Role of glycosylation in EGFR ectodomain interactions : a molecular dynamics study

Azimzadeh Irani, Maryam

2017

Azimzadeh Irani, M. (2017). Role of glycosylation in EGFR ectodomain interactions : a molecular dynamics study. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/71976>

<https://doi.org/10.32657/10356/71976>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

**ROLE OF GLYCOSYLATION IN EGFR ECTODOMAIN
INTERACTIONS; A MOLECULAR DYNAMICS STUDY**

MARYAM AZIMZADEH IRANI

SCHOOL OF BIOLOGICAL SCIENCES

1/1/2017

**ROLE OF GLYCOSYLATION IN EGFR ECTODOMAIN
INTERACTIONS; A MOLECULAR DYNAMICS STUDY**

MARYAM AZIMZADEH IRANI

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2017

Acknowledgements

There are many people that have been involved in my PhD and I would like to express my sincere gratitude to them. Top most of them are:

My supervisor Dr. Chandra Verma for his encouragement, guidance and patience over the past four years. I feel very lucky to know him and to work under his mentorship. His impact on my insights is fundamental and I am constantly inspired by his integrity, abilities to work in collaboration and approaching scientific questions without any default answer.

My parents for always believing in me. Their belief inspires me to make my best effort every day. There are no words that describe the love I feel for them. I could not be more loved and supported in my entire life. They have stood by me in all of life's obstacles however; the most precious gift I received from my parents is that they never discriminated me as a girl compare to my brothers. The equality I experienced at home empowered me to dream big and think freely in life.

My brothers, Dr. Maziar Azimzadeh Irani and Dr. Arash Azimzadeh Irani for encouraging me to look at the bright side of life. Being the younger sister results in being continuously teased by them and prevents any unnecessary ego boost. The hard work they put into this is responded by my deep affection, love and respect for them both.

All other family members and friends who have encouraged me through hard times and celebrated with me in happy times. To mention them all by name would be impossible but their companionship is most appreciated and for those I will not have the chance to see often, they will be missed dearly.

The majority of my time in the last four years was spent in the Bioinformatics Institute, A*STAR and Nanyang Technological University of Singapore. Where I have learned the basics of research and scientific discussions and I came across several nice individuals. Among these I would like to especially acknowledge:

All members of Verma group in the Bioinformatics Institute, A*STAR, Singapore, Dr. Srinivasaraghavan Kannan for insightful discussions, Dr. Stephen John Fox for critical reading of chapter 5 of this thesis, for his help in the binding free energy calculations in chapter 3 and for several helpful advises, Biruhalem Taye Beyene for his full support as my friend and colleague. I was always certain that he will help me in any needed occasion. I could not thank him enough with words however; I try to continue his tradition by being an amazing friend to someone in need and Dr. Roland Huber for his suggestions in the alignment of the thesis.

Admin and IT team members of the Bioinformatics Institute, A*STAR, Singapore especially: YEE Guo Wei Robin for his continuous help with the technical and IT issues I encountered. Many of the analyses used in this thesis were not possible to complete at this speed and with this quality without Robin's help. YONG Tai Pang for his continuous effort in providing computational resources for my time consuming simulations. Here I wish to thank my supervisor again for directing all of us (me, Robin and Tai Pang) in this path of optimizing: the simulations time scale, storage issues and analysis methods.

There are others that had a less direct input into my scientific journey however I value their support:

My professors and admin team members of postgraduate studies in the School of Biological Sciences Nanyang Technological University.

Singapore International Graduate Award for funding my research.

Fellows of the Graduate Student Council Nanyang Technological University for engaging me in several adventures.

Members of my Thesis Advisory Committee; Associate Professors Mu Yuguang and Newman Sze from the School of Biological Sciences Nanyang Technological University and Dr. Robert Robinson from the Institute of Molecular and Cell Biology, A*STAR, Singapore for helpful suggestions.

Table of Contents

Acknowledgements	iii
List of figures	x
List of tables	xiii
Abbreviations	xiv
Abstract	xvi
Chapter 1. Epidermal Growth Factor Receptor	1
1.1 Introduction	1
1.2 ErbB proteins family	1
1.3 EGFR structure	4
1.3.1 EGFR extracellular domain	4
1.3.2 Extracellular domain architecture	4
1.4 EGFR ligands	6
1.4.1 Growth factors	7
1.4.2 EGF-EGFR interactions	7
1.4.3 Monoclonal antibodies	11
1.4.4 Dimeric partner	14
1.5 EGFR activation	16
1.5.1 Ligand induced dimerization	16
1.5.2 Preformed dimer	17
1.6 EGFR glycosylation	19
1.6.1 Protein glycosylation	19
1.6.2 Challenges in glycoprotein analysis and role of computer simulations	22
1.6.3 EGFR glycosylation	23
1.6.4 Functions of EGFR with attached glycans	24
1.7 Summary	25
Chapter 2. Computational methods	28
2.1 Introduction	28
2.2 Molecular mechanics force field	28
2.2.1 Bonded interactions	29
2.2.2 Non-bonded interactions	31

2.2.5 The GLYCAM force field	33
2.3 Solvent model.....	34
2.3.1 Explicit solvent model	35
2.3.2 Implicit solvent model	36
2.4 Energy minimization	38
2.4.1 Derivative minimization method	38
2.5 Molecular Dynamics	40
2.5.1 Periodic Boundary Conditions.....	41
2.5.2 Particle Mesh Ewald.....	42
2.5.3 Bond constraints	43
2.5.4 Different ensembles in molecular dynamics	44
2.6 accelerated Molecular Dynamics (aMD)	47
2.7 Molecular Dynamics simulations of the EGFR extracellular domain	48
2.7.1 Crystallographic structures	48
2.7.2 Capping the protein	48
2.7.3 Protonation states of the Histidines	50
2.7.4 Disulfide bonds.....	51
2.7.5 Construction of glycosylated EGFR models	52
2.7.6 Neutralizing and solvating the system.....	53
2.7.7 Molecular Dynamics simulation protocol	54
Chapter 3. Role of N-glycosylation in EGFR ectodomain ligand binding	57
3.1 Introduction	57
3.2 Methods	60
3.2.1 Model building of glycosylated and unglycosylated EGFR	60
3.2.2 Molecular Dynamics simulations and analysis methods.....	60
3.3 Results	62
3.3.1 Dynamics of the glycosylated monomer of EGFR extracellular domain	62
3.3.2 N-glycosylation increases EGFR-EGF binding affinity	72
3.3.3 N-glycosylation contribution to EGFR-antibody interactions	88
3.3.4 N-glycosylation maintains the dimeric interfaces of EGFR-ed	94
3.4 Discussion	97

3.5 Conclusions.....	101
Chapter 4. Glycosylation modulates the tethered-extended equilibrium of the EGFR extracellular domain.....	102
4.1 Introduction	102
4.2 Materials and methods	104
4.2.1 Model building of glycosylated and unglycosylated monomeric EGFR in the tethered and untethered states.....	104
4.2.2 Accelerated Molecular Dynamics simulations	105
4.2.3 Principal Component Analysis.....	106
4.3 Results	107
4.3.1 Dynamics of glycosylated EGFR in the untethered state	107
4.3.2 Dynamics of glycosylated EGFR in the tethered state.....	118
4.3.3 Alteration of the tethering interactions upon glycosylation	129
4.4 Discussion	138
4.5 Conclusions.....	142
Chapter 5. Elucidating EGFR activation: interplay between glycosylation and growth factor binding	143
5.1. Introduction	143
5.2. Materials and methods	144
5.2.1 Model building of glycosylated and unglycosylated EGFR dimers.....	144
5.2.2. Molecular Dynamics simulations	145
5.2.3. Intradimeric twist angle and Intramonomeric torsion angle calculation	146
5.3. Results	148
5.3.1 Dynamics of dimeric EGFR upon glycosylation.....	148
5.3.2 Glycosylation Increases the intradimeric twist motion and intramonomeric turn motion.....	156
5.3.3 Glycosylated EGFR dimer could adopt the flush conformation in the apo and	160
EGF-bound states	160
5.4 Discussion	167
5.5 Conclusions	169
Chapter 6. Conclusions and future directions	170

Appendices.....	176
Appendix A Contribution of subdomains I-IV to EGF, monoclonal antibodies and the dimeric partner binding to EGFR.....	176
Appendix B Global dynamics of tethered monomeric form of EGFR-ed upon glycosylation	219
Appendix C Dynamics of Module1-8 of subdomain II in the dimeric state	234
References.....	240
List of publications	247

List of figures

Figure 1.1 Structures of the ErbB receptors	3
Figure 1.2 EGFR extracellular domain consists of four subdomains	6
Figure 1.3 Subdomain I of the EGFR-EGF interaction site	9
Figure 1.4 Subdomain III of the EGFR-EGF interaction site	10
Figure 1.5. Hydrogen bonds between Subdomains I and III of EGFR and EGF	10
Figure 1.6. Binding epitopes of monoclonal antibodies bound to EGFR extracellular domain	13
Figure 1.7. Dimerization arm dimeric interface of EGFR extracellular domain	15
Figure 1.8. Tip of subdomain IV dimeric interface of EGFR extracellular domain	15
Figure 1.9. Ligand induced dimerization of EGFR	17
Figure 1.10. Flexible preformed EGFR dimer is stabilized by EGF binding	18
Figure 1.11. N- and O- linked glycans structures	21
Figure 1.12. Summary of consensus N-glycosylation sites of EGFR extracellular domain	24
Figure 1.13. Outline of the thesis chapters	27
Figure 2.1. The Lennard-Jones potential	32
Figure 2.2. Presentation of explicit and implicit solvent models	35
Figure 2.3. TIP3P water model	36
Figure 2.4. Periodic boundary condition	42
Figure 2.5. Sampling the phase space in MD simulations	46
Figure 2.6. Protonation states of Histidine	50
Figure 2.7. Construction of glycosylated EGFR model	53
Figure 3.1. The backbone RMSD plots	65
Figure 3.2. Binding of three groups of ligands to the EGFR-ed	66
Figure 3.3. EGFR-ed subdomain orientations	67
Figure 3.4. Increased volume of the cavity between subdomains I and III	68
Figure 3.5. Distance distribution between each oligosaccharide and its respective subdomain	69
Figure 3.6. Interaction between the two glycans attached to subdomain I and III	70
Figure 3.7. Interaction between the two glycans attached to subdomain I and III-apo	71
Figure 3.8. Average SASA of the EGF binding site from simulations	72
Figure 3.9. The binding free energy of EGF to EGFR-ed	78

Figure 3.10. Distance distribution plots between the last monosaccharides	80
Figure 3.11. Subdomain I per-residue contribution to the binding free energy	80
Figure 3.12. Subdomain I – the most favorably contributing residues	81
Figure 3.13. Subdomain III per-residue contribution to the binding free energy	82
Figure 3.14. Subdomain III - most favorably contributing residues	83
Figure 3.15. EGF per-residue contribution to the binding free energy	84
Figure 3.16. EGF - most favorably contributing residues	85
Figure 3.17. Sequence alignment of the ErbB family members	86
Figure 3.18. Sequence alignment of the ErbB family members	86
Figure 3.19. Structural alignment of the ErbB family members	87
Figure 3.20. Binding epitopes of monoclonal antibodies bound to EGFR	92
Figure 3.21. Conformational changes of mAb806 binding epitope	93
Figure 3.22. Effect of N-glycosylation on EGFR-ed dimerization	96
Figure 4.1. The C α RMSD plots	108
Figure 4.2. The C α RMSD plots of subdomains I-IV	109
Figure 4.3. The C α RMSF plots	110
Figure 4.4. Percentage of variance	112
Figure 4.5. Projection of the untethered structures along PC1 and PC2	113
Figure 4.6. Average RMSD values	114
Figure 4.7. Visualization of the untethered EGFR-ed dynamics	115
Figure 4.8. Visualization of the untethered gEGFR-ed dynamics	116
Figure 4.9. Projection of the tethered monomer trajectories	117
Figure 4.10. The C α RMSD plots	119
Figure 4.11. The C α RMSD plots of subdomains I-IV	120
Figure 4.12. The C α RMSF plots	121
Figure 4.13. Percentage of variance	122
Figure 4.14. Projection of the tethered structures along PC1 and PC2	124
Figure 4.15. Average RMSD values	125
Figure 4.16. Visualization of the tethered EGFR-ed dynamics	126
Figure 4.17. Visualization of the tethered gEGFR-ed dynamics	127
Figure 4.18. Projection of the untethered monomer trajectories	128
Figure 4.19. Cross-correlation motions of the tethered EGFR-ed	133

Figure 4.20. Cross-correlation motions of the tethered gEGFR-ed	134
Figure 4.21. The distance between center of mass of subdomain IV-dimerization arm	135
Figure 4.22. EGFR-ed tethered by interaction of dimerization arm and two inhibitory loops	136
Figure 4.23. Distance distributions between the interacting pairs involved in tethering	137
Figure 5.1. Intradimeric twist angle and intramonomeric torsion angle	147
Figure 5.2. Average RMSD plots of the first simulation	150
Figure 5.3. Average RMSD plots of the second simulation	150
Figure 5.4. Average RMSD plots of the third simulation	151
Figure 5.5. Average RMSD plots of subdomains I-IV of the first simulation	152
Figure 5.6. Average RMSD plots of subdomains I-IV of the second simulation	153
Figure 5.7. Average RMSD plots of subdomains I-IV of the third simulation	154
Figure 5.8. Dynamics of glycosylated EGFR-ed dimers	155
Figure 5.9. Intradimeric twist angle distribution plot	157
Figure 5.10. Cross-correlation motions of the EGF-EGFR-ed dimers	158
Figure 5.11. Cross-correlation motions of the EGFR-ed dimers	159
Figure 5.12. Distribution plots of intermonomer1 (A) and intermonomer2 (B) torsion angles	160
Figure 5.13. Staggered and flush dimeric interfaces of EGFE-ed	162
Figure 5.14. Intradimeric distance distribution plot	164
Figure 5.15. Interaction between the glycans	165
Figure 5.16. Distribution plots of intermonomer1 (A) and intermonomer2 (B) torsion angles	166

List of tables

Table 1.1. Description of available structural data of monoclonal antibodies bound to EGFR	12
Table 2.1. ErbB receptors PDB entries list	49
Table 2.2. Bonded Cysteine residue numbers of EGFR and EGF	51
Table 2.3. Description of MD simulations performed on the extended monomer of EGFR-ed	55
Table 2.4. Description of MD simulations performed on the tethered monomer of EGFR-ed	56
Table 2.5. Description of MD simulations performed on the extended dimer of EGFR-ed	56
Table 3.1. The binding free energy of EGF to EGFR-ed	79

Abbreviations

μVT	Grand canonical ensemble
ADAM	A Disintegrin And Metal protease
aMD	accelerated Molecular Dynamics
AR	Amphiregulin
BSA	Buried Surface Area
BTC	Betacellulin
CA	Carbon Alpha
cDNA	complementary Deoxyribonucleic Acid
CE	Capillary Electrophoresis
CR	Cysteine-Rich
C-terminal	carboxy-terminal
ED	Extracellular Domain
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EGFR-ed	Epidermal Growth Factor Receptor extracellular domain
Ep	Energy potential
EPG	Epigen
EPR	Epiregulin
ER	Endoplasmic Reticulum
Erb	Erythroblastic
ERK	Extracellular signal-Regulated Kinases
GalNAc	N-acetylglucosamine
GBSA	Generalized Born Surface Area
gEGFR-ed	glycosylated Epidermal Growth Factor Receptor extracellular domain
HB-EGF	Heparin Binding Epidermal Growth Factor
Her	Human epidermal growth factor receptor
JMD	Juxtamembrane Domain
kDa	kilo Dalton
LR	Leucine-Rich
mAb	monoclonal Antibody

Man	Mannose
MD	Molecular Dynamics
N-linked	Nitrogen linked
NMR	Nuclear Magnetic Resonance
NPT	Isobaric-Isothermal ensemble
Nrg	Neuregulin
N-terminal	Amino-terminal
NVE	Microcanonical ensemble
NVT	Canonical ensemble
O-linked	Oxygen linked
OTase	Oligosaccharyl Transferase
PBC	Periodic Boundary Conditions
PCA	Principal Component Analysis
PI3K	Phosphatidyl Inositol3 Kinase
PLCγ	Phospholipase C
PME	Particle Mesh Ewald
RMSD	Root Mean Squared Deviation
RMSF	Root Mean Squared Fluctuation
SASA	Solvent Accessible Surface Area
TGF	Transforming Growth Factor
TKD	Tyrosine Kinase Domain
TMD	Transmembrane Domain
vDW	van der Waals
MMGBSA	Molecular Mechanics Generalized Born Surface Area

Abstract

The Epidermal Growth Factor Receptor (EGFR) is a tyrosine kinase protein, overexpressed in several cancers. The extracellular domain of EGFR is heavily glycosylated with sugar moieties of varying compositions. Growth factor (mostly Epidermal Growth Factor or EGF) binding activates EGFR. This occurs by inducing the transition from the autoinhibited tethered conformation to an extended conformation of the monomeric form of EGFR and by stabilizing the flexible pre-formed dimer. Several antibodies inhibit EGFR by targeting the growth factor binding site or the dimeric interfaces. Activated EGFR adopts a back-to-back dimeric conformation after binding of another homologous receptor to its extracellular domain as the dimeric partner. Atomistic MD simulations show that glycosylation of the EGFR extracellular domain play critical roles in the binding of growth factors, monoclonal antibodies and the dimeric partners to the monomeric EGFR extracellular domain. N-glycosylation favors growth factor binding to EGFR by a combinatorial mechanism of hydrogen bonding, non-bonded interactions and several intramolecular contacts. Glycosylation also plays distinct roles in binding of antibodies to spatially separated epitopes of the EGFR extracellular domain and assist in maintaining the dimeric interfaces. In the activation process of EGFR, N-glycosylation reduces the flexibility of the tethered and extended monomeric forms of EGFR extracellular domain. By supporting the glycosylated form of the extended state and weakening the tethering interactions in the tethered state, glycosylation supports the extended conformation and facilitates the tethered-extended transition. Dimerization of the EGFR extracellular domain is regulated by attached N-glycans which decrease the flexibility of the dimeric conformation. The decrease in flexibility results in increases in

intramonomeric turn motions and intradimeric twist motions which likely induces the asymmetric arrangements of the transmembrane and subsequently the intracellular kinase domains which triggers the signaling cascade in the cytosol.

Chapter 1. Epidermal Growth Factor Receptor

1.1 Introduction

Abnormalities in expression, structure or activity of proto-oncogenes contribute to the production of malignant phenotypes [1]. One such proto-oncogene is the Epidermal Growth Factor Receptor (EGFR) whose mutation or overexpression is associated with several cancers [1, 2]. Due to its significant role in detection and treatment of human cancer, understanding EGFR biology, function and interactions has been one of the most fascinating areas of interest in cancer therapy over the past few decades.

1.2 ErbB proteins family

EGFR is a key member of the ErbB/HER glycoprotein family. ErbB stands for avian erythroblastic leukemia viral oncogene to which these receptors are homologous and HER refers to the human EGFR receptor [3]. Human ErbB genes are found on four different chromosomes; absence of any of the ErbB genes in mice results in embryonic or prenatal lethality [4-6]. The ErbB family has four identified homologous members including EGFR (Her1, ErbB1), ErbB2 (Her2), ErbB3 (Her3) and ErbB4 (Her4). ErbB family members bind to different growth factors on the cell surface and trigger downstream signaling pathways mediated by their tyrosine kinase activity (Figure 1.1). Among the family members, ErbB2 lacks ligand (growth factor) binding ability, and ErbB3 lacks tyrosine kinase activity [1]. ErbB receptors are known to form homo/hetero dimers and higher order multimers [7]. Hence, ErbB2 and ErbB3 can make heterodimers with other family members and participate in cell signalling. Protein tyrosine kinase receptors play

key roles in multiple cellular processes such as proliferation, differentiation and migration. The ErbB/HER downstream signaling pathways include the three major pathways: the Phosphatidylinositol 3 Kinase (PI3K)/ Akt(PKB) pathway, the PhospholipaseC (PLC γ) pathway and the Ras/Raf/ MEK/ERK1, ERK2 pathway [3]. The PI3K/Akt pathway engages in cell survival while the Ras/ERK1, ERK2 and PLC γ pathways are involved in cell proliferation. Together the ErbB pathway is implicated in angiogenesis, cell adhesion, motility, development and organogenesis [8, 9].

Figure 1.1 Structures of the ErbB receptors. The extracellular domain of the ErbB receptors consists of four subdomains. These subdomains are shown with blue, green, yellow and magenta squares. Subdomains I and III of the extracellular domain are the growth factor binding subdomains. These are presented with the stop sign in ErbB2 as this receptor is not known to bind to any ligand. Transmembrane and juxtamembrane domains are shown with brown and dark blue squares. The intracellular kinase domain and the tyrosine phosphorylation C-terminal tail are shown in red circles and lines respectively. Stop sign was used to represent the kinase domain of ErbB3 as this receptor lacks tyrosine kinase activity. The numbers depict the amino acid residues of EGFR. The arrow boxes present specific growth factors which bind to each receptor. The boxes show overall sequence homology of ErbB2, ErbB3 and ErbB4 to EGFR. ED: Extracellular domain, TMD: Transmembrane domain, JMD: Juxtamembrane domain, TKD: Tyrosine kinase domain. EGF: Epidermal growth factor, EPG: Epigen, TGF: Transforming growth factor, AR: Amphiregulin, BTC: Betacellulin, HB-EGF: heparin binding epidermal growth factor, EPR: Epiregulin, Nrg: Neuregulin.

1.3 EGFR structure

cDNA analysis and the primary amino acid structure of EGFR suggests that the receptor consists of an extracellular domain of 640-650 amino acid residues, a single hydrophobic transmembrane alpha helix that is 19-25 amino acid residues in length and an intracellular domain of around 550 amino acid residues that are made up of a juxtamembrane segment, a tyrosine kinase subdomain and a C-terminal tyrosine phosphorylation tail [10] (Figure 1.1).

1.3.1 EGFR extracellular domain

The extracellular domain, based on amino acid homology, is divided into four subdomains [1]. subdomains I and III are Leucine-rich segments and participate in growth factor binding, and subdomains II and IV which contain multiple conserved Cysteine residues that participate in disulfide bond formation [3]. These two subdomains perform a role in maintaining the tertiary structure of the extracellular domain (Figure 1.2). EGFR extracellular domain adopts various conformations that appear to be associated with different biological functions. These conformational changes could range from large tethered (inactive) to extended (active) transitions upon growth factor binding to the extracellular domain [7], to modulate the dynamics of dimeric interfaces or exposure of cryptic antibody binding epitopes [11, 12].

1.3.2 Extracellular domain architecture

Leucine-rich subdomains (LR1/EDI and LR2/EDIII) are comprised of six β -helices. Each of these β -helices is capped with a disulfide bond and an α -helix at each end [2] (Figure 1.2). Both Leucine-rich subdomains contain a conserved stack of three Glycine residues.

This stack is located after the second β -sheet and is formed by Glycine residues 39, 63, 85 and 122 in LR1 and 343, 379, 404 and 435 in LR2. The fourth turn of the helix in each Leucine-rich subdomain forms a loop that protrudes from the second β -sheet (Figure 1.2) [2]. A long V-shaped excursion in subdomain LR1 (residues 8 to 18) sits near the large β -sheet and makes an important part of the growth factor binding site. This excursion is replaced by a loop in subdomain LR2. This loop is formed by residues 316-326 and interacts with the growth factor [2]. Another loop located in LR2 (residues 351-369) forms an epitope that elicits interactions with antibodies LA22, LA58 and LA90; these antibodies interfere with growth factor binding. These three loops together with a conserved loop in the fourth turn of the solenoid form important interactions with the growth factors [2]. Cysteine rich subdomains (CR1/EDII and CR2/EDIV) consist of 11 and 10 disulfide bridges respectively [13, 14]. The arrangement of 11 disulfide bonds in CR1 results in a straight rod bent at its C-terminal segment [2]. CR1 and LR1 subdomains make contact via one side of the solenoid. The C-terminal segment of CR1 also makes a contact with LR2, burying 860 Å² of the solvent accessible surface area [2]. The CR1 subdomain C-terminal segment acts as a hinge in the ligand free tethered EGFR. The most significant feature of CR1 is a large loop formed by residues 239-269 that projects away from the growth factor binding site. This loop contains an antiparallel β -helix ribbon. This loop is a major contributor to the receptor-receptor contacts in the dimeric form of EGFR and is mainly referred to as the dimerization arm [2]. CR2 extends from the base of LR2 and curves towards the opposite side of the growth factor binding site [14]. The N-terminal segment of CR2 is tucked against LR2. This segment includes Trp492 which plugs into LR2 and is conserved in the related receptors [14].

Figure 1.2 EGFR extracellular domain consists of four subdomains. Two Leucine-rich subdomains (LR1, LR2) participating in growth factor binding to EGFR and two Cysteine-rich subdomains (CR1, CR2) which maintain the tertiary structure of the extracellular domain. Helices are shown with light pink ribbons. β -strands of subdomains LR1 and LR2 are presented with blue broad arrows. CR1's dimerization arm is shown in orange. Disulfide-linked Cysteines are presented with red lines. Blue circles represent LR1 and LR2 and green circles show CR1 and CR2.

1.4 EGFR ligands

Binding of various ligands to EGFR extracellular domain is followed by either relaying signals to the cytosol and activation of the downstream cellular pathways or by suppressing signals. The most common ligands that bind to EGFR include growth factors,

monoclonal antibodies and the dimeric partners (ErbB family members involved in homo/hetero dimer formation). The binding mechanism of each group of ligands to EGFR extracellular domain plays critical roles in EGFR function and regulation.

1.4.1 Growth factors

EGFR is known to bind to seven different growth factors [3]. A list of the specific ligand-receptor interactions can be found in Figure 1.1. All the EGFR binding growth factors are expressed as single-pass integral membrane proteins [15]. The precursors of these ligands are made up of an extracellular region, a transmembrane region and a small intracellular region. The extracellular region of this precursor is released by proteolysis by ADAM (A Disintegrin And Metal protease) proteins. This proteolytic process is known as “ectodomain shedding” and involves cutting off the extracellular segment and releasing the modified protein [3]. EGF and the other members of the growth factor family (EPG, TGF- α , AR, BTC, HB-EGF and Nrg) contain six cysteine residues that form disulfide bonds. Neuregulins (Nrg) 1-3 are large peptides that contain an EGF-like domain [3]. EGFR binding growth factors are thought to act close to their sites of generation [16]. They may act on the cell of release (autocrine), on an adjacent cell (juxtacrine) or on a nearby cell (paracrine) [3]. Among all EGFR binding growth factors, EGF is the most common stimulator of EGFR and its interaction with subdomains I and III of EGFR extracellular domain is fully characterized in X-ray crystal structures [14, 17]. Details of this interaction are explained in the following section.

1.4.2 EGF-EGFR interactions

The interface between EGF and EGFR is extensive with total buried surface area of about 720 and 730 Å² for subdomains I and III respectively [17]. Two major interacting sites

have been suggested for EGF-EGFR (Figure 1.3-1.5). In site one, the side chains of Leu14, Tyr45, Leu69 and Leu98 in subdomain I of EGFR extracellular domain engage in hydrophobic interactions with Met21, Ile23 and Leu26 in EGF [17] (Figure 1.3). The involvement of these hydrophobic residues of EGF in receptor binding is consistent with site-directed mutagenesis studies on EGF [17]. Replacement of Ile23 with Ala, Val, Leu, Phe, Trp and several other amino acids have shown that Ile23 of EGF is crucial for tight binding with the receptor [17]. The interface with Ile23 is formed from contributions of the side chains of Leu14, Tyr45 and Leu69 of EGFR. In site two, Val350 and Phe357 side chains in subdomain III of EGFR extracellular domain make hydrophobic interactions with Leu15 and Tyr13 of EGF [18]. This is in good agreement with other observations that aromatic side chains of Phe350 (EGFR) and Tyr13 (EGF) are stacked upon each other (Figure 1.4) [17]. In addition, the long amphipathic sidechain of Arg41 (EGF) forms van der Waals contacts with the sidechains of Tyr13 (EGF) and of Phe357 (EGFR) side chains. The side chains of Leu382, Phe412 and Ile438 (EGFR) make hydrophobic interactions with Leu47 (EGF). Several hydrogen bonds are involved in the EGF-EGFR interactions in both interaction sites (Figure 1.5) [17]. In site one, the side chain of Gln16 of EGFR makes a hydrogen bond with the Asn32 side chain of EGF. It is known that replacement of Asn32 with Asp significantly reduces the receptor binding affinity of EGF [17]. In site two, the sidechain of Gln384 of EGFR makes hydrogen bonds with the main chain amide and carbonyl groups of Arg45 and Gln43 of EGF respectively [17] while, the sidechain of Asp355 of EGFR makes a salt bridge with the sidechain of Arg41 of EGF. This salt bridge is a critical determinant for EGFR binding to all growth factors [18, 19].

Figure 1.3 Subdomain I of the EGFR-EGF interaction site. Subdomains I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The box shows an enlarged view of site 1. Hydrophobic interactions between Leu14 (EGFR)-Ile23 (EGF), Tyr45 (EGFR)-Met21 (EGF), Leu69 (EGFR)-Leu26 (EGF) and Leu98 (EGFR)-Leu26 (EGF) are shown as sticks.

Figure 1.4 Subdomain III of the EGFR-EGF interaction site. Subdomain I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The left box shows hydrophobic interactions between Val350 (EGFR)-Leu15 (EGF) and Phe357 (EGFR)-Tyr13 (EGF) with stick presentation. Right box shows hydrophobic interactions between Leu382 (EGFR)-Leu47 (EGF) and Phe412 (EGFR)-Leu47 (EGF) with stick presentation.

Figure 1.5. Hydrogen bonds between Subdomains I and III of EGFR and EGF. Subdomains I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The box shows enlarged view of hydrogen bonds between Gln16 (EGFR)-Asn32 (EGF), Gln384 (EGFR)-Gln43 (EGF), Gln384 (EGFR)-Arg45 (EGF) and the salt bridge between Asp355 (EGFR)-Arg41 (EGF) with stick presentation.

1.4.3 Monoclonal antibodies

Several antibodies have been designed to target the EGFR extracellular domain and have shown promising results in colorectal and head and neck cancer therapy [3]. These antibodies are designed to target EGFR by blocking the growth factor binding site or the dimeric interfaces [20-22]. Detailed antibody-EGFR interactions are characterized in X-ray crystal structures for C22, IMC-11F8, EMD72000, MEHD7946 and GC1118 antibodies (Table 1.1). C225/Cetuximab/Erbitux is a well known FDA-approved human-mouse chimeric IgG1 antibody that inhibits EGF binding to subdomain III of the extracellular domain [3] . IMC-11F8 targets an almost overlapping epitope of C225 although by forming a different set of interactions [20]. Another example is EMD72000/Matuzumab that blocks the interactions required for dimer formation [22]. MEHD7946 is a dual antibody designed to inhibit both EGFR and ErbB3 and binds to subdomain III of the EGFR extracellular domain [23]. GC118 targets a different epitope from C225 and 11F8 in subdomain III of the EGFR extracellular domain and showed encouraging results in inhibitory activity against high-affinity EGFR ligands [24]. The binding epitopes of the above mentioned antibodies are shown in Figure 1.6.

Monoclonal Antibody	PDB ID	FDA-approved	Target
C225 Cetuximab	1YY9	✓	Growth factor binding
IMC-11F8	3B2V	×	Growth factor binding
EMD72000 Matuzumab	3CO8 3CO9	×	Dimerization
MEHD 7946	3POV	×	Growth factor binding
GC118	4UV7	×	Growth factor binding

Table 1.1. Description of available structural data of monoclonal antibodies bound to EGFR extracellular domain.

Figure 1.6. Binding epitopes of monoclonal antibodies bound to EGFR extracellular domain.

A) Epitope recognized by C225 and IMC-11F8 overlap and is presented in blue surface. MEHD7946 binding epitope is shown in yellow surface and EMD72000 binding epitope is presented in red surface. EGFR and EGF are shown in grey and pink surface respectively. B) GC1118 epitope is located in subdomain III shown in orange surface. EGFR and EGF are shown in grey and pink surface respectively.

1.4.4 Dimeric partner

EGFR extracellular domain contains two dimeric interfaces: the dimerization arm located in subdomain II and the tip of subdomain IV (Figure 1.7 and 1.8). These two interfaces are features of dimeric conformations that need to be formed and maintained [14]. The major dimeric interface is formed by the dimerization arm which is a β -hairpin (Figure 1.7) protruding from the center of subdomain II towards the opposite side of the growth factor binding site [17]. The dimerization arm from one monomer interacts with the dimerization arm from the partner monomer (EGFR or another ErbB receptor) and forms a back to back dimeric configuration. Details of this interaction are fully characterized in X-ray crystal structures for the EGFR-EGFR homo-dimer [17]. In this dimeric interface Tyr246, Pro248 and Tyr251 of the dimerization arm of monomer1 make hydrophobic interactions with Phe263, Ala265, Tyr275 and Arg285 of subdomain II of monomer2. These hydrophobic interactions are accompanied by hydrogen bonds between Tyr251 (monomer1)-Arg285 (monomer2), Gln252 (monomer1)-Ala286 (monomer2) and Tyr4246 (monomer1) –Cys283 (monomer2) (Figure 1.7 and 1.8) [17]. The tip of subdomain IV dimeric interface is modest in size (434 Å² of buried surface area compared to the dimerization arm interface of 1270 Å²) [14]. However, the aromatic and hydrophobic nature of the interface residues (Leu582, Trp584 and Tyr602) is important for dimer formation [14].

Figure 1.7. Dimerization arm dimeric interface of EGFR extracellular domain. Subdomains I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces for monomer1 and blue, cyan, yellow, dark pink and red for monomer2. The boxes are the enlarged views of the dimerization arm interactions. The β -hairpin structure of dimerization arm of monomer1 is shown with green cartoon. Subdomain II of monomer2 is shown with grey cartoon for simplicity. Left and right boxes show the hydrophobic and hydrogen bond interactions respectively. Monomer1 and monomer2 interacting residues are shown with green and cyan sticks.

Figure 1.8. Tip of subdomain IV dimeric interface of EGFR extracellular domain. Subdomains I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces for monomer1 and blue, cyan, yellow, dark pink and red for monomer2. The box is the enlarged view of subdomain IV's tip interactions. subdomain IV tips of monomer1 and monomer2 are shown with magenta and dark pink cartoon respectively. Monomer1 and monomer2 interacting residues are shown with magenta and dark pink sticks.

1.5 EGFR activation

Biological functions of EGFR are dependent on receptor activation. EGFR activation starts with conformational changes in the extracellular domain. Two models were proposed to explain these changes; the “ligand-induced dimerization” model and the “preformed dimer” model. Both models are supported by biochemical assays and 3D structures; however in light of new findings, the second one is currently dominant and more widely accepted.

1.5.1 Ligand induced dimerization

In this model the EGFR monomer is thought to be initially tethered by subdomains II and IV through extensive intramolecular contacts. At the center of this interface the dimerization arm, which is a 20 residue loop (residues 240-260), projects from subdomain II between the two loops from subdomain IV (residues 561-569 and 572-585) [25]. After EGF binds to the monomer, rearrangement of subdomains I, II and III occur, the dimerization arm becomes unburied and interacts with the dimerization arm of another monomer and a back to back dimer of the two extracellular domains are formed (Figure 1.9) [25]. This tethered-extended transition is a large conformational change and was believed to be induced by EGF binding to the extracellular domain.

Figure 1.9. Ligand induced dimerization of EGFR. Growth factor binding rearranges the tethered extracellular domain; extended monomer interacts with the other monomer and forms a dimer. Subdomains I, II, III, IV of EGFR extracellular domain and EGF are shown with blue, green, yellow, magenta and red surface presentation. The transmembrane and intracellular domains are shown with pink surface. Cell membrane is shown as lines.

1.5.2 Preformed dimer

The second model, and the currently favored one, suggests that EGFR exists as an active dimer even in the absence of ligands [26-28]. This dimer is flexible and is stabilized by binding of EGF. Intradimeric twist motion of the stabilized extracellular dimers induces transmembrane domain rotation and consequently rearrangement of the intracellular kinase domains and receptor activation. In this model the tethered-extended transition is not prerequisite for EGFR activation [13] (Figure 1.10).

Figure 1.10. Flexible preformed EGFR dimer is stabilized by EGF binding and induces rotation in the transmembrane domain. Subdomains I, II, III, IV of EGFR extracellular domain and EGF are shown with blue, green, yellow, magenta and red surface presentation. The transmembrane domains of receptors are shown with black and orange surface. Cell membrane is shown as lines.

1.6 EGFR glycosylation

Activation, control and fine tuning of interactions in biology depend on post-translational modifications such as phosphorylation, glycosylation etc. [29]. It is well known that the EGFR extracellular domain is heavily glycosylated with oligosaccharides [30, 31]. Glycosylation is known to play critical roles in EGFR dynamics and function [32, 33] but the limitation has been that the detailed mechanisms at an atomic level are not well understood. The EGFR activation models are all based on structural interpretations of the unglycosylated/underglycosylated receptor.

1.6.1 Protein glycosylation

Protein glycosylation modulates diversity and biological activity of proteins [29]. This establishes an important role in disease and a need to understand molecular mechanisms [34-36]. Glycans have essential roles in the cell including molecular trafficking, protein recognition and structural roles [29]. Glycoconjugate molecules are the most informative biomolecules including glycolipids (lipids with a carbohydrate attached), glycoproteins (proteins that contain oligosaccharide chains covalently attach to polypeptide sidechains) and proteoglycans (proteins that are heavily glycosylated) [29]. Glycosylation is the covalent attachment of glycans to the protein by forming a glycosidic bond. It involves 13 different kinds of monosaccharides that can be attached to eight types of amino acids, thus resulting in great complexity. There are thought to be two major forms of glycosylation: N-linked (nitrogen of asparagines or arginine) and O-linked (hydroxy oxygen of serine, threonine, tyrosine, hydroxyserine or hydroxyproline) [37]. N-glycosylation sites are usually recognized by the presence of a sequon or the sequence Asn-Xaa-Ser/Thr (where Xaa is any amino acid except proline) [38, 39]. For N-linked glycosylation, the

attachment of sugars (Glc3Man9GlcNAc₂) to the sidechain of Asparagine residues takes place in the endoplasmic reticulum (ER) by the enzymes oligosaccharyl transferases (OTase) [40]. The glycan is processed and modified in the ER and in the Golgi, generating three types of N-glycans found in mature glycoproteins: high mannose, complex and hybrid (Figure 1.11) [29]. Evidence from studies on N-glycosylation sites suggests that the consensus sequon is neither necessary nor sufficient for N-glycosylation [41]. Various factors are involved in the glycosylation of Asn including the role of Asn in local folding, binding affinity of the Asn proximal region for OTase, secondary structures, disulfide bonds etc. [41]. In contrast, O-glycans are usually found attached to OH group of serine and threonine although in principle they can be attached to any hydroxyl group in amino acids (Figure 1.11); serines and threonines have been found to be modified with N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), xylose, mannose and fucose [42]. In addition, extended assemblies of GalNAc called mucin-type O-glycans are also seen; their biosynthesis takes place after protein N-glycosylation, folding and oligomerization [29].

Figure 1.11. N- and O- linked glycans structures. A) Complex type N-linked glycans attached to Asn in the conserved glycosylation sequon shown with cartoons. B) High mannose N-linked glycans attached to Asn in the conserved glycosylation sequon shown with cartoons. C) O-linked glycan attached to Ser/Thr shown with cartoons.

1.6.2 Challenges in glycoprotein analysis and role of computer simulations

Understanding the functions of glycans has been one of the most challenging areas in biological sciences because it involves a complex machinery that, through a multistep process, produces a mixture of glycosylated variants (glycoforms) for each glycoprotein [29]. In addition, they are often present at low levels, thus making detection difficult. The picture is further made complex by a microheterogeneity arising from differential modifications of the various sequons by a diverse collection of enzymes during biosynthesis [29]. Experimental techniques used to examine glycosylation are encumbered with problems such as for example in using CE (Capillary Electrophoresis with laser induced fluorescence), a very commonly used and efficient technique for separating charged molecules, which usually fail as the glycans lack ionizable charged groups [29]. Similarly, the technique of X-ray crystallography, which has proved to be a successful and powerful technique in the determination of the 3D structures of biomolecules has largely failed here because glycans are very flexible [29]. Carbohydrates have two naturally active nuclei (C and H), thus suggesting possible structural analysis by NMR (Nuclear Magnetic Resonance) and yet has not succeeded because NMR requires highly purified and large amounts of sample (typically > 1mg) which has been difficult to achieve [29]. Hence, currently there exists no experimental method available for the rapid and reliable detailed structural analysis of the proteo-glycan. It is here that computer modelling can play a significant role in establishing structural models. Computer modelling and simulations have been invaluable in complementing the understanding of properties of molecules and their assemblies in terms of their structures and the microscopic interactions between them [43]. Computer simulations bridge the

microscopic length and time scales with the macroscopic world of the laboratory. In addition, they provide a link between theory and experiment. In particular, they prove extremely useful in conducting experiments that are difficult or impossible to carry out in the laboratory [43]. They are playing an increasingly useful role in understanding the complexity of glycobiology through developments in parameterizing, modeling and subsequently studying the functional and structural effects of glycosylation of proteins.

1.6.3 EGFR glycosylation

Several studies have reported that glycosylation is involved in the regulation of EGFR [44-46]. The EGFR extracellular domain (henceforth to be referred to as EGFR-ed) is thought to be N-glycosylated with sugar moieties making up nearly 25% of the 180 kDa net mass [33]. In the extracellular domain there are 10 Asparagines located within the consensus N-glycosylation sequence. These are: Asn104, Asn151, Asn172, Asn328, Asn337, Asn420, Asn504, Asn544, Asn579 and Asn599 (Figure 1.12). Although these Asparagines are part of the known glycosylation sequence, there is no clear consensus about their modifications in the literature. In the study from the laboratory of Daisuke Kolhda, only four of these sites (Asn328, Asn337, Asn544 and Asn579) were reported to be modified with glycans [47]. In another study, from the laboratory of Elizabeth F. Hounsell, five other sites (Asn104, Asn151, Asn389, Asn420 and Asn504) were reported to be glycosylated [48]. Together these studies reported that 8 of the 10 likely Asparagines (except Asn172 and Asn599) are glycosylated. A separate study on fucosylation (addition of fucose to the oligosaccharide glycan) of the EGFR-ed revealed that 9 Asn residues are glycosylated (Asn151, Asn328, Asn337, Asn389, Asn420, Asn504, Asn544, Asn579 and Asn599) [30]. There is another glycosylated Asparagine in the EGFR-ed which is known

to be an atypical glycosylation site. This Asparagine residue is Asn32 and is part of an Asn-Asn-Cys sequence [47] and has been reported to be glycosylated and fucosylated; it is possible that since Ser and Cys differ only by an Oxygen/Sulphur, may account for this observation.

Figure 1.12. Summary of consensus N-glycosylation sites of EGFR extracellular domain.

Subdomains I, II, III and IV of the extracellular domain of EGFR are shown with blue, green, yellow and magenta squares and the amino acid numbers of each subdomain labeled with the same colors. Asparagines within the conserved sequon are labeled in black and the attached glycans are shown with cartoons.

1.6.4 Functions of EGFR with attached glycans

Glycosylation affects translocation and maturation of EGFR [33]. If N-linked glycosylation is inhibited, then the resulting EGFR does not show detectable ligand binding or kinase activity [33]. Studies of specific N-glycans in EGFR have shown that carbohydrate attachments influence local protein conformations and dynamics near glycosylation sites [33]. Glycosylation of subdomain III sites (Asn389, Asn420, Asn328, and Asn337) may affect proper receptor conformation and ligand binding. Asn579,

located at the tip of the auto inhibitory tether loop in domain IV may modulate the kinetics of transition between tethering and untethering [33]. Another glycosylation site, Asn420, is known to play an important role in controlling ligand independent oligomerization [49] and the attached oligosaccharide may assist in maintaining the tethered form [32]. The addition of N-glycan may affect the torsional dynamics of the asparagine side chains, thus affecting flexibility and protein folding [50]; it is known that N-glycosylation is required for the proper sorting of EGFR to the membrane as well as for ligand binding [44]. Modification of N-glycans has been shown by Partridge et al. [51] to affect the rate of EGFR endocytosis and internalization, thus affecting signaling. Several groups have applied molecular modelling and simulations to understand the links between atomistic dynamics of the glycosylated EGFR extracellular domain and function. The group of David E. Shaw showed that asymmetrical interactions of one of the ligands with the membrane in the dimeric form of EGFR could lead to the observed negative binding cooperativity of EGFR (Negative cooperativity is known to be the critical feature of EGFR ligand binding [52, 53]. When one of the growth factor binding sites gets occupied with one growth factor, the affinity of the other monomer for growth factors decreases largely [53, 54]) irrespective of glycosylation [52]. The group of Ilpo Vattulainen suggested that glycosyl units appear to mediate conformational rearrangements of EGFR- by interactions with the membrane and play a role in the binding of the antibody 2E9 to EGFR [55].

1.7 Summary

EGFR is a transmembrane glycoprotein associated with several cellular functions in health and several diseases, including cancers, once mutated/overexpressed. EGFR is

translocated to the cell membrane only after glycosylation. The glycosidically bonded sugars play roles in EGFR activation, ligand binding and dimerization. However, due to the complex structural organization of glycosylation, cell and tissue specific heterogeneity and the difficulties in obtaining structural data, detailed effects of glycosylation on EGFR structural and functional dynamics are not well understood. In this thesis, molecular modeling and simulation techniques are used to investigate the roles of glycosylation on the dynamics and interactions of the EGFR ectodomain (EGFR-ed). These roles were addressed in two parts over three chapters:

I. Effects of glycosylation on EGFR ligand binding: binding of three groups of ligands (growth factor, monoclonal antibodies, dimeric partner) to EGFR-ed was investigated in the absence and presence of glycosylation.

II. Effects of glycosylation on EGFR activation: tethered-extended equilibrium of monomeric EGFR-ed and the dynamics of dimeric EGFR-ed was modeled with and without attached glycans to develop hypotheses regarding the most physiologically plausible activation model.

The above were investigated by using a combination of conventional and accelerated atomistic MD simulations. These simulations ranged from 50 ns to 1800 ns. Models of EGFR-ed in various forms of glycosylation included tethered/untethered monomers, dimers, apo and EGF-bound states; the attached glycans included monosaccharide units of N-acetylglucosamines, and complex type of oligosaccharides. These findings shed light on ligand binding, antibody interactions and dimerization of all the ErbB receptors mediated by glycosylation sites, potentially leading to new understanding on regulation

and also for the design and development of more efficient cancer therapeutics such as antibodies. A summary of the chapters can be found in Figure 1.13.

Figure 1.13. Outline of the thesis chapters.

Chapter 2. Computational methods

2.1 Introduction

Computer simulations have proven to be a reliable approach to probe the mechanics underlying biological phenomena [56]. The most commonly used method, conventional molecular dynamics (MD) simulations, are based on the numerical solution of Newton's equation of motion and provide atomistically detailed information of biomolecules with respect to time. Although this method has been very successful in structural and molecular biology [57], it still has several drawbacks. Biological time scales are much longer than simulation time scales due to limited computational resources. This issue was addressed by the development of enhanced sampling techniques such as: accelerated molecular dynamics, steered molecular dynamics and replica exchange molecular dynamics [56]. The other major drawback is the size of biomolecules and of their higher order assemblies/complexes. This issue is addressed by coarse graining the representations so that groups of atoms can be represented by a large atom [58]. In this chapter the basics of MD simulations, enhanced sampling techniques and analytical methods used in this thesis along with the model building and simulations of unglycosylated and glycosylated EGFR are described.

2.2 Molecular mechanics force field

Molecular Dynamics (MD) simulation is based on solutions of Newton's equation of motion for an atomic system (Equation 2.1) [43].

$$f_i = m_i \ddot{r}_i \quad \text{Equation 2.1}$$

$$f_i = - \frac{\partial}{\partial r_i} U \quad \text{Equation 2.2}$$

Calculation of forces on every atom (f_i) can be obtained by calculating the derivative of the potential energy $U(r^N)$, where r^N represents the 3N atomic coordinates (Equation 2.2) [43]. In MD simulations the potential energy function is known as a force field. The force field describes the atomic composition of a molecule to be a collection of masses interacting with each other via forces described by empirical potential energy functions [59]. The potential energy function describes both bonded and non-bonded molecular interactions (Equation 2.3). Bonded interactions are: bond stretching, angle bending and torsion terms. Non-bonded interactions include electrostatics and Van der Waals interactions. Each term will be explained in the following section respectively.

$$U(r_i) = \sum_i \frac{k}{2} (l_i - l_0)^2 + \sum_i \frac{k}{2} (\theta_i - \theta_0)^2 + \sum_{n=0}^N \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\ + \sum_{i,j=1}^N \frac{q_i q_j}{4\pi \epsilon_0 r_{ij}} \sum_{i,j=1}^N 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6 \right] \quad \text{Equation 2.3}$$

2.2.1 Bonded interactions

2.2.1.1 Bond stretching

The bond potential energy (first term in equation 2.3) is represented by Hooke's law in which the changes in energy arise from the displacement of each bond length (l_i) from the reference bond length l_0 (the average distance between two fixed types of atoms that is

measured empirically) and is independent of the molecular composition) [60]; the bond stretching constant is k

$$v(l) = \frac{k}{2} (l_i - l_0)^2 \quad \text{Equation 2.7}$$

2.2.1.2 Angle bending

The angle potential energy (second term in equation 2.3) is also represented by Hook's law to describe the angle deviation from the reference point [60];

$$v(\theta) = \frac{k}{2} (\theta_i - \theta_0)^2 \quad \text{Equation 2.8}$$

Contribution of each angle in the potential energy function is characterized by a force constant k and a reference value (the average bond angle value between three connected atoms that is measured empirically using NMR and FRET techniques [61]). It is obvious that the amount of energy needed for angle bending is less than for stretching/compressing the bonds.

2.2.1.3 Torsional term

Since the bond stretching and angle bending require substantial amounts of energy for significant deformation from the reference value (see section 2.2.1.1 and 2.2.1.2), most of the variations in structure and relative energies arise from the complex changes of torsions and non-bonded interactions [60]. Torsional energy (the third term in equation 2.3) depends on the dihedral angle made by two planes incorporating the first and last three atoms involved in the torsion (if A-B-C-D are 4 atoms and – represents bonds between the atoms, then the torsion refers to the angle made between the planes defined by A-B-C

and B-C-D) [59]. Torsional potentials are almost always expressed as cosine series expansions:

$$v(\omega) = \sum_{n=0}^N \frac{V_n}{2} (1 + \cos (n\omega - \gamma)) \quad \text{Equation 2.9}$$

ω is the torsion angle, V_n is the barrier height [60].

2.2.2 Non-bonded interactions

2.2.2.1 Electrostatic interactions

Each atom in a molecule is modelled with a point charge [60]. These point charges account for the electrostatic properties of a molecule. The net interactions between the charges, represented by Coulomb's law, is the electrostatic energy of the molecule (the fourth term in equation 2.3):

$$v = \sum_{i,j=1}^N \frac{q_i q_j}{4 \pi \epsilon_0 r_{ij}} \quad \text{Equation 2.10}$$

N is the number of point charges, q_i and q_j are the charges on atoms i and j respectively, ϵ_0 is the permittivity of free space and r_{ij} is the distance between the charges [43].

2.2.2.2 van der Waals interactions

van der Waals forces are the attraction and repulsive forces between the atoms that are not arise from ionic/covalent bonds [59]. For modeling of each atomic system, a function that can calculate several number of van der Waals interactions is needed [60].

The Lennard-Jones 12-6 function is most frequently used to model the van der Waals energies (the fifth term in equation 2.3), which takes the following form for the interactions between two atoms:

$$v(r) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad \text{Equation 2.11}$$

Where σ is the collision diameter and ϵ is the well depth between atoms i and j in Figure 2.1.

Figure 2.1. The Lennard-Jones potential [60].

Several molecular mechanics force fields have been developed [57] of which AMBER [62], GROMOS [63, 64], OPLS and CHARMM [65, 66] are the most widely used.

2.2.5 The GLYCAM force field

GLYCAM06 is a widely used force field for modeling carbohydrates and glycoconjugates [67-69]. GLYCAM06 parameters are self contained i.e. no default or generic parameters are required for simulations of carbohydrates [67]. The torsion term in the GLYCAM06 is derived from quantum mechanical calculations of internal rotational energy for highly symmetrical small molecules and one set of parameters are used for α and β anomers of all monosaccharides [67]. The bond lengths and angle variations in GLYCAM06 are derived from one set of bond force constant, angle force constant, equilibrium bond length and equilibrium angle for both α and β anomers. These values are derived from the stereoelectronic effects at the pyranoses anomeric carbon [70, 71]. Calculation of the intermolecular electrostatic interactions of carbohydrates in GLYCAM06 is carried out by assigning partial charges to different atoms similar to every other force field. Magnitude of charges carried by aliphatic hydrogen atoms is small due to the low polarity of the C-H bond. Thus the charges on all aliphatic carbons are set to zero in GLYCAM06 [67]. Intramolecular interactions and solution properties are defined by the average charge values obtained from conformational ensembles [67]. For simulation of glycoproteins, GLYCAM06 is combined with the AMBER force field. Orthogonality of the GLYCAM06 parameters when combining with AMBER is obtained by the unique atom names for the carbohydrate atoms. The carbohydrate atoms names starts with an uppercase letter followed by a lower case letter or a number. As an example the “CG” atom type used in GLYCAM06 will be changed to “Cg”. The list of newly generated parameters for different glycans or any alteration in the letter codes of the current parameters is regularly updated and accessible at the GLYCAM online facility [72].

GLYCAM06 has shown promising results in reproducing the rotamer population of biomolecules in explicit waters or in crystal lattice [67]. In the current study, GLYCAM06 in combination with AMBER ff99SB force field was used for modeling of the glycosylated EGFR extracellular domain.

2.3 Solvent model

MD simulations of a molecule are normally carried out in explicit aqueous solvent to mimic the cellular environment. There are two types of models to include solvent in the simulations: explicit solvent model and implicit solvent model (Figure 2.2). Explicit solvent model provides a more realistic understanding of solvation thermodynamics and includes atomic details of every water molecule. Implicit solvent models replace the solvent with a dielectric continuum [73-77].

Figure 2.2. Presentation of explicit and implicit solvent models. A) Epidermal growth factor is shown with blue cartoon surrounded by explicit water molecules shown as spheres. B) Epidermal growth factor is shown with blue cartoon in implicit solvent. Continuum solvent is schematically presented with the red box. The blue shade around the protein presents the cavity formed by solute in the continuum solvent.

2.3.1 Explicit solvent model

Explicit water models are classified into different groups based on the number of interaction sites used to model the water molecule (3-site, 4-site, 5-site and 6-site), flexibility of the water molecules and integration of the polarization effect in the model [78]. The most widely used and simple water model is the TIP3P. This is a 3-site model and represents high computational efficiency and computed properties that are

comparable to thermodynamic values derived from experiments [79] with charges and geometry as shown in Figure 2.3.

Figure 2.3. TIP3P water model. Oxygen and hydrogen atoms are shown with red and grey circles respectively.

2.3.2 Implicit solvent model

Explicit solvent models provide more accurate solvation energies but can be computationally expensive and not suited for very large systems or for high throughput modeling [73]. Implicit solvent models are developed to treat solvent based on continuum dielectrics, dielectric screening function and solvent-accessible surface area [80]. The dielectric continuum models are able to estimate ligand-receptor free energy of solvation in a matter of seconds hence they are extensively used in ligand-receptor binding energy calculations. A commonly used model is the Generalized Born Surface Area (GBSA) implicit solvent model and is explained here, as it is used in the binding free energy calculations and per-residue free energy decomposition analysis in this thesis. In the GBSA model, estimation of solvation free energy is carried out by splitting this energy into non-electrostatic (ΔG_{nonele}) and electrostatic (ΔG_{ele}) components (Equation 2.13).

$$\Delta G_{Sol} = \Delta G_{nonele} + \Delta G_{ele} \quad \text{Equation 2.13}$$

ΔG_{nonele} is the solvation energy of a molecule where all charges are set to zero. This provides an estimate of the solvation energy due to favorable VDW interactions between the solute and the solvent and the cost of breaking the solvent structure around the solute. The components are estimated base on the solvent accessible surface area of the solute. ΔG_{ele} is the energy required to transfer the charged solute molecule from vacuum to the particular solvent. ΔG_{ele} is obtained by the Generalized Born (GB) method which is a computationally effective approximation of the more rigorous Poisson-Boltzmann approach shown below:

$$\Delta G_{ele} = \frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{GB}(r_{ij}, R_i, R_j)} \left(1 - \frac{\exp[-kf_{GB}]}{\epsilon} \right) \quad \text{Equation 2.14}$$

where r is the distance between atoms i and j , R is the radius of the sphere assigned to each atom with charge q at its center, f_{GB} is a function that interpolates between R and r_{ij} over short distances and r_{ij} itself in long distances and K is the parameter incorporating the extension of GB model to

low salt concentrations at the Debye-Huckel limit [81, 82]. A common expression of f_{GB} is:

$$f_{GB} = \left[r_{ij}^2 + R_i R_j \exp \left(-r_{ij}^2 / 4 R_i R_j \right) \right]^{1/2} \quad \text{Equation 2.15}$$

R reflects the extent of burial of each atom in the molecule. For an isolated ion, R is equal to the VDW radius ρ and $K=0$ is for pure water, resulting in a simple form of ΔG_{ele} :

$$\Delta G_{ele} = - \frac{q_i^2}{2\rho_i} \left(1 - \frac{1}{\epsilon}\right) \quad \text{Equation 2.16}$$

2.4 Energy minimization

Before the MD simulation is carried out, the initial structure, that may originate in the crystal structure or may have been constructed by homology modeling, or may have modifications such as the sugar moieties added in this study, may consist of clashes between atoms and these need to be relieved. This is carried out by energy minimizations to find the closest local minima in the potential energy surface. There are many approaches for energy minimization including non-derivative minimization and derivative minimization methods. Here the derivative minimization methods used in this study are explained.

2.4.1 Derivative minimization method

Molecular mechanics potential energy is a function of the Cartesian coordinates of the atoms as its variable:

$$f = f(x_1, x_2, x_3, \dots) \quad \text{Equation 2.17}$$

The values of variables for which f is minimum are obtained when the first derivative of f is zero and the second derivative is greater than zero:

$$\frac{\partial f}{\partial x_i} = 0, \quad \frac{\partial^2 f}{\partial x_i^2} > 0 \quad \text{Equation 2.18}$$

Different algorithms such as steepest descents and conjugate gradient have been developed to perform the above calculation and shift the structure towards the local minima through several steps of minimization.

2.4.1.1 Steepest descent method

Steepest descent method moves in the direction parallel to the net force, which is the geographical analogy corresponds to walking straight downhill on the energy surface. This direction is presented by a 3N dimensional vector:

$$S_k = \frac{-g_k}{|g_k|} \quad \text{Equation 2.19}$$

where the gradient is obtained by the Hessian matrix of the potential energy function:

$$g_k = \frac{\partial V}{\partial x_k} \quad \text{Equation 2.20}$$

This method improves quickly due to the direction of movement which is straight downhill in the energy surface. However this method is mainly implemented in the initial steps of minimization where the gradient norm values are large [59].

2.4.2.2 Conjugate gradient method

The conjugate gradient method moves by considering the direction of the net force in both the current step and the previous step:

$$s_k = -g_{k-1} + \gamma_k s_{k-1} \quad \text{Equation 2.21}$$

Where s_k and s_{k-1} are the current and the previous steps directions respectively and γ_k is a constant scalar:

$$\gamma_k = \frac{g_k \cdot g_k}{g_{k-1} \cdot g_{k-1}} \quad \text{Equation 2.22}$$

2.5 Molecular Dynamics

After the clashes are removed by energy minimization, the system is ready to be subject to MD simulations. MD simulations compute the step by step solution of Newton's equation of motion for a system of atoms interacting with the potential energy function as defined and explained above (see section 2.2). These equations of motion are solved numerically through integration. An initial set of velocities corresponding to the average temperature of the system are assigned to each atom and the system is advanced by a small time step (Δt) and the forces and velocities on each atom are computed and this process is repeated iteratively. If Δt is small enough, this produces an acceptable approximate solution to the continuous equation of motion [59]. The Verlet algorithm is widely used to perform this integration in molecular dynamics [43]. The basic idea is to write two third-order Taylor expansions for the positions $r(t)$, one forward and one backward in time:

$$r(t + \Delta t) = r(t) + v(t) \Delta t + \left(\frac{1}{2}\right) a(t) \Delta t^2 + \left(\frac{1}{6}\right) b(t) \Delta t^3 + O(\Delta t^4) \quad \text{Equation 2.23}$$

$$r(t - \Delta t) = r(t) - v(t) \Delta t + \left(\frac{1}{2}\right) a(t) \Delta t^2 - \left(\frac{1}{6}\right) b(t) \Delta t^3 + O(\Delta t^4) \quad \text{Equation 2.24}$$

Where v is velocity, a is acceleration and b is the third derivative of r with respect to time t . Addition of the two above equations together results in the basic form of the Verlet algorithm:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + a(t) \Delta t^2 + O(\Delta t^4) \quad \text{Equation 2.25}$$

In the above equation the velocities are not generated directly. However, the velocities are required for the kinetic energy and consequently for the total energy calculations (total energy = kinetic energy + potential energy). This matter was overcome by developing variants of the Verlet algorithm. The “leap-frog” algorithm is one of these variants where positions and velocities of particles are successively leap-frogged over each other using the acceleration [59]:

$$r(t + \delta t) = r(t) + v \left(t + \frac{1}{2} \delta t \right) \delta t \quad \text{Equation 2.26}$$

$$v \left(t + \frac{1}{2} \delta t \right) = v \left(t - \frac{1}{2} \delta t \right) + a(t) \delta t \quad \text{Equation 2.27}$$

When the velocities are calculated at the time $t + \frac{1}{2} \delta$, and are used to calculate the positions r at the time $t + \delta t$. In this way the velocities are calculated explicitly and the positions and the velocities are successively leap-frogged over each other. The calculations are carried out for a period of time which is decided by the question and by the computing resources available.

2.5.1 Periodic Boundary Conditions

To mimic real systems, multiple copies of the biomolecule to be simulated are needed. However their explicit consideration will be computationally very expensive. This

problem is overcome by placing the biomolecule in a box of explicit water molecules and applying periodic boundary conditions (PBC) [43] to simulate an array of biomolecules. Often the simulation box is large enough to maintain about 10-12 Å distance between the biomolecule and the box edges. As one atom leaves the simulation box its periodic image enters the box through the opposite face (Figure 2.4). In the application of PBC a "minimum image convention" is set that every atom in the simulation box interacts with only the closest image of all other atoms in the box [43].

Figure 2.4. Periodic boundary condition. The central box is the simulation box and the transparent boxes are its replicates. Once a particle leaves the simulation box its periodic image enters the box through the opposite face.

2.5.2 Particle Mesh Ewald

To treat long-range electrostatics interactions in the infinite array of molecules that characterize the PB conditions in MD simulations, the Particle Mesh Ewald method is used. In this method the interaction potential between two interacting particles i and j is

split into two terms: sum of the short-range interactions are treated in real space and sum of the long-range interactions are treated in the Fourier space:

$$E_{tot} = \sum_{i,j} \varphi(r_j - r_i) = E_{sr} + E_{lr} \quad \text{Equation 2.28}$$

$$E_{sr} = \sum_{i,j} \varphi_{sr}(r_j - r_i) \quad \text{Equation 2.29}$$

$$E_{lr} = \sum_k \tilde{\Phi}_{lr}(k) |\tilde{\rho}(k)|^2 \quad \text{Equation 2.30}$$

$\tilde{\Phi}$ and $\tilde{\rho}(k)$ are potential and charge density Fourier transforms. The assumption of Ewald summation is the periodic symmetry of the physical system. Hence it is implemented in the periodic boundary condition in MD simulations.

2.5.3 Bond constraints

The smallest time step used in MD simulations must reflect the highest frequency motion which is the motion of hydrogens attached to their heavy atoms; this motion occurs at 10^{-15} s. Since this is usually not of interest, this motion is constrained to enable a longer time step (multiples of 10^{-15} s) to be used. Therefore, the bond lengths between the hydrogen atoms and the atoms they are attached to are constrained at their equilibrium values during MD simulation. Several algorithms are developed to apply bond constraints such as SHAKE and its variants [83-87], SETTLE[88] and LINCS [89] with SHAKE being the most commonly used method.

2.5.4 Different ensembles in molecular dynamics

MD simulations provide information at the microscopic level (pressure, energy, heat capacities etc. of the system). The transformation of these microscopic properties to macroscopic/thermodynamic ones is carried out by statistical mechanics. For any thermodynamic observable an ensemble average can be calculated with a partition function:

$$P_i^B = \frac{\exp \left[-E_i / K_B T \right]}{Z} \quad \text{Equation 2.31}$$

Where P_i^B is the Boltzmann Probability, K_B is the Boltzmann constant, T is the absolute temperature and Z in the partition function.

$$Z = \sum_i \exp \left[-E_i / K_B T \right] \quad \text{Equation 2.32}$$

$$\langle A \rangle = \sum_i A_i \exp \left[-E_i / K_B T \right] / Z \quad \text{Equation 2.33}$$

$\langle A \rangle$ is the ensemble average of the observable A .

Depending on the favorable thermodynamic properties, different ensembles are used in MD simulations such as:

- Microcanonical ensemble (NVE) is characterized by fixed number of atoms N , fixed volume V and fixed energy E .

- Canonical ensemble (NVT) is characterized by fixed number of atoms N , fixed volume V , and fixed temperature T .
- Isobaric-Isothermal ensemble (NPT) is characterized by fixed number of atoms N , fixed pressure P and fixed temperature T .
- Grand canonical ensemble (μ VT) is characterized by fixed potential μ , fixed volume V and fixed temperature T .

To obtain an ensemble average during MD simulations, all the possible states of the system are considered for calculation of that particular thermodynamic property. This issue is addressed by defining a time average for any observable:

$$\langle A \rangle_{time} = \frac{1}{\tau} \sum_{\tau \rightarrow \infty} A(P^N(t), r^N(t)) \approx \sum_{t=1}^M A(t) \quad \text{Equation 2.34}$$

Where t is the simulation time, M is the number of time steps in the simulation and

$A(P^N, r^N)$ is the instantaneous value of A .

If the simulation time is long enough, it is assumed that the time average is equal to the ensemble average. This assumption is one of the fundamentals of statistical mechanics known as the ‘‘Ergodic hypothesis’’. The length of simulation to obtain the time average is defined based on the physical properties of the variable of interest and implementation of enhanced sampling techniques to overcome the energy barriers.

MD simulations usually contain around 10^6 steps which corresponds to nanoseconds of real time [43]. Reliable statistics are only obtained from MD simulations if the molecular system has reached the equilibrium state. Conventional MD simulations are limited in sampling the phase space sufficiently. Long time scale MD simulations to explore

multiple energy wells are computationally expensive and time consuming and is overcome by using biased MD methods such as accelerated MD or steered MD. Figure 2.5 explains the goal of enhanced MD simulations for efficient sampling of the phase space.

Figure 2.5. Sampling the phase space in MD simulations. The minimization steps lead the conformation to the closest local energy minima. Over the equilibration period, the conformation escapes the local minima separated by short energy barriers. Enhanced sampling techniques are required for the transition between the intermediate states.

2.6 accelerated Molecular Dynamics (aMD)

Accelerated Molecular Dynamics (aMD) is an enhanced sampling technique demonstrated to be efficient for protein simulations [90, 91]. aMD enhances the sampling of different basins in the energy landscape by increasing the chance of escaping the local minima without prior knowledge of the potential energy surface [90]. This technique raises the potential surface near the minima by introducing a bias boost potential. If the value of the potential is below the chosen boost energy (E), the simulation is carried out by adding a biasing boost potential function to the true potential:

$$V^*(r) = V(r), \quad V(r) \geq E$$

$$V^*(r) = V(r) + \Delta V(r), \quad V(r) < E$$

$V^*(r)$ is the modified potential, $V(r)$ is the true potential, $\Delta V(r)$ is the biasing boost potential function and E is the boost energy. The biasing potential is proposed as [90]:

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \quad \text{Equation 2.37}$$

α is the parameter defined by how deep the modified potential energy surface is. Selection of E and α are critical for the efficiency of sampling. E should be greater than V_{min} , that is the energy of the initial structure often trapped in the local minima. The most suitable choice of α is $\alpha = E - V_{min}$ that provides a smooth echo of the true potential surface. Here a dual potential boost was used to modify the potential energy and the torsional term.

2.7 Molecular Dynamics simulations of the EGFR extracellular domain

2.7.1 Crystallographic structures

The Protein Data Bank entries containing crystal structures of ErbB receptors with detailed descriptions of extended/tethered, monomeric/dimeric and chemically modified with glycans or unmodified states are listed in Table 2.1. Of these, two structures describe the extracellular domain of extended human EGFR bound to EGF: PDB entries 1IVO (resolution 3.30 Å, [17]) and 3NJP (resolution 3.30 Å, [14]) and one structure depicts the tethered extracellular domain: PDB entry: 1NQL (resolution 2.8 Å, [25]). In this study, 3NJP was chosen as the initial structure used to represent the extended monomeric and dimeric states since it is the only full length extended dimeric structure of the EGFR extracellular region available. 1NQL has been chosen to represent the tethered monomeric EGFR initial structure. Since most computational studies have reported on unglycosylated EGFR, it is of interest to examine the differences with the glycosylated form, since the latter is most likely the natural form. This is the advantage that computational models bring by allowing such differences to be examined.

2.7.2 Capping the protein

Since the crystal structures used are not the full length proteins but truncated versions, neutral states of the N-terminus and C-terminus are maintained by adding capping residues at each end of the structures. The tleap module of AMBER12 (Assisted Model Building with Energy Refinement) package has been used for capping of EGFR and EGF; the N-terminus is capped with the ACE moiety while the C-terminus is capped with the NME moiety.

PDB ID	Protein	Extended	Tethered	Monomer	Dimer	Underglycosylated (not fully glycosylated)	Unglycosylated
1IVO	EGFR	✓			✓	✓	
1M6B	HER3		✓		✓	✓	
1MOX	EGFR	✓			✓	✓	
1N8Y	HER2	✓		✓		✓	
1N8Z	HER2	✓		✓		✓	
1NQL	EGFR		✓		✓	✓	
1S78	HER2	✓		✓		✓	
1YY9	EGFR	✓		✓		✓	
2AHX	HER4		✓		✓	✓	
3B2U	EGFR (subdomain III)	✓		✓		✓	
3MZW	HER2	✓		✓		✓	
3N85	HER2	✓				✓	
3NJP	EGFR	✓			✓	✓	
3QWQ	EGFR	✓		✓		✓	
3U7U	HER4	✓		✓		✓	
3U9U	HER4	✓		✓			✓
3B2V	EGFR	✓		✓		✓	
3U2P	HER4 (N-terminal bound to ligand)	✓				✓	
3BE1	Her2	✓		✓		✓	
3C09	EGFR (subdomain III)	✓		✓		✓	
3P0Y	EGFR (subdomain III)	✓		✓		✓	
3P11	HER3 (subdomains I-III)	✓		✓		✓	

Table 2.1. ErbB receptors PDB entries list.

2.7.3 Protonation states of the Histidines

Histidine residues can be protonated at the δ (delta) position, the ϵ (epsilon) position or at both positions (Figure 2.6), depending on the environment around the Histidine. PROKA3.1 was used for the prediction of protonation states [92] and the output was evaluated manually for each Histidine. The tleap module of AMBER12 was used to assign the protons to their positions.

Figure 2.6. Protonation states of Histidine with hydrogens on the delta nitrogen (A) and on the epsilon nitrogen (B).

2.7.4 Disulfide bonds

Twenty eight disulfide bonds occur in each monomer of EGFR bound to EGF with twenty five in each EGFR monomer and three in EGF. The tleap module of AMBER12 package is used for specifying these bonds. Detailed description of disulfide bonds are mentioned in Table 2.2.

EGFR bonded Cys residue no.		EGF bonded Cys residue no.	
7	34	6	20
133	163	14	31
166	175	42	34
170	183		
191	199		
195	207		
208	216		
212	224		
227	236		
240	267		
271	283		
287	302		
305	309		
313	338		
446	475		
482	491		
486	499		
502	511		
515	531		
534	547		
538	555		
558	567		
571	593		
596	604		
600	612		

Table 2.2. Bonded Cysteine residue numbers of EGFR and EGF.

2.7.5 Construction of glycosylated EGFR models

Atomistic models of glycosylated EGFR extracellular domain were constructed by using the ‘GLYCAM online glycoprotein builder’ [67, 72] to attach the glycans at Asn32, Asn49, Asn151, Asn172, Asn328, Asn337, Asn420, Asn504 and Asn579 (for extended details on the glycosylation sites selection see chapter 3, section 3.1). This process was carried out on monomeric and dimeric extended EGFR extracellular domain in the apo and EGF-bound states as well as the tethered monomeric state of EGFR extracellular domain. Attached glycans were monosaccharides of N-Acetylglucosamine to resemble the glycosylation pattern of crystal structures and complex hybrid type oligosaccharides (GlcNAc-GlcNAc-man-man-man-GlcNAc), that is a common complex type in mammalian cells [30, 31]. Details of glycosidic linkages within the oligosaccharides are mentioned in Figure 2.7.

Figure 2.7. Construction of glycosylated EGFR model. Four subdomains of Monomeric EGFR extracellular domain are shown with I (blue), II (green), III (yellow), IV (magenta) and EGF (red) surface and cartoon representation. Glycosylated Asparagines are shown with atom name spheres and are labeled. Attached oligosaccharides are shown with atom name sticks. Structure of glycosylated EGFR presented here is the final snapshot of one of the simulations. The box shows the schematic presentation of the attached oligosaccharide and the glycosidic bond features specified according to GLYCAM06 force field.

2.7.6 Neutralizing and solvating the system

Charge calculation and solvation of glycosylated and unglycosylated EGFR extracellular domains were performed by using the tleap module of AMBER12 package. Na⁺ ions were added to the protein for neutralization followed by solvation of the protein with TIP3 water molecules.

2.7.7 Molecular Dynamics simulation protocol

All the molecular dynamics simulations were performed with the AMBER12 package [93]. The parameters and topology files were created using the tleap module of AMBER12. The Amber ff99SB force field was used to represent the protein, with parameters for the glycans taken from the GLYCAM06 force field [67]. All systems were minimized in a three-step process to relieve the systems of any steric clashes: the systems were first subject to 1000 steps of Steepest Descent/Conjugate minimization, holding the receptor-ligand fixed to relax the added waters/ions; this was followed by 1000 steps of Steepest Descent/Conjugate gradient minimization with restraints on the water molecules. Finally the whole system was subject to another 2000 steps of Steepest Descent/Conjugate minimizations, with no restraints. Then the systems were heated up to 300 K in 100 ps followed by equilibration of the whole complex in 250 ps. Finally production runs of 50 ns (extended dimeric state) and 100 ns (extended and tethered monomeric state) were carried out on each system using the isothermal–isobaric ensemble (NPT) scheme. The Particle Mesh Ewald method was applied to simulate the electrostatics. The SHAKE algorithm was used for constraining the bonds involving hydrogens, enabling a timestep of 2fs to be used in the simulations. The above cycle was carried out in triplicates for every system of interest. Description of the MD simulations carried out in this study can be found in Tables 2.3-2.5. Simulation details and analyses methods used are explained in the assigned methods section of each chapter.

Extended monomer			
MD Simulation description		Duration (ns)	No. of replicas
Fully glycosylated	Glycosylated with oligosaccharides EGFR-ed	100	3
	Glycosylated with oligosaccharides EGFR-ed bound to EGF	100	3
	Glycosylated with monosaccharides EGFR-ed bound to EGF	100	3
Unglycosylated	EGFR-ed	100	3
	EGFR-ed bound to EGF	100	3
aMD simulation description		Duration (ns)	No. of replicas
Fully glycosylated	Glycosylated with oligosaccharides EGFR-ed bound to EGF	100	3
Unglycosylated	EGFR-ed bound to EGF	100	3

Table 2.3. Description of MD simulations performed on the extended monomer of EGFR extracellular domain (EGFR-ed).

Tethered monomer			
MD Simulation description		Duration (ns)	No. of replicas
Fully glycosylated	Glycosylated with oligosaccharides EGFR-ed bound to EGF	100	3
Unglycosylated	EGFR-ed	100	3
	EGFR-ed bound to EGF	100	3
aMD simulation description		Duration (ns)	No. of replicas
Fully glycosylated	Glycosylated with oligosaccharides EGFR-ed bound to EGF	100	3
Unglycosylated	EGFR-ed bound to EGF	100	3

Table 2.4. Description of MD simulations performed on the tethered monomer of EGFR extracellular domain (EGFR-ed).

Extended dimer			
MD Simulation description		Duration (ns)	No. of replicas
Fully glycosylated	Glycosylated with oligosaccharides EGFR-ed	50	3
	Glycosylated with oligosaccharides EGFR-ed bound to EGF	50	3
Unglycosylated	EGFR-ed	50	3
	EGFR-ed bound to EGF	50	3

Table 2.5. Description of MD simulations performed on the extended dimer of EGFR extracellular domain.

Chapter 3. Role of N-glycosylation in EGFR ectodomain ligand binding

3.1 Introduction

EGFR is activated either by the binding of growth factor (ligand induced dimerization, see chapter 1, section 1.5.1) to stabilize its monomeric form in an extended state, followed by dimerization, or by the binding of the growth factor to a pre-formed dimer where each monomer is in an extended state (see chapter 1, section 1.5.2). The extracellular domain of EGFR is post-translationally modified extensively by the attachment of sugar (glycosyl) units to Asn residues and these are known to play various functional roles in the regulation of the receptor [30, 32, 33]. In addition, the EGFR is a target of therapeutic intervention including its inhibition by small molecules that target the intracellular kinase domain and by antibodies that target the extracellular region [3]. Hence it is likely that glycosylation of the extracellular domain will modulate the binding to EGFR of the growth factor, dimeric partner and antibody (we refer to these 3 as ligands) at the very least. Understanding the mechanistic effects of glycosylation on the interactions of EGFR with these ligands can thus shed light on plausible activation mechanisms and leads to efficient computational design of inhibitors. The EGFR extracellular domain has 10 Asparagines that are part of the motif that is thought to undergo glycosylation (Asn-x-Ser/Thr, see chapter 1, section 1.6.1). These Asparagine residues are: Asn104, Asn151, Asn172, Asn328, Asn337, Asn420, Asn504, Asn544, Asn579 and Asn599. However, although these Asparagines are part of the known glycosylation sequence, data on their modification in the literature is not consistent. In a study from the laboratory of Daisuke Kolhda, only four of these sites (Asn328, Asn337, Asn544 and Asn579) were reported to

be modified [94]. In a study from the laboratory of Elizabeth F. Hounsell, five other sites (Asn104, Asn151, Asn389, Asn420 and Asn504) were reported to be glycosylated [48]. Together these studies reported that 8 of the 10 likely Asparagines (except Asn172 and Asn599) with the known sequons are glycosylated. A separate study on fucosylation (addition of fucose to the oligosaccharide glycan) of the EGFR extracellular domain revealed that 9 Asn residues are glycosylated (Asn151, Asn328, Asn337, Asn389, Asn420, Asn504, Asn544, Asn579 and Asn599) [30]. There is another glycosylated Asparagine in the EGFR extracellular domain which is known to be an atypical glycosylation site. This Asparagine residue is Asn32 and is part of an Asn-Asn-Cys sequence [94] and has been reported to be glycosylated and fucosylated. Indeed, evidence from studies on N-glycosylation sites suggests that the consensus sequon is neither necessary nor sufficient for N-glycosylation [41]. Various factors are involved in the glycosylation of Asn including the role of Asn in local folding, the influence of regions proximal to Asn for the OTase, secondary structures, disulfide bonds etc. Role of glycosylation in dynamics of EGFR extracellular domain has recently attracted attention in molecular modeling studies. The group of David E. Shaw used glycosylation to investigate “negative binding cooperativity” of EGFR [52]. It was proposed that EGFR dimer rests on the cell membrane and asymmetrical interactions of one EGF with the membrane attenuates binding affinity of the other monomer to EGF and found that this occurred irrespective of whether the dimers were glycosylated or not [52]. The interactions of glycans with each other and with the protein surface was hinted at, but details remained unexplored. In another study from the group of Ilpo Vattulainen, the role of glycosylation in the arrangement of monomeric EGFR within the membrane and in

binding to antibodies C225 and 2E9 antibodies was investigated [55]. This study provided experimental validation for the simulation observations and in this regard is the first and most convincing study on the critical role of glycans in regulation of EGFR. The simulations show that glycosyl units mediate conformational rearrangements of the EGFR extracellular domain by interactions with the membrane. Biochemical binding assays show that glycosylation does not affect binding of C225 to EGFR monomer but increases binding affinity of 2E9 by up to 75% [55]. These observations were consistent with the findings from the simulations where the solvent accessibility changes obtained from the MD trajectory before and after glycosylation were noticeable for the 2E9 antibody [55]. These studies are now complemented with atomistic MD simulations in this chapter by exploring the effects of glycosylation on the binding of the extracellular domain of monomeric EGFR to ligands EGF, the dimeric partner EGFR and monoclonal antibodies C225, IMC-11F8, MEHD7946, EMD72000, GC1118 and 806 (these were chosen because the crystal structures of these ligands in complex with the monomeric form of EGFR are available). Considering the uncertainty in the literature about the glycosylation sites, we chose Asn residues that are glycosylated in the available crystal structures and include Asn32, Asn49, Asn151, Asn172, Asn328, Asn337, Asn420 and Asn504. In addition, Asn579 was also included because there is strong evidence that its glycosylation affects the regulation of EGFR [33].

3.2 Methods

3.2.1 Model building of glycosylated and unglycosylated EGFR

The PDB entry 3NJP [14] which represents the extended extracellular domain of EGFR (residues 1-614) in its homodimeric form bound to two molecules of the growth factor EGF was used as the initial structure from which chain A bound to a molecule of EGF was extracted. As the structure is not a full-length protein, neutral states of the C-termini is maintained by adding capping residue at the ends of the structure using the tleap module of the AMBER12 package [93]. Each monomer of EGFR extracellular domain contains 25 disulfide bonds while the ligand EGF contains 3. The associated Cysteines were covalently linked using the tleap module of the AMBER12 package. The ‘GLYCAM online glycoprotein builder’ [72] was used to attach the glycans to the extended monomers of the EGFR extracellular domain at residues Asn32, Asn49, Asn151, Asn172, Asn328, Asn337, Asn420, Asn504 and Asn579. The glycans used were complex oligosaccharides made up of mannose and N-Acetylglucosamine (GlcNAc-GlcNAc-man-man-man-GlcNAc). This protocol was also used for the addition of monosaccharides of N-Acetylglucosamine to the EGFR extracellular domain. The only difference is that in this construct Asn579 is not glycosylated as this site was not observed to be glycosylated in the crystal structures.

3.2.2 Molecular Dynamics simulations and analysis methods

All the molecular dynamics simulations were performed with the AMBER12 package software [93]. The systems constructed above were immersed in a box of TIP3 waters with at least a 10Å boundary around any atom and the systems were neutralized with 6 NA⁺ ions. The parameters and topology files were created using the tleap module of

AMBER12. The Amber ff99SB force field was used to represent the protein, with parameters for the glycans taken from the GLYCAM_06h force field [67]. All systems were minimized, equilibrated and subject to production MD in triplicates using the protocol outlined in Chapter 2 (section 2.7.7). All MD trajectories were analyzed using the ptraj module of the AMBER12 package and Bio3D [95]. All analyses (except the binding free-energy calculations and per-residue decompositions) were performed on the 300ns trajectory obtained from combining the triplicates of 100ns MD for each system. Binding free energy and per-residue decomposition calculations were carried out on the last 50ns of each simulation and the average values were used. Water density maps were calculated to explore the role of water in mediating interactions of EGFR with its ligands and were generated by using grid analysis of ptraj module of AMBER12. The grid cell dimension was set at $0.5\text{\AA}\times 0.5\text{\AA}\times 0.5\text{\AA}$ and the number of water molecules found in each grid cell was counted. Water density maps presented in XPLOR formatted counter file were visualized by VMD [96]. The cutoff isovalue was chosen to be 2.5 times of the value with no observed water clusters in the bulk. Volume estimates of the EGF cavity were carried out with the *trj_cavity* tool implemented within the GROMACS package [97]. The energies of interactions between EGF and EGFR-ed were computed using the Molecular Mechanics Generalized Born Surface Area (MMGBSA) method [98]. The single trajectory approach, whereby the difference between the total energy of the complex and the sum of the energies of EGFR-ed and EGF separately, extracted from each snapshot of the MD simulation, were used to derive the approximation to the binding free energy. The large size of the systems and the uncertainties in computations precluded the computation of the entropic term, thus the free energy of binding is approximated by the enthalpic

term. Contributions of residues to these energies were computed using the AMBER GB models [98]. All figures were made with VMD and PYMOL [96, 99].

3.3 Results

3.3.1 Dynamics of the glycosylated monomer of EGFR extracellular domain

The unglycosylated and glycosylated forms of monomeric EGFR extracellular domain will be referred to as EGFR-ed and gEGFR-ed respectively. Backbone RMSD plots show that the systems are clearly more stable during the MD when there is either EGF-bound or the system is glycosylated, or both (Figure 3.1). The RMSD plots of the backbones of individual subdomains (Appendix A, Figure 1-3) shows that the “instability” of the apo form arises largely from subdomain IV and to some extent from subdomain II. It is clear that the presence of EGF and sugars both result in an attenuation of the motions. Analysis of the fluctuations of the C α atoms again show that glycans result in reducing the flexibility of the receptor both in the apo and EGF-bound states (Figure 3.2.A-F). Interestingly, when this is decomposed into subdomain-wise fluctuations, it is clear that the effects of EGF are more complex. It can be seen in Figure 3.2.C-F that glycosylation does reduce fluctuations in all four subdomains while EGF does not, even though EGF interacts directly with subdomains I and III suggesting a complex global effect. With the glycosylated apo EGFR being the most affected, with a ~ 5 Å decrease in the overall fluctuations compared to the unglycosylated form (Figure 3.2.B), subdomains II and IV are the most affected subdomains in the apo glycosylated receptor. This is due to the highly flexible nature of the unliganded monomer which leads to a dramatic swinging movement of subdomain IV towards subdomain II (Appendix A, Figure 5) observed and has also been reported in μ s simulation studies of the same system [11]. The global

secondary structure of all subdomains is conserved upon glycosylation (Appendix A, Figures 6-9). However, glycosylation has a significant effect on the tertiary structure of monomeric EGFR-ed (Figure 3.3). Upon glycosylation, the two growth factor binding subdomains (I and III) are more distant, subdomain IV adopts an extended tertiary structure and the relative position of subdomains II and IV with respect to each other is the closest (RMSD value) to the conformation of the monomer in the extended active dimer crystal structure [14]. Given that these two subdomains form the EGF binding site, the volume of the site is explored next. The cavity between subdomains I and III increases by $\sim 1000 \text{ \AA}^3$ in the volume after the attachment of the glycans (Figure 3.4). This is paralleled by an increase in the distance between the centers of masses of subdomains I and III in the apo and EGF-bound glycosylated receptor (Figure 3.3.B). This increase appears to result from the interactions of the two glycans attached to subdomain I (Asn151) and subdomain III (Asn328) with each other (Figures 3.5-3.7) that keeps the two subdomains distant and prevent them from collapsing on to each other as was seen in the unglycosylated simulations (Figure 3.3.A) and in longer simulations of EGFR-ed carried out by others [11]. The increase in the distance between centers of masses of subdomains I and III and volume of the cavity between the two subdomains raises the issue of possible alterations in the hydration characteristics upon glycosylation. The water density maps of apo and EGF-bound states with and without glycosylation show that in both glycosylated complexes there are higher density of hydration waters around the two growth factor binding subdomains (Appendix A, Figure 10). These clusters are located in the proximity of the glycans. This is not surprising as the presence of the glycans introduces several polar groups which would attract the water molecules. Along with the

dynamics of EGFR-ed upon glycosylation, dynamics of the sugars was investigated. Simulations show that most glycans are highly flexible and protrude from the receptor surface towards the solvent during the simulations (Figure 3.5), except for two: attached to Asn151 (in subdomain I) and Asn328 (in subdomain III) as it was mentioned above. Most of the glycans tend to make highly persistent intramolecular hydrogen bonds mainly bridging polar groups between the monosaccharides. Intramolecular hydrogen bonds in the glycans attached to Asn32, Asn172, Asn328, Asn337, Asn420, Asn504 and Asn 579 appear to impart stability to the tertiary structure of the complex (Appendix A, Tables 1 and 2). This has been known for several glycoprotein complexes [29]. However, the glycans attached to Asn172 and Asn328 make some hydrogen bonds with EGFR-ed (residue Cys191, Gln193 and Asp323) probably contributing to the reduced flexibility of the monomeric EGFR-ed upon glycosylation (Appendix A, Tables 1 and 2). The glycans attached to Asn151 and Asn328 are the least flexible of all the glycans (Figure 3.5) and are stabilized by interactions with each other mediated by various intramolecular hydrogen bonds (Figures 3.6 and 3.7). However, these hydrogen bonds are not persistent as those formed within each oligosaccharide and between oligosaccharides and EGFR-ed. A series of formation and breakage of different hydrogen bonds was observed during the simulation showing that the two glycans (at Asn151 and Asn328) frequently interact with each other (Figures 3.6 and 3.7). This conformational arrangement is formed in the proximity of the EGF binding site (Figure 3.6).

Overall, it was shown that the apo and EGF-bound states of monomeric EGFR-ed are less flexible and the tertiary structure of the receptor is close to the extended active form upon glycosylation. The conformational arrangement between the glycans attached to Asn151

and Asn328 could affect EGF binding to EGFR-ed and could be the reason for the increase in the volume of the cavity between subdomains I and III as well as the reason for the large decrease in fluctuations that was observed in subdomain III of EGFR-ed upon glycosylation. Next the mechanism and energetics of the binding of EGF to the monomeric EGFR-ed was explored.

Figure 3.1. The backbone RMSD plots of the EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. A-C presents the values from three 100ns simulations of each system

Figure 3.2. A) Binding of three groups of ligands to the EGFR-ed monomer. Subdomains I-IV of EGFR-ed are shown in blue, green, yellow and magenta tube respectively. EGF is shown in red tube. Ca atoms of glycosylated Asn are presented with light blue spheres. Antibody binding epitopes and the dimeric interfaces are shown with purple and blue spheres respectively. B) Ca atoms fluctuations of the EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. The amino acid residues of subdomains I, II, III and IV of EGFR-ed are marked with blue, green, yellow and magenta transparent boxes respectively. Each trajectory was superimposed on to the first conformation of the corresponding simulation. C-F) The fluctuations of the Ca atoms of subdomains I-IV.

Figure 3.3. EGFR-ed subdomain orientations. A) Orientation of four subdomains of EGFR ectodomain. Subdomain I (blue), II (green), III (yellow), IV (magenta) and EGF (red) are shown in surface and cartoon representation; the oligosaccharides are shown in sticks. B) Distance plots between the centres of mass of the two growth factor binding subdomains (subdomains I and III) of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively.

Figure 3.4. Increased volume of the cavity between subdomains I and III upon glycosylation.

A) The cavity between subdomains I and III in EGFR-ed and gEGFR-ed is shown in black and green mesh respectively. EGFR-ed is shown in pink surface. B) The volume fluctuation plots for the EGFR-ed and gEGFR-ed are shown in black and green respectively.

Figure 3.5. A-B) Distance distribution plots between each oligosaccharide and its respective subdomain in the apo (A) and EGF-bound (B) states.

Figure 3.6. Interaction between the two glycans attached to subdomain I (Asn151) and subdomain III (Asn328) of EGFR-ed bound to EGF. Subdomains I, II, III, IV of EGFR-ed and EGF are shown in blue, green, yellow, magenta and red surfaces. Glycans attached to Asn151 and 328 are shown with atom name sticks. Glycosylated Asparagines are shown with atom name spheres. Boxes show the enlarged views of the final snapshots from first simulation (left box) and its first replicate (right box), presenting the hydrogen bonds formed between the glycans. Residue number and atom name of each monosaccharide involved in the hydrogen bond and the respective Asparagines are labeled.

Figure 3.7. Interaction between the two glycans attached to subdomain I (Asn151) and subdomain III (Asn328) of EGFR-ed in the apo state. Subdomains I, II, III, IV of EGFR-ed are shown in blue, green, yellow and magenta surfaces. Glycans attached to Asn151 and 328 are shown with atom name sticks. Glycosylated Asparagines are shown with atom name spheres. Boxes show the enlarged views of the final snapshots from first simulation (left box) and its first replicate (right box), presenting the hydrogen bonds formed between the glycans. Residue number and atom name of each monosaccharide involved in the hydrogen bond and the respective Asparagines are labeled.

3.3.2 N-glycosylation increases EGFR-EGF binding affinity

Accessible surface area of the residues in EGFR-ed that are known to interact with EGF (see chapter 1, section 1.4.2) were examined in the glycosylated and unglycosylated simulations. In the apo simulations the average solvent exposure of the EGF binding site decreases by 63 Å² after glycosylation and in the EGF-bound state a decrease of 10 Å² was observed in gEGFR-ed (Figure 3.8). It is clear that EGF does not occlude the binding site, since even when EGF is bound, the SASA of the binding site is only reduced by about 60-70%. This decrease is expected as the two glycans attached to Asn151 and Asn328 interact with each other close to the EGF binding site (see section 3.3.1). However, this decrease is not noticeable compare to the total surface area of the EGF binding site (~ 800 Å²). Given the limits of ASA in developing a deeper understanding of the differences in EGF binding upon glycosylation, the energetics of binding of EGF to EGFR as a function of glycosylation was calculated.

Figure 3.8. Average Solvent Accessible Surface Area (SASA) of the EGF binding site from simulations.

The energetics of the interactions between EGF and EGFR-ed was investigated by calculation of the apparent binding free energies using the MMGBSA approximations [98]. A set of control simulations was also carried out to examine the energetics of binding of EGF to EGFR when only the monosaccharide unit was attached to the Asn sidechains in EGFR-ed (refer to as mgEGFR-ed). The binding energies are: -88.1 kcal/mol, -91.3 kcal/mol and -94.3 kcal/mol for the binding of EGF to EGFR-ed, mgEDGR-ed and gEGFR-ed respectively (Figure 3.9 and Table 3.1) clearly suggesting that an increase in the length of the attached glycans from monosaccharide to oligosaccharide leads to increase in the binding free energy. This is due to the better interactions between the full length glycans leading to stronger reduction in the flexibility of the EGF binding site (Figure 3.10). Although the overall energetics are driven by van der Waals interactions, it is clear that electrostatic interactions are the discriminating feature as the desolvation penalty for the fully glycosylated system is minimal (from 73.9 kcal/mol in EGFR-ed to 68.8 kcal/mol and 67.1 kcal/mol in mgEGFR-ed and ogEGFR-ed). Next, the contribution of each residue in the binding free energy was calculated for subdomains I and III of EGFR-ed and gEGFR-ed residues located within 4 Å of EGF and for all the EGF residues bound to EGFR-ed and gEGFR-ed (Figures 3.11-3.16 and Appendix A Tables 3-6 and Figures 11-17). Only residues with at least kT (0.6 kcal/mol) increase in their total energy contribution to the binding free energy upon glycosylation are discussed here. In subdomain I of EGFR-ed Thr15 and Glu90 are the most favorably contributing residues upon glycosylation with 1.1 kcal/mol and ~1.6 kcal/mol increase in their total energy respectively (Figure 3.11). The trajectory of the EGF-EGFR complex shows that both in the absence and presence of the attached glycans Thr15 makes a

hydrogen bond with backbone Oxygen of Glu40 of EGF. However, upon glycosylation, the interaction is well maintained due to the decreased flexibility of the EGF binding site (Figure 3.12). A similar mechanism was observed for Glu90 making a more persistent salt bridge with Lys28 of EGF upon glycosylation. In subdomain III, Asp355, His409 and Phe412 with 3.4 kcal/mol, 1.12 kcal/mol and 0.6 kcal/mol increases are the most favorably contributing residues respectively (Figure 3.13). Asp355 forms a salt bridge with Arg41 of EGF that is known to be the determinant of EGFR binding to all growth factors [18]. In particular the electrostatic stabilization of the salt bridge is noticeable (Appendix A, Figure 14). Note that the 3.4 kcal/mol is the highest increase observed in the total energy among all residues of subdomain I and subdomain III of EGFR-ed. This increase occurs in the pair of residues forming the critical salt bridge required for bonding of all growth factors to EGFR. Several mutagenesis studies show that loss of this salt bridge decreases the binding affinity up to about 75% [17]. His409 makes a stronger hydrogen bond with the backbone oxygen of Arg45 of EGF upon glycosylation and Phe412 forms more favorable non-polar interaction with Leu47 of EGF (Figure 3.14). Arg353 shows ~1.8 kcal/mol decrease in its total energy upon glycosylation. The drop results from destabilization of the electrostatics (~0.08 kcal/mol; Appendix A, Tables 5 and 6 and 0.2 and 0.4 kcal/mol for non-polar solvation and vdw energies respectively). Visualization of the dynamics shows that Arg353 is more exposed to the solvent in the glycosylated system and its interaction with His16 of EGF is destabilized (Appendix A, Figure 15). However the overall increase in stabilization in EGF-EGFR interactions from other subdomain III residues (5.1 kcal/mol) compensate for this decrease. Among the EGF residues, Tyr44, Arg45, and Trp50 contribute the most favorably to the binding

energy (Figure 3.15) with 0.6 kcal/mol, 0.6 kcal/mol and 0.9 kcal/mol increase in their total energy upon glycosylation (Appendix A, Tables 7 and 8 and Figures 16 and 17). Tyr44 forms a strong hydrogen bond with His346 from subdomain III of EGFR-ed upon glycosylation. Arg45 forms a salt bridge with Gln384 of EGFR-ed; this results in a strong electrostatic contribution from Gln384 upon glycosylation (Figure 3.16). Simulations show that sidechains of Trp49 and Trp50 are engaged in stacking interactions upon glycosylation.

The pairs of EGF-EGFR interacting residues observed in the crystal structures (see chapter 1, section 1.4.2) were also investigated for the changes in their energetic components upon glycosylation. The hydrophobic interactions between subdomain I and EGF (Leu14 (EGFR)-Ile23 (EGF), Tyr45 (EGFR)-Met21 (EGF), Leu69 (EGFR)-Leu26 (EGF) and Leu98 (EGFR)-Leu26 (EGF)) do not show a change upon glycosylation (Figure 3.11). Hydrophobic interactions between subdomain III and EGF (Val350 (EGFR)-Leu15 (EGF) and Phe357 (EGFR)-Tyr13 (EGF) and Leu382 (EGFR)-Leu47 (EGF)) do not show a change upon glycosylation (Figure 3.13). However Phe412 (EGFR)-Leu47 (EGF) interaction is strengthened upon glycosylation (Figure 3.14). The hydrogen bonds Between EGF-EGFR (Gln16 (EGFR)-Asn32 (EGF), Gln384 (EGFR)-Gln43 (EGF) and Gln384 (EGFR)-Arg45 (EGF)) are all strengthened upon glycosylation (Figures 3.11 and 3.13). The critical salt bridge between Asp355 (EGFR)-Arg41 (EGF) is the most favorably contributing pair upon glycosylation. This salt bridge is well-maintained and present stabilized electrostatic contribution upon glycosylation (Figure 3.14).

Overall the above mentioned observations show that binding affinity of EGF to EGFR-ed is higher upon glycosylation of the receptor mainly due to the decreased flexibility of the complex and stabilizing electrostatic interactions. Both the growth factor binding subdomains contribute to this increase in favorable binding energy, with subdomain III contributing most (5.1 kcal/mol) compared to subdomain I (2.7 kcal/mol). It is worth mentioning that among all the above mentioned pair of interactions, none were altered upon glycosylation. However, the stabilizing effect of the glycans results in improved interactions between the pairs of interacting residues. The fact that glycosylation does not alter the nature of ligand-receptor interactions suggests that the role of glycosylation as a post-translational modification is to stabilize and maintain the biological function of the complex.

Testable hypothesis: To examine whether the two Asparagines (151 and 328) from subdomain I and III of EGFR-ed that interact with each other and stabilize the growth factor binding site are conserved amongst the four family members of the ErbB receptors and are part of the glycosylation sequon, the sequence alignment was generated (Figures 3.17 and 3.18). It shows that the Asn151 glycosylation site located in subdomain I is conserved between EGFR and ErbB4 (numbers are Asn156 and Asn156). The Asn328 glycosylation site located in subdomain III in EGFR and is not conserved among. However the neighboring glycosylation site Asn337 in EGFR is conserved between ErbB3 (Asn334) and ErbB4 (Asn333). ErbB2 has a sequon at Asn166 located in subdomain I (in the proximity of Asn151) and none in subdomain III (Figure 3.19). It is well known [3] that EGF binds only to EGFR and ErbB4. This could mean that the sugars at Asn151 and Asn328 gate the binding of EGF in EGFR and the sugars at Asn156 and

Asn333 gate the binding of EGF to ErbB4; however the structures of EGFR and ErbB4 extracellular domains (Figure 3.19) show that while glycosylation sites of Asn151 and Asn328 in EGFR are separated by 33.3 Å, the sites at Asn156 and Asn333 in ErbB4 are separated by 48.3 Å suggesting that the gating in ErbB4 may be more rate limiting in EGFR. It can be hypothesized that introducing sequons in ErbB2 at locations that are homologous to glycosylation sites in EGFR may enable ErbB2 to bind EGF. For this, the sequence 160PVD162 in ErbB2 could be mutated to 160NVS162 and 334AIT336 could be mutated to 334NIT336. Similarly mutation of 151DAE153 in ErbB3 to 151NAS153 could generate a sequon. This could speculate that the glycosylation of 151N and the glycosylation of 334N will together enhance EGF binding.

The desire to inhibit overexpressed EGFR as a cancer therapeutic strategy has seen several successes in the clinic as a result of the several monoclonal antibodies that have been designed to inhibit EGFR by binding to its extracellular domain as exogenous ligands. Since EGFR is heavily glycosylated in-vivo, it is crucial to examine the structural role of glycans in the antibody-EGFR interactions, something that has been reported [55] and yet whose detailed mechanisms remain unexplored. The ideal approach in this regard is to carry out atomistic simulations of each antibody-EGFR complex in the glycosylated and unglycosylated forms to obtain the free energy and mechanism of binding. However, such simulations were not possible due to the unavailability of the vast computational resources required. Thus this question was enquired by examining the the structural dynamics of EGFR in the presence/absence of glycosyl units. The simulations of apo and EGF-bound monomer in the glycosylated and unglycosylated forms that was mentioned earlier were used for calculations of solvent exposure of the antibody binding epitopes.

The results and interpretation of supportive/unsupportive role of glycans in each antibody binding to EGFR-ed are discussed in the following section.

Figure 3.9. The binding free energy of EGF to unglycosylated EGFR-ed, glycosylated with monosaccharides mgEGFR-ed and glycosylated with oligosaccharides ogEGFR-ed are shown in blue, pink and magenta respectively. Represented values of each component are the average values calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Table 1.1. The energetic components are: van der Waals (ΔE_{vdw}), electrostatic (ΔE_{ele}), electrostatic energy required for solvation (ΔG_{GB}), non-polar contribution to the solvation energy (ΔG_{SA}) and binding free energy (ΔG_{MMGBSA}).

	EGFR-ed + EGF	err	mgEGFR-ed + EGF	err	gEGFR-ed + EGF	err
First simulation						
ΔE ele	-123.1	0.8	-179.4	3.7	-161.7	2.8
ΔE vdw	-170.7	4.6	-135.6	0.8	-132.9	0.7
ΔGGB	232.0	4.1	244.4	3.5	236.3	2.5
ΔGSA	-21.2	0.1	-21.9	0.1	-22.3	0.1
ΔG MMGBSA	-83.0	0.8	-92.5	0.9	-80.6	0.8
First replicate						
ΔE ele	-191.2	9.8	-168.6	3.5	-217.3	2.9
ΔE vdw	-116.9	4.6	-144.0	0.9	-140.6	0.9
ΔGGB	242.8	11.4	242.7	3.4	270.1	2.6
ΔGSA	-19.7	0.8	-22.7	0.1	-23.0	0.1
ΔG MMGBSA	-84.9	3.4	-92.6	0.8	-110.8	1.0
Second replicate						
ΔE ele	-186.5	3.6	-167.9	3.3	-169.5	4.2
ΔE vdw	-134.8	0.9	-134.5	0.9	-142.3	0.8
ΔGGB	247.4	3.2	235.2	3.2	243.4	3.6
ΔGSA	-22.6	0.1	-21.8	0.1	-23.1	0.1
ΔG MMGBSA	-96.5	0.9	-89.0	0.8	-91.5	0.8
Average						
ΔE ele	-166.9	4.7	-172.0	3.5	-182.8	3.3
ΔE vdw	-140.8	3.3	-138.0	0.9	-138.6	0.8
ΔGGB	240.8	6.3	240.8	3.3	249.9	2.9
ΔGSA	-21.2	0.3	-22.1	0.1	-22.8	0.1
ΔG MMGBSA	-88.1	1.7	-91.4	0.8	-94.3	0.8

Table 3.1. The binding free energy of EGF to unglycosylated EGFR-ed, glycosylated with monosaccharides mgEGFR-ed and glycosylated with oligosaccharides ogEGFR-ed calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔGGB), non-polar contribution to the solvation energy (ΔGSA) and binding free energy (ΔG MMGBSA).

Figure 3.10. Distance distribution plots between the last monosaccharide at Asn151 and Asn328 of mgEGFR-ed and gEGFR-ed are shown with black and magenta respectively.

Figure 3.11. Subdomain I of EGFR-ed per-residue contribution to the binding free energy.

Only residues within 4 Å of EGF are shown. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 3 and 4.

Figure 3.12. Subdomain I – the most favorably contributing residues to the EGF-EGFR binding free energy. A) Subdomain I-IV of EGFR-ed and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The box shows the enlarged view of the hydrogen bond formed between Thr15(EGFR-ed)- Glu40(EGF) and the salt bridge between Glu90(EGFR-ed)-Lys28(EGF) with stick presentation. B and C show the distance between the interacting atoms of each pair of residues. Values from the EGF-EGFR-ed and EGF-gEGFR-ed complex are shown in blue and magenta respectively.

Figure 3.13. Subdomain III of EGFR-ed per-residue contribution to the binding free energy.

Only residues within 4 Å of EGF are shown. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 4 and 5.

Figure 3.14. Subdomain III - most favorably contributing residues to the EGF-EGFR binding free energy. A) Subdomain I-IV of EGFR-ed and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The box shows the enlarged view of the salt bridge formed between Asp355(EGFR-ed)- Arg41(EGF), hydrogen bond between His409(EGFR-ed)-Arg45(EGF) and non-polar interaction between Phe412(EGFR-ed)-Leu47(EGF) with stick presentation. B-D show the distance between the interacting atoms of each pair of residues. Values from the EGF-EGFR-ed and EGF-gEGFR-ed complex are shown in blue and magenta respectively.

Figure 3.15. EGF per-residue contribution to the binding free energy. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 7 and 8.

Figure 3.16. EGF - most favorably contributing residues to the EGF-EGFR binding free energy. A) Subdomain I-IV of EGFR-ed and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The box shows the enlarged view of the hydrogen bond formed between His346(EGFR-ed)- Tyr44(EGF), salt bridge between Gln384(EGFR-ed)-Arg45(EGF) and the stacking interaction between Trp49(EGF)-Trp50(EGF) with stick presentation. B and C show the distance between the interacting atoms of each pair of residues. D shows the distance between all non-hydrogen atoms of interacting pair. Values from the EGF-EGFR-ed and EGF-gEGFR-ed complex are shown in blue and magenta respectively.

Figure 3.17. Sequence alignment of the ErbB family members. Shades of blue in the alignment shown the least (light blue) and most (dark blue) conserved sites. Asn151 glycosylation site is marked with red boxes.

Figure 3.18. Sequence alignment of the ErbB family members. Shades of blue in the alignment shown the least (light blue) and most (dark blue) conserved sites. Asn328 and Asn337 glycosylation sites are marked with red boxes.

Figure 3.19. Structural alignment of the ErbB family members.

A) Subdomains I-III of EGFR-ed and EGF are shown with grey and red cartoon respectively. Asn151 and Asn328 are shown with atom name spheres and the two interacting glycans are shown with atom name sticks. B) Superimposition of the ErbB2 crystal structure (PDB 1N8Y) onto the EGFR glycosylated model. Subdomains I, II, III of ErbB2 and EGFR are shown with blue, green, yellow and grey cartoon respectively. Asn166 glycosylation site in subdomain I of ErbB2 is shown with atom name spheres. C) Superimposition of the ErbB3 crystal structure (PDB 3P11) onto the EGFR glycosylated model. Subdomains I, II, III of ErbB3 and EGFR are shown with blue, green, yellow and grey cartoon respectively. Asn334 glycosylation site in subdomain III of ErbB3 is shown with atom name spheres. D) Superimposition of the ErbB4 crystal structure (PDB 3U7U) onto the EGFR glycosylated model. Subdomains I, II, III of ErbB4 and EGFR are shown with blue, green, yellow and grey cartoon respectively. Asn156 glycosylation site in subdomain I and Asn333 glycosylation site in subdomain III of ErbB4 are shown with atom name spheres.

3.3.3 N-glycosylation contribution to EGFR-antibody interactions

Next the effect of glycosylation on EGFR-ed and its interactions with the second class of ligands, antibodies, was examined. Several antibodies have been used to target EGFR-ed (see Table 1.1). These include mAbC225 (Cetuximab/Erbitux) which blocks ligand binding to EGFR-ed [20], Matuzumab (EMD72000) which inhibits EGFR by blocking dimerization [22] and mAb806 which targets a cryptic pocket in the proximity of the dimerization arm [12, 100]. To get an insight into the role of glycosylation on EGFR-antibody interactions, the available crystal structures of EGFR-ed bound to antibodies [20-24] were analyzed in the context of simulations presented here. While there is only one crystal structure of mAb806 bound to the epitope peptide [100], a previously generated model of this antibody bound to the EGFR-ed was used here [12]. These antibodies were divided into two groups; antibodies which target subdomain III of EGFR-ed (C225, IMC-11F8, EMD72000, GC1118, MEHD7954A) and the antibody which targets subdomain II of EGFR-ed (mAb806). To examine the likely effects of glycosylation on antibody binding, the exposure (solvent accessible areas) of the regions of interactions of each antibody on EGFR-ed (epitope) were analyzed. The solvent exposure in the in the crystal structure of each antibody-EGFR complex was compared to the initial crystal structures used in our simulations and the changes (if any) during the simulations of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed (Figure 3.20 and Appendix A, Figures 20-24). It is worth mentioning that in the initial crystal structure used here (PDB entry: 3NJP), the binding eptiopes of EMD72000, GC1118 and mAb806 are solvent exposed. The epitopes of C225 and IMC-11F8 are exposed for all residues except Gln384, Phe412, Val417, Ile438 and Ser440, but these also become exposed

during the course of the simulations. In the MEHD7946 binding epitope, Gln462 is not accessible and remains buried in the simulations (Appendix A, Figure 21). As a general trend for all of the antibodies presented here the epitopes are either reduced (but only about 5%) in exposure in the presence of sugars (C225/ MC-11F8 and GC1118), or unaffected (MEHD7946, EMD7200 within error); the only one for which the epitope appears to get more exposed (only about 5%) upon glycosylation is that of mAb806. It is clear that the presence of EGF does not really affect the exposure (the small changes visible in Figure 3.20 are well within the fluctuations). Visual inspection of the simulations showed that even the glycans that lay in the vicinity of the epitopes do not appear to affect the epitopes exposure (for example in Appendix A, Figures 18 and 19 it can be seen that while the sugar at Asn579 is highly flexible, nevertheless it does not cover the epitopes of C225/ MC-11F8). This is in agreement with experimental assays showing that glycosylation does not affect Cetuximab binding affinity to EGFR [55]. Of course the sugar at Asn579 was hypothesized to interact with the membrane [46, 101] which is not included in our models. This may also begin to explain why it is that antibodies may be raised by presenting the epitope regions as peptides where there are no sugars, and yet these antibodies can bind to the full receptors which are glycosylated resulting in the conclusion that the sugars in general (see below for mAb806) do not affect antibody binding. It is of course possible that the sugars can interact with and enhance the binding of some antibodies. Long MD simulations will be needed to examine this in greater detail but they are very expensive and hence not part of this thesis. The antibody mAb806 targets a cryptic pocket in subdomain II and is known to inhibit wild type EGFR in cancerous cells [12, 100]; it is hypothesized that this cryptic pocket may become

exposed in an intermediate state [12]. The crystal structure of mAb806 had been computationally docked to the untethered and tethered EGFR-ed and analyzed with MD simulations, suggesting more favorable binding to the untethered state [12]. Interestingly, the interplay of glycans on mAb806 binding to EGFR-ed seems to be different from other mentioned antibodies (Appendix A, Figure 24). Among the seven exposed residues of 806 binding epitope, Glu293, Glu295 and Arg300 show an increase in their solvent accessibility in the glycosylated apo state. The increase in the average SASA values are: Glu293 ($12\text{\AA}^2$), Glu295 ($15\text{\AA}^2$) and Arg300 ($18\text{\AA}^2$). The overall 806 binding epitope solvent exposure increases by about $30\text{\AA}^2$ upon glycosylation of the apo EGFR-ed (Figure 3.20). Although this increase is only about 5%, nevertheless, it is interesting that this is the only epitope that undergoes an increase upon glycosylation. Dynamics of the glycans in the proximity of mAb806 shows that the oligosaccharide attached to Asn337 stabilizes a loop (resid 305-309) close to this glycosylation site (Appendix A, Figure 19). This loop is located next to the antibody binding epitope (resid 287-302, Figure 3.21). Comparison between the orientation of the loop and the epitope in the model of mAb806 bound to EGFR-ed with the orientations in the unglycosylated and glycosylated EGFR-ed, shows that while the epitope is more exposed upon glycosylation, nevertheless the conformation is quite different from that adopted in the presence of the antibody (Figure 3.21). However this arises because there are induced conformational changes in the receptor upon antibody binding [12] and is beyond the scope of the current simulations. Considering that the glycan attached to Asn337 is a common glycosylation site and that this oligosaccharide protrudes from the protein (Figure 3.5), simulations presented here also suggests the possibility that in the dominant dimeric form of EGFR-ed, the glycan

attached to Asn337 will interact with the other glycan attached to Asn337 from the counterpart monomer (see Figure 3.22 B in section 3.3.4). It is possible that this interaction could result in the exposure of the mAb806 cryptic epitope. Currently coarse grained simulations exploring this are underway.

The third class of ligands that interact with EGFR-ed is EGFR-ed itself (or the monomer of one of the other three members of the family i.e. ErbB2, ErbB3 and ErbB4), as they exist in the active state, largely as dimers (and higher order oligomers [7]). The two monomers interact via the dimeric interface [14]. The role of glycosylation in binding of the dimeric partner to EGFR-ed was investigated, as in the case of antibody binding, by calculation of the average solvent exposure of the dimeric interface in the glycosylated and unglycosylated monomers in the apo and EGF-bound states.

Figure 3.21. Conformational changes of mAb806 binding epitope in the glycosylated EGFR.

A) Computationally docked model of mAb806 to the extended monomer to EGFR-ed. Subdomains I, II, III, IV of EGFR-ed and mAb806 are shown with blue, green, yellow, magenta and grey cartoon respectively. The binding epitope (residue 287-302) and the neighboring loop (residue 305-309) affected by the glycan attached to Asn337 are shown in red and purple tube respectively. B and C show the orientation of the binding epitope and the loop at the end of glycosylated apo and glycosylated EGF-bound simulations respectively.

3.3.4 N-glycosylation maintains the dimeric interfaces of EGFR-ed

EGFR-ed contains two dimeric interfaces: the dimerization arm located in subdomain II and the tip of subdomain IV (Figure 3.22). These two interfaces are critical features of the dimeric conformations that need to be formed and maintained. Subdomain II contains the dimerization arm which interacts with the dimerization arm from the partner monomer and forms the back to back dimeric configuration of EGFR-eds. Subdomain II has been divided into eight modules [13]. In the simulations presented here, among the eight modules only modules 1, 2 and 8 undergo a decrease in their solvent exposure upon glycosylation (Appendix A, Figure 25). Although module1 (resid 166 to 186) contains the glycosylated Asn172 and is partly stabilized by the attachment of sugars, Module2 (resid 187-207) shows less solvent exposure and higher stability after glycosylation. Superposed structures of the glycan attached to Asn172 from the simulations shows that this glycan explores the surface of module2 extensively (Figure 3.22). However this is anchored by long-lived hydrogen bonds between the attached oligosaccharide and Cys191 and Gln193 of module2. It is clear from Figure 3.22.B that these sugars are likely to interact with the homologous sugars in the dimeric partner and stabilize the dimer. Module8 is the other module which shows reduced exposure upon glycosylation. This module is a loop consisting of only five amino acids (resid 305-309). This loop is dramatically stabilized by the glycan attached to Asn337. Again it is apparent from Figure 3.22.B that these sugars will interact with homologous sugars from the dimeric partner and further stabilize the dimer. In addition, these sugars also appear to stabilize the conformation of the dimerization arm, thereby further strengthening the monomer-monomer interactions in the dimer. The other parts of subdomain II, i.e. modules 3, 4, 6 and 7 do not undergo a

noticeable change in solvent exposure upon glycosylation during the timescale of the simulations presented here (Appendix A, Figure 25). The other major dimeric interface is the tip of subdomain IV (Figure 3.22.A). This part connects EGFR-ed to the transmembrane helix and the intracellular domain and may interact with the membrane lipids. This part contains one site of glycosylation, Asn579, and is less exposed to solvent upon glycosylation (Figure 3.22.C). Overall, the glycans attached to Asn172 and Asn337 appear to stabilize the dimeric interface and indeed likely enhance the dimerization by interacting with the dimeric partner through the protein or homologous sugars or both. It is likely that this stabilization is what accounts for the presence of dimers in the absence of growth factors.

3.4 Discussion

Since the discovery of the Epidermal Growth Factor Receptor three decades ago, understanding its activation and function has been riddled with complex discoveries [3]. EGFR binds to growth factors on the cell surface, which leads to activation of the signaling cascade in the cytosol [3]. Overexpression or mutation of EGFR is associated with multiple cancers and thus makes EGFR an important target for cancer therapy. EGFR was the first receptor against which antibodies were successfully raised with a therapeutic outcome. To date various clinical antibodies and small molecules have been designed to inhibit EGFR by targeting the extracellular domain or the intracellular kinase domain [20-22]. EGFR activation is believed to be regulated by a “ligand-induced” process where the binding of the growth factor to subdomains I and III of EGFR-ed untethers the monomer of EGFR and exposes the dimerization arm. The dimerization arm then interacts with a similarly exposed dimerization arm on another EGFR monomer, resulting in a back-to-back active dimer [25]. Recent evidence suggests that EGFR also exists as pre-formed ligand free dimers and the binding of growth factor to EGFR-ed stabilizes the dimeric form and induces a rotation in the transmembrane domains which result in the formation of asymmetric kinase domains followed by activation [28, 102]. In both of the above mentioned mechanisms, the binding of growth factor and the other interacting monomer are crucial aspects of EGFR activation. Of course to understand the interactions of EGFR with EGF, with other receptors, and with antibodies, details of the chemical nature of EGFR (post-translational modifications) are needed. N-glycosylation is one of the most important post-translational modifications that occur in the EGFR-ed. These attached N-glycans account for nearly 40 KDa of the total 180 kDa mass of EGFR

[33]. Several studies have shown the importance of glycosylation of EGFR-ed on the conformational dynamics of EGFR and on downstream signaling effects [47, 103], including the complex interplay between glycosphingolipids of the cell membrane and the glycosyl units on EGFR [101]. Experimental and computational studies have shown unambiguously that glycosylation is required for the integrity and stability of the protein [32, 33, 55]. The simulations presented here show that the monomeric form of EGFR is less flexible upon glycosylation. A combination of hydrogen bonding, non-bonded interactions and intramolecular contacts contribute to this stability. Here effect of N-glycosylation on EGFR-ed binding to growth hormone, antibodies and another ErbB monomer (dimeric partner) was investigated. From the different growth factors, EGF was chosen as it is the most common stimulator of EGFR. Multiple MD simulations were carried out in the unglycosylated and glycosylated forms of apo and EGF-bound states. It was shown that interactions of two glycans attached to Asn151 and Asn328 with each other stabilizes the EGF binding site and leads to stronger electrostatic interactions between EGF and EGFR-ed. The conformational arrangement between the two glycans is supported by persistent hydrogen bonds within the glycans and between the glycans and the receptor. Role of the glycans attached to growth factor binding subdomains in regulation of EGFR is suggested by several biochemical studies. These findings suggest that Asn151 (in subdomain I) and Asn328 (in subdomain III) are often glycosylated [103] and may be important for regulating the conformation and function of EGFR. In a separate study, mutations of all glycosylation sites in subdomain III suggested that a combination of N-glycans in subdomain III might be crucial for maintaining the functional conformation of the receptor [32]. In the current study presented in this chapter,

a specific mechanism for the role of glycans in binding of EGF to EGFR-ed was observed. It was shown that in the simulations of EGF-gEGFR-ed, interactions of the two glycans at Asn151 (in subdomain I) and Asn328 (in subdomain III) stabilizes the EGF binding site, resulting in robust interactions between EGF and EGFR (Figures 3.11-3.13 and 3.15) when compared to the unglycosylated EGF-EGFR-ed complex simulations. Sequence alignment of all the ErbB receptors showed that Asn151 and Asn328 glycosylation sites are conserved among EGFR and ErbB4, and both of these receptors are known to be activated by binding to EGF (Figures 1.1 and 3.19). In ErbB3, which has not been reported to bind EGF, there is a conserved glycosylation site only in subdomain III. Based on the observations of this study, the glycan attached to subdomain III contributes the most in stabilizing the growth factor binding site and strengthens the EGF-EGFR interactions. Thus the presence of one conserved glycosylation site in subdomain III could play a role in stabilizing the growth factor binding site in ErbB3. This suggests that the interaction between the two glycans attached to the growth factor binding subdomains (I and III) could be the regulating mechanism for binding of all growth factors to ErbB receptors (Figures 3.17-3.19). The more favorable EGF-EGFR binding free energy upon glycosylation that was observed here is driven by favorable electrostatic interactions between EGF and EGFR-ed. The stronger non-bonded interactions between EGF-EGFR-ed results from the stability of the EGF binding site induced by the two glycans attached to Asn151 and Asn328 interacting with each other. It was shown that the interaction between the two glycans is supported by several intramolecular hydrogen bonds within each glycan. The estimation of the EGF-EGFR-ed binding free energy based on the interface generated by PDBePISA [104] supports these findings. These

computations show that the binding energy of EGF to EGFR-ed in the structures obtained at the end of each MD simulation is more favorable upon glycosylation (average $\Delta G = -4.2$ kcal/mol) compared to the unglycosylated form (average $\Delta G = -2.9$ kcal/mol) (Appendix A Table 9). Again these calculations show that glycosylation induces stronger non-bonded interactions between EGF and EGFR-ed. We hypothesize that this effect can be generalized and that glycosylation likely results in increased binding affinity of all growth factors sharing the EGF-like motif [18]. The role of glycosylation on the other EGFR-ligand interaction, that with antibody binding was investigated by analyzing the available structural data of bound antibodies to EGFR. It was shown that glycosylation does not alter antibody binding to subdomain III of EGFR-ed. It is only in the case of the antibody bound to subdomain II of EGFR-ed (the antibody mAb806) that the mechanism seems to be different and the binding epitope becomes more exposed as a result of the presence of the glycan attached to Asn337; the glycan protrudes from the protein surface and triggers the exposure of the cryptic binding epitope. When we examine the dimer simulations (see chapter 5, section 5.3.3), we find that the interaction of two glycans attached to Asn337 in the dimeric form results in the exposure of this epitope to a higher extent. This arrangement could be disturbed in the mutants EGFR and may partly explain the inhibitory effects of mAb806 only on wild type EGFR. Finally, the interactions of the third class of ligands, i.e. the dimeric partner of EGFR, show that glycosylation of EGFR-ed results in the maintenance of the dimeric conformation of the extracellular domain regardless of the presence of EGF. EGFR is also known to form higher order multimers; a recent study combined computational and biochemical assays to hypothesize the region of EGFR-ed involved in multimerization [7]. This region is located in subdomain IV (resid

526-556) and contains three glycosylation sites (Asn544, Asn579 and Asn599). The solvent exposure of this region from the simulations presented here showed that this region's exposure does not change after glycosylation in the EGF-bound state and decreases in the glycosylated apo state (Appendix A, Figures 26 and 27). With the combination of glycosylated Asparagines presented here, this region's exposure does not alter or decrease and if involved in tetramerization and hence multimerization is unlikely to be affected by the glycans; however this may differ upon glycosylation of Asn544 and Asn599. Together, the simulations and analyses presented in this chapter, shed light on the effects of glycosylation on growth factor binding, antibody interactions and dimerization of EGFR-ed. The findings could be generalized to the effects of glycosylation of Asparagine residues conserved across the family of ErbB receptors.

3.5 Conclusions

Glycosidically bonded sugars on EGFR-ed play critical roles in the binding of growth factors, monoclonal antibodies and the dimeric partners. N-glycosylation favors growth factor binding to EGFR through a combination of hydrogen bonding, non-bonded interactions and several intramolecular contacts, plays distinct roles in the binding of antibodies to spatially separated epitopes of EGFR-ed, and maintains the dimeric interfaces to stabilize the functional dimeric states of EGFR.

Chapter 4. Glycosylation modulates the tethered-extended equilibrium of the EGFR extracellular domain

4.1 Introduction

Biological functions of EGFR are dependent on receptor activation. EGFR activation starts with conformational changes in the extracellular domain. Two models were proposed to explain these changes; the “ligand-induced dimerization” model and the “preformed dimer” model. The first model proposes that the EGFR monomer is initially tethered by subdomains II and IV through extensive intramolecular contacts. After EGF binds to the monomer, rearrangement of subdomains I, II and III occur, the dimerization arm becomes unburied and interacts with the dimerization arm of a neighboring monomer resulting in a back-to-back dimer of the two extracellular domains [14, 17]. This tethered-extended transition is characterized by a large conformational change and was believed to be induced by EGF binding to the extracellular domain. The latter model suggests that EGFR exists as a dimer even in the absence of ligands [28, 102]. This dimer is flexible and is stabilized by binding of EGF to the extracellular domains. The stabilized extracellular dimers undergo twist and turn motions that are believed to induce transmembrane domain rotation and consequently a rearrangement of the intracellular kinase domains takes place, which finally results in receptor activation [28, 102]. Although both models are supported by biochemical assays and 3D structures, they were proposed without accounting for the effects of post-translational modifications of EGFR, most importantly glycosylation. Glycosylation of EGFR is known to play various functional roles in the regulation of the receptor [32, 33]. Hence, its role in modulation of the tethered monomeric/extended dimeric structure could shed light on the most

physiologically plausible activation model and the accompanying structural dynamics of the extracellular domain. The role of glycosylation in the “preformed dimer” activation model is investigated in chapter 5. In the current chapter, a combination of computational modeling and analysis techniques is used to investigate the role of glycosylation in “ligand-induced dimerization” of the EGFR extracellular domain. Based on the X-ray crystal structures of the two extreme ends of this transition [14, 25] it was also suggested that EGF binding may bias the tethered-extended equilibrium towards the extended conformation rather than allosterically removing the tethering intramolecular interactions between subdomains II and IV [25]. This means that both tethered and extended monomeric conformations are sampled by EGFR in vivo. However, growth factor binding requires interactions with subdomains I and III simultaneously which is spatially possible only in the extended conformation. Growth factor binding to this extended conformation biases the equilibration towards one extreme end and facilitates homo/hetero dimerization. Similar to other transmembrane receptors, EGFR is only translocated to the cell surface after the addition of glycans in the ER. Thus, whether the tethered-extended communication is regulated by transition (removal of subdomains II/IV tethering interactions) or equilibration (biasing the conformational population towards the extended conformation), glycosylation likely plays a structural role in this process. Several experimental assays support this assumption by investigating point mutations of various glycosylation sites [32, 33]. The observations from mutations at Asn579 resulting in increased phosphorylation suggest a functional role for the glycan attached to this residue in strengthening the tethering interactions [33]. Asn579 is located within the inhibitory loop II of subdomain IV that is a tethering interaction site. Another example is the study

by M.Takahashi et al. which suggested that deletion of all N-glycans in subdomain III reduces cell surface expression of ErbB3 [32]. They also suggested that in the Asn418 mutants of ErbB3 (Asn420 in EGFR), the loss of N-glycan results in growth factor-independent dimerization. In this chapter, atomistic accelerated Molecular Dynamics (aMD) simulations were used to examine the dynamics of the two extreme ends of this conformational change (the tethered monomer and the extended monomer), both in the absence and presence of glycosylation, to examine the role of glycans in maintaining monomeric EGFR in its tethered/extended state.

4.2 Materials and methods

4.2.1 Model building of glycosylated and unglycosylated monomeric EGFR in the tethered and untethered states

PDB entry 3NJP [14] was used as the initial structure for the extended state. The full-length monomeric ectodomain of EGFR (residues 1-614) bound to EGF was taken from the complex. 1NQL [25] has been chosen as EGFR in the tethered state (residues 3-614); EGF is bound to this crystal structure. 3NJP was capped only at its C-terminus while 1NQL was capped at both ends using the tleap module of the AMBER12 package [93]. The 1NQL structure was explored with the first two residues missing as this represented the crystallographic conformation. Each monomer of the EGFR extracellular domain contains 25 disulfide bonds while the ligand EGF contains 3. The associated Cysteines were covalently linked using the tleap module of the AMBER12 package. The ‘GLYCAM online glycoprotein builder’ [72] was used to attach the glycans to the extended and the tethered monomers of the EGFR extracellular domain at Asn32, Asn49, Asn151, Asn172, Asn328, Asn337, Asn420, Asn504 and Asn579. The glycans used were

GlcNAc-GlcNAc-man-man-man-GlcNAc (For the extended details see chapter 3, section 3.1). The systems were immersed in a box of TIP3 waters with at least a 10Å boundary around any atom and the systems were neutralized with counter ions.

4.2.2 Accelerated Molecular Dynamics simulations

All the molecular dynamics (MD) simulations were performed with the AMBER12 package [93]. The parameters and topology files were created using the tleap module of AMBER12. The Amber ff99SB force field was used to represent the protein, with parameters for the glycans taken from the GLYCAM_06 force field [67]. All systems were minimized and equilibrated using the protocol outlined in Chapter 2 (section 2.7.7). Given the large size of the systems and the associated conformational change, the long time scale classical MD simulations required would be prohibitively expensive, hence it was decided to explore the transition using aMD. Four sets of 100ns aMD simulations, each in triplicate, using the isothermal–isobaric ensemble (NPT), were carried out on monomeric EGFR extracellular domain bound to EGF in the following states: unglycosylated extended, glycosylated extended, unglycosylated tethered and glycosylated tethered. The average total and dihedral energies obtained from a 100ns classical MD of the extended and tethered monomers were used for calculations of the boost energy (E) and acceleration parameter (α) for the aMD simulations. A dual boost in the total and the torsional energy has been added to all the systems using the protocol suggested in the AMBER12 manual and which has been used in several studies [105]. The SHAKE algorithm was used for constraining the bonds involving hydrogens, enabling a timestep of 2fs to be used in the simulations.

4.2.3 Principal Component Analysis

Two sets of PCA were performed: on the combined trajectory of 600ns aMD (glycosylated 300ns and unglycosylated 300ns) simulations of the extended monomer and on the combined trajectory of 600ns aMD (glycosylated 300ns and unglycosylated 300ns) simulations of the tethered monomer. All trajectories were fit on to the starting structures. The crystal structures of extended (PDB ID 3NJP) and tethered (PDB ID 1NQL) states were projected onto the PC space as the reference points for each analysis. Estimation of the overlapping conformations between the two states was carried out by projection of the combined trajectories of the tethered and extended monomers into the PC space of each other. The PCA was carried out on the coordinates of the CA atoms of the EGFR extracellular domain (residues 3-614); in order to compare the 1-614 extended state with the 3-614 tethered state, we used only the 3-614 residues of both systems and EGF and the sugar units were excluded from the actual analysis of PCA although their effects would be implicit in the dynamics being analyzed. GROMACS 5.1.1 [97] was used to construct the covariance matrices, the diagonalization and the projections of eigenvalues along the eigenvectors. Superimposition of the structures onto the reference points was performed by structural alignment in VMD [96]]. Cross-correlations of the motions of the CA atoms were carried out using the Bio3D package [95]. PYMOL was used for the visualization of the correlated and anti-correlated motions [99]. RMSD, RMSF, distance distribution, and dynamic cross-correlation analysis was performed on the 300ns trajectory obtained from combining triplicates of 100ns aMD for each system.

4.3 Results

4.3.1 Dynamics of glycosylated EGFR in the untethered state

The EGFR extracellular domain and the glycosylated EGFR extracellular domain will be referred to as EGFR-ed and gEGFR-ed. Untethered and extended terms are used synonymously. The C α RMSD plots of the untethered glycosylated and unglycosylated monomer shows large scale changes in both states of the monomer in its untethered form (Figure 4.1 and Appendix B Figures 1 and 2). The dramatic increase of the RMSD values in the last 100ns of the unglycosylated monomer simulation (Appendix B Figure 2) is due to the shift towards a more compact conformation that was expected to be explored by implementing aMD (compare with Figure 3.1 showing the RMSD plots of the extended monomer from cMD simulations in chapter 3). At the level of subdomains I-IV of the extracellular domain, a similar dynamical pattern was observed (Figure 4.2.A-D and Appendix B Figures 3 and 4). The apparent paradox between Figure 4.1 and Fig 4.2 where all subdomains move less than the overall movement arises from the superposition of each subdomain individually only in Figure 4.2. Subdomains I and III are the two growth factor binding subdomains. These subdomains move less due to the presence of the EGF in the simulations and yet a noticeable decrease was observed in both subdomains I and III upon glycosylation. Subdomains II and IV generally are more mobile. These two subdomains contain the two dimeric interfaces of EGFR-ed. As the simulations were carried out in the monomeric state this mobility is not surprising. However both of these subdomains are less mobile upon glycosylation. The C α fluctuations of untethered monomer shows an overall decrease of about 10 Å in the gEGFR-ed compare to the EGFR-ed (Figure 4.3). This is consistent with the observed

RMSD values and shows the role of the bonded glycans in reducing the receptor's flexibility. It is also clear that the changes that are captured are largely inter-domain motions as we can see that there is no significant effect on the secondary structures (Appendix B Figures 5-8). The aim of this study is to implement aMD as an enhanced sampling technique to capture the tethered-extended transition in the glycosylated and unglycosylated forms of EGFR-ed. Having mentioned the decreased flexibility of the extended gEGFR-ed and the noticeable increase in the flexibility of the EGFR-ed, PCA analysis was used to investigate the plausible transition towards the tethered state.

Figure 4.1. The C α RMSD plots of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. Each trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation. The data for other two replicates can be found Appendix B Figures 1 and 2.

Figure 4.2. The C α RMSD plots of subdomains I-IV (A-D) of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. Each trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation. The data for other two replicates can be found Appendix B Figures 3 and 4.

Figure 4.3. The C α RMSF plots of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The amino acid residues of subdomains I, II, III and IV of EGFR-ed and EGF are marked on the x axis with blue, green, yellow, magenta and red respectively. Combined trajectory of 300ns was aligned on the starting crystal structure PDB 3NJP.

The percentage of variance captured by PC1 and PC2 of the extended monomer are 60.5% and 12.7% respectively (Figure 4.4). The PC1 and PC2 projections show that the fluctuations of the untethered monomer are dampened upon glycosylation (Figure 4.5). The narrow range of sampled conformations around the initial structure supports the suggested stabilizing role of the glycans. Projection of the untethered (3NJP) and tethered (1NQL) crystal structures onto the PC space shows that the glycosylated untethered EGFR remains close to the untethered crystal structure across the simulations (Figure 4.5). RMSD values for the two extreme structures of PC1 and PC2 from the untethered crystal structure are: (11.4 Å and 7.06 Å) and (7.6 and 9.4 Å) respectively (Figure 4.6). These values are higher (19.1 Å and 9.7 Å) and (11.1 Å and 10.5 Å) in the unglycosylated

monomer. As is shown in Figure 4.5, the structure of the unglycosylated monomer at one extreme of PC1 adopts a compact conformation with subdomain IV being very close to the dimerization arm of subdomain II. This is the key feature of the tethered state that is obtained by the interaction of subdomains II and IV to shape the tethered conformation. The overall RMSD values of the extended monomer from the tethered crystal structure are almost similar in the glycosylated and unglycosylated forms. These values are: untethered EGFR-ed (24.3 Å) and untethered gEGFR-ed (24.5 Å). This suggests that the extended state samples conformations that are not compact and can be observed in Figure 4.5 by comparing the size of the two clusters of blue spots (small cluster to the left and the large diffuse cluster in the centre) formed close to the projection of untethered and tethered crystal structures. The structures at the extremes of PC2 of both untethered EGFR-ed and gEGFR-ed mainly represent the flexibility of subdomain IV in the dynamics of monomeric EGFR. The PC2 extreme structure of untethered EGFR-ed is more flexible compared to the glycosylated monomer.

Figure 4.4. Percentage of variance captured by each principal component from aMD simulations of glycosylated and unglycosylated untethered monomer.

Visualization of the combined trajectories of EGFR-ed and gEGFR-ed shows that the average difference in the RMSD values from the untethered state crystal structure is 3.5 Å higher in the unglycosylated monomer (Figures 4.7 and 4.8). This increase is mainly observed in the second 50ns of each simulation in the dynamical trajectory. Note that the combined 300ns trajectory was obtained by merging the three replicates of 100ns simulations for each of the glycosylated and unglycosylated monomers as explained in section 4.2.3. Thus, 300ns here simply represents the sampling time scale and not the time-dependent evolution of the dynamics.

Figure 4.5. Projection of the untethered structures along PC1 and PC2. Structures obtained from aMD simulations of glycosylated and unglycosylated untethered monomer are shown with magenta and blue spots respectively. Projection of the untethered (PDB ID 3NJP) and tethered (PDB ID 1NQL) states crystal structures are shown with green and red spots respectively. Superimposition of PC1 and PC2 extreme structures on the untethered state crystal structure from the glycosylated and unglycosylated simulations are shown with magenta and blue tubes respectively. The untethered (PDB ID 3NJP) and tethered (PDB ID 1NQL) states crystal structures are shown with green and red tubes respectively. Black arrows show the location of each conformation in the PC space.

Figure 4.6. RMSD values calculated from the glycosylated and unglycosylated untethered monomer simulations. 3NJP and 1NQL are PDB IDs for the untethered and tethered crystal structures.

Projection of the tethered state simulation trajectory in the PC space of the extended state simulations shows that the tethered state ensemble contains some of the states sampled in PC1 in the untethered state, but the variance in fluctuations reduces only a little upon glycosylation, compared to the untethered state where clearly glycosylation does tend to dampen the fluctuations (Figure 4.9).

Figure 4.7. Visualization of the untethered EGFR-ed dynamics. Seven frames of aMD trajectory are shown separated by 50ns. All structures were superimposed onto the untethered crystal structure (PDB ID 3NJP) and their assigned RMSD value is indicated above each structure. Subdomains I, II, III and IV of EGFR-ed are shown in blue, green, yellow and magenta tubes respectively. The untethered crystal structure is shown with the grey tube.

Figure 4.8. Visualization of the untethered gEGFR-ed dynamics. Seven frames of aMD trajectory are shown separated by 50ns. All structures were superimposed onto the untethered crystal structure (PDB ID 3NJP) and their assigned RMSD value is indicated above each structure. Subdomains I, II, III and IV of gEGFR-ed are shown in blue, green, yellow and magenta tubes respectively. The untethered crystal structure is shown with the grey tube.

Figure 4.9. Projection of the tethered monomer trajectories along the untethered state PC1 and PC2. Structures obtained from aMD simulations of untethered EGFR-ed, untethered gEGFR-ed, tethered EGFR-ed and tethered gEGFR-ed are shown with blue, magenta, black and orange spots respectively. Projection of the untethered (PDB ID 3NJP) and tethered (PDB ID 1NQL) states crystal structures are shown with green and red spots respectively.

4.3.2 Dynamics of glycosylated EGFR in the tethered state

Similar to the untethered monomer simulations, RMSD plots of the tethered glycosylated monomer show a noticeable decrease compared to the unglycosylated one (Figure 4.10 and Appendix B Figures 9 and 10). At the level of subdomains I-IV of the extracellular domain, a similar dynamical pattern was observed (Figure 4.11.A-D and Appendix B Figures 11 and 12). The C α fluctuations of the tethered monomer also show large decreases in the gEGFR-ed compare to the EGFR-ed (Figure 4.12). This is like what was observed in the untethered monomer (see section 4.3.1). In contrast to the lack of secondary structural changes seen in the untethered states, in the tethered state, we see some changes, especially in the helical regions of subdomains III and IV (Appendix B Figures 13-16). This is most likely because the glycosylation “density” across the smaller exposed surface of the tethered state is higher, thus likely reducing the solvation of the protein. PCA analysis was used to investigate whether the plausible conformational shift towards the untethered state occurs in the glycosylated/unglycosylated tethered monomer.

Figure 4.10. The $C\alpha$ RMSD plots of the tethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. Each trajectory of 100ns was aligned on the starting crystal structure PDB 1NQL. Data shown here is obtained from one simulation. The data for other two replicates can be found Appendix B Figures 9 and 10.

Figure 4.11. The $C\alpha$ RMSD plots of subdomains I-IV (A-D) of the tethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. Each trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation. The data for other two replicates can be found Appendix B Figures 11 and 12.

Figure 4.12. The C α RMSF plots of the tethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The amino acid residues of subdomains I, II, III and IV of EGFR-ed and EGF are marked on the x axis with blue, green, yellow, magenta and red respectively. Combined trajectory of 300ns was aligned on the starting crystal structure PDB 1NQL.

The dynamics of glycosylated EGFR in its tethered state are captured mostly by the first 3 PCs with PC1 and PC2 accounting for ~50% (39.9% and 19.2%) of the total variance (Figure 4.13). It is clear that glycosylation results in reduced flexibility (Figure 4.14), as we also saw for the untethered state. Interestingly, the motions of glycosylated tethered monomer are restricted to the vicinity of the tethered crystal structure (Figures 4.14 and 4.15) while the system explores regions away from the crystal structure in the absence of glycosylation. This could either result from insufficient sampling of the larger more complex glycosylated state or could also reflect the fact that the crystallographic structure may be glycosylated but the positions of the sugars was not easily resolved due to their fluctuations. Projection of the untethered crystal structure onto the PC space shows that the unglycosylated tethered monomer could likely adopt an extended conformation.

However, such transition was not observed in the simulation time of this study (Figures 4.16 and 4.17). In contrast, we saw above that the unglycosylated extended monomer did adopt a compact conformation which was maintained for the last 50 ns of the simulation time (Figure 4.7).

Figure 4.13. Percentage of variance captured by each principal component from aMD simulations of glycosylated and unglycosylated tethered monomer.

Average RMSD values from the untethered crystal structure calculated from all simulations of the tethered monomer are: EGFR-ed (23.0 Å) and gEGFR-ed (23.1 Å) respectively (Figure 4.15). The PC1 first and last extreme structures of gEGFR-ed show more resemblance to the tethered crystal structure compared to the EGFR-ed. RMSD values of the PC extreme structures from the tethered crystal structure are: gEGFR-ed (6.8 Å and 4.3 Å) and EGFR-ed (7.4 Å and 13.4 Å) respectively. The PC2 extreme structures of the two systems show a similar trend with RMSD values from the tethered crystal being gEGFR-ed (3.3 Å and 6.4 Å) and EGFR-ed (5.0 Å and 14.1 Å) respectively.

Projection of the untethered state simulations trajectory in the PC space of the tethered state simulations present the overlap of the unglycosylated tethered monomer with the untethered state. This overlap is more significant in the projection of gEGFR-ed untethered monomer on to the tethered state PC space (Figure 4.18). Unlike the untethered monomer simulations, here the variation of unglycosylated monomer does not necessarily shift the conformation towards the opposite end of the tethered-extended equilibrium. This is clear from the extreme structures of the two first principal components (Figure 4.14). Considering the large conformational difference between the tethered and extended states even aMD is unable to capture this transition in the time scale of this study. Thus, for clarifying whether the wider range of sampled conformations in the unglycosylated tethered monomer does, in fact, supports the tethered-extended transition, dynamic cross-correlation motions and the strength of the detailed tethering interactions of EGFR-ed and gEGFR-ed were studied.

Figure 4.14. Projection of the tethered structures along PC1 and PC2. Structures obtained from aMD simulations of glycosylated and unglycosylated tethered monomer are shown with magenta and blue spots respectively. Projection of the untethered (PDB ID 3NJP) and tethered (PDB ID 1NQL) states crystal structures are shown with green and red spots respectively. Superimposition of PC1 and PC2 extreme structures on the tethered state crystal structure from the glycosylated and unglycosylated simulations are shown with magenta and blue tubes respectively. The untethered (PDB ID 3NJP) and tethered (PDB ID 1NQL) states crystal structures are shown with green and red tubes respectively. Black arrows show the location of each conformation in the PC space.

Figure 4.15. RMSD values calculated from the glycosylated and unglycosylated tethered monomer simulations. 3NJP and 1NQL are PDB IDs for the untethered and tethered crystal structures.

Figure 4.16. Visualization of the tethered EGFR-ed dynamics. Seven frames of aMD trajectory are shown separated by 50ns. All structures were superimposed onto the tethered crystal structure (PDB ID 1NQL) and their assigned RMSD value is indicated above each structure. Subdomains I, II, III and IV of EGFR-ed are shown in blue, green, yellow and magenta tubes respectively. The tethered crystal structure is shown with the grey tube.

Figure 4.17. Visualization of the tethered gEGFR-ed dynamics. Seven frames of aMD trajectory are shown separated by 50ns. All structures were superimposed onto the tethered crystal structure (PDB ID 1NQL) and their assigned RMSD value is indicated above each structure. Subdomains I, II, III and IV of EGFR-ed are shown in blue, green, yellow and magenta tubes respectively. The tethered crystal structure is shown with the grey tube.

Figure 4.18. Projection of the untethered monomer trajectories along the tethered state PC1 and PC2. Structures obtained from aMD simulations of tethered EGFR-ed, tethered gEGFR-ed, untethered EGFR-ed and untethered gEGFR-ed are shown with blue, magenta, black and orange spots respectively. Projection of the tethered (PDB ID 1NQL) and untethered (PDB ID 3NJP) states crystal structures are shown with red and green spots respectively.

4.3.3 Alteration of the tethering interactions upon glycosylation

The transition of EGFR from its tethered to the extended state occurs by the removal of the interactions between the dimerization arm of subdomain II (residue 239 to 269) and the two inhibitory loops of subdomain IV (residue 561 to 569 and 572 to 585 respectively) [1]. The Cross-correlation motions of subdomain IV relative to the dimerization arm of subdomain II, the position of the dimerization arm relative to the two inhibitory loops and the distance distribution between each pair of interacting residues involved in the tethering interactions was used as the criteria for estimation of the conformational shift towards/away from the extended state in this study. The correlation between the fluctuations of the alpha carbon atoms was mapped for the dynamics of the tethered EGFR-ed and gEGFR-ed (Figures 4.19 and 4.20). They reveal the complex changes in the dynamical pattern of the tethered monomer upon glycosylation which could trigger the conformational transition to the untethered state. As was seen earlier, glycosylation results in dampened fluctuations and this results in subdomains I-IV getting more structured. In addition, the loss of correlated motions of subdomains III and IV (residue 380 to 500) with subdomains I and II (residue 110 to 240) and the appearance of noticeable anti-correlated motions in residues 180 to 240 of subdomain II with respect to subdomain IV (residue 481 to 614) were observed upon glycosylation. The dissociation of the two first subdomains from the C-terminal segment of subdomain III and the whole subdomain IV observed in the simulation presented here, is the critical trigger of the tethered to untethered state transition. This is further complemented by visualization of the correlated and anti-correlated motions on the structure (see Appendix B Figure 17 and 18).

To expand on the above observations, the distance distributions between the N-terminus segment of subdomain IV (residue 481 to 546) - dimerization arm and the C-terminus segment of subdomain IV (residue 547 to 614) - dimerization arm were calculated (Figure 4.21). The N-terminus segment of subdomain IV - dimerization arm distance increases upon glycosylation and persistently adopts values around 34 Å whereas, in the unglycosylated tethered monomer the distance distribution adopt three peaks of 28 Å, 31 Å and 34Å (Figure 5.21.A). The increase in the distance values after glycosylation is even more noticeable in the C-terminus segment of subdomain IV- dimerization arm of glycosylated monomer (18 Å) compared to the unglycosylated monomer (14 Å) (Figure 21.B). Overall, the global dynamics of the tethered monomeric EGFR-ed show that the receptor is more likely to shift towards the extended state in its glycosylated form. In the next step, the distance between the center of mass of each inhibitory loop in subdomain IV and the dimerization arm was calculated (Figure 4.22). The distance between the inhibitory loop I and dimerization arm increases upon glycosylation (the values are distributed ~25 Å in the tethered gEGFR-ed and ~19-21 Å in the unglycosylated state (Figure 4.22.C). Similarly, the distance between the inhibitory loop II and the dimerization arm also undergoes an increase upon glycosylation (between ~10-13 Å in EGFR-ed to ~15 Å in gEGFR-ed; Figure 4.22.D). This increase in the distance between inhibitory loops of subdomain IV and the dimerization arm of subdomain II results in a weakening of the tethered state. Based on the tethered state crystal structure (PDB ID: 1NQL), the tethering interaction is maintained by five hydrogen bonds formed between the dimerization arm and the inhibitory loops of subdomain IV (see Appendix B Figure 3) [25]. Two hydrogen bonds are formed between the dimerization arm and the Inhibitory

loop I: Tyr246 (dim arm)-Asp563 (loop I) and Thr250 (dim arm)-His566 (loop I). Distribution plots (Figure 4.23) show that the Tyr246-Asp563 distance slightly decreases upon glycosylation (Figure 4.23.A). Distance distribution of the Thr250-His566 shows smaller range of values in the glycosylated monomer (2-18 Å). This distance increases to 27 Å in the unglycosylated monomer (Figure 4.23.B). The two hydrogen bonds between dimerization arm and inhibitory loop II are: Tyr246 (dim arm)-Met576 (loop II) and Asn247 (dim arm)-Glu578 (loop II) (see Appendix B Figure 3). Distance distribution of the Tyr246-Met576 show a decrease from 16 Å in the unglycosylated EGFR-ed to 12 Å in the gEGFR-ed (Figure 4.23.C). The Asn247-Glu578 distance also assumes higher values in the unglycosylated monomer (two peaks at 11 Å and 23 Å) compared to the glycosylated system (peak at 11 Å) (Figure 4.23.D). The hydrogen bond between the two inhibitory loops is formed between Asp563 (loop I) and Lys585 (loop II) (see Appendix B Figure 19). The distribution plots of this interaction do not show a change upon glycosylation (Figure 4.23.E). Although the distance between the residues suggested to be involved in the tethering interaction in the X-ray crystal structure, they do not alter or undergo a decrease in the glycosylated monomer; the values calculated from the glycosylated simulations reflects the decreased flexibility of the EGFR-ed as a result of glycosylation. The decreased flexibility of the glycosylated monomer is also seen earlier in the RMSD and PC plots. While the whole structure is less flexible it is expected that the distance between the above mentioned pairs of interacting residues follow the same trend. However, the global dynamics of the receptor observed in the dynamic cross correlation maps and the N- and C-terminal segments of subdomain IV towards the dimerization arm, support a higher chance of untethering in the glycosylated system.

Among all common glycosylation sites of EGFR-ed, Asn579 is located within the inhibitory loop II of subdomain IV. In the simulations presented here, Asn579 attached glycan prevents the tethering interactions (Appendix B, Figure 20). Visualization of the dynamics of Asn579 attached glycan shows that by sampling a wide conformational phase space, this sugar weakens the association of the dimerization arm with subdomain IV. The simulations did not show any aggregation or extensive contacts between any of the oligosaccharides and the oligosaccharide attached to Asn579. This suggests that the tethered conformation is not maintained by the interactions between the glycans. The distance between all atoms in the oligosaccharide attached to Asn579 and all atoms of every other oligosaccharide in the glycosylated EGFR-ed model used in this study supports this hypothesis (Appendix B, Figure 21). However, the glycan attached to Asn172, located in subdomain II, becomes less distant (34Å) from the Asn579 attached glycan in the timescale of the current simulations. This could possibly, over longer simulation times, result in an interaction between these two glycans and hence reduce the distance between subdomains II and IV but only if the oligosaccharides structure is tri/multifurcated.

Figure 4.19. Cross-correlation motions of the tethered EGFR-ed. Only C α atoms were kept for the calculation. Shades of red and blue spots present the correlated and anti-correlated motions of atoms. The amino acid residues of subdomains I, II, III and IV of EGFR-ed and EGF are marked on the x and y axes with blue, green, yellow, magenta and red respectively. The values of the correlations range from -1 to 1 as shown in the bar on the right.

Figure 4.20. Cross-correlation motions of the tethered gEGFR-ed. Only C α atoms were kept for the calculation. Shades of red and blue spots present the correlated and anti-correlated motions of atoms. The Amino acid residues of subdomains I, II, III and IV of EGFR-ed and EGF are marked on the x and y axes with blue, green, yellow, magenta and red respectively. The values of the correlations range from -1 to 1 as shown in the bar on the right.

Figure 4.21. A) The distance between the center of mass of the N-terminus segment of subdomain IV (residue 481 to 546) - dimerization arm. Values for tethered EGFR-ed and gEGFR-ed are shown with blue and magenta respectively. B) The distance between the center of mass of the C-terminus segment of subdomain IV (residue 547 to 614) - dimerization arm. Values for tethered EGFR-ed and gEGFR-ed are shown with blue and magenta respectively.

Figure 4.22. A) EGFR-ed tethered by the interaction of the dimerization arm and two inhibitory loops of subdomain IV. EGFR, dimerization arm and inhibitory loops are shown in grey, blue and red tube respectively. Glycosylated Asparagines are shown with purple spheres. B) Location of the dimerization arms and inhibitory loops in the untethered dimer. C) The distance between the center of mass of the dimerization arm and inhibitory loop I; values for tethered EGFR-ed and gEGFR-ed are shown with blue and magenta respectively. D) The distance between the center of mass of the dimerization arm and inhibitory loop II; values for tethered EGFR-ed and gEGFR-ed are shown with blue and magenta respectively.

Figure 4.23. Distance distributions between the interacting atoms of each pair of residues involved in the tethering interaction. Values for the tethered EGFR-ed and gEGFR-ed are shown with blue and magenta respectively.

4.4 Discussion

Biological functions of EGFR are dependent on its activation and hence a robust understanding of the activation mechanism is necessary. EGFR activation starts by binding of the extracellular domain to growth hormones on the cell surface leading to its dimerization and activation of the signaling cascade in the cytosol [3]. Various studies have shown that EGFR exists in an autoinhibitory conformation known as the tethered state [7, 21, 25]. The tethered state is characterized by extensive intramolecular contacts between subdomains II and IV of EGFR-ed. Growth factor binding to subdomains I and III of the extracellular domain is believed to allosterically remove the tethering interactions between subdomains II and IV [25]. After the removal of the tethering interactions, EGFR-ed adopts an extended conformation. In the extended conformation, the principal axis of the protein is perpendicular to the membrane [11]. The dimerization arm of subdomain II and the tip of subdomain IV which were involved in the intramolecular tethering interaction will act as the intermolecular dimeric interfaces with another ErbB monomer to form the active extended dimer. EGFR-ed is known to exist as both inactive pre-formed dimers [26, 28, 102] and inactive tethered monomers [7, 21, 25] prior to ligand binding. An identical tethered conformation was reported in X-ray crystal structures of ErbB3 and ErbB4, the homologous family members of EGFR [106, 107]. Hence, undoubtedly, both tethered and extended conformations are functional forms of EGFR-ed. Based on the structural data on the EGFR-ed [14, 17, 25], it was suggested that both tethered and extended monomeric conformations are sampled in the dynamics of EGFR in vivo. However, growth factor binding requires interactions with subdomains I and III simultaneously which is spatially possible only in the extended conformation.

Thus it has been hypothesized that EGF binding is associated with a bias of the EGFR population towards the extended conformation. These interpretations are made by not accounting for post-translational modifications of EGFR, most importantly, of glycosylation. Glycosylation is the crucial post-translational modification that occurs in EGFR [47, 103]. Several biochemical assays have suggested functional roles of site-specific glycosylation in the tethered to extended state transition [32, 33]. Here atomistic aMD simulations were used for modeling of the two extreme ends of this conformational change (the tethered monomer and the extended monomer) with and without glycosylation. Simulations showed that both tethered and extended states are less flexible upon glycosylation. This is in good agreement with several experimental [32, 33] and computational [52, 55] studies that have shown unambiguously that glycosylation is required for the integrity and stability of the protein. However, the dynamical patterns that characterize the two states are mediated by glycosylation in entirely different ways. The extended monomer is less flexible and is maintained in the extended form by the attached oligosaccharides over the timescale of all simulations reported here; no other reports on this are available. This gives rise to the hypothesis that glycosylation supports the extended monomer. This is in good agreement with the other observations in this thesis suggesting maintenance of the two dimeric interfaces by glycans in the apo and EGF bound extended monomer (see chapter 3 section 3.3.5) and on the increase of twist and turn motions required for dimer activation in the glycosylated dimers (see chapter 5 section 5.3.2). In contrast, in the absence of attached glycans, the flexible extended state adopt a compact conformation which is stable for a noticeable portion of the second half of the simulation time (200ns; in the third 100ns replicate, the conformation does shift

towards the compact state but doesn't attain the same degree of "compaction" as in the other 2 replicates)(Fig 4.7). All the simulations of the extended state were carried out in the EGF bound form which is known to be more stable [11] and yet the conformation becomes very close to the tethered state in the unglycosylated monomer, reflecting the critical role of glycans in maintaining the extended conformation. On the other end of the equilibrium, the tethered state simulations showed that although unglycosylated tethered monomer is more flexible, it does not adopt/shift towards the extended conformation. Interestingly, it was shown that upon glycosylation the decreased flexibility of the tethered monomer provides the possibility of the transition towards the extended state. The appearance of novel anti-correlated motions and increase of the distance between the dimerization arm and the inhibitory loops of domain IV occurs in the tethered glycosylated EGFR-ed. Among all the attached glycans, the ones at Asn151, Asn328 and Asn337 are relatively rigid. Asn151 and Asn328 attached glycans interact with each other and decrease the flexibility of subdomains I and III and facilitate the binding of the growth factor to EGFR-ed (see chapter 3 section 3.3.1 and 3.3.2). Asn337 maintains the dimeric interface by interacting with its counterpart glycan from the other interacting monomer in the extended dimer, obviously after the completion of the tethered-extended transition (see chapter 5 section 5.3.1 and Appendix C Figure 1). Asn32 and Asn420 attached glycans were neither rigid nor highly flexible in the simulations. Glycan attached to Asn420 is suggested to play a role in the regulation of EGFR autodimerization. M.Takahashi et al. suggested that in the Asn418 mutants of ErbB3 (Asn420 in EGFR) the loss of N-glycan is the reason for autodimerization [32]. Although the authors did not comment on any possible involvement of the Asn420 attached glycan in the tethering

interactions, it seems that the mechanism behind the above notion could be the following: the removal of Asn420 glycan strengthens the tethering interaction and hence its removal would result in the onset of the tethered-extended transition and consequently, autodimerization. In the simulations presented here, Asn420 attached glycan is not engaged in the tethering interactions as it extrudes from EGFR-ed surface. This could be limited by the heterogeneity in length and complexity of the attached glycans and the interactions of the cell membrane with the glycans; the experiments do not detail these points. In the glycosylated tethered model presented here, it is highly unlikely for the Asn420 attached glycan to engage in the tethering interactions and to reach out to the glycan attached to Asn579 (located in the inhibitory loop II of subdomain IV) to strengthen the tethering interaction (see Appendix B Figure 21). The glycans at Asn49, Asn172, Asn504 and Asn579 are all flexible in the simulations; of course these are observations from simulations here that do not include the cell membrane and again contain only one model of a complex oligosaccharide. In a recent 1 μ s simulations study of glycosylated and unglycosylated full-length tethered monomer embedded in the membrane [55] it was suggested that glycans at Asn151, Asn172, Asn389, Asn420 and Asn504 interact with the membrane lipids and act as cushions to lift EGFR-ed up from the membrane. However the Asn579 attached glycan was not reported to act as a cushion between the EGFR-ed and the membrane. This is in good agreement with the simulations presented here that Asn579 attached glycan weakens the tethering interaction by preventing the interactions between the dimerization arm and inhibitory loops. Asn579 is located in the inhibitory loop II of subdomain IV that interacts with the dimerization arm of subdomain II as a tethering interaction site. However, the site mutation of Asn597 to

Gln was suggested to increase EGFR phosphorylation [33], suggesting a role for the Asn579 glycan in strengthening the tethering interaction. This could be explained by the cell and tissue specificity of glycosylation as a non-template process [30] and the length and complexity of branches in the attached oligosaccharide and the possibility of fucosylation or sialylation of the glycan [30]. In the above-mentioned assay, the authors did not specifically mention these features of the glycan attached to Asn579. However, the oligosaccharide used here is commonly seen in the wild-type mammalian cell. Hence it is unlikely that Asn579 glycan strengthens the tethered conformation. Overall, aMD was implemented to bridge the sampling gap between the two known monomeric conformational states of EGFR-ed with and without glycosylation. PCA of the trajectories showed the overlapping dynamics of the two states, reflecting the enhanced sampling of the conformations in the aMD simulations. The two first principal components of the extended and tethered monomer revealed that glycosylation plays different roles in the dynamics of each state. In the extended state, glycosylation reduces the flexibility of EGFR-ed and maintains the conformation. In the tethered state, glycosylation reduces the flexibility and weakens the tethering interaction, thus facilitating the transition towards the extended form.

4.5 Conclusions

Glycosylation reduces the flexibility of tethered and extended monomeric forms of EGFR-ed. By maintaining the glycosylated form of the extended state and weakening the tethering interactions in the tethered state, glycosylation supports the extended conformation and facilitates the tethered-extended transition.

Chapter 5. Elucidating EGFR activation: interplay between glycosylation and growth factor binding

5.1. Introduction

Biological functions of EGFR are dependent on receptor activation. EGFR activation starts with conformational changes in the extracellular domain. Two models were proposed to explain these functional changes: the “ligand-induced dimerization” model [7, 25] and the “preformed dimer” model [28, 102]. The first model states that EGFR tethered-extended transition is induced by binding of the growth factor to the extracellular domain [25]. This model is fully explained in chapter 4, section 4.1. The second model of EGFR activation suggests that EGFR exists as a dimer even in the absence of ligands [26-28]. This dimer is flexible and is stabilized by binding of EGF to the extracellular domain. Twist and turn motions of the stabilized extracellular dimers induce transmembrane domain rotation and consequent rearrangement of the intracellular kinase domains followed by receptor activation. Although both models are supported by biochemical assays and 3D structures, they were proposed without accounting explicitly for post-translational modifications of EGFR, including glycosylation of the EGFR-ed. Glycosylation is a crucial post translational modification that occurs in EGFR and is known to play various functional roles in the regulation of the receptor [32, 33]. Hence, its supporting role towards the tethered monomeric/extended dimeric structure could validate the most physiologically plausible activation model and enlighten structural dynamics of the extracellular domain. Role of glycosylation in “Ligand induced dimerization” was investigated earlier (chapter 4) by implementing enhanced sampling (aMD) and analysis (PCA) techniques to overcome the energy barriers in the tethered-extended transition path

and to reduce the dimensionality of motions to focus on the major dynamical motions of the monomeric EGFR upon glycosylation. It was shown that glycosylation of the EGFR extracellular domain by oligosaccharides clearly contributes to maintenance of the extended state of monomeric EGFR, weakens intramolecular tethering interactions and supports the transition from the tethered to extended conformation. Now the role of glycosylation on modulating the rigid body motions proposed for the activation of EGFR in the “preformed dimer” model are explored using atomistic simulations. The glycosylated and unglycosylated dimers of the EGFR extracellular domain were modeled in the apo and EGF-bound states to investigate the roles of attached glycans in the activation process and to investigate the combinatorial effect of glycans and growth factor by comparing the two states.

5.2. Materials and methods

5.2.1 Model building of glycosylated and unglycosylated EGFR dimers

The PDB entry 3NJP [14] was used as the initial structure for modeling. This structure is obtained at 3.3 Å resolution and contains four chains: two extended monomers of EGFR extracellular domain in full length (residues 1 to 614), each bound to a monomer of EGF (residues 5 to 51). All four chains were used for the simulation of EGF-bound dimers and the apo dimers were obtained by removing EGFs from the complex. The EGFR extracellular domain was capped at its C-terminal end, while the EGF was capped by adding capping residues at the N- and C-terminal ends of the structures, and capping was carried out using the tleap module of the AMBER12 package [93]. Each monomer of the EGFR extracellular domain contains 25 disulphide bonds while the ligand EGF contains 3 disulphide bonds. The disulphide bonds were constructed by bridging the Cysteines using

the tleap module of AMBER12 package. The ‘GLYCAM online glycoprotein builder’ [72] was used to attach the glycans to each monomer of the EGFR extracellular domain at Asn32, Asn49, Asn151, Asn172, Asn328, Asn337, Asn420, Asn504 and Asn579 (Figure 5.1.A) (for extended details on the glycosylation site selection see chapter 3, section 3.1). The glycans used were GlcNAc-GlcNAc-man-man-man-GlcNAc, a commonly characterized complex unit [30, 31]. The systems were immersed in a box of TIP3 waters with at least a 10Å boundary around any atom and the systems were neutralized with Na⁺ ions.

5.2.2. Molecular Dynamics simulations

All the MD simulations were performed with the AMBER12 package [93]. The parameters and topology files were created using the tleap module of AMBER12. The Amber ff99SB force field was used to represent the protein, with parameters for the glycans taken from the GLYCAM06 force field [67]. All systems were minimized, equilibrated and subject to production MD in triplicates using the protocol outlined in Chapter 2 (section 2.7.7). Using the above protocol, four systems were simulated: unglycosylated dimer of EGFR extracellular domain, glycosylated dimer of EGFR extracellular domain, unglycosylated dimer of EGFR extracellular domain bound to EGF and glycosylated dimer of EGFR extracellular domain bound to EGF. All analyses were performed using the ptraj module of the AMBER12 package on the 150ns trajectory obtained from combining triplicates of 50ns MD for each system. Figures were made with VMD and PYMOL [96, 99].

5.2.3. Intradimeric twist angle and Intramonomeric torsion angle calculation

Intradimeric twist angle is defined as the angle between the axis joining the two subdomains I (residue 1 to 165) and the axis joining the two subdomains III (residue 311 to 480). This angle was calculated using the vectormath action of ptraj module of AMBER12 package where each axis was defined as a vector between the centers of masses of two subdomains (Figure 5.1.B). Intramonomeric turn motion has been calculated by the torsion angles between subdomains I and III of each monomer. Dihedral action of ptraj module of AMBER12 was used to calculate this angle and the torsional plane was defined between the center of mass of subdomain I (residues are 1 to 165), the center of mass of the N-terminal segment of subdomain II (residues 166 to 237), the center of mass of the C-terminal segment of subdomain II (residues 238 to 309) and the center of mass of subdomain III (residues 311 to 480) (Figure 5.1.C). These definitions follow those outlined by Poger and Mark [13].

Figure 5.1. Intradimeric twist angle and intramonomeric torsion angle of EGFR extracellular domain dimers. A) EGFR extracellular domain bound to EGF dimer shown in surface. Subdomains I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red for monomer1 and blue, cyan, yellow, dark pink and red for monomer2. For simplicity, glycosylated Asparagines are shown with spheres and black labels only in monomer1. The attached oligosaccharides are shown as sticks. B) Top view from section A. Intradimeric twist angle (Ω) within the dimer, presenting the twist motion between the two axes connecting centers of masses of subdomains I and III of the two monomers. C) Intramonomeric torsion angle (ϕ) between subdomain I, N-terminal segment of subdomain II, C-terminal segment of subdomain II and subdomain III of each monomer.

5.3. Results

5.3.1 Dynamics of dimeric EGFR upon glycosylation

The EGFR extracellular domain and glycosylated EGFR extracellular domain will be referred to as: EGFR-ed and gEGFR-ed. The average RMSD plots show that glycosylation reduces the flexibility in both apo and EGF-bound dimers. The average RMSD values obtained from the last 30ns and 10ns of the apo and EGF-bound replicates of each system are: EGFR-ed (4.6Å, 5.9 Å, 4.5Å), gEGFR-ed (3.8Å, 5Å, 4.4Å), EGF-EGFR-ed (3.4Å, 3.1Å, 3Å) and EGF-gEGFR-ed (4.2Å, 6Å, 4.9Å) (Figures 5.2-5.4). It is worth mentioning that these RMSD values may be significant for a single domain protein but dimeric EGFR-ed consists of two monomers of four subdomains each and these values are expected due to the relative positions of the subdomains in the complex. The decrease in flexibility upon glycosylation is more noticeable in the apo dimer based on the difference in the average RMSD and the range of the sampled values in the unglycosylated dimer. When EGF is bound, although the flexibility of the EGF bound state is less than that of the apo state (as expected), the effects of glycosylation are not as pronounced as they are in the apo state. At the level of the individual subdomains (Figures 5.5-5.7), the growth factor binding subdomain I and subdomain II showed decrease in RMSD upon glycosylation. The average RMSD of these subdomains are: EGFR-ed, subdomain I (1.7Å), gEGFR-ed subdomain I (1Å), EGF-EGFR-ed subdomain I (1.2Å) and EGF-gEGFR-ed subdomain I (0.9Å). EGFR-ed, subdomain II (2.8Å), gEGFR-ed subdomain II (2.3Å), EGF-EGFR-ed subdomain II (1.8Å) and EGF-gEGFR-ed subdomain II (2.6Å). The decrease in the flexibility of the growth factor binding subdomain I is due to the interaction between the glycans attached to Asn151 and Asn328

which stabilize the EGF binding pocket (Figure 5.8; for greater details see Chapter 3 section 3.3.1). Subdomains II is involved in the dimeric interface and present less flexible dynamics upon glycosylation in the apo and EGF-bound states. This is linked to the interaction of the glycans attached to Asn172 and Asn337 of each monomer with the other monomer of EGFR-ed that maintains the dimeric interface in Subdomain II (see Appendix C Figures 1-3). The C α RMSF pattern is comparable in all four systems with a noticeable decrease (5.0 Å) in the apo dimer upon glycosylation (Figure 5.8.A). The final snapshot from one of the simulations of each system is shown in Figure 5.8.B. Unsurprisingly, the EGFR-ed dimer is the most flexible, the subdomain IV of each monomer moves closer to subdomain II, resulting in increasing compaction of the whole complex. Upon glycosylation, the gEGFR-ed dimer shows more extended tertiary structure and the orientation of each subdomain is well ordered. EGF-EGFR-ed and EGF-gEGFR-ed dimers show similar dynamics although in the latter, the interaction between the glycans attached to Asn151 and Asn328 are more persistent due to the presence of EGF and the increased stability due to its binding.

Figure 5.2. Average RMSD calculated of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values were obtained from the last 30ns and 10ns of the apo and EGF-bound first simulation of each system.

Figure 5.3. Average RMSD calculated of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values were obtained from the last 30ns and 10ns of the apo and EGF-bound second simulation of each system.

Figure 5.4. Average RMSD calculated of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values were obtained from the last 30ns and 10ns of the apo and EGF-bound second simulation of each system.

Figure 5.5. Average RMSD calculated for subdomains I-III (A) and IV (B) of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values were obtained from the last 30ns and 10ns of the apo and EGF-bound first simulation of each system.

Figure 5.6. Average RMSD calculated for subdomains I-III (A) and IV (B) of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values were obtained from the last 30ns and 10ns of the apo and EGF-bound second simulation of each system.

Figure 5.7. Average RMSD calculated for subdomains I-III (A) and IV (B) of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values were obtained from the last 30ns and 10ns of the apo and EGF-bound third simulation of each system.

Figure 5.8. Dynamics of glycosylated EGFR-ed dimers.

A) The Ca RMSF plots for EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. The values are shown in black, green, blue and magenta respectively. Amino acid residues of each subdomain of monomer1 and monomer2 are marked above the x-axis with the same color codes used in the B section. RMSF values are from the last 30ns and 10ns of the first replicate of the apo and EGF-bound dimers. B) Final frame from one of the production runs is shown for each system. Subdomain I, II, III, IV of EGFR and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surface for monomer1 and blue, cyan, yellow, dark pink and red for monomer2.

5.3.2 Glycosylation Increases the intradimeric twist motion and intramonomeric turn motion

In the pre-formed dimer activation model [13, 26, 28, 102], the intradimer twist motion is perquisite. This motion is mechanically defined by a twist angle between the axis joining subdomains I of the two monomers and the axis joining subdomains III of the two monomers (Figure 5.1.B). Distributions of the twist angles show that the twist angle increases upon glycosylation by about $\sim 2^\circ$ and $\sim 5^\circ$ in the absence and presence of EGF (Figure 5.9). The larger increase in the twist motion of EGF-bound dimer upon glycosylation suggests that glycosylation indeed supports the twist motion required for EGFR activation but it acts as a secondary force in this mechanism with EGF binding being the primary. Covariance maps show that this is indeed the case and one can see more correlated behavior appearing between the two subdomains I and the two subdomains III from the two monomers are more correlated in the presence of EGF (Figure 5.10). It is also obvious that the anti-correlated motions between the domains I and the domains III is also increased upon EGF binding and these are even more pronounced upon glycosylation. This is what is expected for twist motion based on the angle distribution plots discussed earlier from Figure 5.8. However, in the absence of EGF the effects are not so clear (Figure 5.11). Given the large sizes of these systems, it is likely that much longer simulations are needed to see converged patterns. The turn motion within each monomer is defined by the torsion angle between subdomains I and III (Figure 5.1.C). This angle was calculated for each monomer in the dimeric complex. The EGF-gEGFR-ed dimer shows an increase in the value of the torsion angle in both monomers. The distribution plot of torsion angle in monomer1 shows that the angle

values are distributed around: EGF-EGFR-ed (4.7°) and EGF-gEGFR-ed (bimodal distributions of 2.6° and 14.8°) and in monomer2 are: EGF-EGFR-ed (9.5°) and EGF-gEGFR-ed (bimodal distributions of 5.2° and 18.1°) (Figure 5.12). The apo dimer does not show a change in the torsion angle values in monomer1 however, in monomer2, an increase in the angle values distributed around: EGFR-ed (16.6°) and gEGFR-ed (20.6°) was observed. The increase in the intramonomeric turn motion upon glycosylation appears to be associated with an increase in the twist motion which in turn enables the twist motion of the transmembrane domains that in turn results in the asymmetric association of the kinase domains, a pre-requisite for functional activation. This is in accord with the model proposed by Poger et al. (Figure 2 in the cited article [13]) Although the starting structure used in their work was 1IVO and 1MOX, here similar motions was observed using a different starting structure (PDB ID 3NJP) and with different glycosylations. This suggests that these motions are a general property of receptor activation.

Figure 5.9. Intradimeric twist angle distribution plot. The values for EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively.

Figure 5.10. Cross-correlation motions of the EGF-EGFR-ed (A) and EGF-gEGFR-ed (B) dimers. Only Ca atoms were used in the calculations. Shades of red and blue spots present the correlated and anti-correlated motions of atoms. The values of the correlations range from -1 to 1 as shown in the bar on the right. The amino acid residues of subdomains I, II, III, IV of EGFR-ed and EGF are marked on the x and y axes shown in blue, green, yellow, magenta and light red for monomer1 and blue, cyan, yellow, dark pink and dark red for monomer2. Based on the definition of twist motion the regions implicated are shown with dashed black boxes. **Box A:** the correlations between the motions of subdomain I in monomer1 with respect to subdomain I in monomer2. **Box B:** the correlations between the motions of subdomain I in monomer1 with respect to subdomain III in monomer1. **Box C:** the correlations between the motions of subdomain III in monomer1 with respect to subdomain III in monomer2. **Box D:** the correlations between the motions of subdomain III in monomer1 with respect to subdomain I in monomer2.

Figure 5.11. Cross-correlation motions of the EGFR-ed (A) and gEGFR-ed (B) dimers. Only C α atoms were kept for the calculation. Shades of red and blue spots present the correlated and anti-correlated motions of atoms. The values of the correlations range from -1 to 1 as shown in the bar on the right. The amino acid residues of subdomains I- IV of EGFR-ed are marked on the x and y axes shown in blue, green, yellow and magenta for monomer1 and blue, cyan, yellow and dark pink for monomer2. Based on the definition of twist motion different regions of the dimeric conformation are shown with dashed black boxes. **Box A:** the correlations between the motions of subdomain I in monomer1 with respect to subdomain I in monomer2. **Box B:** the correlations between the motions of subdomain I in monomer1 with respect to subdomain III in monomer1. **Box C:** the correlations between the motions of subdomain III in monomer1 with respect to subdomain III in monomer2. **Box D:** the correlations between the motions of subdomain III in monomer1 with respect to subdomain I in monomer2.

Figure 5.12. Distribution plots of intermonomer1 (A) and intermonomer2 (B) torsion angles.

The values for EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively.

5.3.3 Glycosylated EGFR dimer could adopt the flush conformation in the apo and EGF-bound states

Based on the available crystal structures of the ErbB receptors, two types of dimeric interfaces have been characterized: “staggered” and “flush” [53] (Figure 5.13). Description of the two conformations was first explained by Liu et al. based on the flush or staggered location of two adjacent loops in subdomain II (residues 190 to 208) across the dimeric interface [53]. These two dimeric interfaces are suggested to be the mechanical representation for regulation of EGFR by negative cooperativity [13, 52, 108]. Negative cooperativity is known to be the critical feature of EGFR ligand binding [52, 53]. When one of the growth factor binding sites gets occupied with one growth factor, the dimeric structure stabilizes in the flush conformation. Flush conformation is the most

stable dimeric conformation in drosophila and human EGFR [53]. This conformation decreases the affinity of the other monomer for growth factors [53, 54]. Molecular dynamics simulations of EGF-EGFR-ed dimers embedded in the membrane suggest that the dimer rests on the membrane and the interactions of one ligand with the membrane could result in negative cooperativity [52]. This asymmetric dimer bound to one growth factor, adopts the staggered conformation after binding of the second growth factor [54]. The staggered conformation is not as stable and was observed in EGF-EGFR and Nrg1 β -ErbB4 binding. Here it was attempted to identify the type of dimeric interface adapted by fully glycosylated EGFR in the apo and EGF bound states in the MD simulations that have been carried out and analyzed in Sections 5.3.1 and 5.3.2 above. Staggered and flush conformations were described with the orientation of the two growth factor binding subdomains (I and III) and the distance between subdomain IV of the two monomers. The observations were then compared with the available crystal structures of both conformations.

Figure 5.13. Staggered and flush dimeric interfaces of EGFE-ed adapted from [53].

The initial structure used in this study (PDB entry: 3NJP) is in the staggered conformation. The PDB entry 1MOX shows the dimer to be in the flush conformation. The distance between the centers of masses of subdomains III and IV from each monomer was calculated and compared with the crystal structure values (Figure 5.14). Despite starting from a state which is close to the staggered state, the distribution of gEGFR-ed and EGF-gEGFR-ed dimer conformations are noticeably centered around the flush crystal structure value (69 Å) (Figure 5.14.A). The EGFR-ed and EGF-EGFR-ed dimers do not sample smooth distance distributions. The EGFR-ed dimer shows the highest distance between subdomain III of monomer1 and subdomain III of monomer2 (71 Å). This is due to the flexibility caused by the absence of EGF and the attached glycans. In the EGF-

EGFR-ed dimer the distance between subdomain III of monomer1 and subdomain III of monomer 2 remains close to the staggered crystal structure value (68 Å). Although the crystal structure of the flush conformation (PDB ID 1MOX) does not contain subdomain IV, MD simulations of modeled structure constructed in this work show extensive contacts between subdomains IV, specifically at the tip of subdomain IV, which is the second critical dimeric interface after the dimerization arm (Figure 5.12). Distance distribution plots show that only in EGF-gEGFR-ed dimer there is a noticeable decrease in the distance between subdomains IV (44 Å) compared to EGFR-ed, gEGFR-ed, EGF-EGFR-ed (47 Å) (Figure 5.14.B). This decrease in the distance shows that the two subdomains IV interact more extensively and adopt conformations that are characteristic of the flush interface. The two glycans attached to Asn337 from each monomer stabilize each other and also the interactions between the subdomains II (Figure 5.15.A) while the two glycans attached to Asn579 from each monomer interact with subdomain IV of the other monomer (Figure 5.15.B) result in extensive interactions of subdomains IV. It is also clear that part of the conformational ensemble of the dynamically fluctuating glycans attached to Asn579 (as seen in Appendix A, Figure 18) are membrane binding states (as can be seen in Figure 5.15.B) and are in accord with the hypothesis suggested by Arkhipov et al. [52].

Figure 5.14. Intradimeric distance distribution plot. Distance between subdomains III (A) and subdomains IV (B) of EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed dimers are shown in black, green, blue and magenta respectively. Distance values obtained from staggered and flush crystal structures are shown with red and yellow dashed lines and their PDB IDs are labels with the same colors. Note that as outlined in Figure 5.12, the flush crystal structure (PDB 1MOX) does not contain subdomain IV. Hence only the value from the staggered crystal structure (PDB 3NJP) is shown in B.

Figure 5.15. Interaction between the glycans from the two monomers stabilizes the dimeric interface in the flush conformation. Subdomain II-IV of EGFR-ed are shown in green, yellow and magenta and surface for monomer1 and cyan, yellow and dark pink for monomer2. A) The oligosaccharides attached to Asn337 from monomer1 and monomer2 are shown with blue and orange sticks. The interacting residues from the two oligosaccharides are shown in grey. B) The oligosaccharides attached to Asn579 from monomer1 and monomer2 are shown with blue and orange sticks.

Comparison of the intramonomeric torsion angle distribution with the values from the crystal structures of flush and staggered conformations shows that the EGF-EGFR-ed dimer remains close to the starting staggered conformation (Figure 5.16). Whereas, upon glycosylation, the displacement of EGF-gEGFR-ed dimer from the staggered conformation and the increase in monomer2 torsion distribution seems to push the system towards the flush interface. The same changes are observed in the glycosylated apo dimer.

Figure 5.16. Distribution plots of intermonomer1 (A) and intermonomer2 (B) torsion angles.

The values for EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Torsion values obtained from staggered and flush crystal structures are shown with red and yellow dashed lines and their PDB IDs are labels with the same colors.

5.4 Discussion

Biological functions of EGFR are dependent on its activation and hence a robust understanding of the activation mechanism is necessary. EGFR activation starts by binding of the extracellular domain to growth factors on the cell surface leading to its dimerization and activation of the signaling cascade in the cytosol [109]. In recent years several studies have shown that EGFR exists as preformed dimers [26-28, 110-114]. The preformed dimer is flexible and is stabilized by binding of the growth factor to the extracellular domain. The stability gained by growth factor binding leads to a series of rigid body motions (twists and turns) in the extracellular domain of EGFR which induces transmembrane domain rotation and consequently rearrangement of the intracellular kinase domains followed by receptor activation [26, 111]. These rigid body motions were structurally defined [13] but there was no accounting for post-translational modifications of EGFR, most importantly, of glycosylation. Glycosylation is the crucial post translational modification that occurs in EGFR [32, 33, 47, 103]. These attached N-glycans account for nearly 40 KDa of the total 180 kDa mass of EGFR [33]. Several studies have shown the importance of glycosylation of EGFR-ed on the conformational dynamics of EGFR and on downstream signaling effects [32, 33, 47, 55, 101, 115]. Experimental [29, 32, 33, 50] and computational [52, 55] studies have shown unambiguously that glycosylation is required for the integrity and stability of the protein. In this study, atomistic simulations were used to investigate the effect of glycosylation on the turn and twist rigid body motions proposed for EGFR activation. Simulations show that glycosylation results in increase in the twist and turn motions of dimeric EGFR-ed and in an overall less flexible state of the conformation. The increase in the twist and turn

motions is more significant in the EGF-bound dimer. The observed values for the turn motion in the unglycosylated systems are comparable with a recent computational study (of timescale similar to the work presented here) carried out on the apo and EGF-bound dimers (40ns MD simulations for the EGF-bound dimer and 65ns of the apo dimer) [13]. Together the studies suggest that although glycosylation is not necessary for the occurrence of turn and twist motions, it promotes them and may actually be a function of cell type. A second explanation is that the absence of subdomain IV in the study by Poger et al. [13] reduces the flexibility of the complex to a higher extent and the turn motions are expected to be smoother. However, the models used here do contain subdomain IV and yet the values are still comparable, thus confirming the critical supporting role of the attached glycans. The twist motion values however were comparable with the same study for all the systems except the apo unglycosylated dimer in which the observed twist angle values were higher ($\sim 20^\circ$) than in the current study. This could have resulted from the flexibility increase caused by the loss of both EGF and subdomain IV dimeric contacts simultaneously. Overall, the increase of stability gained by glycosylation promotes turns and twist motions. Accordingly, it is expected that the glycosylated dimer interface adapts the most stable flush conformation. The simulations support this assumption by presenting a noticeable shift towards the flush conformation upon glycosylation although all simulations were started from the staggered conformation. A Recent $3\mu\text{s}$ simulation study carried out on the one-liganded and two-liganded glycosylated and unglycosylated EGFR dimers embedded in the membrane showed that upon glycosylation the two liganded staggered dimer gets closer to the flush conformation [52]. This is in good agreement with the observations in this study. In summary, the simulations from this work and from

those of others suggest that the currently available structural data (ie staggered conformations) may well be the result of un/underglycosylation during the preparation of the systems. Experiments where the terminal sugars in the attachment at Asn337 are removed and where Asn579 either also has a truncated glycan or is mutated to remove the ability to be glycosylated should remove the ability of the systems to adopt the flush states (using a well controlled glycosylated state to achieve the flush states) and/or remove the negative cooperativity attributed to the membrane interactions. Indeed, experiments have shown that mutations at Asn579 result in increased phosphorylation [33] which may well reflect the stabilization of the functional staggered state.

5.5 Conclusions

Molecular dynamics simulations presented here suggest that the dimeric form of EGFR is less flexible upon glycosylation. The decrease in the flexibility gained by glycosylation results in increases in the intramonomeric turn motion and the intradimeric twist motion which likely is communicated to the transmembrane and kinase domains, resulting in functional activation hence trigger the signaling cascade in the cytosol. Glycosylation also keeps the dimeric interface close to the flush conformation regardless of the presence of EGF which likely results in membrane interactions and the observed negative cooperativity of EGF binding.

Chapter 6. Conclusions and future directions

The EGFR extracellular domain is known to be heavily glycosylated [33, 47, 94]. Role of these glycan moieties in modulating the structure and dynamics of EGFR are found to be critical but not well understood at the atomic level [52, 55]. In this thesis extensive MD simulations have been carried out, analyzed and discussed to address the role of glycans in the biological functions of the EGFR extracellular domain. The subject is introduced in Chapter 1 with an extensive literature survey. Chapter 2 outlines the theory and methods of modelling and simulations used. In chapter 3 the effects of glycosylation on binding of three types of ligands (EGF, the dimeric partner EGFR, antibodies) to EGFR were investigated. These ligands are crucial stimulators (EGF and the dimeric partner) and inhibitors (antibodies) of EGFR. It was shown that upon glycosylation, interactions of two glycans attached to the growth factor binding subdomains (Asn151 in subdomain I and Asn328 in subdomain III) increase the binding affinity of EGF to EGFR by stabilizing the EGF binding site which leads to stronger non-bonded interactions between EGF and EGFR. This interaction could begin to provide part explanation of the observations in the literature that glycosylation of Asn151 and Asn328 is often seen in EGFR and hence must be functionally crucial [47] and that the mutation of glycosylation sites of subdomain III reduces the receptor's activation levels [32]. This mechanism is further supported by the existence of the same conserved glycosylation sites in the homologous family member ErbB4 which is also known to bind to EGF (the other two family members ErbB2 and ErbB3 have not been reported to bind EGF and bind instead to other growth factors, as do EGFR and ErbB4 also). As the above mentioned interaction between the two glycans does not alter the nature of the EGF-EGFR interactions and only stabilizes the EGF binding

site, the uncertainty about site and tissue-specific type of glycans was considered to be negligible. The finding that the interactions between the two glycans from the two growth factor binding subdomains modulates the stability and function of EGFR has not been reported elsewhere; however during the writing of this conclusion chapter, a manuscript appeared in the journal *Proteins* (Taylor S et al Conformational stability of the epidermal growth factor (EGF) receptor as influenced by glycosylation, dimerization and EGF hormone binding *Proteins: Str Fn and Bioinf*, appeared online Dec 26, 2016) which has carried out an examination of the effect of glycosylation on the stability of EGFR compared to the binding of EGF. However their work is only limited to using trisaccharides and only at six Asparagine sites and the total simulation time was only 50ns without replicates. While some of their findings are similar to the findings in this thesis, nevertheless, the work in this thesis involves a more complete representation of the EGFR-ed (the work here uses the crystal structure PDB ID 3NJP which contains all four subdomains of the EGFR-ed resolved while the recent study uses the structure PDB ID 1IVO which has only three subdomains of EGFR-ed resolved) more complex oligosaccharides, longer simulations and more extensive sampling (total simulation time for each system in this thesis is five times longer than reported in the recent paper; in addition, there was no study of the implications for antibody binding).

Future work on chapter 3 could explore the effects by modeling site-specific glycosylated EGFR at Asn151 and Asn328 as opposed to the fully glycosylated model used in this study and additionally test different plausible sequences of oligosaccharides at each site. In the experimental assays, the above mentioned mechanism could be tested by removing the glycosylation motif of the Asn151 and

Asn328 via point mutations or by introducing the glycosylation motif in other homologous family members (ErbB2 and ErbB3) which lack these sites (details can be found in chapter 3 section 3.2.2).

In the binding of the dimeric partner to EGFR, this thesis finds that glycans maintain the dimeric interfaces by stabilizing the regions in the neighborhood of the dimeric interfaces, by making stabilizing interactions between sugar units from the two monomers in the dimeric state.

Future work could explore the role of glycans in maintaining the dimeric interface of ErbB heterodimers by considering the variation of the glycosylation sites within the family members.

The thesis also finds that antibody-EGFR interactions seem to be mediated by the glycans. Modelling suggests that binding of antibodies to subdomain III of EGFR-ed does not change upon glycosylation. This may partly explain the efficiency of these antibodies currently in successful clinical trials such as those of Cetuximab [20] given the diversity of glycosylation patterns in different cell types [55]. However, for the case of mAb806, which is known to bind to subdomain II of EGFR-ed, models in this thesis suggest that glycosylation leads to slightly more exposed binding epitope (a separate study is currently ongoing and will be published at a later date). This is in agreement with the suggested role of glycans in exposure of the mAb806 cryptic epitope [100].

Future work can explore the role of glycans in mAb806 binding in greater detail by performing MD simulations of the full-length EGFR bound to mAb806 in the glycosylated and unglycosylated forms.

In chapter 4, the role of glycosylation in the tethered-extended equilibrium of monomeric EGFR was explored. It is known that the auto inhibited tethered form of EGFR transforms to the extended form upon binding of the growth factor [25]. The role of sugars in this process was unexplored. aMD simulations of fully glycosylated and unglycosylated EGFR monomer in the tethered and extended state showed that glycosylation stabilizes both the extended and tethered state of the receptor. However, in the transition of the tethered form towards the extended, glycosylation does not play an inhibitory role. The tethering interaction occurs between the dimerization arm of subdomain II and the two inhibitory loops of subdomain IV. The Asn579 glycosylation site is located in the second inhibitory loop. This site is known to be commonly glycosylated in EGFR [33]. Its mutation is known to increase the phosphorylation rate of the receptor, which was suggested to result from the removal of the strengthening role of the glycan moiety in the tethering interaction [33]. Such a mechanism was not observed in the simulations presented here.

Future work could extend the exploration of this mechanism by carrying out very long simulations as well as coarse-grained simulations. Furthermore, the possibility that Asn579 glycan interacts with the cell membrane [55] due to its proximity to the transmembrane domain, and hence being distant from the tethering interaction site, could be tested by coarse-grained simulations of membrane embedded EGFR.

Chapter 5 explores the mechanism [13] behind the activation of preformed dimers of EGFR [28, 102]. The mechanism of how the ligand-free dimer of EGFR is activated by binding of the growth factors was recently explained by the group of Allan Mark through simulations and proposed the "turn and twist" mechanism [13]. The effects of

glycosylation on this mechanism were explored in Chapter 5 using MD simulations of glycosylated apo and EGF-bound dimers. The simulations suggest that glycosylation promotes the intradimeric twist motions and inter-monomeric turn motions which occur between monomers and between subdomains I and III of each monomer in the dimeric construct respectively. Thus, glycosylation supports the twist and turn motion required for activation of the receptor. Combining these observations together with those from chapter 4, it can be concluded that by considering post-translational glycosylation, EGFR-ed is more likely to adopt the extended conformation and is activated by twist and turn motions of the stabilized dimer upon binding of the growth factors.

Future work could extend this study by extending the models to incorporation of the membrane as well as the kinase domains, perhaps using coarse grained models. In addition, the study could also be extended to the oligomeric states of EGFR which have recently been hypothesized to be required for function [7]. Long MD simulations and/or experiments of Asn337 and Asn579 with shorter glycans in the presence of membranes should remove the flush conformation and also the membrane associated negative cooperativity.

In summary, this thesis utilizes molecular modelling techniques and finds that the role of glycosylation as a crucial post-translational modification, occurring in half of the synthesized proteins and all of the transmembrane proteins [29], is very important in modulating the functional dynamics and interactions of the extracellular domain EGFR (and by extension the whole of EGFR pathway). The area remains largely unexplored at the atomistic level, although interest is slowly growing. The insights provided here for EGFR as one of the most pursued targets in drug design could be used as a general frame

work to explore other oncogenic receptors. Furthermore, these findings highlight the growing importance of computer simulations in the field of glycobiology to help in analyzing the large amounts of structural data related to glycoconjugates that is emerging.

Appendices

Appendix A Contribution of subdomains I-IV to EGF, monoclonal antibodies and the dimeric partner binding to EGFR

Appendix A Figure 1. The backbone RMSD plots of subdomains I-IV (A-D) of the EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Values belong to first simulation for the replicate simulations values see Appendix A, Figures 2 and 3.

Appendix A Figure 2. The backbone RMSD plots of subdomains I-IV (A-D) of the EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Data is from the first replicate simulation.

Appendix A Figure 3. The backbone RMSD plots of subdomains I-IV (A-D) of the EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed are shown in black, green, blue and magenta respectively. Data is from the second replicate simulation.

Appendix A Figure 4. The backbone RMSD plots of the EGF bound to unglycosylated EGFR and glycosylated EGFR are shown in blue and magenta respectively. A-C presents the values from three 100ns simulations of each system

Appendix A Figure 5. Swinging motion of subdomain IV towards subdomain II of EGFR-ed in the apo state. Subdomains I- IV of EGFR-ed are shown in blue, green, yellow and magenta cartoons and transparent surfaces. A shows the starting conformation. B and C are the final snapshots from two simulations.

Appendix A Figure 6. A-D) Subdomain I secondary structure modifications calculated over combined trajectory of three 100ns simulations of the EGFR-ed, EGF-EGFR-ed, gEGFR-ed and EGF-gEGFR-ed respectively. Glycosylated Asparagines are labeled with their number on the left side of each scheme. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix A Figure 7. A-D) Subdomain II secondary structure modifications calculated over combined trajectory of three 100ns simulations of the EGFR-ed, EGF-EGFR-ed, gEGFR-ed and EGF-gEGFR-ed respectively. Glycosylated Asparagine is labeled with its number on the left side of each scheme. Colors in the schemes present different type of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix A Figure 8. A-D) Subdomain III secondary structure modifications calculated over combined trajectory of three 100ns simulations of the EGFR-ed, EGF-EGFR-ed, gEGFR-ed and EGF-gEGFR-ed respectively. Glycosylated Asparagines are labeled with their number on the left side of each scheme. Colors in the schemes present different type of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix A Figure 9. A-D) Subdomain IV secondary structure modifications calculated over combined trajectory of three 100ns simulations of the EGFR-ed, EGF-EGFR-ed, gEGFR-ed and EGF-gEGFR-ed respectively. Glycosylated Asparagines are labeled with their number on the left side of each scheme. Colors in the schemes present different type of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix A Figure 10. Water density maps of EGF-EGFR-ed, EGF-gEGFR-ed (A) and EGFR-ed, gEGFR-ed (B). Highly dense water clusters of the glycosylated and unglycosylated receptors are shown with red and blue mesh respectively. EGFR-ed and EGF are shown in grey and green cartoon respectively.

EGFR residue	Oligosaccharide residue
Cys 191	Man 3-Asn172
Gln 193	Man 3-Asn 172
Oligosaccharide residue	Oligosaccharide residue
GalNac 1-Asn 32	GalNac 2-Asn 32
GalNac 1-Asn 172	GalNac 2-Asn 172
GalNac 1-Asn 328	GalNac 2-Asn 328
GalNac 1-Asn 337	GalNac 2-Asn 337
GalNac 1-Asn 420	GalNac 2-Asn 420
GalNac 1-Asn 579	GalNac 2-Asn 579

Appendix A Table 1. Hydrogen bonds formed between the glycans and EGFR-ed and within the oligosaccharide glycans in the EGF-bound state. All hydrogen bonds listed here are present in > 75 % of the snapshots of the 300ns combined trajectory. Residue numbers of the glycans are mentioned according to the number of the favored sugar in the specific oligosaccharide bound to the specific Asparagine as explained in Figure 3.6.

EGFR residue	Oligosaccharide residue
Asp 323	Man 3-Asn328
Oligosaccharide residue	Oligosaccharide residue
GalNac 1-Asn 151	GalNac 2-Asn 151
GalNac 1-Asn 172	GalNac 2-Asn 172
GalNac 1-Asn 337	GalNac 2-Asn 337
GalNac 1-Asn 420	GalNac 2-Asn 420
GalNac 1-Asn 504	GalNac 2-Asn 504

Appendix A Table 2. Hydrogen bonds formed between the glycans and EGFR-ed and within the oligosaccharide glycans in the apo state. All hydrogen bonds listed here are present in > 75 % of the snapshots of the 300ns combined trajectory. Residue numbers of the glycans are mentioned according to the number of the favored sugar in the specific oligosaccharide bound to the specific Asparagine as explained in Figure 3.6.

Appendix A Figure 11. Subdomain I of EGFR-ed per-residue vdW contribution to the binding free energy. Only residues within 4 Å of EGF are shown. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 3 and 4.

Appendix A Figure 12. Subdomain I of EGFR-ed per-residue ele contribution to the binding free energy. Only residues within 4 Å of EGF are shown. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 3 and 4.

DI in 4 Å of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First simulation										
N12	-28	0.0	-28	0.1	3.8	0.1	-0.3	0.0	-2.0	0.1
K13	-1.2	0.1	-36.9	1.0	36.0	0.8	-0.3	0.0	-2.5	0.1
L14	-3.2	0.1	0.1	0.0	1.0	0.0	-0.5	0.0	-2.6	0.1
T15	-3.9	0.1	-5.3	0.2	6.3	0.1	-0.6	0.0	-3.5	0.1
Q16	-3.5	0.1	-3.4	0.2	3.5	0.1	-0.4	0.0	-3.8	0.1
L17	-3.2	0.0	-1.6	0.1	2.2	0.0	-0.4	0.0	-3.0	0.1
G18	-1.3	0.0	-1.8	0.2	1.7	0.1	-0.3	0.0	-1.7	0.1
D22	-0.5	0.0	10.0	0.4	-7.5	0.5	-0.2	0.0	1.8	0.1
R29	-1.6	0.1	-27.9	0.7	26.0	0.6	-0.6	0.0	-4.2	0.2
Y45	-1.4	0.0	-0.1	0.0	1.1	0.0	-0.2	0.0	-0.6	0.0
L69	-1.8	0.0	-0.1	0.0	0.4	0.0	-0.3	0.0	-1.8	0.0
E90	0.0	0.0	0.0	0.9	0.4	0.9	-0.2	0.0	0.2	0.1
L98	-1.0	0.0	-0.3	0.0	0.5	0.0	-0.1	0.0	-0.9	0.0
S99	-0.7	0.1	-4.2	0.1	5.6	0.1	-0.2	0.0	0.5	0.1
Y101	-0.9	0.0	0.1	0.0	0.8	0.1	-0.2	0.0	-0.2	0.0
First replicate										
N12	-24	0.1	-2.2	0.1	3.1	0.1	-0.2	0.0	-1.7	0.1
K13	-0.9	0.1	-35.4	1.3	34.1	1.2	-0.3	0.0	-2.5	0.1
L14	-2.9	0.1	-0.2	0.0	1.1	0.1	-0.5	0.0	-2.5	0.1
T15	-3.3	0.1	-4.3	0.2	5.3	0.2	-0.5	0.0	-2.8	0.1
Q16	-2.9	0.1	-1.4	0.1	2.3	0.1	-0.3	0.0	-2.4	0.1
L17	-2.7	0.1	-1.3	0.1	1.7	0.1	-0.3	0.0	-2.6	0.1
G18	-0.7	0.1	-2.7	0.2	1.7	0.1	-0.2	0.0	-1.9	0.1
D22	-0.3	0.0	11.9	0.3	-10.6	0.3	0.0	0.0	1.0	0.0
R29	-2.2	0.1	-17.2	0.9	17.6	0.8	-0.5	0.0	-2.3	0.2
Y45	-1.2	0.1	-0.2	0.0	1.0	0.0	-0.1	0.0	-0.5	0.0
L69	-1.5	0.1	-0.1	0.0	0.4	0.0	-0.3	0.0	-1.5	0.1
E90	0.0	0.0	3.3	0.9	-2.8	0.8	-0.1	0.0	0.4	0.1
L98	-0.8	0.0	-0.3	0.0	0.4	0.0	-0.1	0.0	-0.8	0.0
S99	-0.6	0.1	-3.7	0.2	4.8	0.2	-0.2	0.0	0.3	0.1
Y101	-0.9	0.0	0.0	0.0	0.7	0.1	-0.2	0.0	-0.3	0.0

DI in 4Å of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
second replicate										
N12	-1.6	0.0	0.5	0.0	0.5	0.1	-0.1	0.0	-0.7	0.0
K13	-1.1	0.1	-40.7	0.5	38.9	0.4	-0.4	0.0	-3.1	0.1
L14	-3.0	0.0	0.2	0.0	1.0	0.0	-0.5	0.0	-2.3	0.1
T15	-3.5	0.1	-4.2	0.2	5.1	0.1	-0.6	0.0	-3.1	0.1
Q16	-3.4	0.1	-3.8	0.2	3.6	0.1	-0.4	0.0	-4.1	0.1
L17	-3.4	0.1	-1.6	0.1	2.0	0.0	-0.4	0.0	-3.4	0.1
G18	-1.1	0.1	-3.0	0.2	2.6	0.1	-0.3	0.0	-1.8	0.1
D22	-0.2	0.0	13.1	0.2	-11.8	0.2	0.0	0.0	1.0	0.0
R29	-1.9	0.1	-31.0	0.3	27.6	0.3	-0.6	0.0	-5.9	0.1
Y45	-1.2	0.0	-0.2	0.0	0.8	0.0	-0.1	0.0	-0.7	0.0
L69	-1.7	0.0	-0.1	0.0	0.5	0.0	-0.3	0.0	-1.6	0.0
E90	0.1	0.1	-4.8	0.8	4.9	0.8	-0.2	0.0	0.0	0.1
L98	-1.0	0.0	-0.4	0.0	0.6	0.0	-0.1	0.0	-1.0	0.0
S99	-0.6	0.1	-4.1	0.2	5.4	0.2	-0.2	0.0	0.5	0.1
Y101	-0.8	0.0	0.1	0.0	0.7	0.1	-0.2	0.0	-0.2	0.0
Average										
N12	-2.3	0.1	-1.5	0.1	2.5	0.1	-0.2	0.0	-1.5	0.1
K13	-1.0	0.1	-37.7	0.9	36.3	0.8	-0.3	0.0	-2.7	0.1
L14	-3.0	0.1	0.0	0.0	1.0	0.0	-0.5	0.0	-2.5	0.1
T15	-3.6	0.1	-4.6	0.2	5.6	0.2	-0.5	0.0	-3.1	0.1
Q16	-3.3	0.1	-2.9	0.2	3.2	0.1	-0.4	0.0	-3.4	0.1
L17	-3.1	0.1	-1.5	0.1	1.9	0.1	-0.4	0.0	-3.0	0.1
G18	-1.0	0.1	-2.5	0.2	2.0	0.1	-0.3	0.0	-1.8	0.1
D22	-0.3	0.0	11.6	0.3	-10.0	0.3	-0.1	0.0	1.3	0.0
R29	-1.9	0.1	-25.4	0.7	23.7	0.5	-0.6	0.0	-4.1	0.2
Y45	-1.3	0.0	-0.2	0.0	1.0	0.0	-0.2	0.0	-0.6	0.0
L69	-1.7	0.0	-0.1	0.0	0.5	0.0	-0.3	0.0	-1.6	0.0
E90	0.0	0.1	-0.5	0.9	0.9	0.8	-0.2	0.0	0.2	0.1
L98	-0.7	0.1	-4.0	0.2	5.3	0.2	-0.2	0.0	0.4	0.1
S99	-0.7	0.1	-4.0	0.2	5.3	0.2	-0.2	0.0	0.4	0.1
Y101	-0.9	0.0	0.1	0.0	0.8	0.1	-0.2	0.0	-0.2	0.0

Appendix A Table 3. Subdomain I of the unglycosylated EGFR-ed per-residue contribution to the binding free energy calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔG GB), non-polar contribution to the solvation energy (ΔG SA) and binding free energy (ΔG MMGBSA).

DI in 4Å of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First simulation										
N12	-29	00	-21	01	3.3	0.1	-0.3	00	-1.9	0.1
K13	-1.5	00	-28.9	08	28.4	0.6	-0.1	00	-2.2	0.2
L14	-3.1	01	01	00	1.1	0.0	-0.5	00	-2.5	0.1
T15	-4.0	01	-6.0	02	6.7	0.1	-0.6	00	-3.9	0.1
Q16	-3.7	01	-2.9	02	3.4	0.1	-0.4	00	-3.6	0.1
L17	-3.5	00	-1.7	01	2.2	0.0	-0.4	00	-3.4	0.1
G18	-1.2	00	-2.1	02	1.8	0.1	-0.2	00	-1.7	0.1
D22	-0.5	00	11.1	02	-9.0	0.3	-0.1	00	1.5	0.1
R29	-2.8	01	-12.0	06	13.7	0.5	-0.6	00	-1.8	0.1
Y45	-1.3	00	-0.2	00	1.0	0.0	-0.2	00	-0.6	0.0
L69	-1.8	00	-0.1	00	0.5	0.0	-0.3	00	-1.8	0.0
E90	0.2	01	-7.4	08	5.7	0.6	-0.2	00	-1.7	0.2
I98	-1.0	00	-0.4	00	0.5	0.0	-0.2	00	-1.0	0.0
S99	-0.8	01	-3.8	02	5.3	0.2	-0.2	00	0.4	0.1
Y101	-1.0	00	0.0	01	1.0	0.1	-0.2	00	-0.2	0.0
First replicate										
N12	-2.6	01	-2.0	02	3.5	0.2	-0.2	00	-1.4	0.1
K13	-0.8	01	-41.1	03	38.1	0.3	-0.2	00	-4.0	0.1
L14	-3.0	01	-0.2	00	1.1	0.0	-0.5	00	-2.6	0.1
T15	-4.3	01	-7.1	02	7.6	0.2	-0.6	00	-4.4	0.2
Q16	-3.3	01	-4.4	02	3.8	0.1	-0.4	00	-4.2	0.1
L17	-3.9	00	-1.3	01	2.4	0.1	-0.5	00	-3.3	0.1
G18	-1.7	01	-1.5	02	1.8	0.2	-0.4	00	-1.8	0.1
D22	-1.3	00	7.6	05	-5.2	0.5	-0.3	00	0.8	0.0
R29	-3.4	01	-28.2	03	24.5	0.3	-0.8	00	-7.9	0.2
Y45	-1.4	00	-0.1	00	0.9	0.0	-0.2	00	-0.8	0.0
L69	-1.5	00	-0.1	00	0.4	0.0	-0.3	00	-1.4	0.0
E90	0.3	01	-7.4	06	5.9	0.4	-0.1	00	-1.3	0.1
I98	-1.1	00	-0.3	00	0.6	0.0	-0.1	00	-1.0	0.0
S99	-1.1	01	-3.0	02	4.5	0.3	-0.2	00	0.1	0.1
Y101	-0.8	00	0.2	01	0.6	0.1	-0.2	00	-0.2	0.0

DI in 4Å of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Second replicate										
N12	-2.9	0.0	-2.2	0.1	3.4	0.1	-0.3	0.0	-2.0	0.1
K13	-1.9	0.0	-25.0	0.7	25.1	0.5	-0.2	0.0	-2.0	0.2
L14	-3.1	0.1	0.2	0.0	1.0	0.0	-0.5	0.0	-2.4	0.1
T15	-4.1	0.1	-6.7	0.2	7.3	0.1	-0.6	0.0	-4.2	0.1
Q16	-3.4	0.1	-2.7	0.2	3.0	0.1	-0.4	0.0	-3.5	0.1
L17	-3.3	0.0	-1.6	0.1	2.3	0.1	-0.4	0.0	-3.0	0.0
G18	-1.2	0.1	-1.9	0.2	1.7	0.1	-0.3	0.0	-1.8	0.1
D22	-0.7	0.0	12.1	0.3	-10.2	0.3	-0.2	0.0	1.0	0.1
R29	-3.8	0.1	-10.6	0.3	13.5	0.3	-0.8	0.0	-1.7	0.1
Y45	-1.3	0.0	-0.2	0.0	1.1	0.0	-0.2	0.0	-0.6	0.0
L69	-1.8	0.0	-0.1	0.0	0.5	0.0	-0.3	0.0	-1.7	0.0
E90	0.1	0.1	-5.0	0.7	2.5	0.6	-0.1	0.0	-2.5	0.2
L98	-0.9	0.0	-0.4	0.0	0.7	0.0	-0.2	0.0	-0.9	0.0
S99	-1.1	0.1	-1.5	0.3	3.0	0.3	-0.3	0.0	0.1	0.1
Y101	-0.2	0.0	0.7	0.0	-0.5	0.0	0.0	0.0	0.0	0.0
Average										
N12	-2.8	0.1	-2.1	0.1	3.4	0.1	-0.3	0.0	-1.8	0.1
K13	-1.4	0.1	-31.7	0.6	30.5	0.4	-0.1	0.0	-2.8	0.2
L14	-3.1	0.1	0.0	0.0	1.1	0.0	-0.5	0.0	-2.5	0.1
T15	-4.1	0.1	-6.6	0.2	7.2	0.1	-0.6	0.0	-4.2	0.1
Q16	-3.4	0.1	-3.3	0.2	3.4	0.1	-0.4	0.0	-3.8	0.1
L17	-3.6	0.0	-1.5	0.1	2.3	0.1	-0.5	0.0	-3.2	0.1
G18	-1.4	0.1	-1.8	0.2	1.8	0.1	-0.3	0.0	-1.8	0.1
D22	-0.8	0.0	10.3	0.4	-8.1	0.4	-0.2	0.0	1.1	0.1
R29	-3.4	0.1	-16.9	0.4	17.2	0.4	-0.7	0.0	-3.8	0.1
Y45	-1.3	0.0	-0.2	0.0	1.0	0.0	-0.2	0.0	-0.6	0.0
L69	-1.7	0.0	-0.1	0.0	0.5	0.0	-0.3	0.0	-1.6	0.0
E90	0.2	0.1	-6.6	0.7	4.7	0.5	-0.1	0.0	-1.8	0.2
L98	-1.0	0.0	-0.4	0.0	0.6	0.0	-0.2	0.0	-0.9	0.0
S99	-1.0	0.1	-2.8	0.2	4.3	0.3	-0.3	0.0	0.2	0.1
Y101	-0.7	0.0	0.3	0.0	0.4	0.0	-0.1	0.0	-0.1	0.0

Appendix A Table 4. Subdomain I of the glycosylated with oligosaccharides EGFR-ed per-residue contribution to the binding free energy calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔG GB), non-polar contribution to the solvation energy (ΔG SA) and binding free energy (ΔG MMGBSA).

Appendix A Figure 13. Subdomain III of EGFR-ed per-residue vdw contribution to the binding free energy. Only residues within 4 Å of EGF are shown. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 5 and 6.

Appendix A Figure 14. Subdomain III of EGFR-ed per-residue ele contribution to the binding free energy. Only residues within 4 Å of EGF are shown. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 5 and 6.

DIII in 4Å of ECF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First simulation										
HB46	-0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0
L348	-1.0	0.0	0.2	0.0	0.4	0.0	-0.2	0.0	-0.6	0.0
R353	-1.0	0.1	-28.1	1.5	28.0	1.3	-0.3	0.0	-1.4	0.3
D855	-0.6	0.1	-2.8	0.3	3.6	0.3	-0.2	0.0	0.0	0.1
S356	-0.6	0.1	-2.8	0.3	3.6	0.3	-0.2	0.0	0.0	0.1
F357	-4.8	0.1	0.6	0.1	1.1	0.1	-0.8	0.0	-3.9	0.1
T358	-0.6	0.0	1.4	0.1	-1.1	0.0	0.0	0.0	-0.3	0.0
L382	-0.4	0.0	0.2	0.0	0.0	0.0	0.0	0.0	-0.3	0.0
Q884	-1.6	0.0	-1.3	0.2	3.4	0.1	-0.4	0.0	0.2	0.1
Q408	-0.9	0.0	-0.5	0.1	1.8	0.1	-0.2	0.0	0.2	0.0
H409	-2.3	0.2	-3.1	0.4	4.4	0.4	-0.5	0.0	-1.5	0.2
Q411	-0.2	0.0	-1.3	0.2	1.5	0.2	-0.1	0.0	0.0	0.0
F412	-0.9	0.0	0.3	0.0	-0.2	0.0	-0.1	0.0	-0.9	0.1
A415	-0.2	0.0	0.2	0.0	-0.1	0.0	0.0	0.0	-0.2	0.0
V417	-0.9	0.0	0.3	0.0	-0.1	0.0	-0.2	0.0	-0.9	0.1
S418	-0.5	0.0	-0.4	0.1	1.0	0.1	-0.1	0.0	0.0	0.0
I438	-0.9	0.0	0.1	0.0	-0.1	0.0	-0.2	0.0	-1.0	0.0
I467	-1.6	0.1	-0.1	0.0	0.2	0.0	-0.3	0.0	-1.8	0.1

First replicate										
HB46	-0.2	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.1	0.0
L348	-1.6	0.1	-0.7	0.0	1.0	0.0	-0.2	0.0	-1.6	0.1
R353	-1.1	0.1	-46.7	1.9	43.3	1.7	-0.6	0.0	-5.1	0.3
D855	0.0	0.1	-4.4	0.5	-0.2	0.3	-0.2	0.0	-4.8	0.2
S356	-0.7	0.0	-1.1	0.1	2.2	0.1	-0.1	0.0	0.2	0.0
F357	-4.2	0.2	0.5	0.1	0.8	0.1	-0.7	0.0	-3.7	0.2
T358	-0.5	0.0	1.2	0.1	-1.0	0.1	0.0	0.0	-0.2	0.0
L382	-0.8	0.0	0.2	0.0	0.1	0.0	-0.1	0.0	-0.6	0.0
Q884	-2.0	0.1	-6.3	0.3	7.0	0.3	-0.4	0.0	-1.8	0.1
Q408	-1.0	0.0	-0.6	0.1	1.7	0.1	-0.2	0.0	-0.1	0.1
H409	-1.2	0.1	-1.7	0.2	2.4	0.2	-0.3	0.0	-0.7	0.1
Q411	-0.1	0.0	-0.9	0.2	1.1	0.2	0.0	0.0	0.1	0.0
F412	-0.8	0.0	0.4	0.0	-0.4	0.0	-0.1	0.0	-0.9	0.1
A415	-0.4	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.2	0.0
V417	-1.8	0.1	-0.7	0.0	1.2	0.1	-0.2	0.0	-1.6	0.1
S418	-1.0	0.0	-1.0	0.1	2.3	0.2	-0.3	0.0	0.0	0.0
I438	-1.3	0.1	0.2	0.0	-0.1	0.0	-0.2	0.0	-1.4	0.1
I467	-1.0	0.1	0.1	0.0	0.0	0.0	-0.2	0.0	-1.1	0.1

DI in 4Å of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Second replicate										
FB46	-0.3	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.1	0.0
L348	-2.0	0.0	-0.3	0.0	1.0	0.0	-0.2	0.0	-1.4	0.0
R353	-0.3	0.0	-19.9	0.5	20.5	0.5	-0.1	0.0	0.2	0.0
D355	0.3	0.1	-5.9	0.3	0.5	0.3	-0.1	0.0	-5.2	0.1
S356	-0.8	0.0	-2.5	0.3	3.5	0.3	-0.2	0.0	0.0	0.1
F357	-5.2	0.1	0.8	0.1	1.0	0.1	-0.9	0.0	-4.2	0.1
T358	-0.5	0.0	1.2	0.1	-1.0	0.1	0.0	0.0	-0.3	0.0
L382	-1.1	0.0	0.1	0.0	0.2	0.0	-0.1	0.0	-0.9	0.0
Q384	-2.8	0.1	-4.7	0.2	6.3	0.1	-0.4	0.0	-1.7	0.1
Q408	-1.6	0.0	-1.9	0.1	3.6	0.1	-0.3	0.0	-0.2	0.1
H409	-2.9	0.1	-4.2	0.6	5.6	0.5	-0.4	0.0	-2.0	0.2
Q411	-0.4	0.1	-6.3	0.3	6.3	0.3	-0.2	0.0	-0.7	0.1
F412	-2.2	0.0	0.2	0.0	0.0	0.0	-0.2	0.0	-2.2	0.0
A415	-0.5	0.0	0.0	0.0	0.2	0.0	0.0	0.0	-0.4	0.0
V417	-1.4	0.0	0.6	0.0	-0.3	0.0	-0.2	0.0	-1.3	0.0
S418	-0.7	0.0	-0.4	0.1	1.1	0.1	-0.2	0.0	-0.2	0.1
I438	-1.4	0.0	-0.1	0.0	0.2	0.0	-0.2	0.0	-1.5	0.0
I467	-0.7	0.0	-0.2	0.0	0.3	0.0	-0.2	0.0	-0.7	0.0

Average										
FB46	-0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.1	0.0
L348	-1.5	0.0	-0.3	0.0	0.8	0.0	-0.2	0.0	-1.2	0.0
R353	-0.8	0.1	-31.6	1.3	30.6	1.2	-0.3	0.0	-2.1	0.2
D355	-0.1	0.1	-4.4	0.4	1.3	0.3	-0.2	0.0	-3.3	0.1
S356	-0.7	0.0	-2.1	0.2	3.1	0.2	-0.2	0.0	0.0	0.0
F357	-4.7	0.1	0.6	0.1	1.0	0.1	-0.8	0.0	-4.0	0.1
T358	-0.5	0.0	1.3	0.1	-1.0	0.1	0.0	0.0	-0.3	0.0
L382	-0.8	0.0	0.1	0.0	0.1	0.0	-0.1	0.0	-0.6	0.0
Q384	-2.1	0.1	-4.1	0.2	5.6	0.2	-0.4	0.0	-1.1	0.1
Q408	-1.2	0.0	-1.0	0.1	2.4	0.1	-0.2	0.0	0.0	0.0
H409	-2.1	0.1	-3.0	0.4	4.2	0.4	-0.4	0.0	-1.4	0.2
Q411	-0.2	0.0	-2.8	0.3	2.9	0.2	-0.1	0.0	-0.2	0.1
F412	-1.3	0.0	0.3	0.0	-0.2	0.0	-0.2	0.0	-1.4	0.0
A415	-0.4	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.2	0.0
V417	-1.4	0.1	0.1	0.0	0.3	0.0	-0.2	0.0	-1.2	0.1
S418	-0.7	0.0	-0.6	0.1	1.5	0.1	-0.2	0.0	-0.1	0.0
I438	-1.2	0.0	0.1	0.0	0.0	0.0	-0.2	0.0	-1.3	0.0
I467	-1.1	0.1	-0.1	0.0	0.2	0.0	-0.2	0.0	-1.2	0.0

Appendix A Table 5. Subdomain III of the unglycosylated EGFR-ed per-residue contribution to the binding free energy calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔG GB), non-polar contribution to the solvation energy (ΔG SA) and binding free energy (ΔG MMGBSA).

DIII in 4A of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First simulation										
HB46	-0.5	0.0	-0.6	0.1	1.0	0.1	-0.1	0.0	-0.2	0.0
L348	-1.6	0.0	0.0	0.0	0.7	0.0	-0.3	0.0	-1.2	0.0
R353	-0.3	0.0	-20.6	0.4	21.3	0.5	-0.1	0.0	0.3	0.0
D355	0.4	0.1	-6.3	0.4	-0.3	0.4	-0.1	0.0	-6.4	0.1
S356	-0.6	0.0	-2.8	0.3	3.7	0.3	-0.2	0.0	0.1	0.0
F357	-4.4	0.1	0.8	0.1	0.9	0.1	-0.7	0.0	-3.5	0.1
T358	-0.5	0.0	1.5	0.1	-1.5	0.0	0.0	0.0	-0.6	0.0
L382	-0.8	0.0	0.1	0.0	0.1	0.0	-0.1	0.0	-0.7	0.0
Q884	-1.7	0.0	-1.5	0.1	3.8	0.1	-0.4	0.0	0.1	0.1
Q408	-1.6	0.0	-1.7	0.1	3.3	0.1	-0.3	0.0	-0.2	0.0
H409	-3.1	0.1	-4.7	0.2	6.6	0.2	-0.7	0.0	-1.8	0.1
Q411	-0.5	0.0	-3.7	0.3	4.2	0.3	-0.2	0.0	-0.2	0.1
F412	-2.0	0.0	0.2	0.0	-0.1	0.0	-0.2	0.0	-2.1	0.0
A415	-0.4	0.0	0.0	0.0	0.1	0.0	0.0	0.0	-0.3	0.0
V417	-0.8	0.0	0.4	0.0	-0.2	0.0	-0.2	0.0	-0.8	0.0
S418	-0.2	0.0	0.3	0.0	0.1	0.0	0.0	0.0	0.2	0.0
I438	-1.2	0.0	-0.1	0.0	0.2	0.0	-0.2	0.0	-1.3	0.0
I467	-0.4	0.0	0.1	0.0	0.0	0.0	-0.1	0.0	-0.5	0.0

First replicate										
HB46	-0.4	0.0	0.0	0.0	0.4	0.1	0.0	0.0	0.1	0.0
L348	-1.8	0.0	-0.1	0.0	0.7	0.0	-0.3	0.0	-1.5	0.0
R353	-0.6	0.0	-25.6	0.5	26.7	0.6	-0.1	0.0	0.4	0.0
D355	0.3	0.1	-4.4	0.3	-2.7	0.3	-0.1	0.0	-6.9	0.1
S356	-0.8	0.0	-1.8	0.3	3.1	0.3	-0.2	0.0	0.2	0.0
F357	-4.8	0.1	1.1	0.1	0.5	0.1	-0.8	0.0	-3.9	0.1
T358	-0.5	0.0	1.7	0.1	-1.5	0.0	0.0	0.0	-0.3	0.0
L382	-1.1	0.0	0.1	0.0	0.2	0.0	-0.1	0.0	-0.9	0.0
Q884	-2.3	0.1	-6.9	0.2	8.0	0.1	-0.5	0.0	-1.7	0.1
Q408	-1.3	0.1	-2.4	0.1	3.2	0.1	-0.2	0.0	-0.8	0.1
H409	-4.0	0.1	-9.6	0.3	9.9	0.2	-0.5	0.0	-4.2	0.2
Q411	-0.2	0.1	-7.3	0.4	6.9	0.3	-0.2	0.0	-0.9	0.1
F412	-1.7	0.0	0.2	0.0	0.0	0.0	-0.2	0.0	-1.8	0.0
A415	-0.5	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.3	0.0
V417	-0.9	0.0	0.1	0.0	0.2	0.0	-0.2	0.0	-0.7	0.0
S418	-0.2	0.0	0.1	0.1	0.2	0.1	-0.1	0.0	0.0	0.1
I438	-1.2	0.0	0.1	0.1	0.0	0.0	-0.2	0.0	-1.3	0.0
I467	-0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0

DI in 4A of EGF	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Second replicate										
FB46	-0.4	0.0	-0.3	0.0	0.7	0.0	-0.1	0.0	0.0	0.0
L348	-1.8	0.0	-0.7	0.0	1.4	0.0	-0.2	0.0	-1.4	0.0
R353	-0.5	0.0	-25.9	0.5	26.9	0.5	-0.1	0.0	0.4	0.0
D355	0.3	0.1	-7.1	0.3	0.0	0.3	-0.1	0.0	-6.9	0.1
S356	-0.6	0.0	-1.1	0.1	2.1	0.1	-0.1	0.0	0.2	0.0
F357	-4.6	0.1	0.8	0.1	0.9	0.1	-0.7	0.0	-3.7	0.1
T358	-0.5	0.0	1.5	0.0	-1.6	0.0	0.0	0.0	-0.6	0.0
L382	-1.0	0.0	0.1	0.0	0.2	0.0	-0.1	0.0	-0.8	0.0
Q384	-2.6	0.1	-6.5	0.2	7.3	0.1	-0.5	0.0	-2.3	0.1
Q408	-1.8	0.0	-1.9	0.1	3.8	0.1	-0.3	0.0	-0.3	0.0
H409	-2.8	0.1	-4.2	0.5	6.1	0.4	-0.6	0.0	-1.5	0.2
Q411	-0.4	0.0	-2.6	0.3	2.9	0.3	-0.1	0.0	-0.1	0.1
F412	-2.0	0.0	0.1	0.0	0.1	0.0	-0.2	0.0	-2.0	0.0
A415	-0.5	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.4	0.0
V417	-1.4	0.0	0.8	0.0	-0.3	0.0	-0.3	0.0	-1.2	0.0
S418	-0.4	0.0	0.3	0.1	0.4	0.1	-0.1	0.0	0.1	0.0
I438	-1.4	0.0	0.0	0.0	0.2	0.0	-0.2	0.0	-1.4	0.0
I467	-0.5	0.0	0.0	0.1	0.0	0.0	-0.1	0.0	-0.6	0.0

Average										
FB46	-0.4	0.0	-0.3	0.1	0.7	0.1	-0.1	0.0	-0.1	0.0
L348	-1.7	0.0	-0.3	0.0	0.9	0.0	-0.3	0.0	-1.4	0.0
R353	-0.4	0.0	-24.0	0.5	24.9	0.5	-0.1	0.0	0.4	0.0
D355	0.3	0.1	-5.9	0.3	-1.0	0.3	-0.1	0.0	-6.7	0.1
S356	-0.7	0.0	-1.9	0.2	2.9	0.2	-0.2	0.0	0.2	0.0
F357	-4.6	0.1	0.9	0.1	0.8	0.1	-0.7	0.0	-3.7	0.1
T358	-0.5	0.0	1.6	0.1	-1.5	0.0	0.0	0.0	-0.5	0.0
L382	-1.0	0.0	0.1	0.0	0.2	0.0	-0.1	0.0	-0.8	0.0
Q384	-2.2	0.1	-5.0	0.2	6.4	0.1	-0.4	0.0	-1.3	0.1
Q408	-1.6	0.0	-2.0	0.1	3.4	0.1	-0.3	0.0	-0.4	0.0
H409	-3.3	0.1	-6.2	0.3	7.6	0.3	-0.6	0.0	-2.5	0.1
Q411	-0.4	0.1	-4.6	0.3	4.7	0.3	-0.2	0.0	-0.4	0.1
F412	-1.9	0.0	0.2	0.0	0.0	0.0	-0.2	0.0	-2.0	0.0
A415	-0.5	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.3	0.0
V417	-1.0	0.0	0.4	0.0	-0.1	0.0	-0.2	0.0	-0.9	0.0
S418	-0.3	0.0	0.2	0.1	0.2	0.1	-0.1	0.0	0.1	0.0
I438	-1.3	0.0	0.0	0.0	0.1	0.0	-0.2	0.0	-1.4	0.0
I467	-1.3	0.0	0.0	0.0	0.1	0.0	-0.2	0.0	-1.4	0.0

Appendix A Table 6. Subdomain III of the glycosylated with oligosaccharides EGFR-ed per-residue contribution to the binding free energy calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔG GB), non-polar contribution to the solvation energy (ΔG SA) and binding free energy (ΔG MMGBSA).

Appendix A Figure 15. Arg353 in subdomain III of EGFR-ed shows unfavorable total energy in the EGF-EGFR binding free energy. A) Subdomain I-IV of EGFR-ed and EGF are shown in blue, green, yellow, magenta and red cartoons and transparent surfaces. The box shows the enlarged view of the hydrogen bond between Arg353(EGFR-ed)-His16(EGF) with stick presentation. B) The distance between the interacting atoms of Arg353(EGFR-ed)-His16(EGF). C) Solvent Accessible Surface Area of Arg353. Values from the EGF-EGFR-ed and EGF-gEGFR-ed complex are shown in blue and magenta respectively.

Appendix A Figure 16. EGF per-residue vdw contribution to the binding free energy. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 7 and 8.

Appendix A Figure 17. EGF per-residue ele contribution to the binding free energy. Values from the calculations of the EGF bound to EGFR-ed and gEGFR-ed are shown in blue and magenta respectively. Represented values for each amino acid is the average value calculated from the last 50ns of all simulations for each complex. Values for the first simulation and its two replicates are listed in Appendix A, Tables 7 and 8.

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First simulation										
C6	-0.1	0.0	-0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0
P7	0.0	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.2	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
S9	-0.3	0.0	0.4	0.0	0.2	0.0	-0.1	0.0	0.2	0.0
H10	-0.7	0.0	0.3	0.0	0.5	0.0	-0.1	0.0	0.0	0.0
D11	-0.6	0.1	9.0	0.3	-7.8	0.3	-0.1	0.0	0.5	0.0
G12	-0.9	0.0	1.6	0.1	-0.6	0.1	-0.1	0.0	-0.1	0.0
Y13	-2.2	0.0	0.2	0.1	0.6	0.1	-0.2	0.0	-1.7	0.0
C14	-0.3	0.0	0.2	0.0	0.3	0.0	0.0	0.0	0.2	0.0
L15	-1.8	0.0	0.7	0.1	0.0	0.1	-0.3	0.0	-1.4	0.0
H16	-2.4	0.1	-0.5	0.2	3.1	0.1	-0.5	0.0	-0.2	0.1
D17	0.0	0.0	-5.2	1.3	5.0	1.2	-0.1	0.0	-0.3	0.1
G18	-0.1	0.0	-0.5	0.0	1.0	0.0	0.0	0.0	0.3	0.0
V19	-0.2	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	-0.2	0.0
C20	-0.1	0.0	-0.1	0.0	0.3	0.0	0.0	0.0	0.1	0.0
M21	-1.7	0.0	-0.2	0.0	0.5	0.0	-0.3	0.0	-1.6	0.0
Y22	-0.2	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.2	0.0
I23	-2.2	0.0	-0.5	0.0	0.8	0.0	-0.3	0.0	-2.2	0.0
E24	-0.4	0.0	11.1	0.3	-10.1	0.3	0.0	0.0	0.7	0.0
A25	-1.5	0.1	-4.8	0.1	4.8	0.1	-0.5	0.0	-2.0	0.1
I26	-3.6	0.1	-0.4	0.0	1.1	0.0	-0.7	0.0	-3.6	0.1
D27	-0.2	0.0	11.5	0.2	-10.7	0.2	0.0	0.0	0.6	0.0
K28	-1.0	0.1	-22.6	1.1	23.3	1.0	-0.3	0.0	-0.6	0.1
Y29	-0.4	0.0	-0.8	0.0	1.5	0.0	0.0	0.0	0.3	0.0
A30	-1.5	0.0	-0.2	0.0	0.5	0.0	-0.2	0.0	-1.3	0.0
C31	-1.9	0.0	-2.7	0.1	2.3	0.0	-0.2	0.0	-2.5	0.1
N32	-3.1	0.1	-4.4	0.3	5.8	0.2	-0.4	0.0	-2.0	0.1
C33	-2.1	0.1	-0.9	0.1	1.5	0.1	-0.3	0.0	-1.8	0.1
V34	-0.6	0.0	-0.8	0.0	1.1	0.0	0.0	0.0	-0.3	0.0
V35	-0.3	0.0	-2.2	0.1	2.3	0.1	0.0	0.0	-0.2	0.0
G36	-0.4	0.0	-2.1	0.1	2.3	0.1	0.0	0.0	-0.2	0.0
Y37	-1.5	0.0	0.6	0.1	1.0	0.1	-0.2	0.0	-0.1	0.0
I38	-2.3	0.0	-1.1	0.0	1.4	0.0	-0.3	0.0	-2.4	0.0
G39	-1.6	0.1	-4.5	0.2	4.0	0.1	-0.3	0.0	-2.4	0.1
E40	-3.2	0.1	-10.8	1.0	13.8	0.9	-0.6	0.0	-0.8	0.1
R41	-1.6	0.1	-39.4	0.4	34.6	0.4	-0.5	0.0	-6.9	0.1
C42	-0.8	0.0	-0.4	0.0	1.0	0.0	0.0	0.0	-0.2	0.0
Q43	-1.9	0.0	-1.5	0.1	3.5	0.1	-0.3	0.0	-0.2	0.1
Y44	-2.8	0.1	-0.3	0.1	2.6	0.2	-0.6	0.0	-1.0	0.1
R45	-1.6	0.1	6.9	0.5	-4.4	0.5	-0.2	0.0	0.7	0.1
D46	-0.8	0.1	-27.6	1.1	27.2	0.9	-0.2	0.0	-1.4	0.2
L47	-4.2	0.1	-1.8	2.5	3.0	0.2	-1.0	0.0	-4.0	0.1
K48	-2.3	0.1	13.5	0.9	-9.4	0.9	-0.5	0.0	1.3	0.1
W49	-1.8	0.1	-3.8	0.3	3.7	0.2	-0.3	0.0	-2.2	0.2
W50	-3.8	0.1	-2.3	0.3	3.9	0.2	-0.7	0.0	-2.8	0.1

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First replicate										
C6	-0.1	0.0	-0.2	0.0	0.3	0.0	0.0	0.0	0.0	0.0
P7	0.0	0.0	0.2	0.0	-0.2	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.4	0.0	-0.2	0.0	0.0	0.0	0.0	0.0
S9	-0.3	0.0	0.6	0.0	-0.1	0.0	-0.1	0.0	0.1	0.0
H10	-0.9	0.1	0.5	0.0	0.5	0.1	-0.1	0.0	-0.1	0.0
D11	-0.3	0.0	10.0	0.2	-9.4	0.2	0.0	0.0	0.3	0.0
G12	-1.1	0.0	1.7	0.1	-0.6	0.1	-0.2	0.0	-0.1	0.0
Y13	-2.0	0.1	-0.1	0.1	0.7	0.1	-0.2	0.0	-1.6	0.1
C14	-0.3	0.0	0.5	0.0	0.1	0.0	0.0	0.0	0.0	0.0
L15	-2.1	0.1	0.5	0.1	0.6	0.0	-0.4	0.0	-1.4	0.1
H16	-2.2	0.1	-2.5	0.2	3.9	0.2	-0.4	0.0	-1.1	0.1
D17	0.2	0.1	-19.4	1.5	17.4	1.3	-0.2	0.0	-2.0	0.1
G18	-0.1	0.0	-0.4	0.0	0.8	0.0	0.0	0.0	0.3	0.0
V19	-0.2	0.0	0.2	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0
C20	-0.1	0.0	-0.2	0.0	0.4	0.0	0.0	0.0	0.1	0.0
M21	-1.4	0.1	0.0	0.0	0.3	0.0	-0.2	0.0	-1.3	0.1
Y22	-0.2	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.2	0.0
I23	-1.9	0.1	-0.3	0.0	0.6	0.0	-0.3	0.0	-1.8	0.1
E24	-0.3	0.0	8.1	0.3	-7.3	0.3	0.0	0.0	0.5	0.0
A25	-1.3	0.1	-4.3	0.2	4.1	0.2	-0.5	0.0	-2.0	0.1
L26	-3.0	0.1	-0.2	0.0	0.8	0.0	-0.6	0.0	-2.9	0.1
D27	-0.2	0.0	11.1	0.3	-10.5	0.3	0.0	0.0	0.5	0.0
K28	-0.9	0.1	-21.3	1.1	22.0	1.1	-0.3	0.0	-0.4	0.1
Y29	-0.4	0.0	-0.3	0.0	1.0	0.1	0.0	0.0	0.2	0.0
A30	-1.2	0.1	-0.2	0.0	0.6	0.0	-0.2	0.0	-1.0	0.0
C31	-1.7	0.1	-2.0	0.1	1.8	0.1	-0.1	0.0	-2.1	0.1
N32	-2.7	0.1	-4.6	0.3	5.7	0.3	-0.4	0.0	-2.0	0.1
C33	-1.7	0.1	0.1	0.1	0.5	0.1	-0.2	0.0	-1.3	0.1
V34	-0.3	0.0	-0.4	0.0	0.6	0.0	0.0	0.0	-0.1	0.0
V35	-0.2	0.0	-1.0	0.1	1.1	0.1	0.0	0.0	-0.1	0.0
G36	-0.2	0.0	-1.1	0.1	1.4	0.1	0.0	0.0	0.0	0.0
Y37	-1.9	0.1	0.3	0.1	1.6	0.1	-0.2	0.0	-0.3	0.0
I38	-2.0	0.1	-0.9	0.1	1.2	0.1	-0.3	0.0	-2.1	0.1
G39	-1.4	0.1	-3.9	0.2	3.4	0.1	-0.2	0.0	-2.1	0.1
E40	-2.7	0.1	-10.9	1.2	13.5	1.2	-0.5	0.0	-0.6	0.1
R41	-1.6	0.1	-37.7	1.0	33.3	0.9	-0.5	0.0	-6.4	0.3
C42	-0.8	0.0	-0.3	0.0	1.1	0.1	0.0	0.0	0.0	0.0
Q43	-2.2	0.1	-0.9	0.1	3.5	0.2	-0.4	0.0	-0.1	0.0
Y44	-3.5	0.1	-1.9	0.2	3.2	0.2	-0.6	0.0	-2.8	0.1
R45	-1.9	0.1	-2.6	0.5	4.1	0.5	-0.4	0.0	-0.9	0.1
D46	-1.1	0.1	-17.4	1.5	18.9	1.4	-0.2	0.0	0.3	0.2
L47	-4.8	0.2	-3.2	0.2	4.5	0.2	-1.0	0.0	-4.5	0.2
K48	-2.4	0.1	12.8	1.2	-9.2	1.1	-0.5	0.0	0.7	0.1
W49	-2.9	0.1	-3.3	0.2	4.3	0.2	-0.5	0.0	-2.3	0.1
W50	-1.2	0.1	-1.1	0.1	2.2	0.1	-0.1	0.0	-0.3	0.0

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Second replicate										
C6	-0.1	0.0	-0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0
P7	-0.1	0.0	0.2	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.3	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
S9	-0.4	0.0	0.5	0.0	0.1	0.0	-0.1	0.0	0.1	0.0
H10	-0.9	0.0	0.2	0.0	0.7	0.0	-0.1	0.0	-0.1	0.0
D11	-0.6	0.1	10.5	0.4	-9.2	0.4	-0.1	0.0	0.6	0.0
G12	-1.0	0.0	1.7	0.1	-0.6	0.1	-0.2	0.0	-0.1	0.0
Y13	-2.5	0.0	0.4	0.1	0.5	0.1	-0.2	0.0	-1.7	0.0
C14	-0.3	0.0	0.3	0.0	0.3	0.0	0.0	0.0	0.2	0.0
L15	-1.9	0.0	0.9	0.1	-0.1	0.1	-0.4	0.0	-1.4	0.0
H16	-1.9	0.1	0.4	0.1	2.3	0.1	-0.3	0.0	0.4	0.1
D17	-0.1	0.0	5.0	0.3	-4.4	0.3	0.0	0.0	0.5	0.0
G18	-0.1	0.0	-0.6	0.0	1.0	0.0	0.0	0.0	0.4	0.0
V19	-0.2	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	-0.2	0.0
C20	-0.1	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0
M21	-1.8	0.0	0.0	0.0	0.4	0.0	-0.3	0.0	-1.7	0.0
Y22	-0.2	0.0	0.3	0.0	0.1	0.0	0.0	0.0	0.2	0.0
I23	-1.9	0.0	-0.4	0.0	0.7	0.0	-0.3	0.0	-2.0	0.0
E24	-0.3	0.0	12.3	0.2	-11.4	0.2	0.0	0.0	0.6	0.0
A25	-1.3	0.1	-4.9	0.2	4.6	0.1	-0.5	0.0	-2.0	0.1
L26	-3.6	0.1	-0.5	0.1	1.2	0.1	-0.7	0.0	-3.7	0.1
D27	-0.2	0.0	13.4	0.2	-12.6	0.2	0.0	0.0	0.6	0.0
K28	-1.1	0.1	-30.0	0.9	30.3	0.8	-0.5	0.0	-1.2	0.1
Y29	-0.5	0.0	-0.9	0.0	1.6	0.0	0.0	0.0	0.3	0.0
A30	-1.4	0.0	0.1	0.0	0.4	0.0	-0.2	0.0	-1.1	0.0
C31	-1.8	0.0	-2.8	0.1	2.5	0.0	-0.2	0.0	-2.3	0.0
N32	-3.0	0.1	-6.5	0.2	7.0	0.1	-0.4	0.0	-2.9	0.1
C33	-2.1	0.1	-1.2	0.1	1.2	0.1	-0.3	0.0	-2.3	0.1
V34	-0.6	0.0	-0.8	0.0	1.2	0.0	0.0	0.0	-0.3	0.0
V35	-0.4	0.0	-1.5	0.0	1.8	0.1	0.0	0.0	-0.2	0.0
G36	-0.3	0.0	-1.7	0.4	1.8	0.0	0.0	0.0	-0.1	0.0
Y37	-1.5	0.1	-0.1	0.1	1.8	0.1	-0.1	0.0	0.1	0.1
I38	-2.2	0.0	0.5	0.0	0.5	0.0	-0.3	0.0	-1.5	0.0
G39	-1.3	0.0	-2.7	0.2	2.8	0.1	-0.3	0.0	-1.5	0.1
E40	-3.1	0.1	-12.2	0.6	15.4	0.6	-0.7	0.0	-0.6	0.1
R41	-2.1	0.1	-39.9	0.4	35.5	0.4	-0.6	0.0	-7.1	0.1
C42	-0.8	0.0	0.2	0.0	0.5	0.0	0.0	0.0	-0.1	0.0
Q43	-2.9	0.0	-1.1	0.1	3.7	0.1	-0.4	0.0	-0.6	0.1
Y44	-4.3	0.1	-2.4	0.1	4.0	0.1	-0.7	0.0	-3.4	0.1
R45	-2.5	0.1	3.9	0.4	-2.1	0.4	-0.4	0.0	-1.1	0.1
D46	-0.5	0.1	-38.6	0.5	35.5	0.4	-0.2	0.0	-3.8	0.2
L47	-6.3	0.1	-2.4	0.2	4.9	0.1	-1.2	0.0	-5.1	0.1
K48	-3.0	0.1	9.0	1.0	-5.7	1.0	-0.7	0.0	-0.3	0.1
W49	-4.2	0.1	-0.7	0.1	2.3	0.1	-0.6	0.0	-3.2	0.1
W50	-1.2	0.0	-1.7	0.1	2.9	0.1	-0.1	0.0	0.0	0.0

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Average										
C6	-0.1	0.0	-0.2	0.0	0.3	0.0	0.0	0.0	0.0	0.0
P7	0.0	0.0	0.2	0.0	-0.2	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.3	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
S9	-0.3	0.0	0.5	0.0	0.1	0.0	-0.1	0.0	0.1	0.0
H10	-0.8	0.0	0.3	0.0	0.6	0.1	-0.1	0.0	0.0	0.0
D11	-0.5	0.1	9.8	0.3	-8.8	0.3	-0.1	0.0	0.5	0.0
G12	-1.0	0.0	1.7	0.1	-0.6	0.1	-0.2	0.0	-0.1	0.0
Y13	-2.2	0.1	0.2	0.1	0.6	0.1	-0.2	0.0	-1.7	0.1
C14	-0.3	0.0	0.3	0.0	0.2	0.0	0.0	0.0	0.2	0.0
L15	-2.0	0.1	0.7	0.1	0.2	0.1	-0.3	0.0	-1.4	0.0
H16	-2.1	0.1	-0.9	0.2	3.1	0.2	-0.4	0.0	-0.3	0.1
D17	0.0	0.0	-6.5	1.0	6.0	0.9	-0.1	0.0	-0.6	0.1
G18	-0.1	0.0	-0.5	0.0	0.9	0.0	0.0	0.0	0.3	0.0
V19	-0.2	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0
C20	-0.1	0.0	-0.1	0.0	0.3	0.0	0.0	0.0	0.1	0.0
M21	-1.6	0.0	-0.1	0.0	0.4	0.0	-0.3	0.0	-1.5	0.1
Y22	-0.2	0.0	0.3	0.0	0.2	0.0	0.0	0.0	0.2	0.0
I23	-2.0	0.1	-0.4	0.0	0.7	0.0	-0.3	0.0	-2.0	0.1
E24	-0.3	0.0	10.5	0.3	-9.6	0.3	0.0	0.0	0.6	0.0
A25	-1.4	0.1	-4.6	0.2	4.5	0.1	-0.5	0.0	-2.0	0.1
I26	-3.4	0.1	-0.4	0.0	1.0	0.0	-0.6	0.0	-3.4	0.1
D27	-0.2	0.0	12.0	0.2	-11.3	0.2	0.0	0.0	0.6	0.0
K28	-1.0	0.1	-24.6	1.0	25.2	1.0	-0.4	0.0	-0.8	0.1
Y29	-0.4	0.0	-0.6	0.0	1.4	0.0	0.0	0.0	0.3	0.0
A30	-1.4	0.0	-0.1	0.0	0.5	0.0	-0.2	0.0	-1.1	0.0
C31	-1.8	0.1	-2.5	0.1	2.2	0.1	-0.2	0.0	-2.3	0.1
N32	-3.0	0.1	-5.2	0.2	6.2	0.2	-0.4	0.0	-2.3	0.1
C33	-2.0	0.1	-0.7	0.1	1.1	0.1	-0.3	0.0	-1.8	0.1
V34	-0.5	0.0	-0.6	0.0	1.0	0.0	0.0	0.0	-0.2	0.0
V35	-0.3	0.0	-1.5	0.1	1.7	0.1	0.0	0.0	-0.1	0.0
G36	-0.3	0.0	-1.6	0.2	1.9	0.1	0.0	0.0	-0.1	0.0
Y37	-1.6	0.1	0.2	0.1	1.4	0.1	-0.2	0.0	-0.1	0.1
I38	-2.2	0.1	-0.5	0.0	1.0	0.0	-0.3	0.0	-2.0	0.1
G39	-1.4	0.1	-3.7	0.2	3.4	0.1	-0.3	0.0	-2.0	0.1
E40	-3.0	0.1	-11.3	1.0	14.2	0.9	-0.6	0.0	-0.7	0.1
R41	-1.8	0.1	-39.0	0.6	34.5	0.5	-0.5	0.0	-6.8	0.2
C42	-0.8	0.0	-0.2	0.0	0.9	0.0	0.0	0.0	-0.1	0.0
Q43	-2.3	0.1	-1.2	0.1	3.6	0.1	-0.4	0.0	-0.3	0.0
Y44	-3.5	0.1	-1.5	0.2	3.3	0.2	-0.6	0.0	-2.4	0.1
R45	-2.0	0.1	2.7	0.5	-0.8	0.5	-0.3	0.0	-0.4	0.1
D46	-0.8	0.1	-27.9	1.0	27.2	0.9	-0.2	0.0	-1.6	0.2
L47	-5.1	0.1	-2.5	0.9	4.1	0.2	-1.1	0.0	-4.5	0.1
K48	-2.6	0.1	11.8	1.0	-8.1	1.0	-0.6	0.0	0.5	0.1
W49	-2.9	0.1	-2.6	0.2	3.4	0.1	-0.5	0.0	-2.6	0.1
W50	-2.0	0.1	-1.7	0.2	3.0	0.2	-0.3	0.0	-1.0	0.1

Appendix A Table 7. EGF bound to the unglycosylated EGFR-ed per-residue contribution to the binding free energy calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔG GB), non-polar contribution to the solvation energy (ΔG SA) and binding free energy (ΔG MMGBSA).

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First simulation										
C6	-0.1	0.0	-0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0
P7	-0.1	0.0	0.2	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
S9	-0.4	0.0	0.2	0.0	0.5	0.1	-0.1	0.0	0.2	0.0
H10	-1.1	0.1	-0.2	0.1	1.4	0.1	-0.2	0.0	-0.1	0.0
D11	-0.4	0.1	9.3	0.4	-8.2	0.4	-0.1	0.0	0.6	0.1
G12	-0.8	0.0	1.4	0.1	-0.5	0.1	-0.1	0.0	0.0	0.0
Y13	-2.9	0.1	0.0	0.1	1.4	0.1	-0.4	0.0	-1.9	0.1
C14	-0.3	0.0	0.0	0.1	0.6	0.1	0.0	0.0	0.2	0.0
L15	-1.7	0.0	1.1	0.1	-0.3	0.1	-0.3	0.0	-1.3	0.0
H16	-1.7	0.1	-0.6	0.2	2.7	0.2	-0.4	0.0	0.0	0.1
D17	-0.1	0.0	1.8	0.6	-1.2	0.6	0.0	0.0	0.5	0.0
G18	-0.1	0.0	-0.7	0.0	1.1	0.0	0.0	0.0	0.3	0.0
V19	-0.2	0.0	0.4	0.0	-0.4	0.0	0.0	0.0	-0.2	0.0
C20	-0.1	0.0	-0.2	0.0	0.4	0.0	0.0	0.0	0.1	0.0
M21	-1.8	0.0	-0.1	0.1	0.5	0.0	-0.3	0.0	-1.7	0.0
Y22	-0.2	0.0	0.3	0.0	0.2	0.0	0.0	0.0	0.2	0.0
I23	-2.2	0.0	-0.4	0.0	0.8	0.0	-0.3	0.0	-2.1	0.0
E24	-0.4	0.0	12.0	0.4	-10.9	0.4	0.0	0.0	0.7	0.0
A25	-1.5	0.1	-3.9	0.2	4.1	0.1	-0.5	0.0	-1.8	0.1
L26	-3.6	0.1	-0.2	0.0	1.1	0.0	-0.7	0.0	-3.5	0.1
D27	-0.2	0.0	13.8	0.4	-12.8	0.4	0.0	0.0	0.8	0.0
K28	-1.6	0.1	-36.7	1.2	37.5	1.1	-0.6	0.0	-1.4	0.2
Y29	-0.7	0.0	-0.2	0.0	1.2	0.1	-0.1	0.0	0.3	0.0
A30	-1.6	0.0	-0.4	0.0	0.8	0.0	-0.2	0.0	-1.4	0.0
C31	-1.9	0.1	-2.6	0.1	2.2	0.1	-0.1	0.0	-2.6	0.1
N32	-3.3	0.1	-4.5	0.3	6.2	0.2	-0.4	0.0	-2.1	0.1
C33	-2.1	0.1	-0.5	0.1	1.0	0.1	-0.3	0.0	-1.9	0.1
V34	-0.5	0.0	-0.4	0.1	0.7	0.1	0.0	0.0	-0.2	0.0
V35	-0.4	0.0	-2.3	0.1	2.5	0.1	0.0	0.0	-0.2	0.0
G36	-0.3	0.0	-1.7	0.1	1.8	0.1	0.0	0.0	-0.2	0.0
Y37	-1.3	0.0	0.2	0.1	1.3	0.1	-0.2	0.0	0.1	0.1
I38	-2.6	0.0	-1.4	0.1	2.0	0.0	-0.4	0.0	-2.3	0.1
G39	-1.8	0.1	-4.8	0.2	4.8	0.1	-0.3	0.0	-2.0	0.1
E40	-3.3	0.1	-9.3	0.7	13.1	0.6	-0.9	0.0	-0.3	0.2
R41	-2.5	0.1	-39.1	0.4	35.6	0.4	-0.7	0.0	-6.7	0.1
C42	-0.8	0.0	-0.8	0.0	1.4	0.0	0.0	0.0	-0.2	0.0
Q43	-2.2	0.1	-1.9	0.1	4.2	0.1	-0.3	0.0	-0.1	0.1
Y44	-4.2	0.1	-0.4	0.1	3.5	0.1	-0.8	0.0	-1.8	0.1
R45	-1.6	0.1	6.9	0.5	-5.5	0.5	-0.3	0.0	-0.5	0.1
D46	-1.3	0.1	-14.2	1.1	15.9	0.8	-0.1	0.0	0.4	0.2
L47	-5.8	0.1	-3.1	0.2	5.6	0.2	-1.2	0.0	-4.5	0.1
K48	-2.0	0.1	3.6	1.4	-0.7	1.4	-0.5	0.0	0.6	0.1
W49	-3.6	0.1	-3.4	0.2	4.6	0.2	-0.6	0.0	-3.0	0.1
W50	-1.0	0.1	-1.8	0.3	2.5	0.3	-0.2	0.0	-0.5	0.1

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
First replicate										
C6	-0.1	0.0	-0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0
P7	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
S9	-0.4	0.0	0.2	0.1	0.4	0.1	-0.1	0.0	0.2	0.0
H10	-0.9	0.0	0.0	0.1	1.1	0.1	-0.1	0.0	0.1	0.0
D11	-0.4	0.0	7.4	0.4	-6.5	0.3	-0.1	0.0	0.5	0.0
G12	-1.1	0.0	1.5	0.1	-0.4	0.1	-0.2	0.0	-0.2	0.0
Y13	-3.1	0.1	0.1	0.1	1.6	0.1	-0.4	0.0	-1.9	0.1
C14	-0.4	0.0	-0.1	0.1	0.8	0.1	0.0	0.0	0.2	0.0
L15	-1.9	0.0	1.2	0.1	-0.5	0.0	-0.4	0.0	-1.5	0.0
H16	-1.1	0.0	0.9	0.1	1.1	0.1	-0.3	0.0	0.6	0.0
D17	-0.1	0.0	4.2	0.3	-3.6	0.4	0.0	0.0	0.4	0.0
G18	-0.1	0.0	-0.7	0.0	1.2	0.0	0.0	0.0	0.4	0.0
V19	-0.2	0.0	0.3	0.0	-0.3	0.0	0.0	0.0	-0.2	0.0
C20	-0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0
M21	-1.5	0.0	-0.2	0.0	0.5	0.0	-0.2	0.0	-1.4	0.0
Y22	-0.2	0.0	0.3	0.0	0.1	0.0	0.0	0.0	0.3	0.0
I23	-2.0	0.0	-0.4	0.0	0.7	0.0	-0.3	0.0	-2.0	0.0
E24	-0.3	0.0	14.0	0.3	-13.1	0.3	0.0	0.0	0.6	0.0
A25	-1.6	0.1	-3.8	0.1	4.1	0.1	-0.5	0.0	-1.9	0.1
L26	-3.9	0.1	-0.8	0.1	1.7	0.1	-0.7	0.0	-3.8	0.1
D27	-0.3	0.0	15.2	0.3	-14.0	0.3	0.0	0.0	0.8	0.0
K28	-1.4	0.1	-36.3	0.7	37.5	0.6	-0.7	0.0	-0.9	0.2
Y29	-0.7	0.0	-0.7	0.0	1.7	0.0	-0.1	0.0	0.3	0.0
A30	-1.7	0.0	-0.5	0.0	0.7	0.0	-0.2	0.0	-1.7	0.0
C31	-2.0	0.1	-3.0	0.1	2.4	0.0	-0.2	0.0	-2.7	0.1
N32	-3.0	0.1	-5.4	0.3	6.5	0.2	-0.4	0.0	-2.3	0.1
C33	-2.5	0.1	-0.6	0.1	1.2	0.1	-0.3	0.0	-2.2	0.1
V34	-0.5	0.0	-0.5	0.0	0.8	0.0	0.0	0.0	-0.2	0.0
V35	-1.1	0.1	-2.0	0.1	2.5	0.1	-0.2	0.0	-0.8	0.1
G36	-0.2	0.0	-1.7	0.0	1.8	0.0	0.0	0.0	-0.1	0.0
Y37	-0.9	0.0	0.3	0.1	1.1	0.1	0.0	0.0	0.5	0.0
I38	-2.1	0.0	0.0	0.1	0.8	0.1	-0.2	0.0	-1.5	0.0
G39	-1.8	0.0	-2.1	0.1	3.2	0.1	-0.2	0.0	-0.9	0.1
E40	-3.1	0.1	-19.0	0.7	22.1	0.6	-0.9	0.0	-0.9	0.1
R41	-2.7	0.1	-34.1	0.6	30.6	0.5	-0.6	0.0	-6.9	0.1
C42	-1.0	0.0	-0.1	0.0	0.8	0.0	0.0	0.0	-0.3	0.0
Q43	-2.6	0.0	-1.9	0.1	4.6	0.1	-0.4	0.0	-0.3	0.0
Y44	-4.1	0.1	-2.9	0.1	4.0	0.1	-0.6	0.0	-3.6	0.1
R45	-1.9	0.1	3.4	0.6	-3.0	0.5	-0.3	0.0	-1.8	0.1
D46	-0.6	0.1	-44.8	0.5	39.2	0.4	-0.2	0.0	-6.4	0.2
L47	-6.5	0.1	-1.5	0.2	3.6	0.1	-1.2	0.0	-5.5	0.1
K48	-2.5	0.1	14.3	0.9	-11.2	0.8	-0.6	0.0	0.0	0.1
W49	-5.2	0.1	-2.4	0.1	4.4	0.1	-0.7	0.0	-4.0	0.1
W50	-2.2	0.1	-7.4	0.4	7.5	0.3	-0.4	0.0	-2.5	0.2

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Second replicate										
C6	-0.1	0.0	-0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0
P7	-0.1	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0
I8	-0.1	0.0	-0.1	0.1	0.3	0.0	0.0	0.0	0.0	0.0
S9	-0.5	0.0	0.2	0.1	0.5	0.1	-0.1	0.0	0.2	0.0
H10	-1.2	0.1	0.3	0.1	1.2	0.1	-0.2	0.0	0.1	0.0
D11	-0.2	0.0	7.6	0.4	-7.0	0.4	0.0	0.0	0.4	0.0
G12	-0.9	0.0	1.4	0.1	-0.5	0.1	-0.2	0.0	-0.2	0.1
Y13	-3.4	0.1	0.1	0.1	1.3	0.1	-0.4	0.0	-2.4	0.1
C14	-0.4	0.0	-0.2	0.1	0.8	0.1	0.0	0.0	0.3	0.0
L15	-2.2	0.0	1.2	0.1	-0.2	0.0	-0.4	0.0	-1.5	0.0
H16	-0.9	0.0	0.6	0.1	0.7	0.1	-0.2	0.0	0.3	0.0
D17	-0.1	0.0	2.4	0.4	-1.7	0.4	0.0	0.0	0.5	0.0
G18	-0.1	0.0	-0.7	0.0	1.1	0.0	0.0	0.0	0.3	0.0
V19	-0.2	0.0	0.4	0.0	-0.4	0.0	0.0	0.0	-0.2	0.0
C20	-0.1	0.0	-0.1	0.0	0.3	0.0	0.0	0.0	0.1	0.0
M21	-1.7	0.0	-0.1	0.0	0.5	0.0	-0.3	0.0	-1.6	0.0
Y22	-0.2	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.2	0.0
I23	-2.1	0.0	-0.6	0.0	0.9	0.0	-0.3	0.0	-2.0	0.0
E24	-0.3	0.0	12.0	0.5	-10.9	0.5	0.0	0.0	0.8	0.0
A25	-1.2	0.1	-3.1	0.2	3.7	0.2	-0.4	0.0	-0.9	0.1
L26	-3.6	0.1	-0.4	0.0	1.3	0.0	-0.7	0.0	-3.5	0.1
D27	-0.2	0.0	11.8	0.3	-10.8	0.3	0.0	0.0	0.7	0.0
K28	-1.3	0.1	-35.3	1.0	36.3	0.9	-0.7	0.0	-1.0	0.1
Y29	-0.6	0.0	-0.2	0.0	1.2	0.0	0.0	0.0	0.3	0.0
A30	-1.6	0.0	-0.6	0.0	0.8	0.0	-0.2	0.0	-1.6	0.1
C31	-1.8	0.1	-3.0	0.1	2.3	0.1	-0.2	0.0	-2.7	0.1
N32	-3.1	0.1	-4.5	0.3	6.1	0.2	-0.4	0.0	-2.0	0.1
C33	-2.1	0.1	-0.7	0.1	1.1	0.1	-0.3	0.0	-2.0	0.1
V34	-0.5	0.0	0.2	0.1	0.2	0.1	0.0	0.0	-0.1	0.0
V35	-0.9	0.1	-1.5	0.1	2.1	0.1	-0.2	0.0	-0.5	0.1
G36	-0.3	0.0	-0.7	0.1	0.9	0.1	0.0	0.0	-0.1	0.0
Y37	-1.5	0.0	0.4	0.1	1.5	0.1	-0.2	0.0	0.2	0.0
I38	-2.4	0.1	-1.3	0.1	1.8	0.0	-0.4	0.0	-2.4	0.1
G39	-1.8	0.1	-4.2	0.1	4.7	0.1	-0.3	0.0	-1.6	0.1
E40	-3.8	0.1	-8.1	1.0	12.5	0.8	-0.8	0.0	-0.3	0.2
R41	-2.8	0.1	-38.1	0.5	34.8	0.5	-0.7	0.0	-6.7	0.1
C42	-0.9	0.0	-1.0	0.0	1.8	0.0	0.0	0.0	-0.1	0.0
Q43	-2.3	0.1	-4.9	0.2	6.4	0.1	-0.4	0.0	-1.2	0.1
Y44	-4.9	0.1	-2.7	0.1	4.8	0.1	-0.8	0.0	-3.6	0.1
R45	-3.0	0.1	8.5	0.6	-5.8	0.6	-0.5	0.0	-0.7	0.1
D46	-1.3	0.0	-11.8	0.6	14.5	0.6	-0.1	0.0	1.4	0.1
L47	-5.9	0.1	-4.0	0.3	6.0	0.2	-1.2	0.0	-5.1	0.1
K48	-2.2	0.1	2.2	1.3	1.3	1.3	-0.5	0.0	0.8	0.1
W49	-4.1	0.1	-3.2	0.2	4.6	0.1	-0.6	0.0	-3.4	0.1
W50	-1.6	0.1	-2.5	0.1	3.1	0.1	-0.3	0.0	-1.3	0.2

Resid no.	vdw	err	ele	err	polar solvation	err	non-polar solvation	err	total	err
Average										
C6	-0.1	0.0	-0.1	0.0	0.2	0.0	0.0	0.0	0.0	0.0
P7	-0.1	0.0	0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0
I8	-0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
S9	-0.4	0.0	0.2	0.0	0.4	0.1	-0.1	0.0	0.2	0.0
H10	-1.1	0.1	0.0	0.1	1.2	0.1	-0.2	0.0	0.1	0.0
D11	-0.3	0.0	8.1	0.4	-7.2	0.4	-0.1	0.0	0.5	0.0
G12	-0.9	0.0	1.4	0.1	-0.5	0.1	-0.2	0.0	-0.1	0.0
Y13	-3.1	0.1	0.0	0.1	1.4	0.1	-0.4	0.0	-2.0	0.1
C14	-0.4	0.0	-0.1	0.1	0.7	0.1	0.0	0.0	0.2	0.0
L15	-1.9	0.0	1.2	0.1	-0.3	0.0	-0.4	0.0	-1.4	0.0
H16	-1.2	0.0	0.3	0.1	1.5	0.1	-0.3	0.0	0.3	0.0
D17	-0.1	0.0	2.8	0.4	-2.2	0.4	0.0	0.0	0.5	0.0
G18	-0.1	0.0	-0.7	0.0	1.1	0.0	0.0	0.0	0.3	0.0
V19	-0.2	0.0	0.4	0.0	-0.4	0.0	0.0	0.0	-0.2	0.0
C20	-0.1	0.0	-0.1	0.0	0.3	0.0	0.0	0.0	0.1	0.0
M21	-1.6	0.0	-0.1	0.0	0.5	0.0	-0.2	0.0	-1.5	0.0
Y22	-0.2	0.0	0.4	0.0	0.1	0.0	0.0	0.0	0.2	0.0
I23	-2.1	0.0	-0.5	0.0	0.8	0.0	-0.3	0.0	-2.1	0.0
E24	-0.3	0.0	12.7	0.4	-11.6	0.4	0.0	0.0	0.7	0.0
A25	-1.4	0.1	-3.6	0.2	4.0	0.1	-0.5	0.0	-1.5	0.1
L26	-3.7	0.1	-0.5	0.0	1.3	0.0	-0.7	0.0	-3.6	0.1
D27	-0.2	0.0	13.6	0.3	-12.6	0.3	0.0	0.0	0.8	0.0
K28	-1.5	0.1	-36.1	1.0	37.1	0.9	-0.7	0.0	-1.1	0.2
Y29	-0.7	0.0	-0.4	0.0	1.4	0.0	0.0	0.0	0.3	0.0
A30	-1.6	0.0	-0.5	0.0	0.8	0.0	-0.2	0.0	-1.6	0.1
C31	-1.6	0.1	-3.9	0.3	4.3	0.2	-0.3	0.0	-1.4	0.2
N32	-3.1	0.1	-4.8	0.3	6.2	0.2	-0.4	0.0	-2.1	0.1
C33	-2.2	0.1	-0.6	0.1	1.1	0.1	-0.3	0.0	-2.0	0.1
V34	-0.5	0.0	-0.2	0.1	0.6	0.1	0.0	0.0	-0.2	0.0
V35	-0.8	0.1	-1.9	0.1	2.3	0.1	-0.1	0.0	-0.5	0.1
C36	-0.3	0.0	-1.4	0.1	1.5	0.1	0.0	0.0	-0.1	0.0
Y37	-1.2	0.0	0.3	0.1	1.3	0.1	-0.1	0.0	0.3	0.0
I38	-2.4	0.0	-0.9	0.1	1.5	0.0	-0.3	0.0	-2.1	0.1
G39	-1.8	0.1	-3.7	0.1	4.2	0.1	-0.3	0.0	-1.5	0.1
E40	-3.4	0.1	-12.1	0.8	15.9	0.6	-0.8	0.0	-0.5	0.2
R41	-2.7	0.1	-37.1	0.5	33.6	0.5	-0.7	0.0	-6.8	0.1
C42	-0.9	0.0	-0.6	0.0	1.3	0.0	0.0	0.0	-0.2	0.0
Q43	-2.4	0.0	-2.9	0.1	5.1	0.1	-0.4	0.0	-0.6	0.1
Y44	-4.4	0.1	-2.0	0.1	4.1	0.1	-0.7	0.0	-3.0	0.1
R45	-2.2	0.1	6.3	0.6	-4.8	0.5	-0.3	0.0	-1.0	0.1
D46	-1.1	0.1	-23.6	0.7	23.2	0.6	-0.1	0.0	-1.6	0.2
L47	-6.1	0.1	-2.8	0.2	5.1	0.2	-1.2	0.0	-5.0	0.1
K48	-2.2	0.1	6.7	1.2	-3.5	1.2	-0.5	0.0	0.5	0.1
W49	-4.3	0.1	-3.0	0.1	4.5	0.1	-0.6	0.0	-3.5	0.1
W50	-1.6	0.1	-3.9	0.3	4.3	0.2	-0.3	0.0	-1.4	0.2

Appendix A Table 8. EGF bound to the glycosylated with oligosaccharides EGFR-ed per-residue contribution to the binding free energy calculated from the last 50ns of each simulation. The energetic components are: van der Waals (ΔE vdw), electrostatic (ΔE ele), electrostatic energy required for solvation (ΔG GB), non-polar contribution to the solvation energy (ΔG SA) and binding free energy (ΔG MMGBSA).

Appendix A Figure 18. C225, IMC-11F8, MEHD7946 and EMD72000 binding epitopes. A) Epitope recognized by C225 and IMC-11F8 is overlapping and presented in blue surface. MEHD7946 epitope is shown in yellow surface and EMD72000 binding epitope is presented in red surface. EGFR and EGF are shown in grey and pink surface respectively. B) Overlaid structures of the glycans in the proximity of antibody binding epitopes from simulations.

Appendix A Figure 19. mAb806 and GC1118 binding epitopes. A) Epitope recognized by mAb806 is located in subdomain II and presented in magenta surface. GC1118 epitope is located in subdomain III shown in orange surface. EGFR-ed and EGF are shown in grey and pink surface respectively. B) Overlaid structures of glycans in the proximity of antibody binding epitopes from simulations.

Appendix A Figure 20. Average SASA of C225 and IMC-11F8 binding epitope located in subdomain III of EGFR-ed from simulations compared with SASA of the crystal structures of C225 bound to EGFR (PDB entry: 1YY9), IMC-11F8 bound to EGFR (PDB entry: 3B2V) and the initial structure of this study (PDB entry: 3NJP).

Appendix A Figure 21. Average SASA of GC1118 binding epitope located in subdomain III of EGFR-ed from simulations compared with SASA of the crystal structures of GC1118 bound to EGFR (PDB entry: 4UV7) and the initial structure of this study (PDB entry: 3NJP).

Appendix A Figure 22. Average SASA of EMD72000 binding epitope located in subdomain III of EGFR-ed from simulations compared with SASA of the crystal structures of EMD72000 bound to EGFR (PDB entry: 3C09) and the initial structure of this study (PDB entry: 3NJP).

Appendix A Figure 23. Average SASA of MEHD7946 binding epitope located in subdomain III of EGFR-ed from simulations compared with SASA of the crystal structures of MEHD 7946 bound to EGFR (PDB entry: 3P0Y) and the initial structure of this study (PDB entry: 3NJP).

Appendix A Figure 24. Average SASA of mAb806 binding epitope located in subdomain II of EGFR-ed from simulations. Compared with SASA of the computationally docked model of mAb806 (PDB entry: 3G5V) to EGFR-ed (PDB entry: 1IVO) and the initial crystal structure of this study (PDB entry: 3NJP).

Appendix A Figure 25. Average SASA of modules 1-8 of subdomain II and tip of subdomain IV.

Structure	ΔG (kcal/mol)
PDB ID 3NJP EGF-EGFR-ed Chains A-C	-14.8
EGF-gEGFR-ed first simulation	- 4.9
EGF-gEGFR-ed first replicate	-3.2
EGF-gEGFR-ed second replicate	-4.6
EGF-EGFR-ed first simulation	-2.4
EGF-EGFR-ed first replicate	-4.9
EGF-EGFR-ed second replicate	-1.4

Appendix A Table 9. The ΔG values estimated based on the interface between EGF-EGFR-ed. PDB 3NJP is the initial crystal structure used for modeling in this study. The other structures are the final conformations of 100ns MD simulations of glycosylated and unglycosylated EGFR-ed bound to EGF.

Appendix A Figure 26. Average SASA of the overall area predicted to be involved in EGFR-ed multimerization.

Appendix A Figure 27. Average SASA of the residues predicted to be involved in EGFR-ed multimerization.

Appendix B Global dynamics of tethered monomeric form of EGFR-ed upon glycosylation

Appendix B Figure 1. The C α RMSD plots of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation.

Appendix B Figure 2. The C α RMSD plots of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation.

Appendix B Figure 3. The $C\alpha$ RMSD plots of subdomains I-IV (A-D) of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation.

Appendix B Figure 4. The $\text{C}\alpha$ RMSD plots of subdomains I-IV (A-D) of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation.

Appendix B Figure 5. A-B) Subdomain I secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Untethered EGFR-ed and Untethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 6. A-B) Subdomain II secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Untethered EGFR-ed and Untethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 7. A-B) Subdomain III secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Untethered EGFR-ed and Untethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 8. A-B) Subdomain IV secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Untethered EGFR-ed and Untethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 9. The $C\alpha$ RMSD plots of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 1NQL. Data shown here is obtained from one simulation.

Appendix B Figure 10. The $C\alpha$ RMSD plots of the untethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 1NQL. Data shown here is obtained from one simulation.

Appendix B Figure 11. The $C\alpha$ RMSD plots of subdomains I-IV (A-D) of the tethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation.

Appendix B Figure 12. The $C\alpha$ RMSD plots of subdomains I-IV (A-D) of the tethered EGFR-ed in the glycosylated (magenta) and unglycosylated (blue) form. The trajectory of 100ns was aligned on the starting crystal structure PDB 3NJP. Data shown here is obtained from one simulation.

Appendix B Figure 13. A-B) Subdomain I secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Tethered EGFR-ed and Tethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 14. A-B) Subdomain II secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Tethered EGFR-ed and Tethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 15. A-B) Subdomain III secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Tethered EGFR-ed and Tethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 16. A-B) Subdomain IV secondary structure modifications calculated over combined trajectory of three 300ns simulations of the Tethered EGFR-ed and Tethered gEGFR-ed respectively. Colors in the schemes present different types of secondary structures - magenta: alpha-helix, dark blue: 3-10 helix, dark green: turn, light green: isolated bridge, yellow: extended configuration.

Appendix B Figure 17. Visualization of the C α atoms whose motions are correlated in the tethered EGFR-ed (A) and gEGFR-ed (B). Correlated motions are shown as red lines between each pair of atoms. Only correlated motions with absolute normalized values > 0.8 from the last 50ns of the combined trajectory are shown for simplicity. Subdomains I, II, III and IV of EGFR-ed are shown in blue, green, yellow and magenta tubes respectively. Glycosylated Asparagines are shown with wheat spheres. The tethering interaction sites ie the dimerization arm in subdomain II and the inhibitory loops I and II in subdomain IV are shown with black and grey tubes respectively.

Appendix B Figure 18. Visualization of the C α atoms whose motions are anti-correlated in the tethered EGFR-ed (A) and gEGFR-ed (B). Anti-correlated motions are shown as blue lines between each pair of atoms. Only anti-correlated motions with absolute normalized values > 0.6 from the last 50ns of the combined trajectory are shown for simplicity. Subdomains I, II, III and IV of EGFR-ed are shown in blue, green, yellow and magenta tubes respectively. Glycosylated Asparagines are shown with wheat spheres. The tethering interaction sites ie the dimerization arm in subdomain II and the inhibitory loops I and II in subdomain IV are shown with black and red tubes respectively.

Appendix B Figure 19. Tethering interactions between the dimerization arm of subdomain II and inhibitory loops of subdomain IV. A) The tethered monomer of EGFR-ed. Subdomain I, II, III and IV are shown in blue, green, yellow and magenta cartoon respectively. B-F) Hydrogen bonds between Tyr246 (dim arm)-Asp563 (loop I), Thr250 (dim arm)-His566 (loop I), Tyr246 (dim arm)-Met576 (loop II), Asn247 (dim arm)-Glu578 (loop II), and Asp563 (loop I)-Lys585 (loop II). Each pair of interacting residues is shown with atom name sticks.

Appendix B Figure 20. Dynamics of Asn579 attached glycan visualization. Subdomains I, II, III and IV of EGFR-ed are shown in blue, green, yellow and magenta cartoons respectively. Red, white and blue licorices represent the glycan dynamics from the beginning, middle and end of the trajectory respectively. The black circle shows the dimerization arm of subdomain II which is kept distant from the inhibitory loops of subdomain IV by the glycan.

Appendix B Figure 21. Distance distribution plots between each oligosaccharide and the oligosaccharide attached to Asn579.

Appendix C Dynamics of Module1-8 of subdomain II in the dimeric state

Appendix C Figure 1. Dimerization interface of subdomains II are maintained by glycans attached to Asn172 and Asn337. EGFR-ed dimer, dimerization arms (residue 239 to 269) and module2 (residue 187 to 207) and module8 (residue 305 to 309) of subdomains II from the two monomers are shown with grey, marine and green surface respectively. Overlaid structures of glycans attached to Asn172 and Asn337 of each monomer from the simulations are shown with line presentation.

Appendix C Figure 2. Buried Surface Area (BSA) between glycan attached to Asn172 and module2 (residue 187 to 207) of subdomain II in monomer1 gEGFR-ed, monomer2 gEGFR-ed, monomer1 EGF-gEGFR-ed and monomer2 EGF-gEGFR-ed.

Appendix C Figure 3. Buried Surface Area (BSA) between glycan attached to Asn337 and module8 (residue 305 to 309) of subdomain II in monomer1 gEGFR-ed, monomer2 gEGFR-ed, monomer1 EGF-gEGFR-ed and monomer2 EGF-gEGFR-ed.

Appendix C Figure 4. The box and whisker RMSD plots for module1 (residue 166 to 186) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the data (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 5. The box and whisker RMSD plots for module2 (residue 187 to 207) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the data (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 6. The box and whisker RMSD plots for module3 (residue 208 to 225) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the data (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 7. The box and whisker RMSD plots for module4 (residue 226 to 238) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the data (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 8. The box and whisker RMSD plots for module5 (residue 239 to 269) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the date (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 9. The box and whisker RMSD plots for module6 (residue 270 to 285) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the date (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 10. The box and whisker RMSD plots for module7 (residue 286 to 304) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the data (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

Appendix C Figure 11. The box and whisker RMSD plots for module8 (residue 305 to 309) of subdomain II in EGFR-ed, gEGFR-ed, EGF-EGFR-ed and EGF-gEGFR-ed. Monomer1 and monomer2 are labeled with 1 and 2 in parenthesis. Statistical quantities are schematically shown on the boxes as: first 25% of the data (blue diamond), 50% of the data (green triangle) and 75% of the data (blue cross). Two ends of the whiskers are: the minimum value (red square) and the maximum value (purple cross).

References

1. Yarden, Y., *The EGFR family and its ligands in human cancer. signalling mechanisms and therapeutic opportunities*. Eur J Cancer, 2001. **37**(4): p. S3-8.
2. Garrett, T.P., et al., *Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor α* . Cell, 2002. **110**(6): p. 763-773.
3. Roskoski Jr, R., *The ErbB/HER receptor protein-tyrosine kinases and cancer*. Biochemical and Biophysical Research Communications, 2004. **319**(1): p. 1-11.
4. Citri, A. and Y. Yarden, *EGF-ERBB signalling: towards the systems level*. Nat Rev Mol Cell Biol, 2006. **7**(7): p. 505-16.
5. Rajagopalan, V., et al., *Cardiac ErbB-1/ErbB-2 mutant expression in young adult mice leads to cardiac dysfunction*. Am J Physiol Heart Circ Physiol, 2008. **295**(2): p. 3.
6. Britsch, S., *The neuregulin-I/ErbB signaling system in development and disease*. Adv Anat Embryol Cell Biol, 2007. **190**: p. 1-65.
7. Huang, Y., et al., *Molecular basis for multimerization in the activation of the epidermal growth factor receptor*. eLife, 2016. **5**: p. e14107.
8. Yarden, Y. and G. Pines, *The ERBB network: at last, cancer therapy meets systems biology*. Nat Rev Cancer, 2012. **12**(8): p. 553-63.
9. Seshacharyulu, P., et al., *Targeting the EGFR signaling pathway in cancer therapy*. Expert Opin Ther Targets, 2012. **16**(1): p. 15-31.
10. Ullrich, A., et al., *Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells*. 1984.
11. Arkhipov, A., et al., *Architecture and Membrane Interactions of the EGF Receptor*. Cell, 2013. **152**(3): p. 557-569.
12. Ng, Y.Z., et al., *mAb806 binding to epidermal growth factor receptor: a computational study*. Proteins: Structure, Function, and Bioinformatics, 2015. **83**(1): p. 153-168.
13. Poger, D. and A.E. Mark, *Activation of the Epidermal Growth Factor Receptor: A Series of Twists and Turns*. Biochemistry, 2014. **53**(16): p. 2710-2721.
14. Lu, C., et al., *Structural evidence for loose linkage between ligand binding and kinase activation in the epidermal growth factor receptor*. Molecular and cellular biology, 2010. **30**(22): p. 5432-5443.
15. Gullick, W., *The Type 1 growth factor receptors and their ligands considered as a complex system*. Endocrine-Related Cancer, 2001. **8**(2): p. 75-82.
16. Blobel, C.P., *ADAMs: key components in EGFR signalling and development*. Nature reviews Molecular cell biology, 2005. **6**(1): p. 32-43.
17. Ogiso, H., et al., *Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains*. Cell, 2002. **110**(6): p. 775-787.
18. Sanders, J.M., et al., *Molecular Determinants of Epidermal Growth Factor Binding: A Molecular Dynamics Study*. PloS one, 2013. **8**(1): p. e54136.
19. Engler, D., et al., *Critical functional requirement for the guanidinium group of the arginine 41 side chain of human epidermal growth factor as revealed by mutagenic inactivation and chemical reactivation*. Journal of Biological Chemistry, 1992. **267**(4): p. 2274-2281.
20. Li, S., et al., *Structural basis for inhibition of the epidermal growth factor receptor by cetuximab*. Cancer cell, 2005. **7**(4): p. 301-311.
21. Li, S., P. Kussie, and K.M. Ferguson, *Structural basis for EGF receptor inhibition by the therapeutic antibody IMC-11F8*. Structure, 2008. **16**(2): p. 216-227.

22. Schmiedel, J., et al., *Matuzumab binding to EGFR prevents the conformational rearrangement required for dimerization*. Cancer cell, 2008. **13**(4): p. 365-373.
23. Schaefer, G., et al., *A two-in-one antibody against HER3 and EGFR has superior inhibitory activity compared with monospecific antibodies*. Cancer cell, 2011. **20**(4): p. 472-486.
24. Lim, Y., et al., *GC1118, an anti-EGFR antibody with a distinct binding epitope and superior inhibitory activity against high-affinity EGFR ligands*. Molecular cancer therapeutics, 2015.
25. Ferguson, K.M., et al., *EGF activates its receptor by removing interactions that autoinhibit ectodomain dimerization*. Mol Cell, 2003. **11**(2): p. 507-17.
26. Moriki, T., H. Maruyama, and I.N. Maruyama, *Activation of preformed EGF receptor dimers by ligand-induced rotation of the transmembrane domain*. Journal of molecular biology, 2001. **311**(5): p. 1011-1026.
27. Tao, R.-H. and I.N. Maruyama, *All EGF (ErbB) receptors have preformed homo-and heterodimeric structures in living cells*. Journal of cell science, 2008. **121**(19): p. 3207-3217.
28. Maruyama, I.N., *Mechanisms of activation of receptor tyrosine kinases: monomers or dimers*. Cells, 2014. **3**(2): p. 304-330.
29. Mariño, K., et al., *A systematic approach to protein glycosylation analysis: a path through the maze*. Nature chemical biology, 2010. **6**(10): p. 713-723.
30. Liu, Y.-C., et al., *Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells*. Proceedings of the National Academy of Sciences, 2011. **108**(28): p. 11332-11337.
31. Cummings, R., A.M. Soderquist, and G. Carpenter, *The oligosaccharide moieties of the epidermal growth factor receptor in A-431 cells. Presence of complex-type N-linked chains that contain terminal N-acetylgalactosamine residues*. Journal of Biological Chemistry, 1985. **260**(22): p. 11944-11952.
32. Takahashi, M., et al., *N-glycan of ErbB family plays a crucial role in dimer formation and tumor promotion*. Biochimica et Biophysica Acta (BBA) - General Subjects, 2008. **1780**(3): p. 520-524.
33. Whitson, K.B., et al., *Functional effects of glycosylation at Asn-579 of the epidermal growth factor receptor*. Biochemistry, 2005. **44**(45): p. 14920-14931.
34. Dennis, J.W., I.R. Nabi, and M. Demetriou, *Metabolism, cell surface organization, and disease*. Cell, 2009. **139**(7): p. 1229-1241.
35. Freeze, H.H. and M. Aebi, *Altered glycan structures: the molecular basis of congenital disorders of glycosylation*. Current Opinion in Structural Biology, 2005. **15**(5): p. 490-498.
36. Peracaula, R., et al., *Altered glycosylation in tumours focused to cancer diagnosis*. Disease markers, 2008. **25**(4-5): p. 207-218.
37. Shental-Bechor, D. and Y. Levy, *Effect of glycosylation on protein folding: A close look at thermodynamic stabilization*. Proceedings of the National Academy of Sciences, 2008. **105**(24): p. 8256-8261.
38. Shakin-Eshleman, S., et al., *N-linked glycosylation of rabies virus glycoprotein. Individual sequons differ in their glycosylation efficiencies and influence on cell surface expression*. Journal of Biological Chemistry, 1992. **267**(15): p. 10690-10698.
39. Kornfeld, R. and S. Kornfeld, *Assembly of asparagine-linked oligosaccharides*. Annual review of biochemistry, 1985. **54**(1): p. 631-664.
40. Nilsson, I. and G. Von Heijne, *Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane*. Journal of Biological Chemistry, 1993. **268**(8): p. 5798-5801.

41. Schulz, B.L., *Beyond the sequon: sites of N-glycosylation*. 2012: INTECH Open Access Publisher.
42. Spiro, R.G., *Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds*. Glycobiology, 2002. **12**(4): p. 43R-56R.
43. Allen, M.P., *Introduction to molecular dynamics simulation*. Computational Soft Matter: From Synthetic Polymers to Proteins, 2004. **23**: p. 1-28.
44. Soderquist, A.M. and G. Carpenter, *Glycosylation of the epidermal growth factor receptor in A-431 cells. The contribution of carbohydrate to receptor function*. Journal of Biological Chemistry, 1984. **259**(20): p. 12586-12594.
45. Slike, L., T. Martensen, and M. Lane, *Synthesis of epidermal growth factor receptor in human A431 cells. Glycosylation-dependent acquisition of ligand binding activity occurs post-translationally in the endoplasmic reticulum*. Journal of Biological Chemistry, 1986. **261**(32): p. 15233-15241.
46. Bremer, E., J. Schlessinger, and S.-I. Hakomori, *Ganglioside-mediated modulation of cell growth. Specific effects of GM3 on tyrosine phosphorylation of the epidermal growth factor receptor*. Journal of Biological Chemistry, 1986. **261**(5): p. 2434-2440.
47. Zhen, Y., R.M. Caprioli, and J.V. Staros, *Characterization of glycosylation sites of the epidermal growth factor receptor*. Biochemistry, 2003. **42**(18): p. 5478-5492.
48. Smith, K.D., et al., *Analysis of the glycosylation patterns of the extracellular domain of the epidermal growth factor receptor expressed in Chinese hamster ovary fibroblasts*. Growth Factors, 1996. **13**(1-2): p. 121-132.
49. Tsuda, T., Y. Ikeda, and N. Taniguchi, *The Asn-420-linked Sugar Chain in Human Epidermal Growth Factor Receptor Suppresses Ligand-independent Spontaneous Oligomerization POSSIBLE ROLE OF A SPECIFIC SUGAR CHAIN IN CONTROLLABLE RECEPTOR ACTIVATION*. Journal of Biological Chemistry, 2000. **275**(29): p. 21988-21994.
50. Petrescu, A.-J., et al., *Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding*. Glycobiology, 2004. **14**(2): p. 103-114.
51. Partridge, E.A., et al., *Regulation of cytokine receptors by Golgi N-glycan processing and endocytosis*. Science, 2004. **306**(5693): p. 120-124.
52. Arkhipov, A., et al., *Membrane interaction of bound ligands contributes to the negative binding cooperativity of the EGF receptor*. 2014.
53. Liu, P., et al., *A single ligand is sufficient to activate EGFR dimers*. Proceedings of the National Academy of Sciences, 2012. **109**(27): p. 10861-10866.
54. Bessman, N.J., D.M. Freed, and M.A. Lemmon, *Putting together structures of epidermal growth factor receptors*. Current Opinion in Structural Biology, 2014. **29**: p. 95-101.
55. Kaszuba, K., et al., *N-Glycosylation as determinant of epidermal growth factor receptor conformation in membranes*. Proceedings of the National Academy of Sciences, 2015. **112**(14): p. 4334-4339.
56. Fox, S., et al., *The multifaceted roles of molecular dynamics simulations in drug discovery*. Current pharmaceutical design, 2016.
57. Karplus, M. and J. Kuriyan, *Molecular dynamics and protein function*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(19): p. 6679-6685.
58. Marrink, S.J., et al., *The MARTINI force field: coarse grained model for biomolecular simulations*. The Journal of Physical Chemistry B, 2007. **111**(27): p. 7812-7824.
59. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*. Vol. 21. 2010: Springer.

60. Andrew, R.L., *Molecular modelling: principles and applications*. 2001, Prentice hall, London.
61. Hillisch, A., M. Lorenz, and S. Diekmann, *Recent advances in FRET: distance determination in protein–DNA complexes*. *Current Opinion in Structural Biology*, 2001. **11**(2): p. 201-207.
62. Hornak, V., et al., *Comparison of multiple Amber force fields and development of improved protein backbone parameters*. *Proteins: Structure, Function, and Bioinformatics*, 2006. **65**(3): p. 712-725.
63. van Gunsteren, W.F., X. Daura, and A.E. Mark, *GROMOS force field*. *Encyclopedia of computational chemistry*, 1998.
64. Oostenbrink, C., et al., *A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6*. *Journal of computational chemistry*, 2004. **25**(13): p. 1656-1676.
65. Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields*. *Journal of computational chemistry*, 2010. **31**(4): p. 671-690.
66. Yu, W., et al., *Extension of the CHARMM general force field to sulfonyl-containing compounds and its utility in biomolecular simulations*. *Journal of computational chemistry*, 2012. **33**(31): p. 2451-2468.
67. Kirschner, K.N., et al., *GLYCAM06: a generalizable biomolecular force field. Carbohydrates*. *Journal of computational chemistry*, 2008. **29**(4): p. 622-655.
68. Salisbury, A.M., et al., *Ramachandran-type plots for glycosidic linkages: Examples from molecular dynamic simulations using the Glycam06 force field*. *Journal of computational chemistry*, 2009. **30**(6): p. 910-921.
69. Tessier, M.B., et al., *Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids*. *Molecular simulation*, 2008. **34**(4): p. 349-364.
70. Jeffrey, G., et al., *Application of ab initio molecular orbital calculations to the structural moieties of carbohydrates. 3*. *Journal of the American Chemical Society*, 1978. **100**(2): p. 373-379.
71. Takagi, S. and G. Jeffrey, *Neutron diffraction refinement of the crystal structures of methyl α -D-galactopyranoside monohydrate and methyl β -D-galactopyranoside*. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 1979. **35**(4): p. 902-906.
72. Woods Group. (2005-XXXX) GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (<http://glycam.org>).
73. Zhang, L.Y., et al., *Solvent models for protein–ligand binding: Comparison of implicit solvent Poisson and surface generalized Born models with explicit solvent simulations*. *Journal of computational chemistry*, 2001. **22**(6): p. 591-607.
74. Cheatham III, T.E. and P.A. Kollman, *Molecular dynamics simulation of nucleic acids*. *Annual Review of Physical Chemistry*, 2000. **51**(1): p. 435-471.
75. Makarov, V., B.M. Pettitt, and M. Feig, *Solvation and hydration of proteins and nucleic acids: a theoretical view of simulation and experiment*. *Accounts of chemical research*, 2002. **35**(6): p. 376-384.
76. Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules*. *Nature structural & molecular biology*, 2002. **9**(9): p. 646-652.
77. Bizzarri, A.R. and S. Cannistraro, *Molecular dynamics of water at the protein-solvent interface*. *The Journal of Physical Chemistry B*, 2002. **106**(26): p. 6617-6633.

78. Skyner, R., et al., *A review of methods for the calculation of solution free energies and the modelling of systems in solution*. Physical Chemistry Chemical Physics, 2015. **17**(9): p. 6174-6191.
79. Jorgensen, W.L., et al., *Comparison of simple potential functions for simulating liquid water*. The Journal of chemical physics, 1983. **79**(2): p. 926-935.
80. Feig, M. and C.L. Brooks, *Recent advances in the development and application of implicit solvent models in biomolecule simulations*. Current Opinion in Structural Biology, 2004. **14**(2): p. 217-224.
81. Still, W.C., et al., *Semianalytical treatment of solvation for molecular mechanics and dynamics*. Journal of the American Chemical Society, 1990. **112**(16): p. 6127-6129.
82. Srinivasan, J., et al., *Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects*. Theoretical Chemistry Accounts, 1999. **101**(6): p. 426-434.
83. Ryckaert, J.-P., G. Cicciotti, and H.J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. Journal of Computational Physics, 1977. **23**(3): p. 327-341.
84. Kräutler, V., W.F. Van Gunsteren, and P.H. Hünenberger, *A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations*. Journal of computational chemistry, 2001. **22**(5): p. 501-508.
85. Gonnet, P., *P-SHAKE: a quadratically convergent SHAKE in $O(n^2)$* . Journal of Computational Physics, 2007. **220**(2): p. 740-750.
86. Andersen, H.C., *RATTLE: A "Velocity" version of the SHAKE algorithm for molecular dynamics calculations*. Journal of Computational Physics, 1983. **52**(1): p. 24-34.
87. Lee, S.-H., K. Palmo, and S. Krimm, *WIGGLE: A new constrained molecular dynamics algorithm in Cartesian coordinates*. Journal of Computational Physics, 2005. **210**(1): p. 171-182.
88. Miyamoto, S. and P.A. Kollman, *SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models*. Journal of computational chemistry, 1992. **13**(8): p. 952-962.
89. Hess, B., et al., *LINCS: a linear constraint solver for molecular simulations*. Journal of computational chemistry, 1997. **18**(12): p. 1463-1472.
90. Hamelberg, D., J. Mongan, and J.A. McCammon, *Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules*. The Journal of chemical physics, 2004. **120**(24): p. 11919-11929.
91. Hamelberg, D., C.A.F. de Oliveira, and J.A. McCammon, *Sampling of slow diffusive conformational transitions with accelerated molecular dynamics*. The Journal of chemical physics, 2007. **127**(15): p. 155102.
92. Søndergaard, C.R., et al., *Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values*. Journal of chemical theory and computation, 2011. **7**(7): p. 2284-2295.
93. Pearlman, D.A., et al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules*. Computer Physics Communications, 1995. **91**(1): p. 1-41.
94. Sato, C., et al., *Characterization of the N-Oligosaccharides Attached to the Atypical Asn-X-Cys Sequence of Recombinant Human Epidermal Growth Factor Receptor*. Journal of biochemistry, 2000. **127**(1): p. 65-72.

95. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures*. Bioinformatics, 2006. **22**(21): p. 2695-2696.
96. Humphrey, W., Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics" J. Molec. Graphics 1996, 14.1, 33-38.
97. Berendsen, H.J., D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation*. Computer Physics Communications, 1995. **91**(1): p. 43-56.
98. Miller III, B.R., et al., *MMPBSA. py: An efficient program for end-state free energy calculations*. Journal of chemical theory and computation, 2012. **8**(9): p. 3314-3321.
99. The PyMOL Molecular Graphics System, V.S., LLC.
100. Sivasubramanian, A., et al., *Structural model of the mAb 806-EGFR complex using computational docking followed by computational and experimental mutagenesis*. Structure, 2006. **14**(3): p. 401-414.
101. Yoon, S.-J., et al., *Epidermal growth factor receptor tyrosine kinase is modulated by GM3 interaction with N-linked GlcNAc termini of the receptor*. Proceedings of the National Academy of Sciences, 2006. **103**(50): p. 18987-18991.
102. Maruyama, I.N., *Activation of transmembrane cell-surface receptors via a common mechanism? The "rotation model"*. BioEssays, 2015. **37**(9): p. 959-967.
103. Stroop, C.J., et al., *Characterization of the carbohydrate chains of the secreted form of the human epidermal growth factor receptor*. Glycobiology, 2000. **10**(9): p. 901-917.
104. Krissinel, E. and K. Henrick. *Detection of protein assemblies in crystals*. in *International Symposium on Computational Life Science*. 2005: Springer.
105. Pierce, L.C., et al., *Routine access to millisecond time scale events with accelerated molecular dynamics*. Journal of chemical theory and computation, 2012. **8**(9): p. 2997-3002.
106. Cho, H.-S. and D.J. Leahy, *Structure of the extracellular region of HER3 reveals an interdomain tether*. Science, 2002. **297**(5585): p. 1330-1333.
107. Bouyain, S., et al., *The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(42): p. 15024-15029.
108. Endres, N.F., et al., *Conformational coupling across the plasma membrane in activation of the EGF receptor*. Cell, 2013. **152**(3): p. 543-556.
109. Roskoski, R., *The ErbB/HER receptor protein-tyrosine kinases and cancer*. Biochemical and Biophysical Research Communications, 2004. **319**(1): p. 1-11.
110. Chung, I., et al., *Spatial control of EGF receptor activation by reversible dimerization on living cells*. Nature, 2010. **464**(7289): p. 783-787.
111. Gadella, T. and T.M. Jovin, *Oligomerization of epidermal growth factor receptors on A431 cells studied by time-resolved fluorescence imaging microscopy. A stereochemical model for tyrosine kinase receptor activation*. The Journal of cell biology, 1995. **129**(6): p. 1543-1558.
112. Bader, A.N., et al., *Homo-FRET imaging enables quantification of protein cluster sizes with subcellular resolution*. Biophysical journal, 2009. **97**(9): p. 2613-2622.
113. Martin-Fernandez, M., et al., *Preformed oligomeric epidermal growth factor receptors undergo an ectodomain structure change during signaling*. Biophysical journal, 2002. **82**(5): p. 2415-2427.
114. Szabó, Á., et al., *Quantitative characterization of the large-scale association of ErbB1 and ErbB2 by flow cytometric homo-FRET measurements*. Biophysical journal, 2008. **95**(4): p. 2086-2096.

115. Wang, X.-q. and A.S. Paller, *Membrane Regulation of EGFR Signaling by Gangliosides*. Open Dermatology Journal, 2009. **3**: p. 159-162.

List of publications

Journal articles

Azimzadeh Irani M, Kannan S and Verma CS (2017) Role of N -glycosylation in EGFR ectodomain ligand binding (Manuscript accepted PROTEINS).

Oral presentation

Azimzadeh Irani M, Verma C (2016) Role of N-glycosylation on EGFR Ectodomain Ligand Binding. 15th International Conference On BioInformatics (InCOB 2016), 21th to 23rd of September, 2016.

Poster presentations in international conferences

Azimzadeh Irani M, Verma C (2016). N-glycosylation Modulates the Tethered-extended Equilibrium of the Extracellular Domain of EGFR. 2016 Annual Meeting of the Society for Glycobiology, November 19-22, 2016 New Orleans, Louisiana, USA.

Azimzadeh Irani M, Verma C (2016) Role of N-glycosylation on EGFR Ectodomain Ligand Binding. 15th International Conference On BioInformatics (InCOB 2016), 21th to 23rd of September, 2016.

Azimzadeh Irani M, Verma C (2015) Glycosylation of EGFR Extracellular Domain Induces Receptor Stability. the 29th Annual Symposium of The Protein Society, 22-25 July 2015.