2001. Increasing protein stability using a rational approach combining sequence homology and structural alignment: Stabilizing the WW domain. **Protein Sci.** 10(7): 1454-1465.

Jones, S. and J.M. Thornton. 1996. Principles of protein-protein interactions. **Proc. Natl. Acad. Sci. U S A.** 93(1): 13-20.

Jorgensen, W.L., J. Chandrasekhar, J.D. Madura, R.W. Impey and M.L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. **J. Chem. Physics.** 79(2): 926-935.

Jura, N., N.F. Endres, K. Engel, S. Deindl, R. Das, M.H. Lamers, D.E. Wemmer, X. Zhang and J. Kuriyan. 2009. Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. **Cell** 137(7): 1293-1307.

Juy, M., H. Lam-Thanh, K. Lintner and S. Fermandjian. 1983. Conformation and mobility of tyrosine side chain in tetrapeptides: Specific effects of cis- and trans -proline in Tyr-Pro- and Pro-Tyr-segments. **Int. J. Pept. Protein Res.** 22(4): 437-449.

Kannan, S. and M. Zacharias. 2014. Role of Tryptophan Side Chain Dynamics on the Trp-Cage Mini-Protein Folding Studied by Molecular Dynamics Simulations. **PLoS ONE** 9(2): e88383.

Kar, P., R. Lipowsky and V. Knecht. 2013. Importance of Polar Solvation and Configurational Entropy for Design of Antiretroviral Drugs Targeting HIV-1 Protease. **J. Phys. Chem. B.** 117(19): 5793-5805.

Karplus, M., A.T. Brunger, R. Elber and J. Kuriyan. 1987. Molecular dynamics: applications to proteins. **Cold Spring Harb Symp. Quant. Biol.** 52: 381-390.

Karplus, M. and J. Kuriyan. 2005. Molecular dynamics and protein function. **Proc. Natl. Acad. Sci. U S A.** 102(19): 6679-6685.

Kawahara, A., K. Azuma, A. Sumi, T. Taira, K. Nakashima, E. Aikawa, H. Abe, T. Yamaguchi, S. Takamori, J. Akiba and M. Kage. 2011. Identification of non-small-cell lung cancer with activating EGFR mutations in malignant effusion and cerebrospinal fluid: rapid and sensitive detection of exon 19 deletion E746-A750 and exon 21 L858R mutation by immunocytochemistry. **Lung Cancer** 74(1): 35-40.

Khandelia, H. and Y.N. Kaznessis. 2007. Structure of the antimicrobial β-hairpin peptide protegrin-1 in a DLPC lipid bilayer investigated by molecular dynamics simulation. **Biochim. Biophys. Acta.** 1768(3): 509-520.

Kholodenko, B.N., O.V. Demin, G. Moehren and J.B. Hoek. 1999. Quantification of short term signaling by the epidermal growth factor receptor. **J. Biol. Chem.** 274(42): 30169-30181.

Kisseleva, T., S. Bhattacharya, J. Braunstein and C.W. Schindler. 2002. Signaling through the JAK/STAT pathway, recent advances and future challenges. **Gene** 285(1-2): 1-24.

Kleywegt, G.J. and T.A. Jones. 1996. Phi/psi-chology: Ramachandran revisited. **Structure** 4(12): 1395-1400.

Knaul, F.M. and B.A. Chabner. 2013. Announcing a New Section: Global Health and Cancer. **The oncologist** 18(5): 485-486.

Kobayashi, S., T.J. Boggon, T. Dayaram, P.A. Janne, O. Kocher, M. Meyerson, B.E. Johnson, M.J. Eck, D.G. Tenen and B. Halmos. 2005. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. **N. Engl. J. Med.** 352(8): 786-792.

Kratochvil, B., H.L. Yeager and V. Gutmann. 1972. **Nonaqueous Chemistry.** Springer berlin Herbberg, New York.

Kumar, A., E.T. Petri, B. Halmos and T.J. Boggon. 2008. Structure and clinical relevance of the epidermal growth factor receptor in human cancer. **J. Clin. Oncol.** 26(10): 1742-1751.

Kumar, S. and R. Nussinov. 2002. Relationship between ion pair geometries and electrostatic strengths in proteins. **Biophys. J.** 83(3): 1595-1612.

Kundrotas, P.J. and E. Alexov. 2006. Electrostatic properties of protein-protein complexes. **Biophys. J.** 91(5): 1724-1736.

Kuriyan, J., X. ZHANG and P. Cole. 2012. **Inhibitors of the EGFR kinase targeting the asymmetric activating dimer interface**. U.S. Patent NO. 8,242,080 B2

Kussie, P.H., S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A.J. Levine and N.P. Pavletich. 1996. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. **Science** 274(5289): 948-953.

Landau, M. and N. Ben-Tal. 2008. Dynamic equilibrium between multiple active and inactive conformations explains regulation and oncogenic mutations in ErbB receptors. **Biochim. Biophys. Acta.** 1785(1): 12-31.

Landau, M., S.J. Fleishman and N. Ben-Tal. 2004. A putative mechanism for downregulation of the catalytic activity of the EGF receptor via direct contact between its kinase and C-terminal domains. **Structure** 12(12): 2265-2275.

Laskowski, R.A. 2009. Structural quality assurance. **Structural Bioinformatics.** John Wiley & Son, Inc, New York.

Laskowski, R.A., M.W. MacArthur, D.S. Moss and J.M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. **J. Appl. Crystallogr.** 26(2): 283-291.

Laskowski, R.A. and M.B. Swindells. 2011. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. **J. Chem. Inf. Model.** 51(10): 2778-2786.

Laurence, C. and M. Berthelot. 2000. Observations on the strength of hydrogen bonding. **Perspect. Drug Discovery Des.** 18(1): 39-60.

Lax, I., A.K. Mitra, C. Ravera, D.R. Hurwitz, M. Rubinstein, A. Ullrich, R.M. Stroud and J. Schlessinger. 1991. Epidermal growth factor (EGF) induces oligomerization of soluble, extracellular, ligand-binding domain of EGF receptor. A low resolution projection structure of the ligand-binding domain. **J. Biol. Chem.** 266(21): 13828-13833.

Leevers, S.J., B. Vanhaesebroeck and M.D. Waterfield. 1999. Signalling through phosphoinositide 3-kinases: the lipids take centre stage. **Curr. Opin. Cell Biol.** 11(2): 219-225.

Leite, F.L., C.C. Bueno, A.L. Da Roz, E.C. Ziemath and O.N. Oliveira. 2012. Theoretical models for surface forces and adhesion and their measurement using atomic force microscopy. **Int. J. Mol. Sci.** 13(10): 12773-12856.

Lemmon, M.A. and J. Schlessinger. 1994. Regulation of signal transduction and signal diversity by receptor oligomerization. **Trends Biochem. Sci.** 19(11): 459-463.

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

Levy, R.M., L.Y. Zhang, E. Gallicchio and A.K. Felts. 2003. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. **J. Am. Chem. Soc.** 125(31): 9523-9530.

Levy, Y. and J.N. Onuchic. 2006. Water mediation in protein folding and molecular recognition. **Annu. Rev. Biophys. Biomol. Struct.** 35: 389-415.

Li, T., M. Froeyen and P. Herdewijn. 2008. Computational alanine scanning and free energy decomposition for E. coli type I signal peptidase with lipopeptide inhibitor complex. **J Mol Graph Model.** 26(5): 813-823.

Lima, T.W.d., A. Caliri, F.L.B.d. Silva, R. Tinós, G. Travieso, I.N.d. Silva, P.S.L.d. Souza, E. Marques, A.C.B. Delbem, V. Bonatto, R. Faccioli, C.R.S. Brasil, P.H.R. Gabriel, V.T.d. Ó and D.R.F. Bonetti. 2009. Some Modeling Issues for Protein Structure Prediction Using Evolutionary Algorithms, pp 156-157. *In* Wellington Pinheiro dos Santos., Ed. **Evolutionary Computation**. T-Tech, Australia.

Lindahl, E.R. 2008. Molecular dynamics simulations. **Methods Mol. Biol.** 443: 3-23.

Liu, P., T. Sudhaharan, R.M. Koh, L.C. Hwang, S. Ahmed, I.N. Maruyama and T. Wohland. 2007. Investigation of the dimerization of proteins from the epidermal growth factor receptor family by single wavelength fluorescence cross-correlation spectroscopy. **Biophys. J.** 93(2): 684-698.

Lynch, T.J., D.W. Bell, R. Sordella, S. Gurubhagavatula, R.A. Okimoto, B.W. Brannigan, P.L. Harris, S.M. Haserlat, J.G. Supko, F.G. Haluska, D.N. Louis, D.C. Christiani, J. Settleman and D.A. Haber. 2004. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. **N. Engl. J. Med.** 350(21): 2129-2139.

Macdonald-Obermann, J.L. and L.J. Pike. 2009. The intracellular juxtamembrane domain of the epidermal growth factor (EGF) receptor is responsible for the allosteric regulation of EGF binding. **J. Biol. Chem.** 284(20): 13570-13576.

Maemondo, M., A. Inoue, K. Kobayashi, S. Sugawara, S. Oizumi, H. Isobe, A. Gemma, M. Harada, H. Yoshizawa, I. Kinoshita, Y. Fujita, S. Okinaga, H. Hirano, K. Yoshimori, T. Harada, T. Ogura, M. Ando, H. Miyazawa, T. Tanaka, Y. Saijo, K. Hagiwara, S. Morita, T. Nukiwa and G. North-East Japan Study. 2010. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. **N. Engl. J. Med.** 362(25): 2380-2388.

Mahoney, M.W. and W.L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. **J. Chem. Physics.** 112(20): 8910-8922.

Makkinje, A., D.A. Quinn, A. Chen, C.L. Cadilla, T. Force, J.V. Bonventre and J.M. Kyriakis. 2000. Gene 33/Mig-6, a transcriptionally inducible adapter protein that binds GTP-Cdc42 and activates SAPK/JNK. A potential marker transcript for chronic pathologic conditions, such as diabetic nephropathy. Possible role in the response to persistent stress. **J. Biol. Chem.** 275(23): 17838-17847.

Marrero, M.B. 2005. Introduction to JAK/STAT signaling and the vasculature. **Vasc. Pharmacol.** 43(5): 307-309.

Martinez-Marti, A. and E. Felip. 2011. PI3K Pathway in NSCLC. **Front. Oncol.** 1: 55.

Masago, K., Y. Togashi, M. Fukudo, T. Terada, K. Irisa, Y. Sakamori, S. Fujita, Y.H. Kim, T. Mio, K.I. Inui and M. Mishima. 2010. Good Clinical Response to Erlotinib in a Non-Small Cell Lung Cancer Patient Harboring Multiple Brain Metastases and a Double Active Somatic Epidermal Growth Factor Gene Mutation. **Case. Rep. Oncol.** 3(2): 98-105.

Massova, I. and P.A. Kollman. 1999. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. **J. Am. Chem. Soc.** 121(36): 8133-8143.

Matthews, B.W. 2001. **Hydrophobic Interactions in Proteins**. John Wiley & Sons, Inc. Chichester

Mattoon, D., P. Klein, M.A. Lemmon, I. Lax and J. Schlessinger. 2004. The tethered configuration of the EGF receptor extracellular domain exerts only a limited control of receptor function. **Proc. Natl. Acad. Sci. U S A.** 101(4): 923-928.

McCubrey, J.A., L.S. Steelman, S.L. Abrams, J.T. Lee, F. Chang, F.E. Bertrand, P.M. Navolanic, D.M. Terrian, R.A. Franklin, A.B. D'Assoro, J.L. Salisbury, M.C. Mazzarino, F. Stivala and M. Libra. 2006. Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. **Adv. Enzyme Regul.** 46: 249-279.

McLaughlin, S., S.O. Smith, M.J. Hayman and D. Murray. 2005. An electrostatic engine model for autoinhibition and activation of the epidermal growth factor receptor (EGFR/ErbB) family. **J. Gen. Physiol.** 126(1): 41-53.

Megy, S., G. Bertho, J. Gharbi-Benarous, N. Evrard-Todeschi, G. Coadou, E. Segeral, C. Iehle, E. Quemeneur, R. Benarous and J.P. Girault. 2005. STD and TRNOESY NMR studies on the conformation of the oncogenic protein beta-catenin containing the phosphorylated motif DpSGXXpS bound to the beta-TrCP protein. **J. Biol. Chem.** 280(32): 29107-29116.

Meiselbach, H., A.H. Horn, T. Harrer and H. Sticht. 2007. Insights into amprenavir resistance in E35D HIV-1 protease mutation from molecular dynamics and binding free-energy calculations. **J. Mol. Model.** 13(2): 297-304.

Merlino, G.T., S. Ishii, J. Whang-Peng, T. Knutsen, Y.H. Xu, A.J. Clark, R.H. Stratton, R.K. Wilson, D.P. Ma, B.A. Roe and et al. 1985. Structure and localization of genes encoding aberrant and normal epidermal growth factor receptor RNAs from A431 human carcinoma cells. **Mol. Cell. Biol.** 5(7): 1722-1734.

Merz Jr, K.M., D. Ringe and C.H. Reynolds. 2010. **Drug design: structure-and ligand-based approaches.** Cambridge University Press, USA.

Michel , J., N. Foloppe and J.W. Essex. 2010. Rigorous Free Energy Calculations in Structure-Based Drug Design. **Molecular Informatics** 29(8-9): 570-578.

Mildvan, A.S. 2004. Inverse Thinking about Double Mutants of Enzymes. **Biochemistry** 43(46): 14517-14520.

Mir, M.M., N.A. Dar, I. Salam and Z.A. Shah. 2008. Mutations in epidermal growth factor receptor gene in esophageal squamous cell carcinoma patients in kashmir- a high incidence area of India. **Int. J. Health Sci.** 2(2): 17-25.

Mlcochova, J., P. Faltejskova, R. Nemecek, M. Svoboda and O. Slaby. 2013. MicroRNAs targeting EGFR signalling pathway in colorectal cancer. **J. Cancer Res. Clin. Oncol.** 139(10): 1615-1624.

Montagut, C., A. Dalmases, B. Bellosillo, M. Crespo, S. Pairet, M. Iglesias, M. Salido, M. Gallen, S. Marsters, S.P. Tsai, A. Minoche, S. Seshagiri, S. Serrano, H. Himmelbauer, J. Bellmunt, A. Rovira, J. Settleman, F. Bosch and J. Albanell. 2012. Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. **Nat. Med.** 18(2): 221-223.

Moreira, I., P. Fernandes and M. Ramos. 2008. Protein–protein recognition: a computational mutagenesis study of the MDM2–P53 complex. **Theor. Chem. Acc.** 120(4-6): 533-542.

Moreira, I.S., P.A. Fernandes and M.J. Ramos. 2007a. Computational alanine scanning mutagenesis--an improved methodological approach. **J. Comput. Chem.** 28(3): 644-654.

_____. 2007b. Hot spots--a review of the protein-protein interface determinant amino-acid residues. **Proteins** 68(4): 803-812.

Moriki, T., H. Maruyama and I.N. Maruyama. 2001. Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain. **J. Mol. Biol.** 311(5): 1011-1026.

Morrison, K.L. and G.A. Weiss. 2001. Combinatorial alanine-scanning. **Curr. Opin. Chem. Biol.** 5(3): 302-307.

Morrow, J.K., L. Du-Cuny, L. Chen, E.J. Meuillet, E.A. Mash, G. Powis and S. Zhang. 2011. Recent development of anticancer therapeutics targeting Akt. **Recent Pat. Anticancer Drug Discov.** 6(1): 146-159.

Mu, X.L., L.Y. Li, X.T. Zhang, M.Z. Wang, R.E. Feng, Q.C. Cui, H.S. Zhou and B.Q. Guo. 2005. Gefitinib-sensitive mutations of the epidermal growth factor receptor tyrosine kinase domain in Chinese patients with non–small cell lung cancer. **Clin. Cancer Res.** 11(12): 4289-4294.

Nicholas, M.K., R.V. Lukas, N.F. Jafri, L. Faoro and R. Salgia. 2006. Epidermal growth factor receptor - mediated signal transduction in the development and therapy of gliomas. **Clin. Cancer Res.** 12(24): 7261-7270.

Nilsson, J., C. Vallbo, D. Guo, I. Golovleva, B. Hallberg, R. Henriksson and H. Hedman. 2001. Cloning, characterization, and expression of human LIG1. **Biochem. Biophys. Res. Commun.** 284(5): 1155-1161.

Nowak, W. 2012. Applications of Computational Methods to Simulations of Proteins Dynamics, pp.1127-1153. *In* J.Leszczynski., ed. **Handbook of Computational Chemistry.** n.p.

Nymeyer, H. and A.E. Garcia. 2003. Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized Born approximation with explicit solvent. **Proc. Natl. Acad. Sci. U S A.** 100(24): 13934-13939.

O'Shea, J.J., M. Pesu, D.C. Borie and P.S. Changelian. 2004. A new modality for immunosuppression: targeting the JAK/STAT pathway. **Nat. Rev. Drug Discov.** 3(7): 555-564.

Ogiso, H., R. Ishitani, O. Nureki, S. Fukai, M. Yamanaka, J.H. Kim, K. Saito, A. Sakamoto, M. Inoue, M. Shirouzu and S. Yokoyama. 2002. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. **Cell** 110(6): 775-787.

Okines, A., D. Cunningham and I. Chau. 2011. Targeting the human EGFR family in esophagogastric cancer. **Nat. Rev. Clin. Oncol.** 8(8): 492-503.

Olayioye, M.A., R.M. Neve, H.A. Lane and N.E. Hynes. 2000. The ErbB signaling network: receptor heterodimerization in development and cancer. **EMBO J.** 19(13): 3159-3167.

Olmez, E.O. and B.S. Akbulut. 2012. Protein-Peptide Interactions Revolutionize Drug Development. *In* Kotb Abddelmohsen., ed. **Binding Protein.** In Tech, Croatia.

Onufriev, A., D. Bashford and D.A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. **Proteins** 55(2): 383-394.

Otzen, D. 2005. Antagonism, non-native interactions and non-two-state folding in S6 revealed by double-mutant cycle analysis. **Protein Eng. Des. Sel.** 18(11): 547-557.

Papakyriakou, A., D. Vourloumis, F. Tzortzatou-Stathopoulou and M. Karpusas. 2009. Conformational dynamics of the EGFR kinase domain reveals structural features involved in activation. **Proteins** 76(2): 375-386.

Park, J.H., Y. Liu, M.A. Lemmon and R. Radhakrishnan. 2012. Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain. **Biochem. J.** 448(3): 417-423.

Pellicena, P. and J. Kuriyan. 2006. Protein-protein interactions in the allosteric regulation of protein kinases. **Curr. Opin. Struct. Biol.** 16(6): 702-709.

Perez-Jimenez, R., R. Godoy-Ruiz, B. Ibarra-Molero and J.M. Sanchez-Ruiz. 2004. The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect? **Biophys. J.** 86(4): 2414-2429.

Perez, M.A., S.F. Sousa, E.F. Oliveira, P.A. Fernandes and M.J. Ramos. 2011. Detection of farnesyltransferase interface hot spots through computational alanine scanning mutagenesis. **J. Phys. Chem. B.** 115(51): 15339-15354.

Petsko, G.A. and D. Ringe. 2004. **Protein Structure and Function**. n.p.

Picksley, S.M. and D.P. Lane. 1993. The p53-mdm2 autoregulatory feedback loop: a paradigm for the regulation of growth control by p53? **Bioessays.** 15(10): 689-690.

Polier, G., J. Neumann, F. Thuaud, N. Ribeiro, C. Gelhaus, H. Schmidt, M. Giaisi, R. Kohler, W.W. Muller, P. Proksch, M. Leippe, O. Janssen, L. Desaubry, P.H. Krammer and M. Li-Weber. 2012. The natural anticancer compounds rocaglamides inhibit the Raf-MEK-ERK pathway by targeting prohibitin 1 and 2. **Chem. Biol.** 19(9): 1093-1104.

Ponder, J.W. and D.A. Case. 2003. Force fields for protein simulations. **Adv. Protein Chem.** 66: 27-85.

Popowicz, G.M., A. Domling and T.A. Holak. 2011. The structure-based design of Mdm2/Mdmx-p53 inhibitors gets serious. **Angew. Chem. Int. Ed. Engl.** 50(12): 2680-2688.

Qian, X., C.M. LeVea, J.K. Freeman, W.C. Dougall and M.I. Greene. 1994. Heterodimerization of epidermal growth factor receptor and wild-type or kinase-deficient Neu: a mechanism of interreceptor kinase activation and transphosphorylation. **Proc. Natl. Acad. Sci. U S A.** 91(4): 1500-1504.

Ramachandran, G.N., C. Ramakrishnan and V. Sasisekharan. 1963. Stereochemistry of polypeptide chain configurations. **J. Mol. Biol.** 7: 95-99.

Rawlings, J.S., K.M. Rosler and D.A. Harrison. 2004. The JAK/STAT signaling pathway. **J. Cell Sci.** 117(Pt 8): 1281-1283.

Red Brewer, M., S.H. Choi, D. Alvarado, K. Moravcevic, A. Pozzi, M.A. Lemmon and G. Carpenter. 2009. The juxtamembrane region of the EGF receptor functions as an activation domain. **Mol. Cell.** 34(6): 641-651.

Ribeiro, J., N.F.S.A. Cerqueira, I. Moreira, P. Fernandes and M. Ramos. 2012. CompASM: an Amber-VMD alanine scanning mutagenesis plug-in. **Theor. Chem. Acc.** 131(10): 1-7.

Roskoski, R., Jr. 2004. The ErbB/HER receptor protein-tyrosine kinases and cancer. **Biochem. Biophys. Res. Commun.** 319(1): 1-11.

Rosti, G., F. Castagnetti, G. Gugliotta, F. Palandri, G. Martinelli and M. Baccarani. 2010. Dasatinib and nilotinib in imatinib-resistant Philadelphia-positive chronic myelogenous leukemia: a 'head-to-head comparison'. **Leuk. Lymphoma.** 51(4): 583-591.

Rubin, A. 2012. **Statistics for evidence-based practice and evaluation.** n.p.

Rubin, C., G. Gur and Y. Yarden. 2005. Negative regulation of receptor tyrosine kinases: unexpected links to c-Cbl and receptor ubiquitylation. **Cell Res.** 15(1): 66-71.

Rudgers, G.W., W. Huang and T. Palzkill. 2001. Binding properties of a peptide derived from beta-lactamase inhibitory protein. **Antimicrob. Agents Chemother.** 45(12): 3279-3286.

Ryckaert, J.-P., G. Ciccotti and H.J.C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **J. Comput. Physics.** 23(3): 327-341.

Saggu, M., N.M. Levinson and S.G. Boxer. 2012. Experimental quantification of electrostatics in X-H...pi hydrogen bonds. **J. Am. Chem. Soc.** 134(46): 18986-18997.

Saiz-Urra, L., M.A. Cabrera and M. Froeyen. 2011. 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.** 29(5): 726-739.

Salomon-Ferrer, R., D.A. Case and R.C. Walker. 2013. An overview of the Amber biomolecular simulation package. **Wiley Interdisciplinary Reviews: Computational Molecular Science.** 3(2): 198-210.

Saxon, M.L. and D.C. Lee. 1999. Mutagenesis reveals a role for epidermal growth factor receptor extracellular subdomain IV in ligand binding. **J. Biol. Chem.** 274(40): 28356-28362.

Scagliotti, G.V., G. Selvaggi, S. Novello and F.R. Hirsch. 2004. The biology of epidermal growth factor receptor in lung cancer. **Clin. Cancer Res.** 10(12 Pt 2): 4227s-4232s.

Schapira, M., M. Totrov and R. Abagyan. 1999. Prediction of the binding energy for small molecules, peptides and proteins. **J. Mol. Recognit.** 12(3): 177-190.

Schreiber, G. and A.R. Fersht. 1995. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. **J. Mol. Biol.** 248(2): 478-486.

Schreiber, G. and S.J. Fleishman. 2013. Computational design of protein–protein interactions. **Curr. Opin. Struct. Biol.** 23(6): 903-910.

Segatto, O., S. Anastasi and S. Alema. 2011. Regulation of epidermal growth factor receptor signalling by inducible feedback inhibitors. **J. Cell. Sci.** 124(Pt 11): 1785-1793.

Shahsavar, A., J.S. Kastrup, E.Ø. Nielsen, J.L. Kristensen, M. Gajhede and T. Balle. 2012. Crystal Structure of *Lymnaea stagnalis* AChBP Complexed with the Potent nAChR Antagonist DHβE Suggests a Unique Mode of Antagonism. **PLoS ONE** 7(8): e40757.

Shangary, S. and S. Wang. 2008. Targeting the MDM2-p53 interaction for cancer therapy. **Clin. Cancer Res.** 14(17): 5318-5324.

_____. 2009. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: a novel approach for cancer therapy. **Annu. Rev. Pharmacol. Toxicol.** 49: 223-241.

Sheinerman, F.B. and B. Honig. 2002. On the role of electrostatic interactions in the design of protein-protein interfaces. **J. Mol. Biol.** 318(1): 161-177.

Shen, F., Q. Lin, Y. Gu, C. Childress and W. Yang. 2007. Activated Cdc42-associated kinase 1 is a component of EGF receptor signaling complex and regulates EGF receptor degradation. **Mol. Biol. Cell.** 18(3): 732-742.

Shoelson, S.E. 1997. SH2 and PTB domain interactions in tyrosine kinase signal transduction. **Curr. Opin. Chem. Biol.** 1(2): 227-234.

Shoichet, B.K., W.A. Baase, R. Kuroki and B.W. Matthews. 1995. A relationship between protein stability and protein function. **Proc. Natl. Acad. Sci. U S A.** 92(2): 452-456.

Sibilia, M., R. Kroismayr, B.M. Lichtenberger, A. Natarajan, M. Hecking and M. Holcmann. 2007. The epidermal growth factor receptor: from development to tumorigenesis. **Differentiation** 75(9): 770-787.

Sinha, N. and R. Nussinov. 2001. Point mutations and sequence variability in proteins: Redistributions of preexisting populations. **Proc. Natl. Acad. Sci. U S A.** 98(6): 3139-3144.

Sokalingam, S., G. Raghunathan, N. Soundrarajan and S.-G. Lee. 2012. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. **PloS one** 7(7): e40410.

Songtawee, N., M.P. Gleeson and K. Choowongkomon. 2013. Computational study of EGFR inhibition: molecular dynamics studies on the active and inactive protein conformations. **J. Mol. Model.** 19(2): 497-509.

Sordella, R., D.W. Bell, D.A. Haber and J. Settleman. 2004. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. **Science** 305(5687): 1163-1167.

Sorkin, A. and L.K. Goh. 2008. Endocytosis and intracellular trafficking of ErbBs. **Exp. Cell Res.** 314(17): 3093-3106.

Sousa, S.r.F., B. Tamames, P.A. Fernandes and M.J.o. Ramos. 2011. Detailed Atomistic Analysis of the HIV-1 Protease Interface. **J. Phys. Chem. B.** 115(21): 7045-7057.

Stamos, J., M.X. Sliwkowski and C. Eigenbrot. 2002. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. **J. Biol. Chem.** 277(48): 46265-46272.

Steelman, L.S., F.E. Bertrand and J.A. McCubrey. 2004. The complexity of PTEN: mutation, marker and potential target for therapeutic intervention. **Expert Opin. Ther. Targets.** 8(6): 537-550.

Stein, A. and P. Aloy. 2008. Contextual specificity in peptide-mediated protein interactions. **PLoS One** 3(7): e2524.

Steiner, T. 2002. The Hydrogen Bond in the Solid State. **Angew. Chem. Int. Ed.** 41(1): 48-76.

Still, W.C., A. Tempczyk, R.C. Hawley and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. **J. Am. Chem. Soc.** 112(16): 6127-6129.

Stoica, I., S.K. Sadiq and P.V. Coveney. 2008. Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. **J. Am. Chem. Soc.** 130(8): 2639-2648.

Strub, C., C. Alies, A. Lougarre, C. Ladurantie, J. Czaplicki and D. Fournier. 2004. Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. **BMC Biochem.** 5: 9.

Tan, C., Y.H. Tan and R. Luo. 2007. Implicit nonpolar solvent models. **J. Phys. Chem. B.** 111(42): 12263-12274.

Tan, C., L. Yang and R. Luo. 2006. How well does Poisson-Boltzmann implicit solvent agree with explicit solvent? A quantitative analysis. **J. Phys. Chem. B.** 110(37): 18680-18687.

Tan, Y.H., C. Tan, J. Wang and R. Luo. 2008. Continuum polarizable force field within the Poisson-Boltzmann framework. **J. Phys. Chem. B.** 112(25): 7675-7688.

Tanizaki, S. and M. Feig. 2005. A generalized Born formalism for heterogeneous dielectric environments: application to the implicit modeling of biological membranes. **J. Chem. Phys.** 122(12): 124706.

Tanner, D.E., K.Y. Chan, J.C. Phillips and K. Schulten. 2011. Parallel Generalized Born Implicit Solvent Calculations with NAMD. **J. Chem. Theory Comput.** 7(11): 3635-3642.

Tao, R.H. and I.N. Maruyama. 2008. All EGF(ErbB) receptors have preformed homo- and heterodimeric structures in living cells. **J. Cell Sci.** 121(Pt 19): 3207-3217.

Telser, A. 2002. Molecular Biology of the Cell, 4th Edition. **Shock** 18(3): 289.

Thiel, K.W. and G. Carpenter. 2007. Epidermal growth factor receptor juxtamembrane region regulates allosteric tyrosine kinase activation. **Proc. Natl. Acad. Sci. U S A.** 104(49): 19238-19243.

Thomas, S. 2003. Electrostatics and dynamics of proteins. **Rep. Prog. Phys.** 66(5): 737.

Tiwari, A. and S.K. Panigrahi. 2007. HBAT: a complete package for analysing strong and weak hydrogen bonds in macromolecular crystal structures. **In Silico Biol.** 7(6): 651-661.

Toyooka, S., K. Kiura and T. Mitsudomi. 2005. EGFR mutation and response of lung cancer to gefitinib. **N. Engl. J. Med.** 352(20): 2136; author reply 2136.

Tue-ngeun, P., K. Kodchakorn, P. Nimmanpipug, N. Lawan, S. Nangola, C. Tayapiwatana, N.A. Rahman, S.M. Zain and V.S. Lee. 2013. Improved scFv anti-HIV-1 p17 binding affinity guided from the theoretical calculation of pairwise decomposition energies and computational alanine scanning. **Biomed Res. Int.** 2013: 713585.

Ullrich, A., L. Coussens, J.S. Hayflick, T.J. Dull, A. Gray, A.W. Tam, J. Lee, Y. Yarden, T.A. Libermann, J. Schlessinger and et al. 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. **Nature** 309(5967): 418-425.

van den Burg, B. and V.G.H. Eijsink. 2002. Selection of mutations for increased protein stability. **Curr. Opin. Biotechnol.** 13(4): 333-337.

Vaughan, C.K., P. Harryson, A.M. Buckle and A.R. Fersht. 2002. A structural double-mutant cycle: estimating the strength of a buried salt bridge in barnase. **Acta Crystallogr., Sect. D: Biol. Crystallogr.** 58(Pt 4): 591-600.

Voldborg, B.R., L. Damstrup, M. Spang-Thomsen and H.S. Poulsen. 1997. Epidermal growth factor receptor (EGFR) and EGFR mutations, function and possible role in clinical trials. **Ann. Oncol.** 8(12): 1197-1206.

Vousden, K.H. and X. Lu. 2002. Live or let die: the cell's response to p53. **Nat. Rev. Cancer.** 2(8): 594-604.

Wagoner, J.A. and N.A. Baker. 2006. Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. **Proc. Natl. Acad. Sci .U S A.** 103(22): 8331-8336.

Wallace, A.C., R.A. Laskowski and J.M. Thornton. 1995. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. **Protein Eng.** 8(2): 127-134.

Wan, S. and P.V. Coveney. 2011. Rapid and accurate ranking of binding affinities of epidermal growth factor receptor sequences with selected lung cancer drugs. **J. R. Soc. Interface.** 8(61): 1114-1127.

Wan, S., P.V. Coveney and D.R. Flower. 2007. **Molecular Dynamics Simulations.** 409: 321-339.

Wang, W.P., K.N. Wang, Q. Gao and L.Q. Chen. 2012. Lack of EGFR mutations benefiting gefitinib treatment in adenocarcinoma of esophagogastric junction. **World J. Surg. Oncol.** 10: 14.

Wang, X., P. Pan, Y. Li, D. Li and T. Hou. 2014. Exploring the prominent performance of CX-4945 derivatives as protein kinase CK2 inhibitors by a combined computational study. **Mol. BioSyst.**:

Wang, Z., L.L. Raines, R.M. Hooy, H. Roberson, D.J. Leahy and P.A. Cole. 2013. Tyrosine phosphorylation of mig6 reduces its inhibition of the epidermal growth factor receptor. **ACS Chem. Biol.** 8(11): 2372-2376.

Weiser, J., P.S. Shenkin and W.C. Still. 1999. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). **J. Comput. Chem.** 20(2): 217-230.

Wheeler, D.L., E.F. Dunn and P.M. Harari. 2010. Understanding resistance to EGFR inhibitors-impact on future treatment strategies. **Nat. Rev. Clin. Oncol.** 7(9): 493-507.

Wilkins, M.R., E. Gasteiger, A. Bairoch, J.C. Sanchez, K.L. Williams, R.D. Appel and D.F. Hochstrasser. 1999. Protein identification and analysis tools in the ExPASy server. **Methods Mol. Biol.** 112: 531-552.

Wong, E.T., D. Na and J. Gsponer. 2013. On the importance of polar interactions for complexes containing intrinsically disordered proteins. **PLoS Comput. Biol.** 9(8): e1003192.

Wu, J.Y., S.G. Wu, C.H. Yang, C.H. Gow, Y.L. Chang, C.J. Yu, J.Y. Shih and P.C. Yang. 2008. Lung cancer with epidermal growth factor receptor exon 20 mutations is associated with poor gefitinib treatment response. **Clin. Cancer Res.** 14(15): 4877-4882.

Wu, W.-J. and D.P. Raleigh. 1998. Local control of peptide conformation: Stabilization of cis proline peptide bonds by aromatic proline interactions. **Biopolymers** 45(5): 381-394.

Wymann, M.P. and L. Pirola. 1998. Structure and function of phosphoinositide 3-kinases. **Biochim. Biophys. Acta.** 1436(1-2): 127-150.

Xu, D., A. Makkinje and J.M. Kyriakis. 2005. Gene 33 is an endogenous inhibitor of epidermal growth factor (EGF) receptor signaling and mediates dexamethasone-induced suppression of EGF function. **J. Biol. Chem.** 280(4): 2924-2933.

Xu, L., H. Sun, Y. Li, J. Wang and T. Hou. 2013. Assessing the Performance of MM/PBSA and MM/GBSA Methods. 3. The Impact of Force Fields and Ligand Charge Models. **J. Phys. Chem. B.** 117(28): 8408-8421.

Xue, W., M. Wang, X. Jin, H. Liu and X. Yao. 2012. Understanding the structural and energetic basis of inhibitor and substrate bound to the full-length NS3/4A: insights from molecular dynamics simulation, binding free energy calculation and network analysis. **Mol. BioSyst.** 8(10): 2753-2765.

Yamasaki, S., T. Terada, K. Shimizu, H. Kono and A. Sarai. 2009. A generalized conformational energy function of DNA derived from molecular dynamics simulations. **Nucleic Acids Res.** 37(20): e135.

Yan, M. and D.J. Templeton. 1994. Identification of 2 serine residues of MEK-1 that are differentially phosphorylated during activation by raf and MEK kinase. **J. Biol. Chem.** 269(29): 19067-19073.

Yan, Z. and J. Wang. 2013. Optimizing Scoring Function of Protein-Nucleic Acid Interactions with Both Affinity and Specificity. **PLoS ONE** 8(9): e74443.

Yang, X.-Q., J.-Y. Liu, X.-C. Li, M.-H. Chen and Y.-L. Zhang. 2014. A key amino acid associated with acephate detoxification by Cydia pomonella carboxylesterase based on molecular dynamics with alanine scanning and site-directed mutagenesis. **J. Chem. Inf. Model.**:

Yang, Y., H. Liu and X. Yao. 2012. Understanding the molecular basis of MK2-p38alpha signaling complex assembly: insights into protein-protein interaction by molecular dynamics and free energy studies. **Mol. BioSyst.** 8(8): 2106-2118.

Yarden, Y. 2001. The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities. **Eur. J. Cancer.** 37 Suppl 4: S3-8.

Ying, H., H. Zheng, K. Scott, R. Wiedemeyer, H. Yan, C. Lim, J. Huang, S. Dhakal, E. Ivanova, Y. Xiao, H. Zhang, J. Hu, J.M. Stommel, M.A. Lee, A.J. Chen, J.H. Paik, O. Segatto, C. Brennan, L.A. Elferink, Y.A. Wang, L. Chin and R.A. DePinho. 2010. Mig-6 controls EGFR trafficking and suppresses gliomagenesis. **Proc. Natl. Acad. Sci. U S A.** 107(15): 6912-6917.

Yokota, A., K. Tsumoto, M. Shiroishi, H. Kondo and I. Kumagai. 2003. The role of hydrogen bonding via interfacial water molecules in antigen-antibody complexation. The HyHEL-10-HEL interaction. **J. Biol. Chem.** 278(7): 5410-5418.

Yoshikawa, S., M. Kukimoto-Niino, L. Parker, N. Handa, T. Terada, T. Fujimoto, Y. Terazawa, M. Wakiyama, M. Sato, S. Sano, T. Kobayashi, T. Tanaka, L. Chen, Z.J. Liu, B.C. Wang, M. Shirouzu, S. Kawa, K. Semba, T. Yamamoto and S. Yokoyama. 2013. Structural basis for the altered drug sensitivities of non-small cell lung cancer-associated mutants of human epidermal growth factor receptor. **Oncogene** 32(1): 27-38.

Yoshimura, A., T. Naka and M. Kubo. 2007. SOCS proteins, cytokine signalling and immune regulation. **Nat. Rev. Immunol.** 7(6): 454-465.

Yun, C.H., T.J. Boggon, Y. Li, M.S. Woo, H. Greulich, M. Meyerson and M.J. Eck. 2007. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. **Cancer Cell** 11(3): 217-227.

Yun, C.H., K.E. Mengwasser, A.V. Toms, M.S. Woo, H. Greulich, K.K. Wong, M. Meyerson and M.J. Eck. 2008. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. **Proc. Natl. Acad. Sci. U S A.** 105(6): 2070-2075.

Zhang, X., J. Gureasko, K. Shen, P.A. Cole and J. Kuriyan. 2006. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. **Cell** 125(6): 1137-1149.

Zhang, Y.W. and G.F. Vande Woude. 2007. Mig-6, signal transduction, stress response and cancer. **Cell Cycle** 6(5): 507-513.

Zhang, Y.W. and G.F. Vande Woude. 2013. MIG-6 and SPRY2 in the Regulation of Receptor Tyrosine Kinase Signaling: Balancing Act via Negative Feedback Loops, pp 199-121. **Future Aspects of Tumor Suppressor Gene**. In Tech, Croatia.

Zhang, Z., A.L. Stiegler, T.J. Boggon, S. Kobayashi and B. Halmos. 2010. EGFR-mutated lung cancer: a paradigm of molecular oncology. **Oncotarget** 1(7): 497-514.

Zhao, J.J. and T.M. Roberts. 2006. PI3 kinases in cancer: from oncogene artifact to leading cancer target. **Sci. STKE.** 2006(365): pe52.

Zhong, H. and H.A. Carlson. 2005. Computational studies and peptidomimetic design for the human p53-MDM2 complex. **Proteins** 58(1): 222-234.

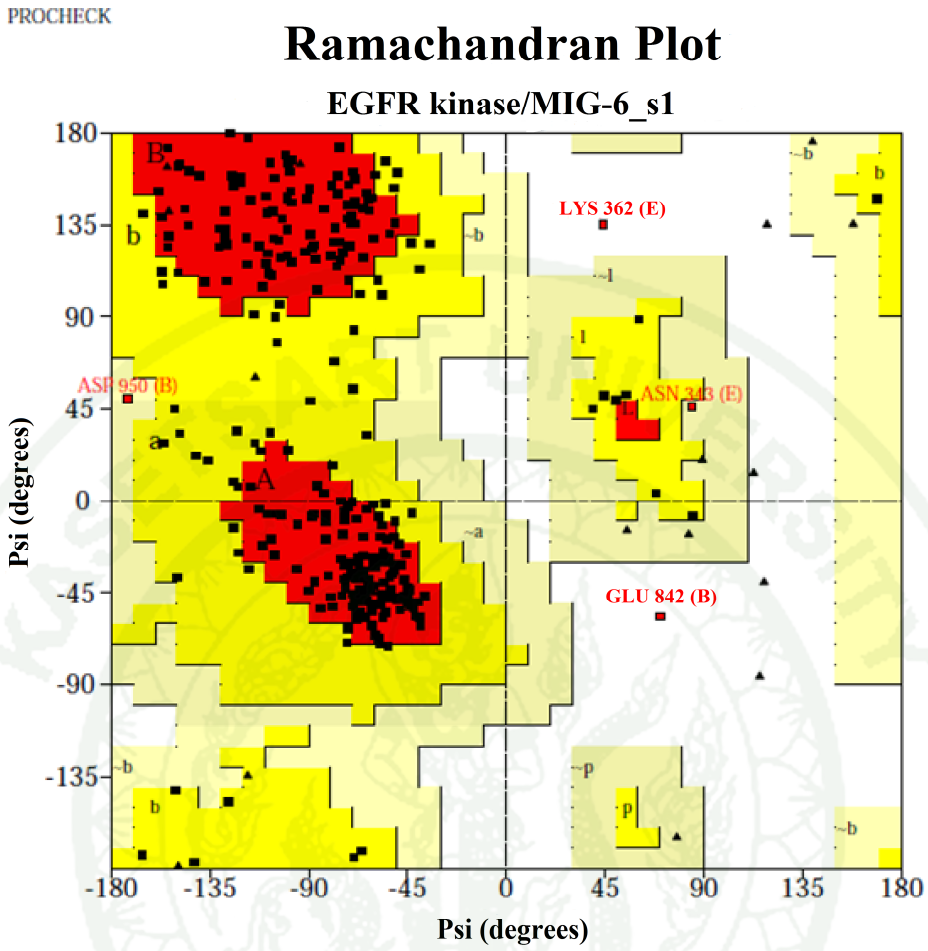
Zimmerman, D.W. 2012. Correcting Two-Sample z and t Tests for Correlation: An Alternative to One-Sample Tests on Difference Scores. **Psicológica** 33(2):

Zuo, Z., N.S. Gandhi, K.M. Arndt and R.L. Mancera. 2012. Free energy calculations of the interactions of c-Jun-based synthetic peptides with the c-Fos protein. **Biopolymers** 97(11): 899-909.

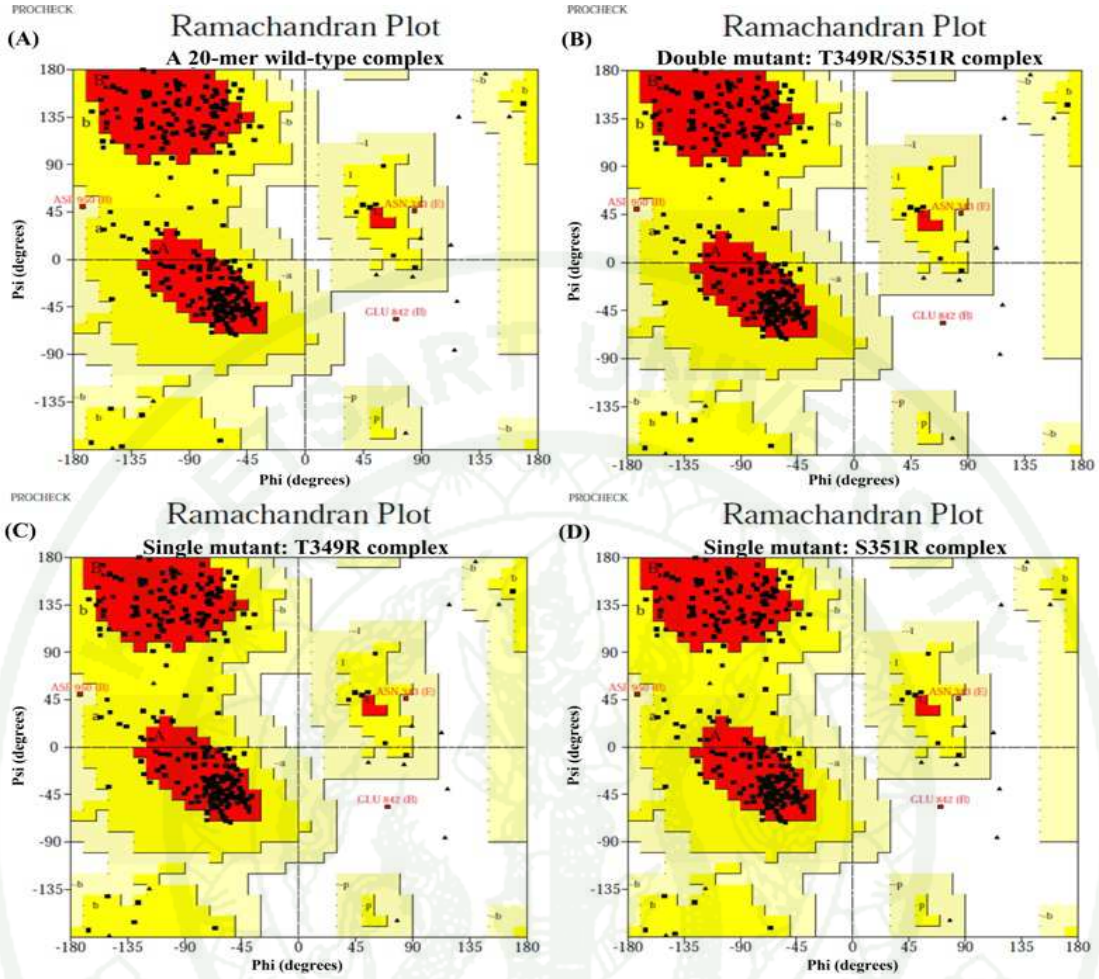
APPENDICES

Appendix A

Ramachandran plots of wild-type and mutant structures produced by PROCHECK



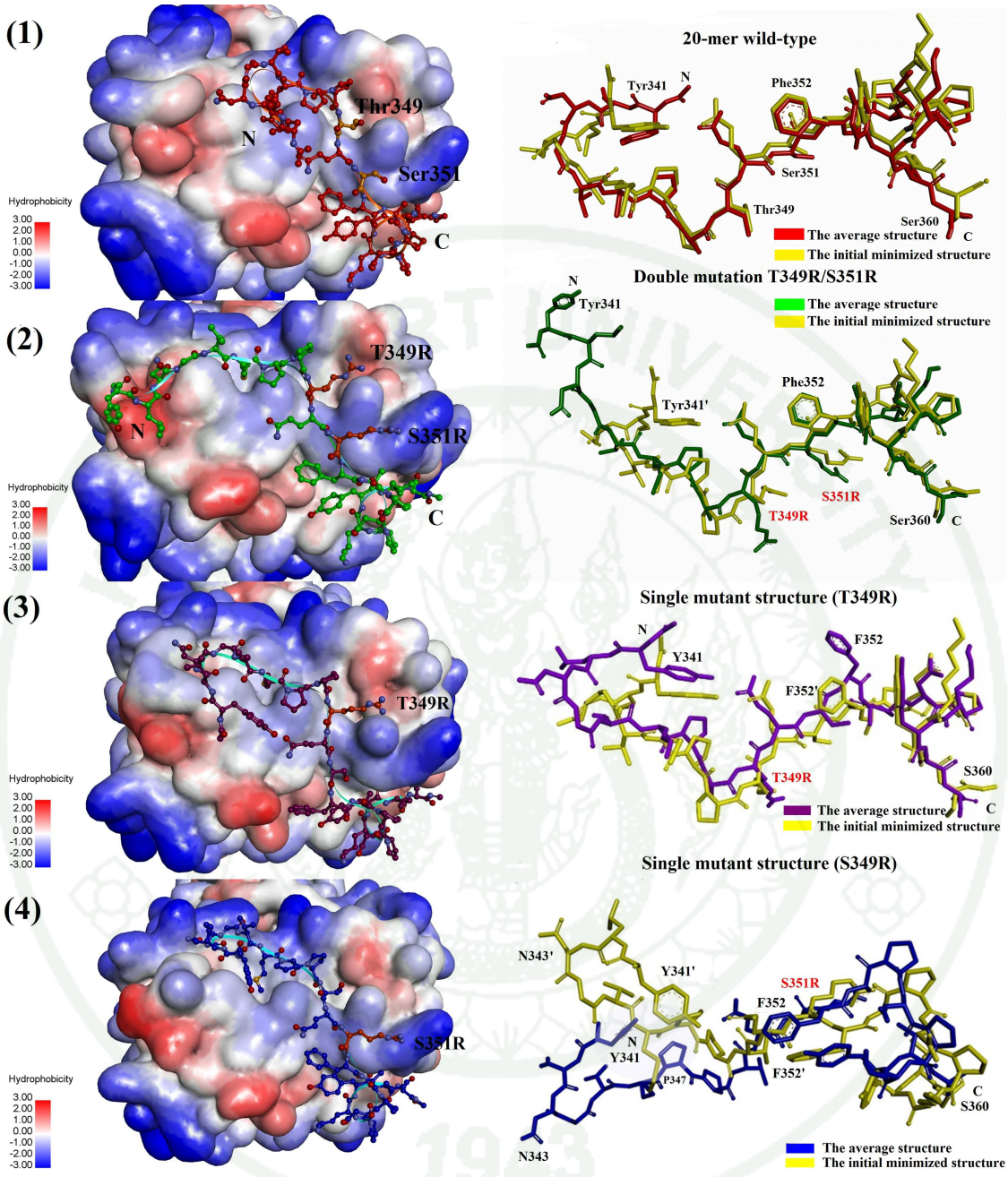
Appendix Figure A1 Ramachandran plots of a 27 mer wild-type of MIG-6_s1. The color areas of red, yellow, cream and white represented the most favored, allowed regions, generously allowed and disallowed regions, respectively.



Appendix Figure A2 Ramachandran plots of a 20-mer wild-type of MIG-6_s1 (A), double mutant T349R/S351 (B) and two single mutants T349R (C), S351R (D) models. The color areas of red, yellow, cream and white represented the most favored, allowed regions, generously allowed and disallowed regions, respectively.

Appendix B

Superimposition between the MD-simulated structures and the initial-minimized structures



Appendix Figure B1 The MD-simulated structures of a 20-mer wild-type of MIG-6_s1 and mutant vatiants. Each of the average structure was extracted from the last 3 ns of a 16-ns MD simulation and superimposed with their initial structures. A 20-mer wild-type (1), double mutation T349R/S351R (2) and two single mutation structures of T349R (3) and S351R (4) were presented.

Appendix C

Z-scores calculation

Z-scores calculation

Two large samples (>30) are normally using Z-scores to observe a significant difference between two groups (Rubin, 2012).

$$Z = \frac{\text{observed difference} - \text{expected difference}}{\text{SE for difference}} \quad ; \quad Z = \frac{(\bar{X_1} - \bar{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where $\bar{X}$ is the mean score on two samples, μ is the hypothesized difference between population mean (0 if expecting for equal means), n is number of two samples and σ is standard deviations of two samples.

The research hypothesis of experiment

a. The null hypothesis (H_0): μ_1-μ_2=0: There is no difference between two groups

b. The alternative hypothesis (H_1): μ_1-$\mu_2 \neq 0$: There is a difference between two groups

p-value

The *p* is the significance level related to the strong between these relationships. It s normally used 0.01, 0.5 and 0.10 these refer to the strong in the relationship at the confidence intervals are 99, 95 and 90 %, respectively.

Interpretation of the findings

If the Z-score is within the distribution area, it can be interpreted to accept of the null hypothesis of no statistically siginificant differences between two groups. If the Z-score is excessed the two-tailed distribution area, it can be interpreted to reject of the null hypothesis and accept of alternative hypothesis of statistically siginificant differences between two groups (Zimmerman, 2012).

CURRICULUM VITAE

NAME : Ms. Ninnutt Moonrin

BIRTH DATE : October 7, 1984

BIRTH PLACE : Chiangmai, Thailand

EMAIL ADDRESS : Moonrin_jear@hotmail.com

EDUCATIONS :

YEAR	INSTITUTE	DEGREE
2004	Chiangmai University	B.S. (Biotechnology)
2014	Kasetsart University	M.S. (Biochemistry)

SCHOLARSHIPS

1. Graduate Scholarships from Thailand Graduate Institute of Science and Technology (TGIST: TG-๒๒-๑๑-๕๕-๐๒๔M) of the National Science and Technology Development Agency (NSTDA), Thailand (2012-2014)

PROCEEDINGS

1. Moonrin, N., W. Mokmak, S. Tongsima and K. Choowongkomon, 2-3 April 2014. The Computational Studies of the Hot Spot Residues of Mig-6 to EGFR kinase interface Based on Molecular Dynamics Simulation, in The 4 th Biochemistry and Molecular Biology (BMB) Conference. Science Society of Thailand, Kasetsart University, Bangkok, Thailand.

RESEARCH TRAINING

1. Workshop on Computer-aided Molecular Design 9-10 May 2013, National e-science infrastructure consortium, Bangkok, Thailand