Spring 5-15-2017

Biophysical studies of the intracellular domains of the EGFR family of Receptor Tyrosine Kinases

Kwabena A N Sarpong
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Biochemistry

Dissertation Examination Committee:
Ron Bose, Chair
Thomas Brett
Carl Frieden
Roberto Galletto
Linda Pike

Biophysical studies of the intracellular domains of the EGFR family of Receptor Tyrosine Kinases

by

Kwabena Amofa Nketia Sarpong

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2017

St. Louis, Missouri

Table of Contents

List of Figures……...……iv

List of Tables……...….vii

Acknowledgements…….viii

Abstract……...……...xi

Chapter 1: Introduction……..……..1

1.1. EGFR Receptor Tyrosine kinases, an overview…….……..2

1.2. Role and significance of the EGFR family in cancer……...……3

1.3. Structural analysis of the kinase domain of the EGFR family……..4

1.4. Inhibition of the kinase domain by small molecules and Mig6……...…….6

1.5. Regulation of EGFR signaling by the C-terminal tail……..6

1.6. Intrinsic disorder and role in diseases…….……..7

1.7. Biophysical approaches for studying intrinsically disordered proteins……...….8

1.8. Site-specific modification of proteins……...…9

1.9. The use of sortase to efficiently label proteins……...……10

1.10. Conclusion……11

Chapter 2: Biophysical Evidence for Intrinsic disorder in the C-terminal tail domains of the Epidermal Growth Factor Receptors……..……13

2.1. Acknowledgement……..……15

2.2. Abstract……...…16

2.3. Introduction…….……17

2.4. Methods……...……20

2.5. Results……27

2.6. Discussion……47

Chapter 3: Efficient Sortase-mediated N-terminal Labeling of TEV Protease Cleaved Recombinant Proteins……48

3.1. Acknowledgement……50

3.2. Abstract……51

3.3. Introduction……52

3.4. Methods……55

3.5. Results……63

3.6. Discussion……70

Chapter 4: Biophysical Characterization of kinase domain dimerization of the Epidermal Growth Factor Receptors……73

4.1. Acknowledgement……74

4.2. Abstract……75

4.3. Introduction……76

4.4. Methods……79

4.5. Results……88

4.6. Discussion……98

Chapter 5: Conclusion and Future Directions……100

5.1. Conclusion and Future Perspectives……101

References……106

List of Figures

Figure 1.1 Structural architecture and mechanism of activation of the EGFR family of Receptor Tyrosine kinases……3

Figure 1.2 Model of the EGFR asymmetric kinase domain dimer……..5

Figure 1.3 Sortase-mediated anchoring of proteins to bacterial cell wall…….11

Figure 2.1 Disorder prediction by PONDR VL-XT algorithm……..28

Figure 2.2 Purification and Phosphorylation of C-Terminal Tail (CTT) constructs…..31

Figure 2.3 Circular dichroism spectroscopy……34

Figure 2.4 Hydrodynamic Properties of EGFR CTT…….38

Figure 2.5 Small angle X-ray scattering analysis of EGFR CTT……...42

Figure 2.6 Small angle X-ray scattering analysis of HER3 CTT in solution ± urea……44

Figure 3.1 Schematic representation of N-terminal labeling of TEV-cleavable proteins using sortase…….60

Figure 3.2 Coomassie-stained SDS-PAGE gel of EGFR kinase domain purification and TEV protease cleavage……63

Figure 3.3 Coomassie-stained SDS-PAGE gel of Sortase A 5M (without a His-tag) purification and TEV protease cleavage……64

Figure 3.4 Sortase activity of 6x-His SrtA 5M and TEV-cleaved SrtA 5M……..65

Figure 3.5 Comparison of labeling of Gly-EGFR, Gly2-EGFR and Gly4-EGFR using an LPETG peptide substrate containing the TAMRA fluorophore…….66

Figure 3.6 Effect of fluorophore label on EGFR kinase domain activity and activation on nickel liposomes……..67

Figure 3.7 **Coomassie-stained gel showing the labeling and purification of EGFR kinase domain using an LPETG peptide substrate containing CruzQuencher maleimide……..68**

Figure 3.8 **Coomassie and fluorescent gels showing the labeling and purification of MSP protein using the H_6-LPETGG peptide substrate containing fluorescein…..69**

Figure 3.9 **Time course N-terminal labeling of EGFR KD using the H_6-LPETGG peptide substrate labeled with fluorescein at two different positions.....69**

Figure 4.1 **Mig6 inhibition of HER2 heterodimers……...……..78**

Figure 4.2 **Multisequence alignment of the kinase domains of EGFR family members.80**

Figure 4.3 **Cysteine residues in the kinase domain of wildtype EGFR……...……..81**

Figure 4.4. **Effect of single cysteine mutation (EGFR C0, C1 and C2) on kinase activity and activation on liposomes. Cysteines at respective residues were mutated to serines……87**

Figure 4.5. **Effect of single cysteine mutation (EGFR C3, C5 and C6) on kinase activity and activation on liposomes. Cysteines at respective residues were mutated to serines……88**

Figure 4.6. **Effect of triple (C0C5C3) and quadruple (C0C5C3C2) cysteine mutations on kinase activity and activation on liposomes. Cysteines at respective residues were mutated to serines……...88**

Figure 4.7. **Effect of single (C1, C-A) and quadruple (C0C5C3C2) cysteine mutations on kinase activity and activation on liposomes……...89**

Figure 4.8 **Effect of cysteine mutation at the C2 position in the quadruple mutant (C0C5C3C2) on kinase activity and activation on liposomes……..89**

Figure 4.9. Effect of cysteine mutation at the C2 position in the quadruple mutant (C0C5C3C2) on kinase activity and activation on liposomes……..90

Figure 4.10. Effect of cysteine mutation at the C1 and C2 positions in two quadruple mutants (C0C5C3C2 and C0C5C3C1) on kinase activity and activation on liposomes……..….90

Figure 4.11. Effect of EGFR kinase domain pentamutant (C0C5C3C2C1) on kinase activity and activation on liposomes……...91

Figure 4.12 Effect of LplA and CuAAC conditions on the kinase activity of EGFR kinase domain……...92

Figure 4.13 Effect of EGFR kinase domain C-terminal modifications on kinase activity and activation on liposomes……93

Figure 4.14 Effect of TEV protease cleavage, labeling and purification conditions on EGFR kinase domain activity……94

Figure 4.15. Effect of fluorophore labeling on EGFR kinase domain activity. EGFR kinase domain was labeled with Alexa488 maleimide……..95

Figure 4.16. Effect of fluorophore labeling on EGFR kinase domain activity. EGFR kinase domain was labeled with Alexa568 maleimide……..95

Figure 4.17. FRET measurements using Alexa488- and Alexa568-labeled EGFR kinase domains…….....97

Figure 4.18 Effect of concentration of the fluorophore-labeled peptide alone on FRET ` measurements……...……....98

List of Tables

Table 2.1 EGFR and HER3 amino acid composition analysis by region…….29

Table 2.2 Structural parameters observed by SAXS……46

Table 4.1 Nomenclature of cysteine residues in the EGFR kinase domain…….80

Table 4.2 FRET ratios for EGFR kinase domain (and dimerization-deficient mutant) and MSYY peptide……...98

Acknowledgments

I am most grateful to the Almighty God for his providence and care throughout my doctoral training at Washington University. I would like to express my sincere gratitude to everyone who has been there through thick and thin to make this thesis a reality. I am wholeheartedly thankful to my research and thesis advisor, Dr. Ron Bose whose mentorship, guidance and support have instilled a high level of discipline in me. His supervision, academic and research expertise have kindled my passion to pursue biomedical research to the highest level and be a valuable asset to humanity. I owe him more than he knows.

I would like to thank Drs. Carl Frieden, Linda Pike, Thomas Brett and Roberto Galletto for serving on my thesis committee. Their suggestions, time and effort invested into my research helped trigger and nourish my intellectual maturity. I wish to express my heartfelt gratitude to Robert Obermann, Drs James Janetka, Gaya Amarasinghe, Jian Zhu, Paolo De bona, Phyllis Hanson, Timothy Lohman and Alex Kozlov for providing instrumental support and explaining other scientific concepts.

To the former and current lab mates of the Bose lab, I am very fortunate to have enjoyed the conducive working environment that existed in the lab and the collaborative work that we engaged in. I am very thankful to John Monsey, for his immense expertise and all-round experience in the projects that I have undertaken in the lab. I thank Runjun Kumar, Vandna Kukshal, Shyam Kavuri, Naveen Jain, Adam Searlemann, Tim Collier, Alaina Boyer, Ted Keppel, Elisa Murray, Wei Shen, Ari Gao for their interactions on both scientific and non-scientific levels.

My graduate school life wouldn't have been fulfilling without the collegial environment provided by friends, especially John Jimah, Shannon Ohlemacher, Araba Budu-Anguah and Kwaku Tawiah.

To my mother and siblings, your love and care, sacrifice and immense support throughout these years is highly appreciated and would be remembered in years to come. This thesis would not have been possible without the enormous input of my father (of blessed memory), who has been a strong backbone in my pursuance of doctoral level training in biomedical research. Thanks Daddy.

To end, I am very grateful for having such a wonderful and supportive wife, Julie, who has been there throughout my PhD training. We have been blessed with a daughter, Afia and a son, Nana, whose energy and curiosity has been a constant motivation for me. My family have accommodated all the joys and griefs throughout my training and I appreciate their love, sacrifice and support.

Kwabena Amofa Nketia Sarpong

Washington University in St. Louis

May 2017

Dedicated to my family.

ABSTRACT OF THE DISSERTATION

Biophysical studies of the intracellular domains of the EGFR family of Receptor Tyrosine Kinases

by

Kwabena Amofa Nketia Sarpong

Doctor of Philosophy in Biology and Biomedical Sciences

Biochemistry

Washington University in St. Louis, 2017

Professor Ron Bose, Chair

The epidermal growth factor receptor and its three orthologues, HER2, HER3 and HER4 have been the subject of intensive basic, clinical and translational research due to their involvement in cancers. These proteins are part of elaborate networks that interact with a myriad of other molecules to effect diverse signaling pathways and affect cellular processes such as migration, apoptosis, cell differentiation and so on. In addition to the formation of preformed dimers, it is well established that ligand engagement leads to receptor dimerization in all family members, except HER2. In the past decade, it has been shown that their intracellular domains dimerize in

an asymmetric fashion where the C-lobe of the donor kinase interacts extensively with the N-lobe of the receiver kinase. This structural model is critical for activation of the receiver kinase and subsequent phosphorylation of the C-terminal tail. While the structures of the different domains have been solved, minimum structural and biophysical studies have been performed on the C-terminal tails. In this thesis, we use multiple approaches to characterize the C-terminal tails of EGFR and HER3, and show that they are intrinsically disordered regions of these receptors. We also developed a strategy that has been employed to site-specifically fluorophore-label the EGFR kinase domain, and would serve as a biological probe to directly measure EGFR kinase domain dimerization on lipid surfaces.

Chapter 1: Introduction

1.1 EGFR Receptor Tyrosine kinases, an overview

The isolation of the epidermal growth factor by Stanley Cohen in the 1950s (1) and its significance in physiology led to the identification of the epidermal growth factor receptor, EGFR, HER or ERBB1 (2). Since then, the EGFR family has grown to include HER2 or ERBB2, HER3 or ERBB3 and HER4 or ERBB4 (3). These family of receptor tyrosine kinases (RTKs) play significant roles in the normal functioning of the cell. Cellular processes that these RTKs regulate include cell migration, proliferation and differentiation, cell survival and metabolism, and cell-cycle control. All four family members share a very similar molecular and structural architecture; a large ectodomain, that binds of ligand, a single alpha helical transmembrane domain, and an intracellular cytoplasmic domain that contains the juxtamembrane segment, the kinase domain and an intrinsically disordered C-terminal tail region as shown in Figure 1.1 (4-7). Except HER2 which is constitutively active and does not have a cognate ligand, the other three members are activated by ligands (epidermal growth factor, transforming growth factor, heregulin, amphiregulin and heparin-binding epidermal growth factor) (8). HER3 on the other hand has little to no kinase activity and studies have shown that HER3 assumes the donor kinase position when in a heterodimeric pair (4,9). Following dimerization of the extracellular domains, their intracellular cytoplasmic domains dimerize in an asymmetric fashion and activates the receiver kinase. This engages the unstructured C-terminal tail and becomes phosphorylated at various tyrosine residues, providing docking sites for Src homology 2 or phosphotyrosine-binding signaling molecules to interact (Figure 1.1) (10). The

result is the recruitment of other adaptor proteins that signal via different pathways (11-14).

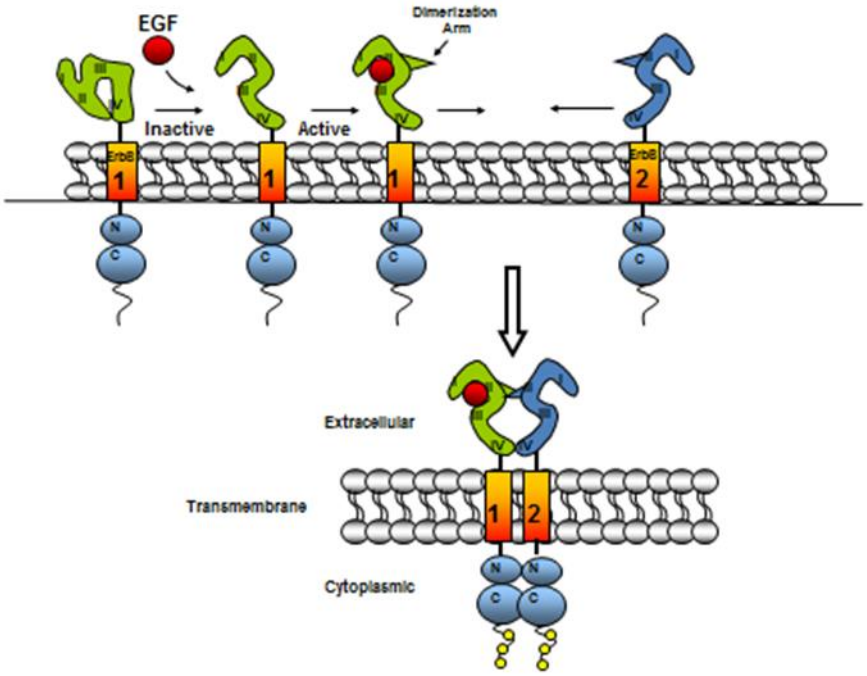


Figure 1.1. Structural architecture and mechanism of activation of the EGFR family of Receptor Tyrosine Kinases. This figure is adapted from Zahnow, C.A., 2006 (10)

1.2 Role and significance of EGFR family in cancer

EGFR involvement in cancer surfaced in the 1980s when the erythroblastosis tumor virus was shown to encode an abnormal form of EGFR. The EGFR family of receptor tyrosine kinases have been implicated in several solid tumors (11). These cancers have often been overexpression of the receptors, gene amplification, autocrine loop activation and the presence of activating somatic mutations in the EGFR family (15-19). It has been widely known that the HER2-HER3 signaling is the most potent signaling pair (20), especially in HER2-amplified breast cancers where HER3 expression is a requirement for the growth of these cancers. A similar consequence is seen in patients with non-small cell lung cancers caused by an activating mutation in EGFR,

where studies have shown that suppression of HER3 signaling results in growth suppression (21, 22). HER3 is now an actively pursued drug target since it provides acquired resistance to therapeutics by activating alternative signaling pathways (23). HER3 achieves this by increasing outputs in processes such as transcription, translation, phosphorylation and the use of other redundant mechanisms which are downstream of its signaling output. Somatic mutations in HER3 have recently been published and various groups are actively pursuing the functional significance of these mutations (18, 24-26). HER4, on the other hand has not been shown to have important roles in cancer, though some studies have reported its involvement in melanoma (27). Currently, there are FDA-approved small molecule tyrosine kinase inhibitors such as lapatinib and antibody-based therapeutics such as trastuzumab and pertuzumab that target EGFR and HER2-driven cancers (28-30).

1.3 Structural analysis of the kinase domain of the EGFR family

The structural organization of EGFR and its three orthologues (HER2, HER3 and HER4) is strikingly similar; they all possess a large extracellular domain that binds to ligand, a single alpha-helical transmembrane domain and an intracellular cytoplasmic domain that contains the kinase domain and the intrinsically disordered C-terminal tail (4-7). The extracellular domain is organized into 4 subdomains. While subdomains I and III are rich in leucine repeat motifs, subdomains II and IV contain C-C disulfide-rich bonds. In the absence of ligand, EGFR, HER3 and Her4 receptors exist in a closed, tethered conformation that is mediated by interactions between domains II and IV. EGF (or other cognate ligands) binding leads to an alteration in the structure of the extracellular domain to create a ligand-binding pocket and the protrusion of a dimerization arm (from subdomain II). Homo- and heterodimerization is facilitated by interactions with the dimerization arm (31). HER2 does not require ligand binding to achieve an

active conformation and its extracellular domain is poised in a fixed, activated state even when monomeric, rendering HER2 permanently available for dimerization (32). There are dimer contacts in the extracellular and intracellular regions as well as the transmembrane helix, though not coupled together (33). The kinase domain consists of the N-terminal and C-terminal lobes. Several lines of evidence show that the kinase domain dimerizes in an asymmetric fashion; where the C-lobe of the donor kinase makes extensive contacts with the N-lobe of the receiver kinase. In this model, (figure 1.2) the receiver kinase becomes activated and phosphorylates tyrosine residues in the C-terminal tail, which serve as docking sites for phosphotyrosine-binding proteins and SH2 domains (4). The activation process is sequential and dependent on the enzymatic kinase activity of the receptor (34). Other critical regions in the kinase domain include the activation loop and the αC helix, both of which adopt different conformations when in an active or inactive state (35). The formation of the asymmetric dimer means that specific mutations at the interface can compromise the catalytic activity of the kinase (4, 36).

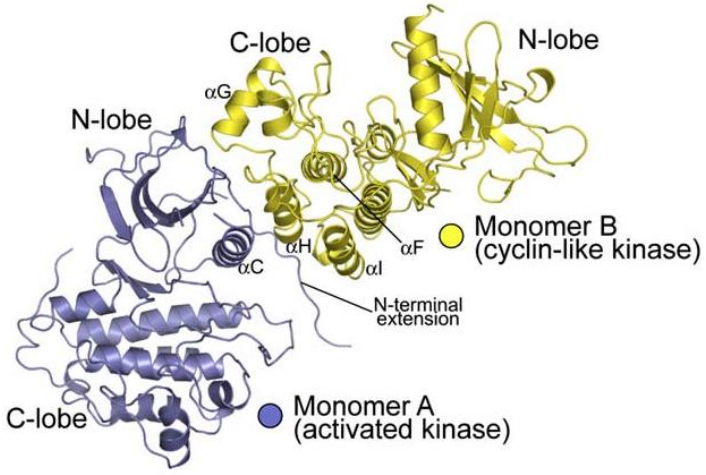


Figure 1.2. Model of the EGFR asymmetric kinase domain dimer. This figure is adapted from Zhang et. al, 2006 (4)

1.4 Inhibition of the kinase domain by small molecules and Mig6

Since the discovery of their involvement in cancer, a variety of therapeutic strategies have been developed to target EGFR family of proteins-driven tumors. This has been achieved by using monoclonal antibodies such as pertuzumab, trastuzumab and others that bind to the ectodomain and interfere with signaling. These antibodies are either used singly or in combination, where their proposed mechanism of action is bind to subdomain II or IV resulting in the inhibition of dimerization and blocking of ligand-dependent and ligand-independent signaling (37). Recently, alternative approaches have been investigated and resulted in the use of ATP-competitive small molecules to inhibit the kinase domain (30). While some of these kinase domain inhibitors possess a reversible mode of action, others such as neratinib attaches irreversibly to the cysteine residue close to the ATP binding domain and results in the complete shutdown of the kinase activity of the receptor (28). Although enormous success has been achieved by targeting the extracellular and kinase domains, the emergence of drug resistance has necessitated the development of new targets and better therapies for these EGFR-driven cancers (38). The formation of an asymmetric dimer is critical for activation of the kinase and downstream signaling. This dimer interface presents a crucial target that can be exploited for development of drug candidates to inhibit EGFR signaling. Indeed, there is an endogenous protein, Mig6 (Mitogen-induced gene 6) that is known to bind to the kinase domain and inhibit its activity. Structure-based methods have shown that Mig6 binds to the distal surface of the C-lobe of the kinase domain and thus prevents the formation of an asymmetric dimer (39). The Bose lab and other research groups are actively pursuing the use of the Mig6 peptide (and its other mutants) to inhibit dimerization and block signaling (unpublished data).

1.5 Regulation of EGFR signaling by the C-terminal tail

The intracellular domain of EGFR (and its three orthologues) consist of the membrane proximal juxtamembrane segment, the kinase domain and a C-terminal tail that contains more than 200 residues. Composite analysis has shown a <25% sequence similarity among the four EGFR C-terminal tails. Upon dimerization and activation, the C-terminal tail interacts with the active site of the kinase domain and results in the phosphorylation of the tyrosine residues (11). The phosphorylated tail serves as a docking site for binding by phosphotyrosine binding proteins and other proteins that contain SH2 domains (40). In addition, phosphatases can dephosphorylate the C-terminal tail (41, 42). This cycle of events compels the C-terminal tail to adopt a variety of conformations that allows it to interact efficiently with its different partners under various conditions. It wasn't until recently that structural and biophysical approaches begun to shed light on our understanding of the C-terminal tail (43-45). Experimental and in silico approaches have shown that the C-terminal tail of EGFR and HER3 are intrinsically disordered, and possess a larger apparent molecular size compared to a globular protein of similar molecular weight (45). Crystal structures containing residues 982-1022 of the C-terminal tail of EGFR have suggested a negative inhibitory role for the C-terminal tail. In this model, the proximal residues of the C-terminal tail form an alpha helix (residues 997-1001) that stabilizes the inactive dimer, while the acidic residues (Electrostatic hook, residues 1003-1014) and distal residues (1015-1022) that form a beta strand prevents the formation of the juxtamembrane latch necessary for activation (46). This finding is buttressed by another study where the authors identified residues 982-1054 in the proximal region of the C-terminal tail that inhibit the formation of an asymmetric dimer (47)

1.6 Protein intrinsic disorder and role in diseases

Intrinsically disordered proteins (IDPs) are regions or domains that do not have stable secondary or tertiary structures. Nonetheless, these regions maintain the ability to function in cellular processes including regulation of transcription, signaling, protein-protein interactions, cell life cycle events etc. They perform these roles due to their ability to adopt different conformations under different conditions and due to their structural flexibility (48-55). This results in IDPs having to fold and assume different conformations when interacting with different partners. Misfolding of these regions ultimately leads to the formation of unwanted aggregates and complexes which do not possess the structural and functional integrities to undertake their roles thereof. The proteomes of eukaryotes are replete with intrinsically disordered regions, where most of these proteins contain long stretches of disordered regions or the whole protein is almost completely disordered (56). Using PONDR VL-XT disorder predictor algorithm to analyze protein datasets, it has been suggested that there is a strong correlation between protein intrinsic disorder and their roles in signaling and cancer (57). Most importantly, in silico and experimental data have revealed many protein kinases contain intrinsically disordered regions, which facilitate their interactions with other proteins and enhance their signaling outputs (58).

1.7 Biophysical approaches for studying intrinsically disordered proteins

Though intrinsically disordered proteins might be accorded unstructured properties in protein chemistry, their functional properties provide enormous suggestion that their structural organization is intimately related to the cellular processes they are involved in. Traditional and classical methods such as Nuclear Magnetic Resonance and X-ray Crystallography were mainly developed to study proteins that form stable structures. Due to their structural and other inherent physical properties, it is often very difficult to obtain high concentrations or representative crystal structures of IDPs to enable their study using these traditional methods (51, 53). In the

past few years, there have been considerable efforts to develop methods to study the behavior of IDPs. Scattering approaches such as Dynamic and Light Scattering, Small Angle X-ray Scattering and Small Angle Neutron Scattering in combination with Analytical ultracentrifugation and Size Exclusion Chromatography have been used extensively to assess the size and shape of proteins (59-63). These methods can provide quantitative measurements such as the hydrodynamic radius, radius of gyration, stoke radius on IDPs under solution conditions. Because of their highly dynamic and flexible nature, experimental and molecular dynamics methods have been and are still being developed to analyze the conformational behavior of IDPs. Lastly, temperature, pH and denaturant-induced changes together with single-molecule Fluorescence Resonance Energy Transfer and Hydrogen-Deuterium exchange with mass spectrometry have spearheaded experimental approaches that are utilized to assess the conformational transitions of IDPs (64-68)

1.8 Site-specific modification of proteins

Biotechnological and biochemical applications are becoming increasingly dependent on derivatized proteins. These modified proteins provide excellent avenues for studying protein-protein and protein-ligand interactions, folding, post-translational modifications, conformational changes and further interrogating the functions of proteins both *in vivo* and *in vitro*. The past few decades have seen the development of several protein engineering strategies to site-specifically modify proteins by employing distinct chemistries and utilizing the specificities of various enzymes (69, 70). Most commonly used chemical handles for protein modification include fluorescent tags, chemical cross linkers, biotin, carbohydrates, nucleic acids and so on. Due to their low abundance in the proteome, cysteines have traditionally been employed via maleimide chemistry to label proteins (71). Other methods have exploited the reactive N-terminal amino

group and the ε-amino group of lysines to selectively attach N-hydroxysuccinimidyl esters to proteins. In addition, chemical labeling strategies have been developed to modify tyrosine, aspartate and glutamate residues. Chemoenzymatic approaches, on the other hand utilize a variety of enzymes to recognize a specific engineered motif on the protein and attach a chemical handle onto the protein. Enzymes such as lipoic acid ligase, biotin ligase, sortase, phosphopantetheinyltransferase among others have been employed to selectively label different proteins (72). Such tag-mediated protein labeling approaches should (i) be fast, robust and quantitative, (ii) be highly selective and specific, (iii) have minimum structural and functional consequences (iv) short reaction times and (v) result in the formation of a stable, covalent bond between the protein and the desired moiety. Although not a single method can satisfy all these requirements, the bottlenecks associated with these labeling methods include the tolerance of the proteins to incorporate the recognition motif and the difficulty in removing unmodified proteins prior to their use.

1.9 The use of sortase to efficiently label proteins

Sortases are membrane-associated transpeptidases that anchor Gram-positive bacterial surface proteins to their cell walls. Due to their significance in virulence, sortases might be considered as strong drug targets for fighting bacterial infections. Although there are four subfamilies of sortases (subfamily A – D), sortase A has been the prototypical member that has been studied extensively and employed for a myriad of biophysical applications. Sortase A is more versatile and accepts a variety of protein substrates as compared to the other family members. Mechanistically, sortase A cleaves the threonine-Glycine bond on any protein that contains the sortase A recognition motif, Leu-Pro-Xxx-Thr-Gly (LPXTG, where X can any amino acid, preferably glutamate). This leads to the formation of an acyl-enzyme intermediate, which is

resolved by the nucleophilic attack of an oligoglycine peptide. In bacteria, the resulting oligoglycine-protein complex becomes incorporated and displayed on the cell wall, as shown in figure 1.3 (73-77). Because the oligoglycine and LPXTG motifs can be easily synthesized or engineered into proteins, sortase A has been used in a variety of ways to modify any protein of interest with different chemical handles. One important advantage in using sortase for protein labeling is its ability to modify proteins at the C- and N-termini, and also on internal loop regions (76, 77). Sortase-mediated reactions are known to be highly specific and result in higher yields of the labeled protein. As a result, sortase has been utilized to label proteins in solution, in cell lysates and on the surface of living cells.

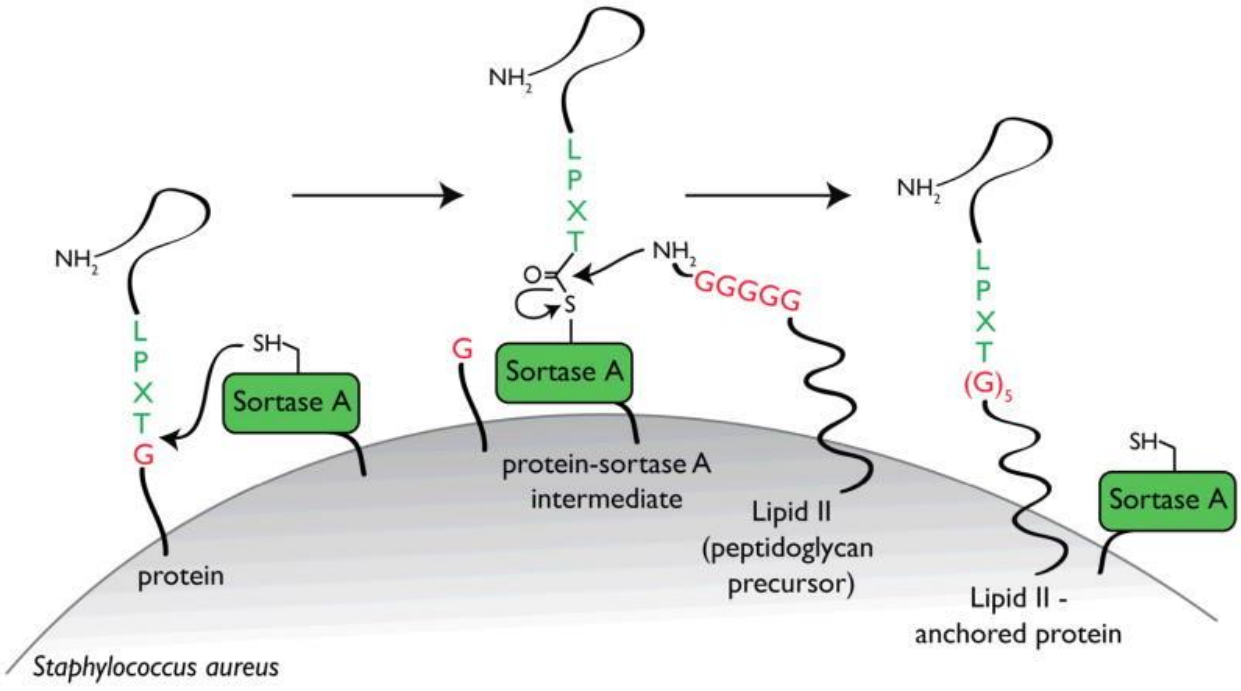


Figure 1.3. Sortase-mediated anchoring of proteins to bacterial cell wall. Adapted from Guimaraes et. al, 2013 (77)

1.10 Conclusion

In conclusion, EGFR and its three orthologues have been primary targets for drug developments for nearly four decades, because of their involvement in tumorigenesis. From a research perspective, the common structural organization of this family of receptor tyrosine kinases allows studying their modular domains in isolation. This helps to build a robust composite structural picture of EGFR, and stimulate investigation of how the various domains function together in signaling. The discovery in 2006 by the Kuriyan group, that the formation of an asymmetric dimer by the kinase domains was needed for activation, was very critical in shaping our understanding of the mechanistic details of EGFR signaling. The field has been fortunate to gain detailed insights on the structural basis of activation and how specific regions such as the juxtamembrane domain and transmembrane contribute to dimerization. Nonetheless, we still do not have a general platform that allows the direct measurement of dimerization of the kinase domains on lipid membranes, which would be instrumental in determining quantitative parameters that govern EGFR kinase domain dimerization. Such studies would also allow us to answer questions pertaining to how cancer-associated mutations affect dimerization. In addition, the C-terminal tail of these proteins have not been extensively studied to determine its significance in dimerization. This is in part due to the absence of enough biophysical and structural studies on the EGFR C-terminal tails. As such, the next three chapters of this thesis attempts to address two pertinent questions in the field: (i) what are the structural and biophysical characteristics of EGFR and HER3 C-terminal tails and (ii) can we develop a robust strategy to directly measure EGFR kinase domain dimerization on lipid surfaces.

Chapter 2: Biophysical Evidence for Intrinsic disorder in the C-terminal tail domains of the Epidermal Growth Factor Receptors

This chapter was adapted from:

Keppel TR*, Sarpong K*, Murray EM, Monsey J, Zhu J, Bose R. (2017) Biophysical Evidence for Intrinsic Disorder in the C-terminal Tails of the Epidermal Growth Factor Receptor (EGFR) and HER3 Receptor Tyrosine Kinases. *J. Biol. Chem.* 292, 597 – 610. *These authors contributed equally to this work.

2.1 ACKNOWLEDGEMENTS

I thank the members of the Bose lab for the extensive support and assistance in completing this work. I acknowledge the following researchers at Washington University for assistance in instrumentation: Daved Fremont and Gaya Amarasinghe for providing dynamic light scatter and multiangle light scatter instrumentation, Roberto Galletto, Paolo De Bona, Timothy Lohman and Alex Kozlov for assistance in analytical ultracentrifugation, Carl Frieden for providing circular dichroism instrumentation and Jian Zhu (Evan Sadler lab) for facilitation of the small angle X-ray scatter experiments. We also thank Linda Pike, Rohit Pappu, Carl Frieden, and Elliot Elson for helpful discussions.

2.2 ABSTRACT

The epidermal growth factor receptor (EGFR)/ErbB family of receptor tyrosine kinases include oncogenes important in the progression of breast and other cancers, and they are targets for many drug development strategies. Each member of the ErbB family possesses a unique, structurally uncharacterized C-terminal tail that plays an important role in autophosphorylation and signal propagation. To determine whether these C-terminal tails are intrinsically disordered regions, we conducted a battery of biophysical experiments on the EGFR and HER3 tails. HDX-MS experiments (performed by Ted Keppel, Bose lab) on the intracellular domain of EGFR, HER2 and HER3 had shown a rapid exchange behavior of the C-terminal tails compared to the kinase domains, which is characteristic of a highly dynamic and/or frequently exchange-competent conformational state. We expressed and purified recombinant EGFR and HER3 C-terminal tail-only constructs. Results from circular dichroism spectroscopy, size exclusion chromatography with multiangle light scattering, dynamic light scattering, analytical ultracentrifugation, and small angle X-ray scattering each provide evidence that the EGFR and HER3 C-terminal tails are intrinsically disordered with extended, non-globular structure in solution. The intrinsic disorder and extended conformation of these tails may be important for their function by increasing the capture radius and reducing the thermodynamic barriers for binding of down- stream signaling proteins.

2.3 INTRODUCTION

The epidermal growth factor receptor (EGFR)/ErbB family of receptor tyrosine kinases (RTKs) contains four member proteins: EGFR/ErbB1, HER2/ErbB2/neu, HER3/ErbB3, and HER4/ErbB4. These RTKs carry out important signaling functions via the sequential process of ligand binding by the extra- cellular domain, homo- or heterodimerization, activation of their intracellular kinase domain, and recruitment of down- stream signaling proteins. These RTKs are also important oncogenic drivers in many breast, lung, and other human cancers (78). Several structural biology studies on the ErbB family have been published, and this has helped advance drug development for HER2-positive breast cancer (79). Protein crystallography studies published in 2004 showed the structure of pertuzumab bound to the extracellular domain of HER2 and lapatinib bound to the kinase domain of EGFR (80, 81). Since that time, growing structural biology-based understanding of how EGFR, HER2, and HER3 function at the atomic level has dramatically reshaped our understanding of RTKs (78, 79). Despite these advances, there is a domain in each EGFR/ErbB family protein for which little structural biology information is available; this domain is the C-terminal tail (CTT) domain. The CTTs contain numerous autophosphorylation sites that are essential for recruiting downstream signaling proteins and initiating intracellular signaling (43, 82). The CTT can also contribute to autoinhibition of the kinase domain of RTK (83). The lack of available crystallographic information on the CTT region of EGFR/ErbB family RTKs led us to examine whether these proteins lack a stable secondary and/or tertiary structure.

Intrinsically disordered regions (IDRs) represent an emerging area of interest in medicine. IDRs are regions within proteins that exhibit high flexibility and may lack a secondary or tertiary structure but are still able to carry out important biological functions (48-55).

Algorithm prediction methods indicate that around 25–30% of eukaryotic proteins can be categorized as having disordered regions (56). These disordered regions provide certain advantages in protein-protein interactions, including a larger hydrodynamic radius (84, 85), faster on- and off-rates of binding (86), high binding specificity (87), and the ability to adopt different conformations depending on the binding partner (49,88). Correlation studies have revealed a high propensity for disordered regions to undergo post-translational modification, particularly phosphorylation (89). Because traditional methods, such as NMR and X-ray crystallography, were developed to study stable protein structures, IDRs have been more difficult to analyze because of difficulties in obtaining high concentrations or representative protein crystals (51, 90). Correlation studies suggest a strong association between IDRs and human cancer-associated proteins (57); therefore, identifying and analyzing IDRs in cancer-related proteins is vital in understanding how they function.

Protein kinases demonstrate a high degree of specificity in facilitating phosphorylation; however, many can perform such interactions with multiple substrate partners (91). An analysis of the human kinome shows that as many as 83% of kinase genes contain IDRs, which could facilitate these multiple interactions. RTKs are involved in more protein-protein interactions than any other kinase group, thus potentially pointing to the involvement of IDRs in their recognition mechanisms (58). Multisequence alignment (92) of EGFR/ErbB family RTKs shows that the kinase domain sequences are highly conserved between all four family members. However, the CTT regions are highly divergent between each EGFR/ErbB family member. One previous study using circular dichroism (CD) spectroscopy indicated that the EGFR CTT is rich in α-helical and β-sheet content (43), but a later study used coarse grained modeling to show many possible

conformations of the EGFR CTT based on the assumption that it is naturally disordered (44). A high degree of flexibility in the CTT would provide distinct advantages to EGFR interaction with downstream Src homology 2 (SH2) and phosphotyrosine-binding domains during signaling functions (43).

In this work, we examine the biophysical properties of the CTT of EGFR and HER3. The motivation of this study is to gather structural information on the CTT region to determine whether the tails are highly dynamic, disordered regions. We expressed and purified EGFR and HER3 CTT-only constructs and demonstrated that they are functional because they are recognized and phosphorylated by EGFR family kinases and once phosphorylated can be bound by the Grb2 SH2 domain. Using these CTT constructs, we performed multiple biophysical analyses, including CD spectroscopy, size exclusion chromatography with multiangle light scattering (SEC-MALS), dynamic light scattering (DLS), analytical ultracentrifugation (AUC), and small angle X-ray scattering (SAXS). The results of these methods support the hypothesis that the EGFR and HER3 CTTs are IDRs with extended, non-globular structure in solution.

2.4 METHODS

Plasmids

DNA encoding residues 696–1022 of the human epidermal growth factor receptor was cloned into pFAST BAC HT (Invitrogen) using the NcoI and HindIII restriction sites. This EGFR wildtype kinase domain construct contains an N-terminal 6-His tag, a linker, and a TEV protease cleavage site (MSYYHHHHHHDY DIPTTENLYFQGAM). For the EGFR CTT cloning, DNA encoding residues 961–1186 of the human EGFR was cloned into a pET30b vector (Addgene) using NdeI and XhoI restriction sites in-frame with a C-terminal His tag in the vector. Similarly, DNA encoding residues 981–1342 of the human epidermal growth factor receptor 3 (HER3) was cloned into a pET30b vector (Addgene) using NdeI and XhoI restriction sites in-frame with a C-terminal His tag in the vector. pGEX Grb2- SH2 (amino acids 58 –159) was purchased from Addgene (plasmid 46440, deposited by Dr. Bruce Mayer). All plasmids were confirmed by DNA sequencing.

Expression and purification of EGFR kinase domain:

The protein expression procedure was adapted from protocols described previously (4, 36). Recombinant bacmid (Bac-to-Bac expression system, Gibco BRL) were transfected into *Spodoptera frugiperda* Sf9 cells to produce recombinant baculovirus, which were used to infect Sf9 cells grown at 27 °C and 120 rpm in IPL-41 medium (Invitrogen) supplemented with 10% bovine calf serum supplemented (Hyclone), yeastolate extract (Invitrogen), penicillin, streptomycin, and Pluronic F-68. Cultures at a density of 1 x 10^6 cells/ml were infected at 1 MOI and cells were harvested 48 – 72hours post infection by centrifugation at 6000 rpm and resuspended in Buffer A (25 mM Tris-HCl, pH 8, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 1 mM DTT) supplemented with phenylmethylsulfonyl fluoride, benzamidine, DNase,

and protease inhibitor mixture tablets lacking EDTA (Roche). The cells were homogenized using a probe sonicator and the lysate was centrifuged at 20000 rpm for 45 mins. 0.5 - 0.75 ml of Ni-NTA beads (Qiagen) were added to the supernatants and incubated on a rotator for 1 h at 4 °C. The beads were poured into a 10-ml disposable column (Bio-Rad) and the flow-through collected. The column was washed with 10 column volumes of Buffer A + 20 mM imidazole, and the proteins were eluted with Buffer A + 50–125 mM imidazole. The eluted fractions were further purified by gel filtration chromatography on a Superdex 200 GL 10/300 column (GE Healthcare) using 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM DTT buffer. Pooled fractions were concentrated and treated with TEV protease (at a mass ratio of 1mg of TEV protease to 40mg of EGFR kinase domain) overnight at 4°C to remove the N-terminal His-tag. This was bound to Ni-NTA beads and the flow-through was concentrated, snap frozen, and then stored 80 °C.

Expression and purification of EGFR CTT and HER3 CTT:

The *E. coli* strain BL21(DE3) was used for expression of the EGFR CTT construct. Transformed cells were grown in Luria-Bertani medium containing 50 μg/ml kanamycin at 37 °C until an OD_{600} of between 0.5 and 0.8 was reached. Expression was induced through the addition of 0.5 mM isopropyl -D-1-thiogalactopyranoside (IPTG) for 3 h at 37 °C. Bacterial cells were centrifuged at 6000 rpm, and the pellets were frozen. Cell lysis was performed at 4 °C with a probe sonicator on frozen pellets using Buffer A (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1mM DTT) supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF), 1mM -mercaptoethanol, Roche Applied Science protease inhibitor mixture tablets, 4 mM benzamidine HCl. The lysate was clarified by centrifugation at 48,000 x *g* for 30 min at 4 °C. Purification was performed using Nickel-chelating chromatography followed by gel filtration

chromatography on a Superdex-75 column. The protein was concentrated to 5 mg/ml in 20 mM phosphate, 150 mM sodium chloride, pH 8.0. Aliquots were snap frozen in liquid N_2 and stored at -80 °C. Protein concentration was determined using UV-visible spectroscopy with a molar extinction coefficient of EGFR CTT at 280nm of $\in$ = 18,910 $M^{-1}cm^{-1}$. For specific subsequent applications, the addition of tris(2-carboxyethyl) phosphine to EGFR CTT samples was critical. A similar procedure was followed to express and purify the HER3 CTT but had trouble using BL21 DE3 cells to express the construct. We used BL21 DE3 Rosetta cells to express milligram amounts of the HER3 CTT. The use of the codon-optimized Rosetta cells was critical in successful expression of the HER3 CTT. This was purified at a concentration of 2 mg/ml in 20mM Tris, 300mM sodium chloride, 5% (w/v) glycerol, pH 8.0. HER3 CTT concentration was determined using a molar extinction coefficient at 280nm of $\in$= 25,245 $M^{-1}cm^{-1}$.

C-terminal Tail Function via Western Blotting Analysis:

EGFR CTT was phosphorylated via a reaction in 0.1 mM ATP, 0.1 mM sodium orthovanadate, 1 mM DTT and 1 mM $MnCl_2$ catalyst. EGFR CTT concentrations of 0.25 μM and 0.50 μM tail were incubated in the presence of 0.50 μM EGFR kinase domain. Two negative controls were also incubated under the same conditions with the following exceptions: 0.50 μM CTT was incubated without kinase domain present, and separately 0.50 μM kinase domain was incubated without CTT present. A positive control of ATP phosphorylation of 0.50 μM EGFR ICH construct, which contains both kinase domain and CTT regions together, was also incubated. Two EGFR tail region-specific phosphotyrosines were analyzed by Western blotting. Following the phosphorylation reaction, EGFR CTT samples were separated by SDS-PAGE on a 10% polyacrylamide gel followed by electroblotting onto nitro-cellulose membrane. Two anti-EGFR phosphotyrosine primary antibodies were used in separate Western blotting analyses: 1:1000

anti-EGFR Tyr(P)-1068 and 1:5000 anti-EGFR Tyr(P)- 1173. Secondary antibody was 1:5000 anti-rabbit in both blots. The results of the Western blotting analyses are shown in figure 2.2. The positive control was also analyzed on the same two Western blots as the EGFR CTT samples. Bands from this positive control were observed on both Western blots but are not shown herein. HER3 CTT was phosphorylated via the same buffer conditions as the EGFR CTT above. HER3 CTT at 0.50 μM was incubated in the presence of 0.50 μM EGFR or 0.50 μM HER2 kinase domains. Negative controls were incubated under the same conditions where only CTT or kinase domain was incubated in each case. Positive controls pairing HER3 ICH with EGFR kinase domain or HER2 kinase domain were also incubated. Samples were separated by SDS-PAGE and electroblotted onto nitrocellulose membrane. 1:1000 anti-HER3 Tyr(P)-1289 primary with 1:10,000 anti-rabbit secondary antibodies were used for the first blot. A 1:1000 general anti-phosphotyrosine anti- body, PY20, was used for the second blot with 1:10,000 anti-mouse secondary antibody. The two Western blots are shown in figure 2.2.

GST-Grb2-SH2 Expression and Purification:

BL21(DE3) cells were transformed, and a colony was picked and used to inoculate a 100 ml LB ampicillin culture. The culture was induced to mid log phase with 1 mM IPTG for 2 hours at 37 °C. Pellets were collected and subsequently frozen at -80 °C. GST-Grb2-SH2 was purified using glutathione-agarose beads (Sigma). Fractions were eluted with reduced glutathione, pooled, dialyzed, and concentrated using an Amicon 10,000 molecular weight cutoff spin concentrator in 20mM Tris-HCl, pH 8 (4 °C), 150mM NaCl, 1mM DTT. Aliquots were flash frozen and stored at -80 °C.

Interaction of Phosphorylated EGFR CTT and GST-Grb2- SH2 by Pulldown Assay:

EGFR CTT was phosphorylated by EGFR KD *in vitro* for 60 min at room temperature. 10 μg of

GST-Grb2-SH2 was prebound to a 10 μl bed volume of glutathione-agarose beads. 2 μg of either phosphorylated CTT or non-phosphorylated CTT was added to the beads in a volume of 100 μl of buffer and incubated for 2 hours at 4 °C. Beads were then spun briefly and the flow-through (FL) was collected. This was followed by two washes with buffer. Beads (bound) and aliquots of the FL were then boiled in SDS-PAGE sample buffer. Western blotting was done on the fractions using anti-His_6-HRP antibody (Thermo Fisher, MA1-21314-HRP) as a probe.

Circular Dichroism Spectroscopy:

EGFR CTT protein was prepared at a concentration of 5 μM in 10 mM sodium phosphate, pH 7.4, 5 mM NaCl. HER3 CTT was also prepared at the same conditions. A 0.1-cm path length quartz cell was used for samples and buffer blanks. Cells were placed in a Jasco CD spectrometer set at room temperature. The cell in the CD spectrometer was allowed to equilibrate for 15 min with nitrogen flowing into the CD spectrometer before scans. The CD spectrometer scanned from 260 to 190nm at 20 nm per min. Five individual scans were averaged, and the buffer blank was subtracted before analysis. The CD spectra were analyzed on the DichroWeb online analysis website (93, 94). The method chosen for this analysis used the CDSSTR algorithm with reference database 7 as the reference protein set (95-98). Secondary structure assignment was divided into six categories: regular α-helix (α_R), distorted α-helix (α_D), β_R,β_D, turns, and unordered. CDSSTR structure assignment results are presented in figure 2.3C.

Size Exclusion Chromatography with Multiangle Light Scattering:

100 μl of EGFR CTT or HER3 CTT (A_{280} ~ 1) was loaded onto a pre-equilibrated Superdex-75 10/300 GL (GE Healthcare) column. Elution buffer of 20 mM sodium phosphate, pH 8.0, 150mM NaCl, 0.5mM DTT was used for EGFR CTT, and 20mM Tris, 300mM sodium chloride, 5% (w/v) glycerol, pH 8.0 was used for HER3 CTT. The buffer flow rate was 0.5 ml/min for the

EGFR CTT and carbonic anhydrase elutions. For the HER3 CTT and EGFR KD elutions, the buffer flow rate was 0.3 ml/min. The column was coupled to a Wyatt Optilab rEX and Dawn Helios II, which are refractive index and multiangle light scatter detectors, respectively. Carbonic anhydrase was used as a globular reference for the EGFR CTT elution because of the similar molecular weights of the proteins. Similarly, the EGFR KD was used as a reference for the HER3 CTT elution. Astra (Wyatt) software was used to calculate the molar mass.

Dynamic Light Scattering:

EGFR and HER3 CTT samples at a concentration of 0.5 mg/ml in 20 mM Tris, 150 mM NaCl, 1 mM DTT, pH 8.0, were centrifuged for 5mins. 12 µl of sample was aliquoted into a clean microcuvette, which was placed into a Wyatt Dynapro MSX. The cuvette was equilibrated for 5 min at 25 °C before data acquisition. BSA was used as a standard at 1.0 mg/ml in the same buffer. Dynamics software was used to determine the radius of the protein.

Analytical Ultracentrifugation:

Samples were dialyzed *against* Buffer A (20 mM sodium phosphate, pH 8.0, 150 mM NaCl, 1 mM DTT) for EGFR CTT or Buffer B (20mM Tris, 300mM sodium chloride, 5% (w/v) glycerol, 1 mM DTT, pH 8.0) for HER3 CTT at 4 °C overnight. The samples and buffer were recovered and briefly centrifuged before loading into centrifuge cells. The cells were balanced and then placed in an eight-position rotor, which was put into a Beckman XLA ultracentrifuge. The speed was 42,000 rpm, and the temperature was 25 °C. The SEDFIT program was used to analyze the data, and the resulting *c*(*s*) analysis is shown in figures 2.4*E* and *F* (99).

Small Angle X-ray Scattering Analysis:

SAXS experiments were performed on the SIBYLS beamline 12.3.1.2 at the Advanced Light Source, a national user facility operated by the Lawrence Berkeley National Laboratory

(Berkeley, CA) and supported by the Director, Office of Science, Office of Basic Energy Sciences of the United States Department of Energy under Contract DE-AC02-05CH11231 (100). To optimize the data quality and minimize radiation damage, exposure series of 0.5, 1, 2, and 5 sec were performed. The concentration of EGFR CTT (20 mM phosphate, pH 8.0, 150 mM NaCl) used for these experiments ranged from 0.9 to 3.7 mg/ml. BSA in 20 mM phosphate, pH 8.0, 150 mM NaCl was used as a standard in the range of 1.0 –3.0 mg/ml. The concentration of HER3 CTT (20mM Tris-HCl, pH 8.0, 300mM NaCl, 5% glycerol, 1 mM DTT) ranged from 0.9 to 1.5 mg/ml. A minimum of three different concentrations for each protein were used for data collection, and five different concentrations were used for EGFR CTT. The data sets were processed using standard procedures for ATSAS programs (101). At low angles, the scattered intensities of EGFR CTT and HER3 CTT were very well approximated by the Guinier law, whereas HER3 CTT showed signs of radiation damage at 5 sec exposures. Before further analysis, the scattering curves, free of radiation damage, from a given type of sample with the same protein concentration but different exposure times were merged with either PRIMUS (101) or SCÅTTER (SIBYLS Beamline at Lawrence Berkeley National Laboratory) and then averaged among the same type of sample after normalization using their individual protein concentration. The average scattering curves are presented in figures 2.5*A* and 2.6*A*. The R_g from data sets free of radiation damage was calculated using AutoRg and is presented in Table 2.2. We evaluated the molecular weight of the sample by comparing the forward scattering $I(0)$ with that from a reference solution of BSA. The R_g and $D_{\max}$ values for EGFR kinase domain (Protein Data Bank code 1XKK) were calculated using the program CRYSOL (102). The pair distribution functions for EGFR CTT and BSA were generated using Datgnom (101).

2.5 RESULTS

Disorder Predictions on EGFR, HER2, and HER3

We used computation algorithms to provide an initial survey of disordered regions in the EGFR/ErbB family of RTKs. Predictor of natural disordered regions (PONDR) is a collection of algorithms that use an amino acid sequence to predict native disorder (103,104). The VL-XT algorithm assigns a disorder propensity score for disorder on a residue by residue basis (101-105). The VL-XT prediction for EGFR, HER2, and HER3 can be seen in figure 2.1. PONDR scores greater than 0.5 indicate predicted disorder, and scores less than 0.5 indicate predicted order. Kinase domains of the three EGFR/ErbB family members show mostly predicted order, whereas the CTT in each shows large regions of predicted disorder. The amino acid composition of the CTT and kinase domains of both EGFR and HER3 is shown in Table 2.1. When comparing the CTTs with the kinase domains in both proteins, we observe that the CTTs are more enriched in polar, uncharged residues and prolines but relatively depleted in hydrophobic residues. For EGFR, polar, uncharged residues make up 30.1% of the CTT *versus* 14.9% of the kinase domain. Prolines make up 10.6% of the EGFR CTT *versus* 5.4% in the kinase domain, and hydrophobic residues (excluding tyrosine, which can become phosphorylated) are 27.4% of the EGFR CTT *versus* 38.9% of the kinase domain (Table 2.1). Similar values are seen with HER3 CTT (Table 2.1) despite its primary sequence divergence from EGFR CTT (22% sequence identity between EGFR and HER3 CTT; calculated with Clustal Omega) (106). The lowered hydrophobicity provides a simple explanation as to why the tails would not form a hydrophobic core and therefore be disordered in an aqueous environment. These predictions are a useful tool for gaining a general view of disordered regions in a protein sequence, but direct empirical evidence would better support the hypothesis of the CTTs as being IDRs.

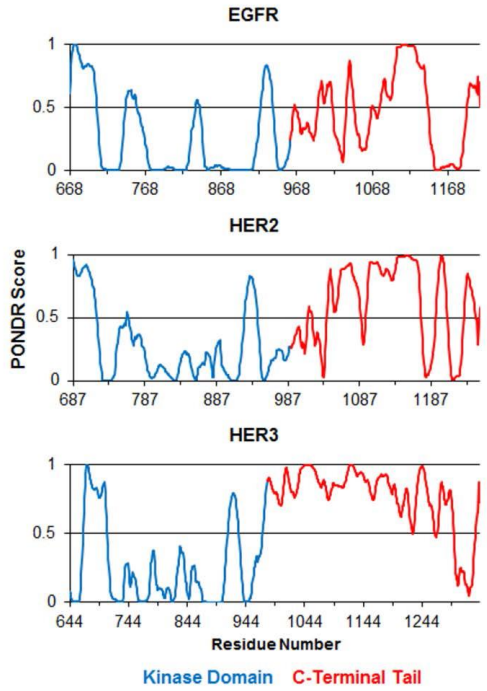


Figure 2.1. Disorder prediction by PONDR VL-XT algorithm. Shown are the disorder predictions for EGFR, HER2, and HER3 ICD construct sequences. In all three graphs, results for the kinase domain residues are colored blue, and the tail domain residues are colored in red. A score above 0.5 indicates predicted disorder while a score below 0.5 indicates predicted order.

		% of Total Sequence			
		C-terminal Tails		Kinase Domain	
	A.A. Residue	EGFR	HER3	EGFR	HER3
Polar Uncharged	S	11.9%	12.4%	4.7%	5.1%
	T	5.8%	5.2%	4.4%	3.5%
	N	6.2%	2.8%	2.2%	2.2%
	Q	6.2%	3.0%	3.5%	4.4%
	subtotal	30.1%	23.5%	14.9%	15.2%
Negative Charge	D	8.4%	5.0%	4.4%	4.4%
	E	5.3%	10.5%	7.9%	6.6%
	subtotal	13.7%	15.5%	12.3%	11.1%
Positive Charge	R	3.5%	6.9%	7.0%	7.0%
	H	2.7%	3.6%	2.5%	3.2%
	K	2.7%	1.7%	7.6%	6.3%
	subtotal	8.8%	12.2%	17.1%	16.5%
Hydrophobic (excluding Y)	A	6.6%	5.8%	5.7%	6.3%
	V	4.9%	3.9%	7.3%	7.9%
	I	3.5%	0.6%	7.9%	5.7%
	L	6.2%	9.7%	11.1%	12.0%
	M	1.3%	3.0%	2.8%	3.2%
	F	4.4%	0.8%	2.2%	2.5%
	W	0.4%	0.3%	1.9%	2.2%
	subtotal	27.4%	24.0%	38.9%	39.9%
Tyrosine	Y	4.0%	3.6%	3.5%	3.5%
Special	C	1.3%	1.7%	1.9%	1.3%
	G	4.0%	9.4%	6.0%	7.3%
	P	10.6%	10.2%	5.4%	5.4%

Table 2.1. EGFR and HER3 amino acid composition analysis by region

Expression of EGFR and HER3 C-terminal Tails

To further characterize the conformational and physical state of the EGFR and HER3 CTTs, we expressed and purified EGFR and HER3 CTT-only constructs in a bacterial *Escherichia coli* system. EGFR residues 961–1186 with a C-terminal His tag were purified using nickel-nitrilotriacetic acid column chromatography and gel filtration. Estimated purity by SDS-PAGE was 95% (figure 2.2.*A*) with a concentration of 5 mg/ml. The HER3 CTT, HER3 residues 981–1342, was also expressed in *E. coli* and purified to a concentration of 2 mg/ml with 95% purity in the same manner (figure 2.2.*B*).

Confirmation of EGFR C-terminal Tail Function

Phosphorylation of the CTT region and subsequent binding of SH2 domain-containing proteins are essential steps in EGFR signaling. We first validated the ability of our CTT constructs to be

both recognized and phosphorylated at specific tyrosine sites by separate EGFR family kinase domains in solution just as they would be if they were part of their respective ICH constructs. To test whether the *E. coli*-expressed EGFR CTT construct is functional, we incubated it with recombinant EGFR kinase domain and measured tail phosphorylation using reaction with ATP and blotting with phosphospecific antibodies. We show, through Western blotting identification of specific phosphorylated tyrosine residues (EGFR Tyr-1068 and Tyr-1173) in the EGFR CTT (figure 2.2.*C*), that the EGFR CTT is recognized and phosphorylated by the EGFR kinase domain even when these two domains are expressed as separated constructs. Further- more, we used anti-HER3 Tyr(P)-1289 and general PY20 anti- bodies to verify that the HER3 CTT was phosphorylated by both the EGFR kinase domain and HER2 kinase domain via Western blotting (figure 2.2.*D*). Next, we tested whether EGFR CTT can be bound by an SH2 domain-containing protein. We recombinantly expressed GST-tagged Grb2 SH2 domain and performed a GST pulldown assay (figure 2.2.*E*). Incubating EGFR CTT with EGFR kinase domain produced a phosphorylated CTT that appeared as multiple bands of higher molecular weight, and this phosphorylated EGFR CTT did not bind on its own to glutathione-agarose (figure 2.2*E*, flow-through in *lane 1* and bound in *lane 2*). Similarly, incubating unphosphorylated EGFR CTT with the GST-Grb2 SH2 domain construct did not result in binding, and EGFR CTT eluted in the flow-through (*lane 3*). However, once phosphorylated, EGFR CTT was bound by GST-Grb2-SH2 and was pulled down by glutathione-agarose (figure 2.2.*E*, *lane 6*). These experiments provide evidence for the functionality of the EGFR and HER3 CTT constructs.

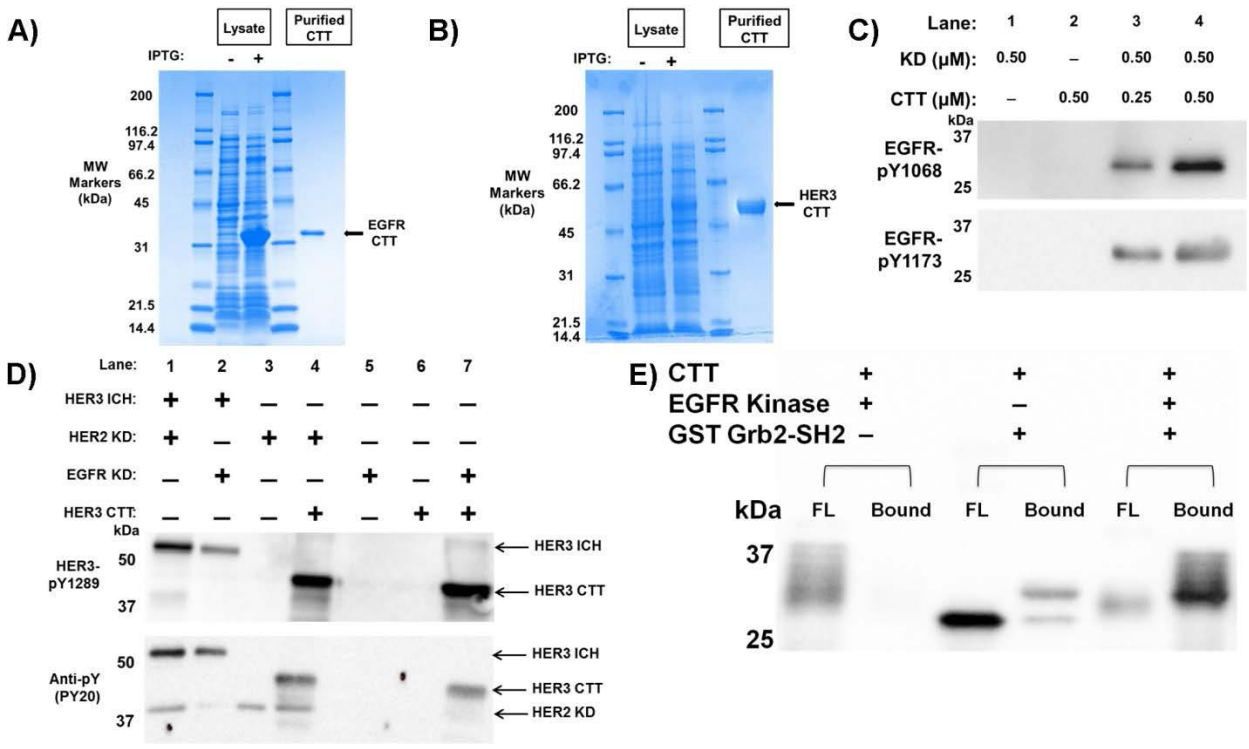


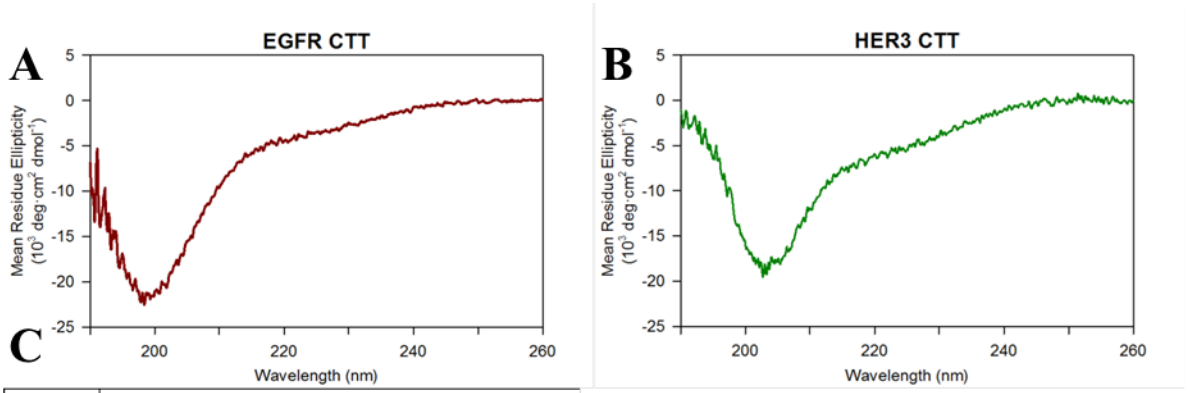
Figure 2.2. Purification and Phosphorylation of C-Terminal Tail (CTT) constructs. A, Coomassie stained SDS-PAGE gel of EGFR CTT purification. Lanes 1 and 4 are molecular weight markers, labeled respectively. Lane 2 is transformed cell lysate pre-induction with IPTG. Lane 3 is lysate post-induction. Lane 5 is purified EGFR CTT following Ni2+ chelating chromatography and gel filtration. EGFR CTT purity determined to be >95%. B, Coomassie stained SDS-PAGE gel of HER3 CTT purification. This gel uses similar lane arrangement as in figure 2.2.A for HER3 CTT expression. C, EGFR CTT is recognized and phosphorylated by the EGFR kinase domain. Phospho-specific antibodies to EGFR pY1068 and EGFR pY1173 were used to detect CTT phosphorylation. D, Western blot analysis of HER3 CTT phosphorylation. The HER3 ICH serves as a positive control showing C-terminal tail phosphorylation via interaction with EGFR KD or HER2 KD. HER3 CTT is recognized and phosphorylated by the EGFR and HER2 kinase domains, just as it would be as if it was part of its ICH construct. Phospho-specific antibodies to HER3 pY1289 and general anti-phosphotyrosine pY20 were used

to detect HER3 CTT phosphorylation. E, Phosphorylated EGFR CTT binds to the Grb2-SH2 domain. GSH-agarose beads were pre-loaded with GST tagged Grb2-SH2 domain protein. Phosphorylated or unphosphorylated EGFR CTT was then incubated with the beads for 2 hours at 4 °C, flow-through (FL) collected, washed, then the beads were boiled in sample buffer (Bound). SDS-PAGE of fractions then transferred to Nitrocellulose. Anti His6-HRP Ab was used to detect EGFR CTT in the FL and bound fractions.

Analysis of EGFR CTT by CD Spectroscopy

Far-UV CD spectroscopy provides information about the secondary structure characteristics of protein molecules in solution based on the absorption of circularly polarized light by amide bonds in the polypeptide backbone. Because secondary structure types absorb circularly polarized light differently and each amide bond contributes to the UV absorption, the resulting spectrum will reflect a global average of the secondary structure content in a protein. Due to a lack of -helix and -sheet content, mostly or fully disordered and unfolded proteins will have CD spectra that are distinct from highly ordered proteins. The CD spectra between 190 and 260 nm wavelengths for EGFR CTT and HER3 CTT are shown in figure 2.3, *A* and *B*, respectively. Distinctive features expected for a mostly -strand protein would include a positive ellipticity maximum at 195nm with a negative ellipticity minimum near 215 nm. For a mostly -helical protein, one would expect two minima, one at 208 nm and the other at 224 nm wavelength, along with a pronounced maximum near 192nm (55, 93). However, the CD spectrum observed for EGFR CTT shows a minimum between 195 and 200 nm, which is a characteristic of disordered proteins (figure 2.3.*A*). However, the slightly negative ellipticity at 222nm indicates the possibility of residual secondary structure. Analysis via the CDSSTR algorithm, an algorithm

used to assign secondary structure composition to CD spectra, shows 77% unordered character in the EGFR CTT spectrum (figure 2.3.*C*). These results suggest that much of the EGFR CTT construct has an unfolded conformation in solution. The CD spectrum for HER3 CTT shows similar features as the EGFR CTT with a minimum between 195 and 200 nm and slightly negative ellipticity at 222 nm (figure 2.3.*B*). The HER3 CTT has mostly unordered content (60%) but slightly higher β-sheet content than EGFR CTT: regular -strand (α_R) 15% and distorted -strand (α_D) 7% for HER3 CTT as compared with β_R 6% and β_D 3% for EGFR CTT (figure 2.3.*C*).



C

	Calculated Secondary Structure Fractions						
Protein	α_R	α_D	β_R	β_D	T	**U**	**Total**
EGFR	0.01	0.04	0.06	0.03	0.08	**0.77**	**0.99**
HER3	0.01	0.04	0.15	0.07	0.13	**0.60**	**1.00**

Figure 2.3. Circular dichroism spectroscopy. A, The CD spectrum for the EGFR CTT construct does not show prominent spectral features of α-helices and β-sheets. B, Similarly, the HER3 CTT CD spectrum shows high unordered content, with slightly higher calculated β-sheet content than in the EGFR CTT spectrum. C, Secondary structure assignment was assigned the following symbols: regular α-helix (αR), distorted α-helix (αD), regular β-strand (βR), distorted β-strand (βD), turns (T), and unordered (U).

Size Exclusion Chromatography with Multiangle Light Scattering

Another property of intrinsically disordered regions is a higher than expected apparent molecular mass during gel filtration separations (59). SEC can be used to separate proteins based on their hydrodynamic size, which is increased in IDRs depending on the degree of conformational extension. Proteins with higher hydrodynamic size elute earlier than smaller proteins, and accurate, absolute measurement of molecular weight can be obtained by coupling SEC to a MALS detector (60). Figure 2.4.*A* shows the SEC elution chromatograms for both the EGFR CTT and carbonic anhydrase with the coupled MALS detection inserted above. Our analysis shows that EGFR CTT and carbonic anhydrase share similar molecular masses at 26 and 29kDa, respectively, and both are monomeric under these conditions. EGFR CTT elutes from the SEC column before 13 ml of buffer volume. Despite having a lower molecular weight, the EGFR CTT elutes earlier than carbonic anhydrase, which elutes just after 15 ml of buffer. The expected Stokes radius of carbonic anhydrase is 2.4 nm (107). This earlier elution indicates that the EGFR CTT conformational ensemble is more extended than that of carbonic anhydrase, which has a globular structure. Figure 2.4.*B* shows the SEC-MALS chromatogram profiles for HER3 CTT and EGFR kinase domain. The molecular masses, determined by MALS, of the HER CTT and EGFR kinase domain were 42 and 38 kDa, respectively. HER3 CTT eluted at about 14 ml of buffer, which is earlier than the EGFR kinase domain, which eluted between 16 and 17 ml of buffer. As with EGFR CTT, no multimeric peaks were observed in the HER3 CTT chromatogram.

Dynamic Light Scattering

The DLS technique is used to analyze hydrodynamic properties of proteins in solution. Using the hydrodynamic radius, or Stokes radius, derived from measurements of translational diffusion

coefficients, we are able to differentiate between a protein ensemble with mostly compact, globular conformations and one with more extended conformations as would be seen for IDRs (61). DLS is complementary to the static light scattering used in SEC-MALS in that DLS provides information on molecular size, whereas MALS provides a molecular weight and verification of a monomeric state. A histogram showing a distribution of measured hydrodynamic radii within the EGFR CTT and bovine serum albumin (BSA) populations is shown in figure 2.4.*C*. The EGFR CTT, which has a molecular mass of 26kDa, was determined to have a hydrodynamic radius of 5.3 nm. For a comparison, we also analyzed BSA, a globular protein with a molecular mass of 66 kDa, and the hydrodynamic radius of BSA was measured to be 4.8 nm. This demonstrates that EGFR CTT has a larger hydrodynamic radius than a globular protein more than 2.5 times its size. Similarly, HER3 CTT has a very large hydrodynamic radius (figure 2.4.*D*). The hydrodynamic radius (R_H) of HER3 CTT is 6.4 nm, although it has higher polydispersity than EGFR CTT.

Analytical Ultracentrifugation

We used sedimentation velocity AUC to determine the sedimentation coefficient, *s*, of EGFR CTT, which provides information about the molecular weight of the tail, M_f, as it relates to its R_H (figure 2.4.*E*). We also determined that our sample is homogenous due to the absence of secondary peaks in figure 2.4*C* as secondary populations or impurities would have different sedimentation velocities. From these AUC measurements, we obtained the frictional coefficient of the tail, *f*, which can be used to relate its molecular shape to that of a globular protein. The frictional ratio, f/f_{min}, provides an indication of the shape of a macromolecule based on the ratio of its measured frictional coefficient to a theoretical minimum frictional coefficient for a particle with minimum Stokes radius at a given molar mass (62). For a globular protein, the expected

f/f_{min}wouldfallbetween1.15and1.3 (108-110). We determined the f/f_{min} value for EGFR CTT to be much greater, at 1.77, which is consistent with this being an IDR. Similarly, HER3 CTT has an f/f_{min} value of 1.52, which is also greater than the expected frictional ratio for globular proteins (figure 2.4*F*).

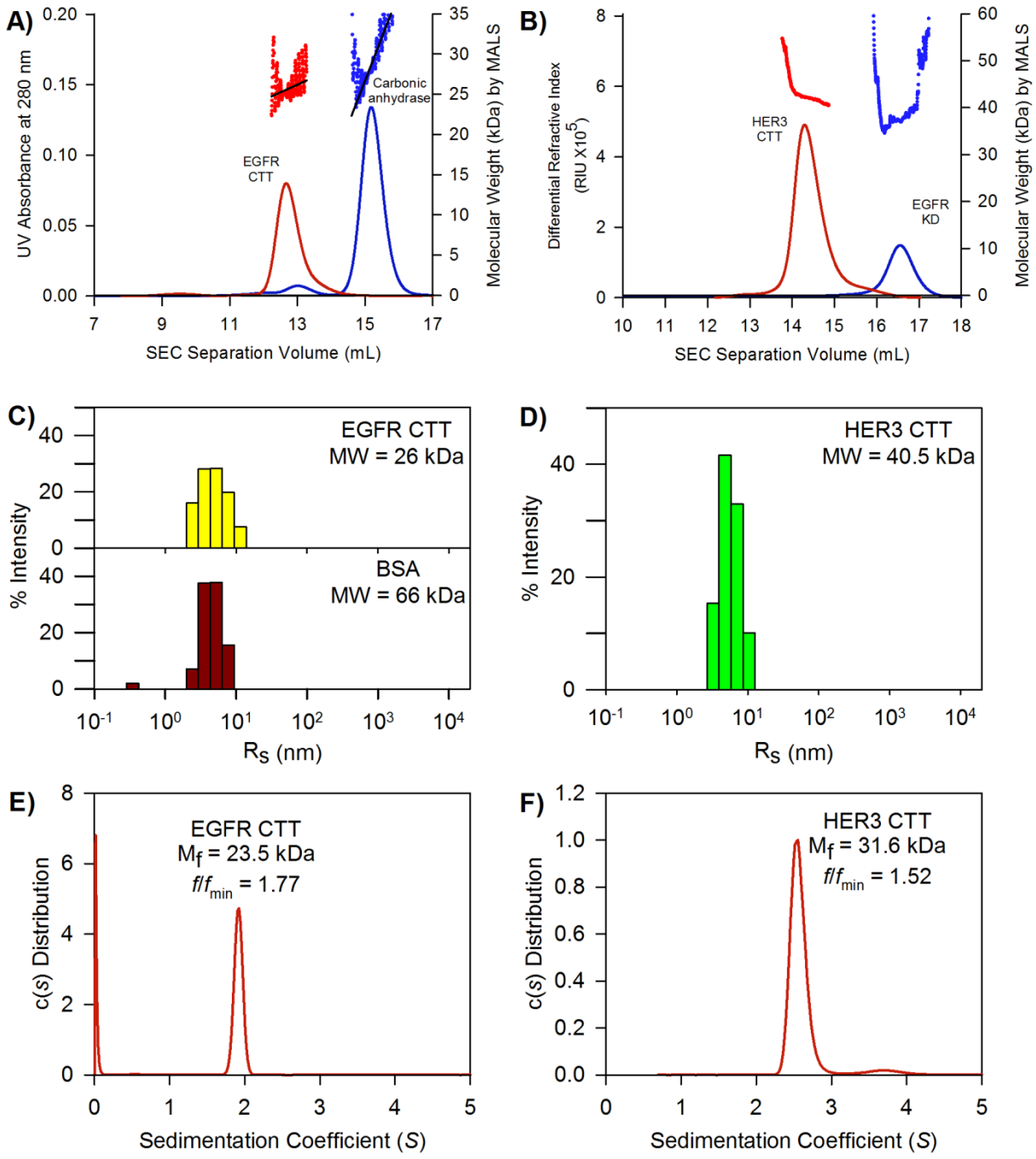


Figure 2.4 Hydrodynamic Properties of EGFR CTT. A, Elution chromatograms and molecule weight determination of EGFR CTT by SEC-MALS. EGFR CTT data is shown in red, carbonic anhydrase data is shown in blue. Despite similar molecular weights, EGFR CTT elutes sooner, indicating a larger hydrodynamic size. B, SEC-MALS analysis of HER3 CTT. HER3 CTT data

is shown in red, EGFR kinase domain data is shown in blue. Again, the two molecules share similar molecular weights, but HER3 CTT elutes sooner than the EGFR KD. C, Hydrodynamic radius determination of the EGFR CTT and BSA by dynamic light scattering. The measured DLS hydrodynamic radius distribution for EGFR CTT is shown as yellow bars. The measured hydrodynamic radius of EGFR CTT is 5.3 nm. The histogram for BSA (monomeric fraction obtained by gel-filtration) is shown as red bars. The RH of monomeric BSA is 4.8 nm. D, Hydrodynamic radius determination of the HER3 CTT. The measured DLS hydrodynamic radius for HER3 CTT is 6.4 nm. E, Frictional coefficient, frictional ratio, and molecular weight determined by analytical ultracentrifugation. Sedimentation coefficient is used to measure an apparent molecular weight of Mf = 23.5 kDa. The frictional ratio for the EGFR CTT is 1.77, indicating a large Stokes radius relative to globular proteins. F, Frictional coefficient, frictional ratio, and molecular weight determined by analytical ultracentrifugation. The measured apparent molecular weight of HER3CTT is 31.6 kDa. The frictional ratio is 1.52.

Small Angle X-ray Scattering

Small angle X-ray scattering methods have been particularly useful in analyzing conformation ensemble properties of IDRs (63). In small angle X-ray scattering analyses, we measured the isotropic scattering intensity, $I(s)$, as a function of the momentum transfer, s, of monochromatic X-ray light diffracted by protein macromolecules in solution. In protein structure studies, SAXS has been used to characterize structure and dynamics of monodisperse molecular species with a dominating structural configuration. IDRs show larger average sizes compared with globular proteins that contain tightly packed cores. By comparing the measured R_g and $D_{\max}$ from SAXS with other known structured globular proteins, together with Kratky plots ($I(s)s^2$ as a function of s), which characterizes the flexibility state of the proteins, one is able to identify the flexibility of the IDRs (111). Figure 2.5.*A* shows the scattering profiles for EGFR CTT and BSA. In figure 2.5.*B*, Guinier plot regions of EGFR CTT in three different concentrations are shown. R_g values calculated by AutoRg (101) are 46.6 ± 2.2 (EGFR CTT) and 33.6 ± 0.6 Å (BSA), and the molecular masses for each of them are 25.6 ± 0.7 and 69.0 ± 2.0 kDa, respectively (Table 2.2.). The Kratky plot (figure 2.5.*C*) of BSA shows a parabolic shape with a well-defined peak. In contrast, EGFR CTT shows a hyperbolic shape on the Kratky plot, indicating an intrinsically disordered region (figure 2.5.*C*). Figure 2.5*D* shows the pair distribution function ($P(r)$) of EGFR CTT generated by Datgnom (100). The $D_{\max}$ of EGFR CTT is 156 ± 6 Å, which is much larger than that of BSA (Table 2.2.). Thus, the R_g and $D_{\max}$ measurements, Kratky plot, and $P(r)$ distribution of EGFR CTT indicate that the conformational ensemble adopted by EGFR CTT is extended and flexible. When we analyzed the HER3 CTT with SAXS, we also included samples containing urea to determine whether a chemical denaturant would have an effect on the conformational ensemble. It is also important to note that, unlike in EGFR CTT buffer, 5%

glycerol was also included in the HER3 CTT buffer. The glycerol was necessary to maintain protein solubility and stability of the HER3 CTT, but it may also act to stabilize conformations with diminished extendedness. We do not directly compare derived R_g values between EGFR CTT and HER3 CTT because of this difference in buffer conditions. The scattering profiles for HER3 CTT without urea and with 4 M urea are shown in figure 2.6*A*, and Guinier plot regions are shown in figure 2.6*B*. From these data, we calculated an R_g of 45.4 ± 0.4 Å in the absence of urea, and that increases slightly to 50.5 ± 1.2 Å upon the addition of 4M urea (Table 2.2.). The derived Kratky profiles for these two conditions are shown in figure 2.6.*C*. Without urea present, the HER3 CTT produces a slightly parabolic Kratky profile, with a defined local maximum observed, at lower momentum transfer values. Moving toward greater momentum transfer values then gives a more hyperbolic increase in $I(s)s^2$. For HER3 CTT with 4 M urea, the Kratky profile has a hyperbolic shape indicative of disorder, similar to the profile observed for the EGFR CTT (figure 2.5.*C*). The difference in Kratky plots between HER3 CTT with or without urea, together with increased R_g, indicates that HER3 CTT possesses characteristics of IDRs but not of random coil structures when in solution.

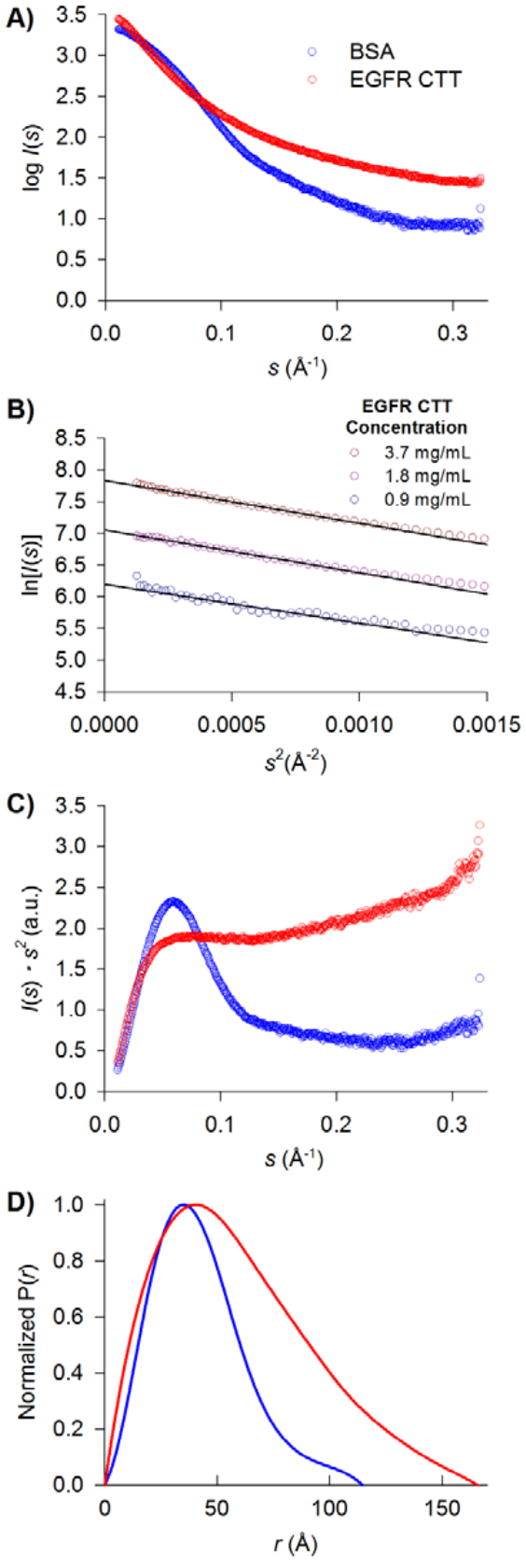

A)
BSA
EGFR CTT
log I(s)
s (Å-1)
B)
EGFR CTT
Concentration
3.7 mg/mL
1.8 mg/mL
0.9 mg/mL
ln[I(s)]
s2(Å-2)
C)
I(s) · s2 (a.u.)
s (Å-1)
D)
Normalized P(r)
r (Å)

Figure 2.5. Small angle X-ray scattering analysis of EGFR CTT. A, Scattering profiles of EGFR CTT (red) and BSA (blue). B, Guinier plots of EGFR CTT in different concentrations, 0.9, 1.8, and 3.7 mg/ml. C, The data is represented in a Kratky Plot with arbitrary units. In folded proteins like BSA (blue), Kratky plot shows parabolic features. While in the EGFR CTT Kratky plot (red), a hyperbolic shape indicates intrinsically disordered protein character. D, Pair distribution functions for the EGFR CTT (red) and BSA (blue) scattering profiles, generated using Datgnom (55).

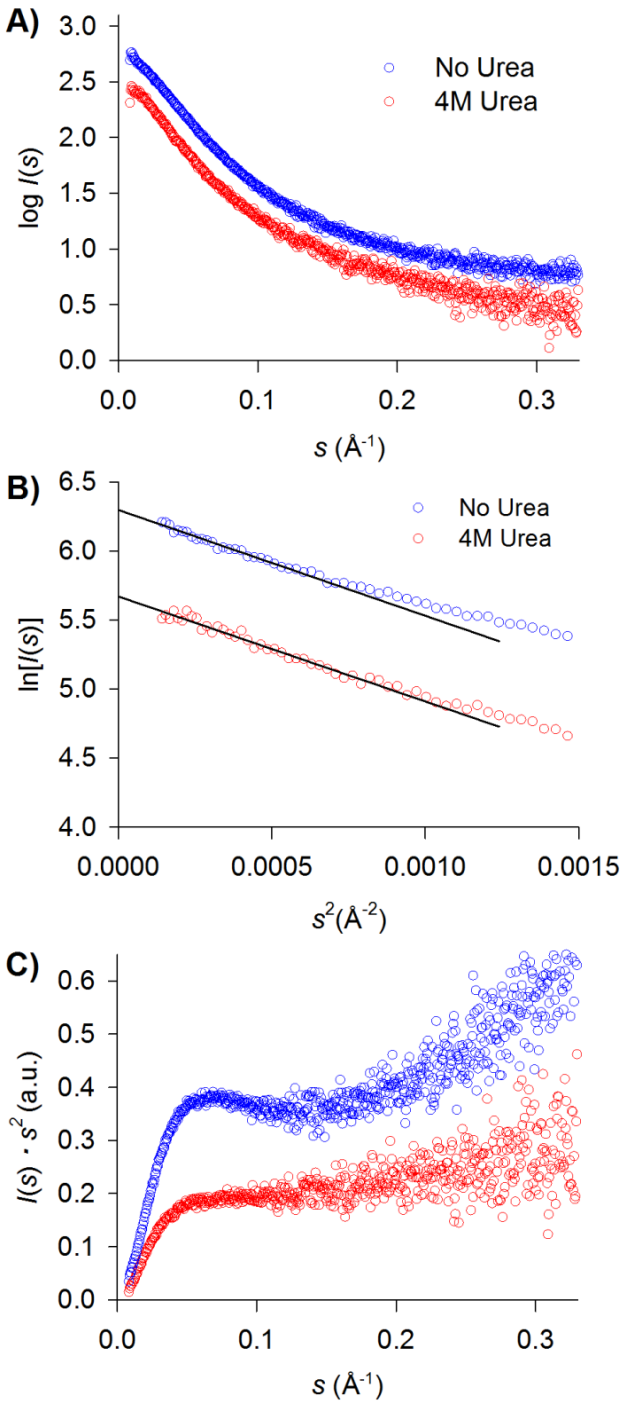


Figure 2.6 Small angle X-ray scattering analysis of HER3 CTT in solution ± urea. A, Scattering profiles of HER3 CTT in solution without urea (blue) and with 4M urea (red). B, Guinier plots of HER3 CTT with or without the presence of 4 M urea. C, The data is represented

in a Kratky Plot with arbitrary units. The Kratky Plot for HER3 CTT with 4M urea additive shows a more hyperbolic profile, indicating greater disorder with the addition of chemical denaturant.

Construct	Predicted Molecular weight[a]	Number of amino acids	Measured Molecular weight[b]	R_g	D_{max}
	kDa		*kDa*	Å	Å
EGFR CTT [c]	26.2	235	25.6 ± 0.7	46.6 ± 2.2	156 ± 6[f]
Her3 CTT [d]	40.5	371	41.5 ± 0.8	45.4 ± 0.4	153 ± 9[f]
Her3 CTT + Urea [d]	40.5	371	36 ± 4	50.5 ± 1.2	164 ± 10[f]
BSA	66.5	583	69 ± 2	33.6 ± 0.6	114 ± 5
EGFR kinase dom. (PDB:1XKK)[e]	41.1	352	41.1	20	64

Table 2.2. Structural parameters observed by SAXS.

[a] Values were calculated based on protein amino acid sequences.

[b] Measured MW was calculated from the X-ray scattering profile.

[c] Buffer for EGFR CTT was 20mM Phosphate pH 8.0, 150mM NaCl

[d] Buffer for Her3 CTT was 20mM Tris-HCl pH 8.0, 300mM NaCl, 5% glycerol, +/- 4M Urea as indicated above.

[e] Values of EGFR kinase domain were calculated from a theoretical scattering profile generated by the Crysol program using the crystal structure of EGFR kinase domain (PDB accession number 1XKK).

[f] Characteristics of intrinsically disordered proteins are shown in the Kratky plots for these constructs.

2.6 DISCUSSION

We have shown that the CTT domains in EGFR family proteins are intrinsically disordered regions that have an extended, highly dynamic conformational state. CD spectroscopy shows that they have high unordered character as expected for IDRs. SEC and DLS demonstrate that hydrodynamic size of the CTT is larger than protein standards of comparable molecular weight, which is another property of IDRs. AUC and the Kratky plot from SAXS demonstrate that the CTTs are IDRs, and R_g and D_{max} measurements from SAXS demonstrate that the CTTs have an extended conformation. The R_g and D_{max} measurements of the EGFR kinase domain were estimated from crystal structure (Protein Data Bank code 1XKK), and we observed that the EGFR CTT had R_g and D_{max} values that were 2.4 times larger than those of the EGFR kinase domain. A striking finding on the EGFR and HER3 CTTs is that despite their low primary sequence homology (22% sequence identity) the structural features of the CTTs are strongly conserved. This implies that the intrinsic disorder of these tails is important for their function, and the multiple conformers that a disordered region can adopt could reduce the thermodynamic barriers for protein binding and thereby increase both the on- and off-rates for binding of SH2 domain-containing proteins. Furthermore, the extended conformation of the CTT could increase the capture radius for recruitment of these signaling proteins (112, 113). Both mechanisms would facilitate signal propagation by these receptor tyrosine kinases. The intrinsically disordered properties of these CTTs also explain why these regions are mostly absent from crystallographic studies on EGFR (47, 82). In conclusion, these methods demonstrate that the EGFR and HER3 CTTs are IDRs with extended, non-globular structure in solution, and this finding may have important implications for the recruitment of downstream signaling proteins and signal propagation from these RTKs.

Chapter 3: Efficient sortase-mediated N-terminal labeling of TEV protease cleaved recombinant proteins

This chapter was adapted from:

Sarpong K, Bose R. (2017) Efficient Sortase-mediated N-terminal Labeling of TEV Protease Cleaved Recombinant Proteins. ***Anal. Biochem.*** **15, 55 – 58.**

3.1 ACKNOWLEDGEMENTS

We thank Dr. Hidde L. Ploegh for providing the SrtA 5M plasmid and Drs. Thomas E. Kraft and Paul W. Hruz for providing purified MSP1E3D1 protein. We also want to thank Robert Obermann and Dr. James W. Janetka for assistance with peptide synthesis.

3.2 ABSTRACT

A major challenge in attaching fluorophores or other handles to proteins is the availability of a site-specific labeling strategy that provides stoichiometric modification without compromising protein integrity. We developed a simple approach that combines TEV protease cleavage, sortase modification and affinity purification to N-terminally label proteins. To achieve stoichiometrically-labeled protein, we included a short affinity tag in the fluorophore-containing peptide for post-labeling purification of the modified protein. This strategy can be easily applied to any recombinant protein with a TEV site and we demonstrate this on Epidermal Growth Factor Receptor (EGFR) and Membrane Scaffold Protein (MSP) constructs.

3.3 INTRODUCTION

Site-specific protein labeling is the method of choice for most biochemical and biophysical applications, as this offers a high level of precision for the attachment of a fluorophore or other chemical moiety (69, 70). Due to the relatively low abundance of cysteines in proteins (71), chemical labeling of proteins using maleimide chemistry is a common strategy for most applications. However, many proteins contain multiple cysteine residues and mutagenesis of these cysteines is time-consuming and may compromise protein function. An alternate approach is to label primary amines with N-hydroxysuccinimide ester-based fluorophores. However, the relatively high abundance of lysines and pKa requirements renders the utility of amino groups for protein modification a less commonly used strategy. These challenges are compounded by long reaction times to ensure complete modification of the protein. With fluorophores having a MW of <1kDa, separation of labeled products from the unlabeled protein can also present challenges. Sub-stoichiometric labeling often results in a diminished signal-to-noise ratio and impacts the utility of fluorophore-labeled proteins for biophysical studies.

Enzymatic approaches for site-specific incorporation of fluorophores are an alternative to these chemical labeling strategies. Sortases are membrane-associated transpeptidases that anchor Gram-positive bacterial surface proteins to their cell walls. Since the discovery of sortases, *Staphylococcus aureus* sortase A (SrtA) has been the prototype for understanding the mechanism of action of these enzymes (73). Proteins anchored to the bacterial cell wall by SrtA possess a C-terminal sorting signal that contains a hydrophobic domain sandwiched between the conserved LPXTG recognition motif and a positively charged tail (73). SrtA catalyzes the hydrolysis of the peptide bond between the threonine and glycine residues to generate an acyl-enzyme

intermediate that is subsequently attacked by an oligoglycine peptide in a nucleophilic attack (74). This results in the formation of a new peptide bond between the incoming nucleophilic glycine-containing peptide and the protein. Seminal work (by Schneewind and coworkers) that laid the ground for its utility in biochemical and biotechnological studies showed a recombinant peptide containing the LPXTG motif alone is sufficient for recognition and catalysis (75). These studies also indicated that a peptide containing 1-3 N-terminal glycines could replace the peptidoglycan involved in the sortase-mediated reaction (75). Current biochemical evidence has suggested that only one additional residue (preferably a glycine) is required at the C-terminus of the LPXTG recognition sequence for efficient sortase binding and catalysis (76).

Most recombinant proteins used for biochemical and biotechnological applications contain affinity tags that ensure their easy and efficient purification (114, 115). Sandwiched between the affinity tags and the proteins are protease recognition sites that offer the cleavage of the affinity tags following purification. Commonly used recognition sites include TEV, Factor Xa, and Thrombin protease cleavage sites. An important requirement for the sortase reaction is the generation of the N-terminal glycine residue which can be done by removing the initial methionine of an expressed protein using methionylaminopeptidase or engineering a thrombin or TEV protease recognition site that exposes an N-terminal glycine following cleavage (116).

Recent years have seen the development and utility of sortase to modify proteins at their carboxyl and amino termini in addition to internal loops (76, 77). Unlike traditional chemical strategies that are easy to use, protein modification employing short genetically encoded tags such as the LPXTG-tag, ACP-tag and LAP-tag offer a high degree of precision. However, back reaction from the final product (containing the LPXTG motif and therefore an efficient substrate for the sortase enzyme) and the reversible nature of the sortase reaction can lead to sub-

stoichiometric protein modification and decreased labeling efficiencies. To address this challenge, the equilibrium of the reaction is driven towards product formation by increasing the fold excess of the fluorophore-containing peptide (76, 77). Recent methodologies to address this issue of irreversibility have included the use of a sortase-tagged expressed protein ligation (STEPL) system that circumvents the removal of unconjugated species (117), dialysis to remove reaction by-products (118) and the introduction of tryptophan-derived zippers around the SrtA recognition motif that induces the formation of a stable β-hairpin (119). Other research groups have solved this problem by utilizing a depsipeptide, which replaces the amide bond between the threonine and glycine residues with an ester linkage (120). These challenges make protein modification using sortase cumbersome and potentially expensive when the fluorophore-containing peptide is needed in many fold excess. The presence of reaction by-products as a result of back reaction and the reversible nature of the reaction affect the purity and degree of labeling of the final product, and subsequently present challenges in the utilization of fluorophore-labeled proteins.

We have developed a simple approach that combines TEV protease cleavage, sortase modification and affinity purification to N-terminally label proteins. To achieve stoichiometrically-labeled product, a short affinity tag is included in the fluorophore-containing peptide so that post-labeling affinity purification of only the labeled protein can be performed.

3.4 METHODS

Plasmids

Staphylococcus aureus sortase A plasmid was generously provided by the Ploegh lab (MIT). This sortase A pentamutant (SrtA 5M) is Ca^{2+} -dependent and had previously undergone directed evolution to be catalytically more efficient (121). SrtA 5M contains five mutations and has a 140-fold increase in transpeptidase activity over wildtype SrtA. We engineered SrtA 5M to have a TEV cleavage site before the C-terminal histidine tag, to facilitate our downstream purification strategy. Sequences of all the plasmids were confirmed by DNA sequencing.

Expression, purification and TEV protease cleavage of EGFR kinase domain constructs:

We expressed and purified the EGFR kinase domain using a protein expression protocol described previously (4, 36). Recombinant bacmid (Bac-to-Bac expression system, Gibco BRL) was transfected into *Spodoptera frugiperda* Sf9 cells to produce recombinant baculovirus, which were used to infect Sf9 cells grown at 27 °C and 120 rpm in IPL-41 medium (Invitrogen) supplemented with 10% bovine calf serum supplemented (Hyclone), yeastolate extract (Invitrogen), penicillin, streptomycin, and Pluronic F-68. Cultures at a density of 1 x 10^6 cells/ml were infected at 1 MOI and cells were harvested 48 – 72hours post infection by centrifugation at 6000 rpm and resuspended in Buffer A (25 mM Tris-HCl, pH 8, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 1 mM DTT) supplemented with phenylmethylsulfonyl fluoride, benzamidine, DNase, and protease inhibitor mixture tablets lacking EDTA (Roche). The cells were homogenized using a probe sonicator and the lysate was centrifuged at 20000 rpm for 45 mins. 0.5 - 0.75 ml of Ni-NTA beads (Qiagen) were added to the supernatants and incubated on a rotator for 1 h at 4 °C. The beads were poured into a 10-ml disposable column (Bio-Rad) and the flow-through collected. The column was washed with 10 column volumes of Buffer A + 20 mM

imidazole, and the proteins were eluted with Buffer A + 50–125 mM imidazole. The eluted fractions were further purified by gel filtration chromatography on a Superdex 200 GL 10/300 column (GE Healthcare) using 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM DTT buffer. Pooled fractions were concentrated and treated with TEV protease (at a mass ratio of 1mg of TEV protease to 40mg of EGFR kinase domain) overnight at 4°C to remove the N-terminal His-tag and generate an EGFR kinase domain with an N-terminal glycine, Gly-EGFR. This was bound to Ni-NTA beads and the flow-through was concentrated, snap frozen, and then stored 80 °C. We also expressed and purified two other EGFR kinase domain constructs, whose TEV cleavage resulted in two and four N-terminal glycines, Gly_2-EGFR and Gly_4-EGFR respectively.

Expression, purification and TEV protease cleavage of Sortase A protein constructs

The protocol we used for expressing and purifying SrtA 5M (with and without the TEV cleavage site) had been previously described (76). Briefly, BL21 (DE3) competent *E. coli* cells were transformed with pET30b *S. aureus* sortase A 5M or pET30b *S. aureus* sortase A 5M (containing an insertion for TEV protease cleavage) onto LB-kanamycin plates. An individual colony was picked and inoculated into 100 mL LB-ampicillin and placed in a shaker at 220 rpm at 37 °C overnight. 10 mL of the overnight culture was diluted into 1 L LB media with the corresponding kanamycin. This was grown at 37°C until OD_{600} of 0.4–0.6 (~3 h) was achieved. Expression was induced through the addition of 0.5 mM isopropyl -D-1-thiogalactopyranoside (IPTG) for 16 h at 25 °C. The bacterial cells were centrifuged at 6,000 x g for 15 min at 4 °C. The pellet was resuspended in 50 ml Nickel-binding buffer (50 mM Tris-HCl, pH 7.5 and 150 Mm NaCl), transferred to a 50 ml centrifuge tube, centrifuged again and the pellet frozen at −80 °C. Cell lysis was performed at 4 °C with a probe sonicator on frozen pellets using 25 ml ice cold lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 10 mM imidazole, 10% (vol/vol)

glycerol, 1 mg ml−1 DNase and 1 mg ml−1 lysozyme). The lysate was clarified by centrifugation at 20,000 x *g* for 30 min at 4 °C. A Poly-prep column was packed with 1–1.5 mL (bed volume) nickel-nitrilotriacetic acid (Ni-NTA) agarose resin and washed with 10 column volumes of wash buffer (by gravity flow) and the bacterial supernatant loaded onto the column (by gravity flow). The column was then washed with 50 column volumes of wash buffer to remove nonspecifically bound proteins and proteins eluted with 3 column volumes ice-cold elution buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 500 mM imidazole and 10% (vol/vol) glycerol). The protein was buffer exchanged and concentrated to 15 mg/ml in Nickel-binding buffer (no imidazole) + 10% imidazole. A fraction was aliquoted into tubes, snap frozen in liquid N_2 and stored at -80 °C. The SrtA 5M construct containing a TEV cleavage site was treated with TEV protease (at a mass ratio of 1mg of TEV protease to 40mg of SrtA) overnight at 4°C to remove the N-terminal His-tag. This was bound to Ni-NTA beads and the flow-through was concentrated, snap frozen, and then stored 80 °C. Protein concentration was determined using UV-visible spectroscopy with a molar extinction coefficient of SrtA 5M (with a TEV cleavage site) at 280nm of $\in$= 15,930.

Enzymatic activity of SrtA 5M constructs

The sortase activity of the wildtype SrtA 5M and SrtA 5M containing a TEV cleavage site was determined using an already published protocol (122). We used a self-fluorescent quenching peptide, *o*-aminobenzoyl-LPETG-2,4-dinitrophenyl as a substrate in the cleavage reaction containing 1.5 μM SrtA enzyme dissolved in assay buffer (20 mM HEPES, pH 7.5, 10mM $CaCl_2$). The *o*-aminobenzoyl-LPETG-2,4-dinitrophenyl substrate was dissolved in dimethyl sulfoxide and added to the reaction to a final concentration of 6.25 μM, for a total reaction

volume of 37.5 µl. The samples were incubated at 37 °C for 1 hr and analyzed in a spectrophotometer using 340 nm for excitation and recording the emission maximum at 420 nm.

Design and fluorophore-labeling of peptides

The LPETGG sortase substrate was made using solid-phase synthesis on a Wang resin. Following removal of the Fmoc on the ε-aminocaproic acid residue, 3 molar equivalents of 5(6)-TAMRA succinimidyl ester was added and the column wrapped in aluminum foil. This was shaken overnight at room temperature. The resin was cleaved in a solution consisting of 95 % TFA, H2O and TIS, and the final crude TAMRA-LPETGG was purified by reversed-phase HPLC on a C18 column using a 10 – 70 % Buffer B (H_2O, Acetonitrile and 0.05% TFA) gradient over 15 mins. The fractions were analyzed by LC/MS and the desired fractions were pooled together and lyophilized to dryness. We also designed a 25-mer peptide, MSYYHHHHHHDYDIPTCENLPETGG (or H_6-LPETGG peptide, hereafter) that was labeled with fluorescein (or CruzQuencher) at the N-terminus using the same protocol described for labeling with TAMRA. For labeling of the H_6-LPETGG peptide at the internal cysteine residue, the lyophilized peptide was reconstituted into water and the pH adjusted to around 8. Three molar equivalents of fluorescein (in DMSO) was added to 5mg of peptide in 50 mM Tris-HCl, pH 7.3 and 150 mM NaCl, and incubated at room temperature overnight. The labeled peptide was purified, analyzed and lyophilized using the method described above. Measurement of the absorbance of the peptide and fluorophore at 280 nm (A_{280} and A_{max}, respectively) and their molar extinction coefficients ($\varepsilon_{protein}$ and ε_{max}, respectively), allowed us to calculate the peptide concentration and degree of labeling using equation 1 below;

$$DOL = \frac{A_{max} \cdot \varepsilon_{peptide}}{(A_{280} - A_{max} \cdot CF_{280}) \cdot \varepsilon_{max}}$$

The correction factor (CF) is included in this equation to account for the absorption of fluorophore at 280 nm and equals the A_{280} of the dye divided by the A_{max} of the dye.

TAMRA-LPETGG labeling of Gly-EGFR, Gly_2-EGFR and Gly_4-EGFR

0.5 mM of TAMRA-LPETGG was mixed with 25 μM of either Gly-EGFR, Gly_2-EGFR or Gly_4-EGFR and 1 μM SrtA 5M in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$. The reaction was incubated at 4 °C with shaking and aliquots taken at 1 min, 5 mins, 15 mins, 30 mins and 1 hr. These, together with their unlabeled counterparts, were boiled in SDS-gel loading buffer for 2 mins and analyzed by SDS-PAGE.

Quenching of sortase activity by EDTA and EGTA

1.5 μM of SrtA 5M was added to a reaction containing 6.25 μM of *o*-aminobenzoyl-LPETG-2,4-dinitrophenyl in 20 mM HEPES, pH 7.5, 10mM $CaCl_2$, in the presence and absence of 10 and 20 μM of EDTA or EGTA. The samples were incubated at 37 °C for 1 hr and analyzed in a spectrophotometer using 340 nm for excitation and recording the emission maximum at 420 nm.

Sortase-mediated labeling and purification of EGFR kinase domain

In our newly developed strategy for labeling EGFR kinase domain with a fluorophore using sortase (Figure 3.1) , 0.5 mM of CruzQuencher-labeled H_6-LPETGG peptide was mixed with 25 μM of Gly-EGFR and 1 μM SrtA 5M in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$. The reaction was incubated at 4 °C for 30 mins and quenched with 10 mM EDTA. This was loaded onto a Superdex 75 10/30 GL column (GE Healthcare Life Sciences, Pittsburg, PA) and pooled fractions containing the EGFR kinase domain was incubated with 200 μl of Nickel NTA beads at 4 °C for 30 mins. CruzQuencher-labeled EGFR was eluted with 125 mM imidazole, buffer exchanged into Buffer A (25 mM Tris-HCl pH 7.5, 150 mM

NaCl, 1 mM DTT) and concentrated to ~ 1 mg/ml. Aliquots were taken at each step of the process and analyzed by SDS-PAGE. We determined the concentration and degree of labeling of the CruzQuencher-labeled EGFR using equation 1, where peptide is replaced with protein in the equation.

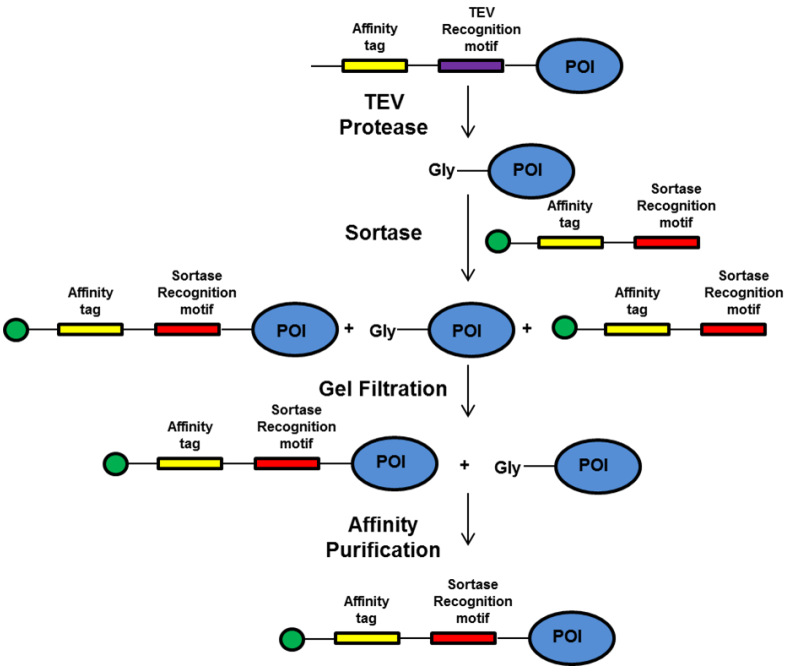


Figure 3.1. Schematic representation of N-terminal labeling of TEV-cleavable proteins using sortase. Green circle at the N-terminus of peptide indicates the fluorophore or any other desired moiety. Affinity tag used in here was 6x-His. However, this tag can be varied and the purification process can be made more stringent by using two different affinity tags. As noted in the text, the TEV recognition motif used must be ENLYFQG rather than ENLYFQS.

Sortase-mediated labeling and purification of MSP1E3D1 protein

MSP1E3D1 protein was treated with TEV protease to generate Gly- MSP1E3D1. 0.5 mM of fluorescein-labeled H_6-LPETGG peptide was mixed with 25 μM of Gly- MSP1E3D1 and 1 μM SrtA 5M in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$. The reaction was incubated at 4 °C for 30 mins and quenched with 10 mM EDTA. This was loaded onto a Superdex 75 10/30 GL column (GE Healthcare Life Sciences, Pittsburg, PA) and pooled fractions containing the MSP1E3D1 protein were incubated with 200 μl of Nickel NTA beads at 4 °C for 30 mins. Fluorescein-labeled MSP1E3D1 was eluted with 125 mM imidazole, buffer exchanged into Buffer A (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT) and concentrated to ~ 1 mg/ml. Aliquots were taken at each step of the process and analyzed by SDS-PAGE. We determined the concentration and degree of labeling of the fluorescein-labeled MSP1E3D1 using equation 1, where peptide is replaced with protein in the equation.

Kinase activity of labeled EGFR kinase domain

The protocol used to measure the kinase activities of labeled EGFR kinase domains, both in solution and on Nickel liposomes (to induce dimerization) have been described in Ref. 36. Wildtype or fluorophore-labeled EGFR kinase domain were incubated in 20 mM Tris, pH 7.5, 10 mM $MgCl_2$, 100 μM Na_3VO_4, 2 mM DTT, 100 μM ATP, 1 μCi of [γ- 32P] ATP, 60 mM NaCl, 5% glycerol in 25-μl total reaction volume. The reaction was initiated by the addition of 1.25 μM kinase, incubated at 30 °C for 6 mins and quenched with EDTA. 100 μg of avidin was added to the samples and transferred to centrifugal filtration units with 30,000 molecular weight cutoff (Millipore). The samples were washed four times with 0.5 M sodium phosphate, 0.5 M NaCl, pH 8.5 and the retained ^{32}P measured by scintillation counting. The kinase activity on liposomes was measured by incubating the wildtype and fluorophore-labeled EGFR kinase domains with nickel

liposomes on ice for 15 mins before initiating the reaction. Total lipid concentration was 0.5 mg/ml.

3.5 RESULTS

Expression and TEV protease cleavage of EGFR kinase domain constructs

The Sf9 insect cell baculovirus system was used to express all EGFR kinase domain (residues constructs. Our two-step purification strategy (Nickel NTA chromatography and Gel filtration) resulted in >95% purity and ~60% yield. Following overnight incubation with TEV protease, we performed a reverse Nickel NTA purification by binding the TEV protease and uncleaved EGFR kinase domain to the Nickel beads. The flow through containing the TEV protease-cleaved EGFR kinase domain was concentrated, resulting in >95% purity and ~30% yield (figure 3.2). Time-dependent incubation of EGFR kinase domain with TEV protease allowed the shortening of the incubation period from 16hours to less than 8 hours and still preserves the activity of the EGFR kinase.

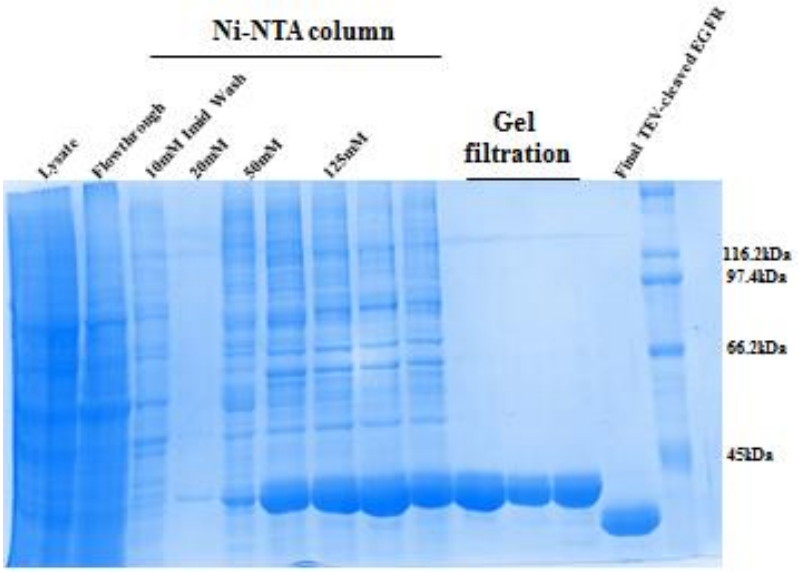


Figure 3.2. Coomassie-stained SDS-PAGE gel of EGFR kinase domain purification and TEV protease cleavage.

Expression and enzymatic activity of Sortase A 5M (with no His-tag)

Bacterial expression of Sortase A 5M (with and without His-tag) produced milligram amounts of the proteins, which were purified using a one-step Nickel NTA chromatography. The protein yield was 75% while SDS-PAGE analysis showed a >95% purity of the TEV protease-cleaved Sortase A 5M (figure 3.3). Enzymatic activity assays using *o*-aminobenzoyl-LPETG-2,4-dinitrophenyl as a substrate showed that Sortase A 5M (without a His-tag) had comparable sortase activity to the wildtype Sortase A 5M, as showed in Figure 3.4.

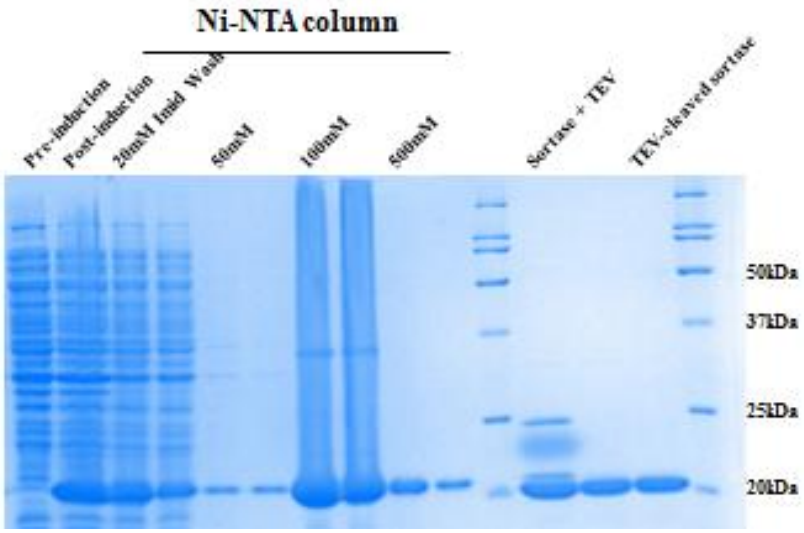


Figure 3.3. Coomassie-stained SDS-PAGE gel of Sortase A 5M (without a His-tag) purification and TEV protease cleavage.

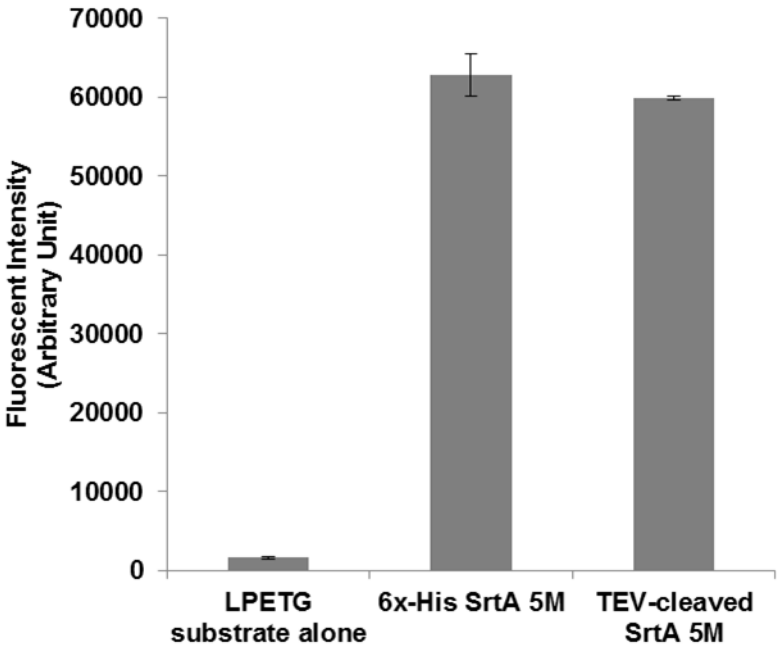


Figure 3.4. Sortase activity of 6x-His SrtA 5M and TEV-cleaved SrtA 5M

One N-terminal glycine is required for sortase-mediated labeling of EGFR KD

We tested whether additional N-terminal glycines (Gly_2-EGFR and Gly_4-EGFR) are better substrates in the sortase-mediated reaction. Our results using the TAMRA-labeled LPETGG peptide showed that the labeling of Gly-EGFR, Gly_2-EGFR, and Gly_4-EGFR were very similar (figure 3.5). The addition of several N-terminal glycine residues had no effect on the efficiency of the reaction and all future experiments utilized one N-terminal glycine for the sortase-mediated labeling reaction

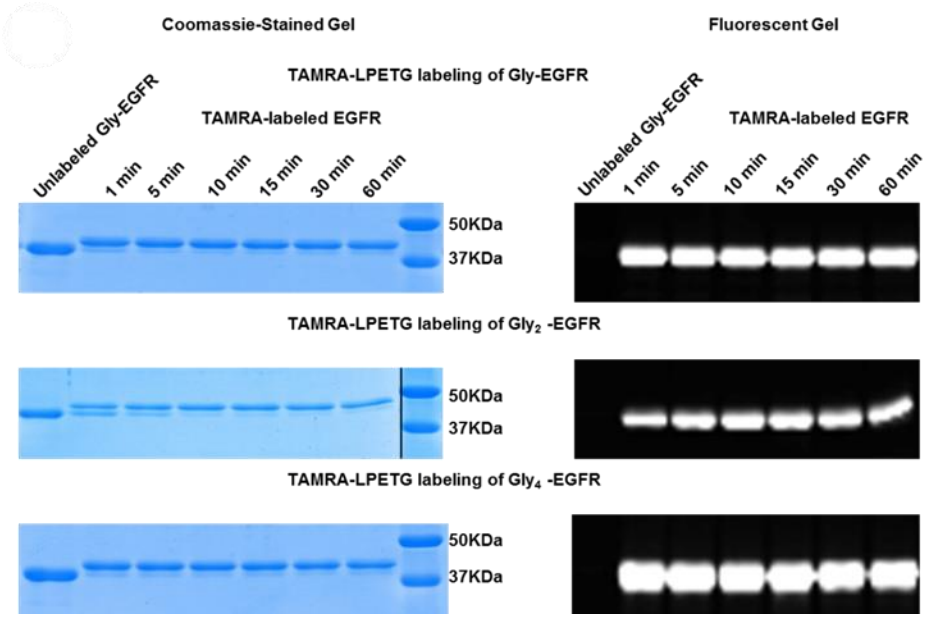


Figure 3.5. Comparison of labeling of Gly-EGFR, Gly2-EGFR and Gly4-EGFR using an LPETG peptide substrate containing the TAMRA fluorophore.

Inhibition of sortase enzymatic activity by metal chelators

We tested the chelating effect of EDTA and EGTA at different concentrations of the metal chelators. Our results showed that both chelators at 10 mM concentration were sufficient in quenching the activity of sortase.

Fluorophore-labeled EGFR KD has comparable kinase activity to wildtype

Comparison of the kinase activity of a fluorophore-labeled EGFR kinase domain (Alexa568-EGFR KD) shows similar fold of activation to the wildtype EGFR kinase domain (figure 3.6). We have consistently achieved a 10-12 fold of kinase activation of the fluorophore-labeled EGFR kinase domain on nickel liposomes, which is comparable to the activation of wildtype EGFR kinase domain.

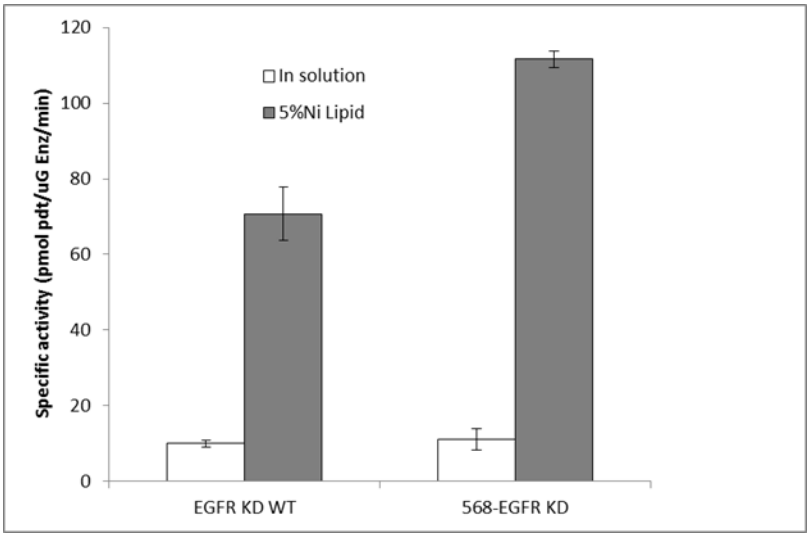


Figure 3.6. Effect of fluorophore label on EGFR kinase domain activity and activation on nickel liposomes

Affinity tagged peptide aids in post labeling purification of labeled EGFR KD and MSP protein

N-terminal labeling of Gly-EGFR using sortase and a short peptide (that contains the fluorophore and the sortase recognition motif) initially resulted in <60% modification of the protein with the labeled peptide. This was due to the presence of unlabeled EGFR kinase domain (which differs from the labeled EGFR kinase, in molecular weight terms by the presence of a fluorophore) and could not be easily removed by gel filtration. We hypothesized that the addition of an affinity tag to the LPETG sortase substrate would enhance the purification of only the labeled protein. We designed a peptide that had a 6× histidine tag (for purification of labeled EGFR from unlabeled EGFR), a cysteine residue for maleimide labeling (we also utilized the NH_2-amino group of the peptide for labeling with NHS-succinimidyl esters) and an LPETGG motif at the C-terminus. Labeling of Gly-EGFR with the H_6-LPETGG peptide results in the regeneration of the sequence N-terminal to the EGFR that was cleaved off during TEV protease digestion (albeit subtle

mutations at the C-terminus of the peptide, ENLPETG instead of the original ENLYFQG). Figure 3.7 shows the use of this strategy to site specifically label TEV protease-cleaved EGFR KD with the H_6-LPETGG peptide labeled with a quencher (CruzQuencherTM1 Maleimide). Our protein recovery from labeling 300 µg of EGFR was 140 µg, resulting in ~50% protein yield. Using equation 1 above, we have consistently achieved >0.95 moles of dye per mole of protein in our purified labeled EGFR kinase domain.

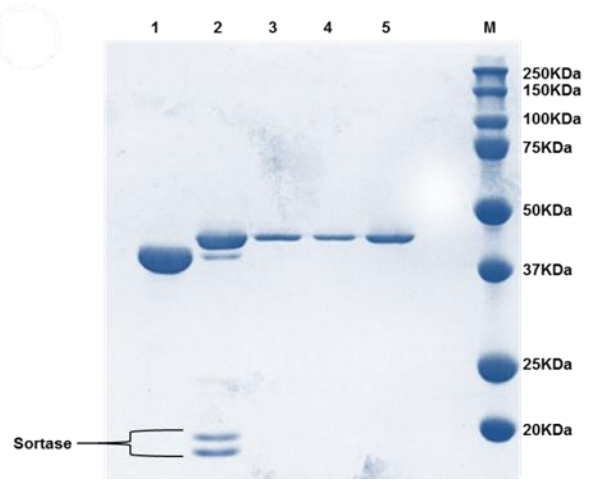


Figure 3.7. Coomassie-stained gel showing the labeling and purification of EGFR kinase domain using an LPETG peptide substrate containing CruzQuencher maleimide. Lanes; 1- Unlabeled EGFR KD, 2- Reaction mix after 30 mins, 3- Fraction from Gel filtration, 4- Eluted protein from Ni-NTA beads, 5- Final concentrated labeled protein

We also used this labeling and purification strategy to modify MSP1E3D1 at the N-terminus using a fluorescein-labeled peptide. As shown in figure 3.8, we achieve >95% purity after purification and the fluorescent image confirms the presence of the fluorophore on the protein.

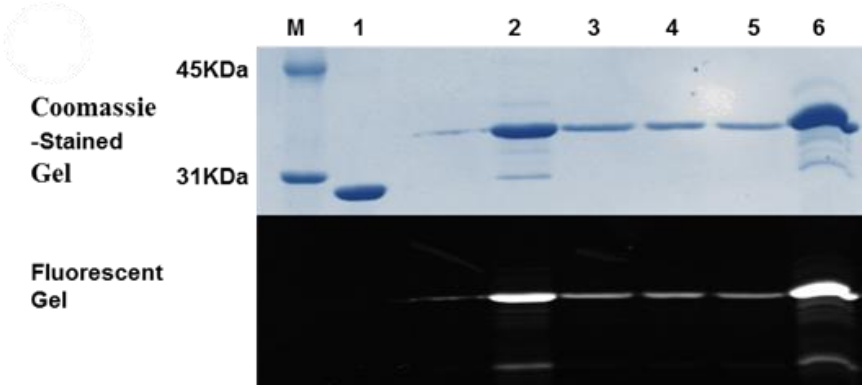


Figure 3.8. Coomassie and fluorescent gels showing the labeling and purification of MSP protein using the H_6-LPETGG peptide substrate containing fluorescein. Lanes; 1- Unlabeled MSP, 2- Reaction mix after 30 mins, 3,4- Fractions from Gel filtration, 5- Eluted protein from Ni-NTA beads, 6- Final concentrated labeled protein.

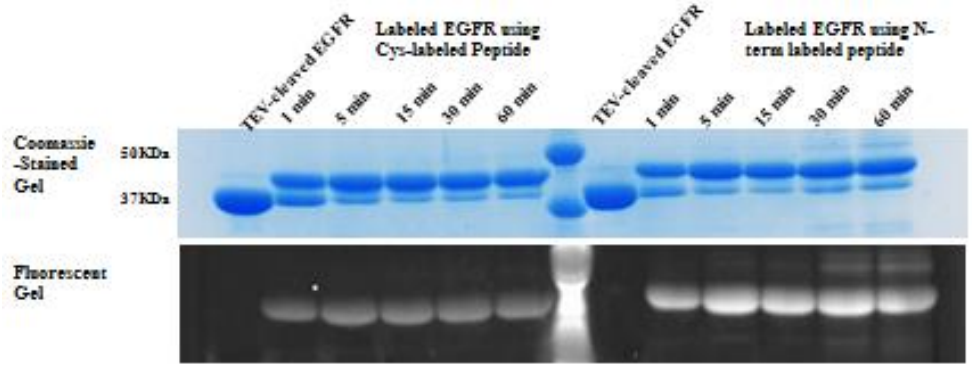


Figure 3.9. . Time course N-terminal labeling of EGFR KD using the H_6-LPETGG peptide substrate labeled with fluorescein at two different positions.

3.6 DISCUSSION

We have demonstrated that TEV-cleaved recombinant proteins are good substrates for SrtA. The use of TEV-cleaved proteins shows greater promise of practicality in biophysical and biochemical applications, but there remained the challenge of sub-stoichiometric protein modification due to back reaction and the reversible nature of sortase-mediated reactions. Addition of an affinity tag to the peptide substrate enables the efficient purification of labeled product, resulting in >95% purity. Although our strategy utilized the 6x histidine tag, which is a commonly used affinity tag for protein purification, we propose that the protocol can be made more stringent by employing two different affinity tags, one that is engineered into the protein of interest (prior to TEV protease cleave) and the other that is part of the fluorophore-containing peptide. The ability to site-specifically modify proteins with fluorophores to stoichiometric proportions is highly desirable for downstream biophysical studies.

In our effort to obtain labeled EGFR kinase domains for biophysical studies, we tested different strategies which included maleimide labeling of cysteine residues, lipoic acid ligase-mediated click chemistry and sortase-mediated C-terminal labeling of EGFR kinase domain. The challenges we faced were two-fold; the temperamental behavior of the EGFR kinase domain where multiple rounds of cysteine-to-serine mutations compromised its enzymatic activity and activation on nickel liposomes, and the tolerance of the kinase domain to reaction conditions. While developing this robust labeling strategy, we realized that the sortase enzyme was still active after loading the labeled products onto a size exclusion chromatography column. Since this sortase A construct is calcium dependent, we demonstrated that molar equivalents of EDTA or EGTA was sufficient to kill the activity of sortase. In addition, while a stretch of N-terminal glycines can be used for the nucleophilic attack of the sortase-peptide intermediate, we have

shown that one glycine residue was enough to allow this reaction to proceed. An interesting discovery was the time needed for TEV protease cleavage of EGFR kinase domain. Standard TEV protease protocols require overnight incubation (at least 16 hrs), which can be detrimental to certain proteins. In our initial experiments, kinase activity assays of TEV-cleaved EGFR kinase domains (incubated overnight) showed diminished enzymatic activities compared to the wildtype EGFR. We have been able to optimize the conditions to minimize the incubation time for TEV protease cleavage (currently 6 – 8 hrs is required to efficiently cleave the EGFR kinase domain). However, we do experience ~50% loss in yield of the cleaved EGFR protein after the reverse nickel purification process, a challenge we are working on to improve the yield. Lastly, the H_6-LPETGG peptide was carefully designed to enable the regeneration of the sequence proximal to the EGFR kinase domain. Though not entirely understood, our data suggests that the location of the fluorophore within this affinity peptide might enhance the fluorescent properties of the labeled protein (Figure 3.9). We realized that the N-terminally labeled fluorophore peptide exhibited greater fluorescence compared to the internal cysteine-labeled fluorophore peptide. This might be attributed to the interaction of the fluorophore with the protein, depending on its position and orientation.

The epidermal growth factor receptor, EGFR has been the archetypical protein for studying receptor tyrosine kinases. EGFR and its orthologues (HER2, HER3 and HER4) have been known to play significant roles in tumorigenesis. Our site-specific fluorophore-labeled EGFR kinase domain offers a biological tool that can be used to answer mechanistic questions regarding protein-lipid and protein-protein interactions by receptor tyrosine kinases. Since these proteins are integrated into the membrane, the major challenge relates to careful selection of a fluorophore that exhibits similar optical properties in the presence and absence of the lipid

environment. Further, the development of the nanodisc technology by Sligar and colleagues has allowed for biophysical studies of proteins in a native bilayer environment. We have demonstrated the generality and applicability of this labeling strategy to other proteins by labeling the membrane scaffold protein MSP1E3D1 using sortase. MSP constructs are used to make nanodiscs and MSP1D1 has been site-specifically labeled using sortase and a fluorescein-labeled depsipeptide (123). MSP1E3D1 in particular generates a 12.1 nm bilayer disc and has been useful in solubilizing and studying membrane proteins in their functional forms (124). Our fluorescein-labeled MSP1E3D1 broadens the utility and applicability of nanodiscs by providing a toolbox to use bulk and single molecule fluorescent spectroscopy to study conformational dynamics and macromolecular interactions between lipids and proteins.

In conclusion, the N-terminal amino group offers unique functional properties for site-specific labeling of proteins. Our N-terminal labeling strategy provides a robust and straightforward protocol for functionalizing TEV protease cleaved recombinant proteins via a sortase-mediated reaction

Chapter 4: Biophysical characterization of EGFR kinase domain dimerization on lipid membranes

4.1 ACKNOWLEDGEMENTS

We thank John Monsey for initial experiments using fluorophore-labeled HER4 kinase domain to directly measure dimerization on Nickel liposomes and Wei Shen for her support with molecular biology. We also thank Dr. Vandna Kukshal for help with FRET experiments and Dr. Jieya Shao for fluorimeter access.

4.2 ABSTRACT

The EGFR family of transmembrane receptor tyrosine kinases consists of EGFR, Her2, Her3 and Her4 that are known to play significant roles in tumorigenesis. Their kinase domains form asymmetric homo- and heterodimers necessary for activation, and leading to many downstream signaling cascades. While these family members share a common architecture with high sequence homology and structural similarity, the signaling potency of the different dimer pairs is dependent on their dimerization partner. Though crystallographic studies have elucidated the structures of EGFR homodimers, our current understanding of how the various heterodimers are organized in three-dimensional space has been the result of cell biological and in vitro biochemical assays. Our lab has shown that inhibition of Her3-Her2 heterodimers by Mig6 (an endogenous protein) is significantly weaker than the other dimer pairs, with Mig6 having a lower binding affinity to Her3. In addition, identification of mutations for all family members has necessitated research into determining how interactions at the molecular level contribute to activation. We hypothesized that structural differences at the dimer interface are one of the factors controlling the functional characteristics of the different dimer pairs. We use a combination of site-directed mutagenesis, *in vitro* kinase assays, biochemical and biophysical approaches especially Fluorescence Resonance Energy transfer to directly measure EGFR kinase domain dimerization on lipid surfaces. Our results using a dimerization-deficient control shows that our current Nickel liposome system may not be a suitable system for measuring dimerization of EGFR kinase domains on lipid membranes.

4.3 INTRODUCTION

Abnormal signaling through the receptor tyrosine kinase of human EGFR family members underlies many cancer tumors (10). This family of proteins includes EGFR, HER2, HER3 and HER4 (3). All the members of the EGFR family share a common architecture: a large extracellular domain, a single alpha helical transmembrane segment, an intracellular domain containing the Tyr kinase domain (except HER3), and an unstructured C-terminal tail with tyrosine autophosphorylation sites (4-7). With the exception of HER2, all three family members are activated by ligand binding. Although HER3 has been regarded as a pseudokinase over the past few years, recent studies have shown that its kinase domain contains a low activity compared with the other three family members (4, 9). Ligand activation in the EGFR family of proteins leads to receptor homo- and heterodimerization, and autophosphorylation of tyrosine residues in the C- terminal region that provide docking sites for Src homology-2 or other signaling molecules that contain phosphotyrosine-binding domains. This initiates a cascade of signaling events that traverses the cytoplasm to communicate with the nucleus and the cytoskeleton (11-14). Recent structural and mutational studies have shown that an allosteric mechanism is necessary for activation by EGFR family dimers. This requires an interaction between the C-lobe of one kinase monomer (activator kinase/donor monomer) and N-lobe of the other monomer (receiver kinase/acceptor monomer). Alteration in the surface contact residues in the N-lobe region in Her3 means that it can only serve as an activator in a heterodimeric pair (4). Despite sequence and structural similarity, homo- and heterodimeric pairs of this family exhibit distinct signaling potencies. Studies have shown that Her2 is the preferred signaling partner for EGFR and HER3 with the HER2-HER3 heterodimers forming a potent signaling pair. Functional

characterization has shown the dependence of the heterodimeric partners for efficient oncogenic transformation (20). Using *in vitro* biochemical approaches, our lab has shown that an endogenous protein, Mig6, inhibits HER2-HER3 heterodimers to a lesser extent compared with its inhibition of other HER2 heterodimers (Figure 4.1). Mig6 binds to the distal surface of the C-lobe of the kinase domain and inhibits dimerization (39). However, how molecular interactions at these dimeric interfaces impact the function of EGFR signaling is not fully understood. Whereas structural data can form the basis of the design of drugs to these receptors, effective therapeutic interventions would necessitate the accurate measurement of binding affinities and kinetic parameters inherent in protein-protein interactions in the EGFR family of kinases. We propose a quantitative approach to measure EGFR family kinase domain dimerization. This would provide direct experimental data regarding thermodynamic and kinetic parameters that govern EGFR dimerization. We will focus mainly on the kinase domain as our starting point since 1) it serves as a link between the extracellular domain and the C-terminal tail for activation and phosphorylation of other proteins, and 2) it gives us a baseline of significant experimental data that would enable us make predictions about how addition of other regions (such as the juxtamembrane domain and C-terminal tail) may impact dimerization and guide us to test these regions. Determination of these parameters with regards to how the kinase domain mutants interact with other family members will inform the field on the mechanism involved in the altered activation and signaling differences exhibited by these mutants.

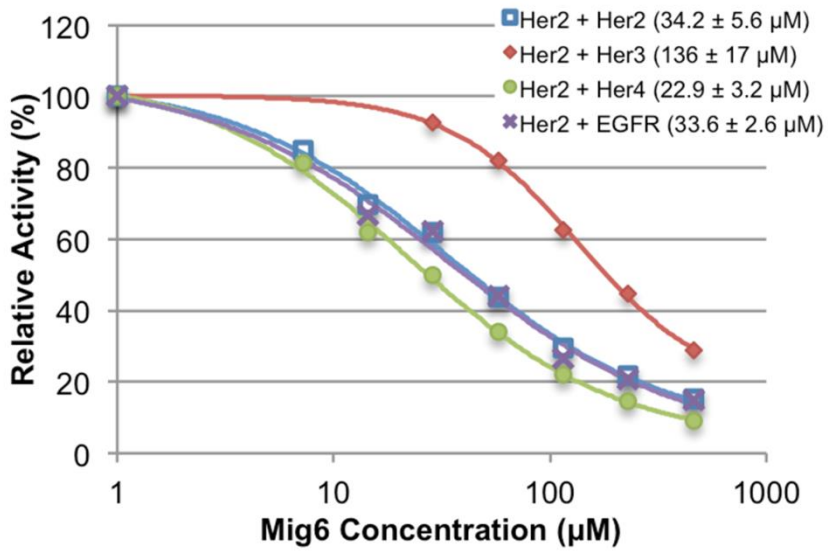


Figure 4.1 Mig6 inhibition of HER2 heterodimers

4.4 METHODS

Plasmids

DNA encoding residues 696–1022 of the human epidermal growth factor receptor was cloned into pFAST BAC HT (Invitrogen) using the NcoI and HindIII restriction sites. This EGFR wildtype kinase domain construct contains an N-terminal 6-His tag, a linker, and a TEV protease cleavage site (MSYYHHHHHHDY DIPTTENLYFQGAM). Multisequence alignment of the EGFR family of proteins shows that their kinase domains contain seven cysteine residues, five of which are conserved (figure 4.2). The six cysteines in the EGFR kinase domain were mutated to serines, either individually or in combination. The nomenclature of the six cysteines was made to reflect conservative cysteines in our prior fluorophore labeling studies of the HER4 kinase domain. Table 4.1 and figure 4.3 provides the nomenclature and position of the cysteines in the structure of the kinase domain, respectively. Site-directed mutagenesis of cysteines was done using the QuikChange kit (Stratagene). Each construct was verified by DNA sequence analysis. We also used an in-fusion method to insert a DNA sequence at the C-terminus of the EGFR kinase domain, the expression of which generates an EGFR kinase domain containing an LPETG motif at the C-terminus. This insertion was also confirmed by DNA sequence analysis.

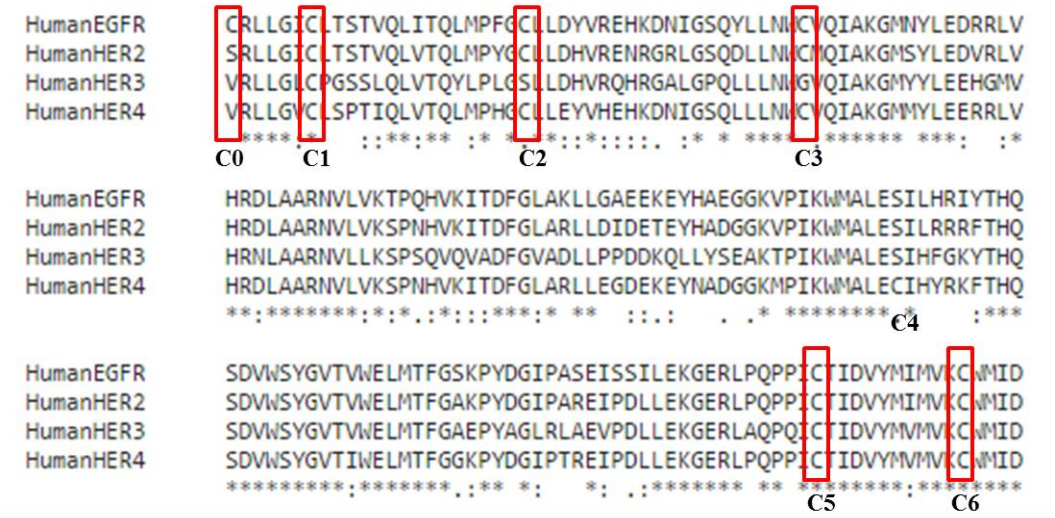


Figure 4.2. Multisequence alignment of the kinase domains of EGFR family members

Cysteine residue	Nomenclature	Notes
Cysteine-775	C0	NA
Cysteine-781	C1	Close proximity to dimer interface
Cysteine-797	C2	Reacts irreversibly with therapeutic inhibitors
Cysteine-818	C3	NA
Cysteine-939	C5	NA
Cysteine-950	C6	NA

Table 4.1. Nomenclature of cysteine residues in the EGFR kinase domain

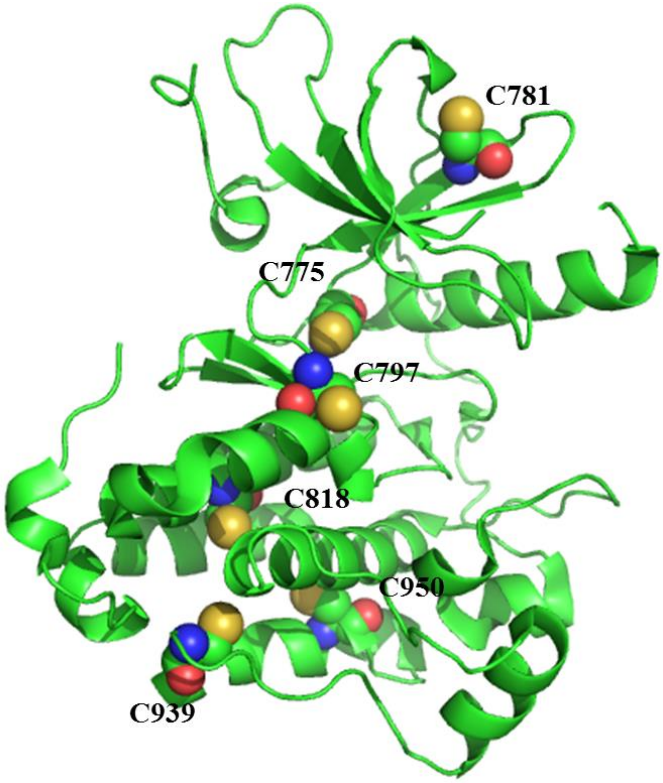


Figure 4.3. Cysteine residues in the kinase domain of wildtype EGFR. This figure was generated using PYMOL and the 2GS2 EGFR crystal structure

Expression and purification of EGFR KD Cysteine-to-Serine mutants

We expressed and purified the EGFR kinase domain cysteine-to-serine mutants using a protein expression protocol described previously (4, 36). Recombinant bacmids (Bac-to-Bac expression system, Gibco BRL) were transfected into *Spodoptera frugiperda* Sf9 cells to produce recombinant baculovirus, which were used to infect Sf9 cells grown at 27 °C and 120 rpm in IPL-41 medium (Invitrogen) supplemented with 10% bovine calf serum supplemented (Hyclone), yeastolate extract (Invitrogen), penicillin, streptomycin, and Pluronic F-68. Cultures at a density of 1 x 10^6 cells/ml were infected at 1 MOI and cells were harvested 48 – 72hours post infection by centrifugation at 6000 rpm and resuspended in Buffer A (25 mM Tris-HCl, pH

8, 300 mM NaCl, 10 mM imidazole, 5% glycerol, 1 mM DTT) supplemented with phenylmethylsulfonyl fluoride, benzamidine, DNase, and protease inhibitor mixture tablets lacking EDTA (Roche). The cells were homogenized using a probe sonicator and the lysate was centrifuged at 20000 rpm for 45 mins. 0.5 - 0.75 ml of Ni-NTA beads (Qiagen) were added to the supernatants and incubated on a rotator for 1 h at 4 °C. The beads were poured into a 10-ml disposable column (Bio-Rad) and the flow-through collected. The column was washed with 10 column volumes of Buffer A + 20 mM imidazole, and the proteins were eluted with Buffer A + 50–125 mM imidazole. The eluted fractions were further purified by gel filtration chromatography on a Superdex 200 GL 10/300 column (GE Healthcare) using 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM DTT buffer.

LplA ligase/Click Chemistry conditions on EGFR KD activity

We used an already published protocol to determine the effect of click chemistry conditions on the enzymatic activity of EGFR kinase domain (125, 126). 50 μM of EGFR kinase domain was mixed with 100 μM $CuSO_4$, 0.5 mM Tris(3-hydroxypropyltriazolyl-methyl)amine, 5 mM Sodium ascorbate, 5 mM Aminoguanidine in buffer containing 100 mM potassium phosphate, pH 7.0 and 100 mM NaCl in a 2ml Eppendorf tube. The tube was closed and attached to a rotisserie and the reaction allowed to proceed for 1 hr. An aliquot of the final reaction was taken and the kinase activity determined, along with positive and negative controls. Positive control included the untreated wildtype EGFR kinase domain while negative controls included reactions that had one or more components omitted.

Design and fluorophore-labeling of oligoglycine peptides for C-terminal labeling

We used solid-phase synthesis to make two different oligoglycine peptides; GGGK and GGGCGGAWSHPQFEK. The eight-amino acid sequence, WSHPQFEK is a strep tag and

interacts with a specially engineered streptavidin construct called strep tactin. Following cleavage from the resin, the final crude oligoglycine peptides were purified by reversed-phase HPLC on a C18 column using a 10 – 70 % Buffer B (H_2O, Acetonitrile and 0.05% TFA) gradient over 15 mins. The fractions were analyzed by LC/MS and the desired fractions were pooled together and lyophilized to dryness. The lyophilized peptides were reconstituted into water and the pH adjusted to around 8. Three molar equivalents of fluorescein or QSY7 maleimide (in DMSO) were added to 5 mg of peptide in 50 mM Tris-HCl, pH 7.3 and 150 mM NaCl, and incubated at room temperature overnight. The labeled peptides were purified, analyzed and lyophilized using the same method described above. Measurement of the absorbance of the peptide and fluorophore at 280 nm (A_{280} and A_{max}, respectively) and their molar extinction coefficients ($\varepsilon_{protein}$ and ε_{max}, respectively), allowed us to calculate the peptide concentration and degree of labeling using equation 1 below;

$$DOL = \frac{A_{max} \cdot \varepsilon_{peptide}}{(A_{280} - A_{max} \cdot CF_{280}) \cdot \varepsilon_{max}}$$

The correction factor (CF) is included in this equation to account for the absorption of fluorophore at 280 nm and equals the A_{280} of the dye divided by the A_{max} of the dye.

C-terminal fluorophore labeling of EGFR KD using Fluorescein-labeled GGGK peptide

A solution containing 25 μM of EGFR kinase domain (containing the LPETG sortase recognition motif at the C-terminus), 6 μM sortase and 0.5 mM fluorescein-labeled GGGK peptide was mixed together in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$. The reaction was incubated at 4 °C for 30 mins in the dark. This was loaded onto a

Superdex 75 GL 10/300 column (GE Healthcare) using 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM DTT buffer. Fractions containing EGFR kinase domain were pooled together, concentrated and snap-frozen at -80 °C. Aliquots were taken at this step and analyzed by SDS-PAGE followed by Coomassie staining. The controls included were: EGFR kinase domain (containing the LPETG sortase recognition motif at the C-terminus) only, sortase only, fluorescein-labeled GGGK peptide only, EGFR kinase domain (containing the LPETG sortase recognition motif at the C-terminus) and sortase, EGFR kinase domain (containing the LPETG sortase recognition motif at the C-terminus) and fluorescein-labeled GGGK peptide, sortase and fluorescein-labeled GGGK peptide.

C-terminal fluorophore labeling of EGFR KD using Fluorescein and QSY7-labeled GGGCGGAWSHPQFEK peptide

A solution containing 25 μM of EGFR kinase domain (containing the LPETG sortase recognition motif at the C-terminus), 6 μM sortase and 0.5 mM fluorescein-labeled GGGCGGAWSHPQFEK peptide was mixed together in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$. The reaction was incubated at 4 °C for 30 mins in the dark. This was quenched with 10 mM EDTA and loaded onto a Superdex 75 GL 10/300 column (GE Healthcare) using 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM DTT buffer. Pooled fractions containing the EGFR kinase domain was incubated with 100 μl of strep tactin II beads at 4 °C for 30 mins. Fluorescein-labeled EGFR was eluted with 2.5 mM desthiobiotin, buffer exchanged into Buffer A (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM DTT) and concentrated to ~ 1 mg/ml. Aliquots were taken at each step of the process and analyzed by SDS-PAGE. We determined the concentration and degree of labeling of the fluorescein-labeled EGFR using equation 1, where peptide is replaced with protein in the equation.

Design and fluorophore-labeling of peptides for N-terminal labeling

We designed, synthesized, and purified a 25-mer peptide, MSYYHHHHHHDYDIPTCENLPETGG (or H_6-LPETGG peptide). The internal cysteine was labeled with different maleimide fluorophores, depending on the downstream application. The maleimide fluorophores used were Alexa350, Alexa488, Alexa568 and fluorescein. Three molar equivalents of fluorophore were made to react with one molar equivalent of peptide at room temperature overnight. The resulting labeled peptides were purified, analyzed and lyophilized using the method already described. The degree of labeling was calculated using equation 1.

N-terminal fluorophore labeling of EGFR KD

The method used to site-specifically fluorophore label the EGFR kinase domains at the N-terminus using a sortase-mediated reaction has been described exhaustively in chapter 3. LPETGG peptides labeled with any of the four fluorophores (Alexa350, Alexa488, Alexa568 or fluorescein) was incubated with Gly-EGFR and sortase in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$ under similar reaction conditions and reagent molar equivalents, as described already. Fluorophore-labeled EGFR kinase domains were concentrated to ~1 mg/ml and the degree of labeling calculated using equation 1.

Kinase activity of EGFR KD mutants and fluorophore-labeled EGFR KD

The protocol used to measure the kinase activities of labeled EGFR KDs, both in solution and on Nickel liposomes (to induce dimerization) have been described in Ref 36. Wildtype or fluorophore-labeled EGFR KDs were incubated in 20 mM Tris, pH 7.5, 10 mM $MgCl_2$, 100 μM Na_3VO_4, 2 mM DTT, 100 μM ATP, 1 μCi of [γ- 32P] ATP, 60 mM NaCl, 5% glycerol in 25-μl total reaction volume. The reaction was initiated by the addition of 1.25 μM kinase, incubated at 30 °C for 6 mins and quenched with EDTA. 100 μg of avidin was added to the samples and

transferred to centrifugal filtration units with 30,000 molecular weight cutoff (Millipore). The samples were washed four times with 0.5 M sodium phosphate, 0.5 M NaCl, pH 8.5 and the retained ^{32}P measured by scintillation counting. The kinase activity on liposomes were measured by incubating the wildtype and fluorophore-labeled EGFR KDs on ice for 15 mins before initiating the reaction. Total lipid concentration was 0.5 mg/ml. Samples tested included EGFR KD wildtype, Gly-EGFR KD, Fluorescein-labeled EGFR KD, Alexa488-labeled EGFR KD, Alexa568-labeled EGFR KD, CruzQuencher-labeled EGFR KD, EGFR KD (with an LPETG motif at the C-terminus), and EGFR KD labeled with either GGGK-fluorescein, GGGCGGAWSHPQFEK-fluorescein or GGGCGGAWSHPQFEK-QSY7 at the C-terminus.

FRET assay using Alexa488-labeled-EGFR and Alexa568-labeled EGFR

1μM Alexa488-labeled EGFR KD was mixed with 1μM or 2μM of Alexa568-labeled EGFR KD in a buffer containing 20 mM Tris-HCl, pH 8, 150 mM NaCl, 1 mM DTT or Nickel liposomes. The total volume of the mixture was 22μl. This was incubated on ice for 20mins, aliquoted into 384 corning plate wells and placed in an Infinite M200 Microplate reader. The mixture was excited at 490nm and fluorescent measurements taken over a wavelength range between 520nm and 620nm. Controls included Alexa488-labeled EGFR KD(V-R dimerization-deficient mutant) in buffer or on nickel liposomes, Alexa568-labeled EGFR KD(V-R dimerization-deficient mutant) in buffer or on nickel liposomes, Alexa488-labeled H_6-LPETGG peptide in buffer or on nickel liposomes, Alexa568-labeled H_6-LPETGG peptide in buffer or on nickel liposomes and nickel liposomes only.

4.5 RESULTS

EGFR kinase domain Cysteine mutants have diminished kinase activity

The enzymatic kinase activity of EGFR kinase domain was to determine the structural and functional integrity of the cys-ser mutants. We measured the phosphorylation of a biotinylated degenerate peptide by the cys-ser mutants. Compared to wildtype EGFR kinase domain, four of the singly mutated cys-ser constructs showed similar kinase activity. However, mutations at two key cysteine residues (C1 for Cys781 and C2 for Cys797) resulted in abrogation of the kinase activity. We made triple, quadruple and quintuple constructs that contained those cys-ser mutants that showed appreciable kinase activity. We also tested mutating C1 to alanine and C2 to aspartate and asparagine. Our data suggests that combination of these cysteine mutations in the quintuple mutant resulted in very low kinase activity compared to the wildtype (Figures 4.4 to 4.11)

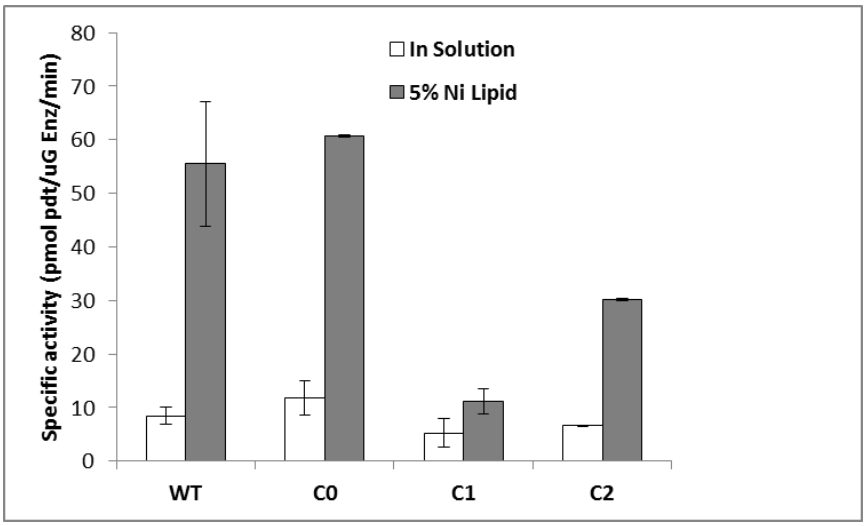


Figure 4.4. Effect of single cysteine mutation (EGFR C0, C1 and C2) on kinase activity and activation on liposomes. Cysteines at respective residues were mutated to serines.

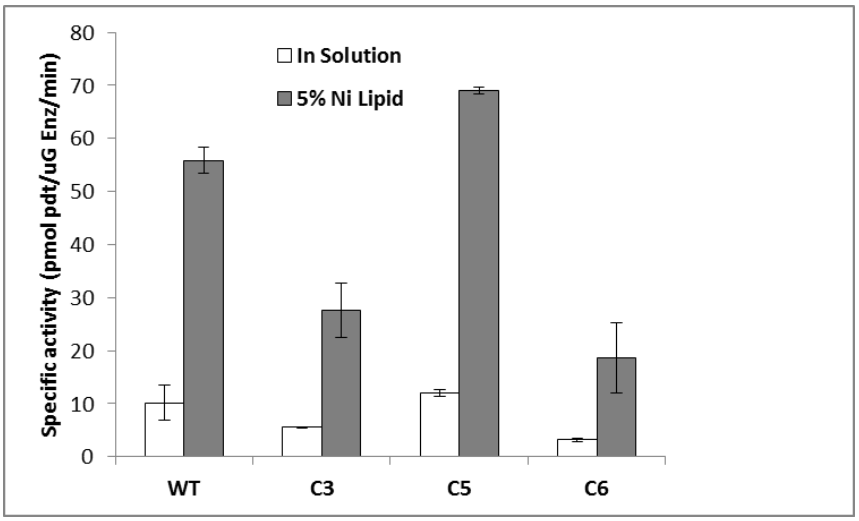


Figure 4.5. Effect of single cysteine mutation (EGFR C3, C5 and C6) on kinase activity and activation on liposomes. Cysteines at respective residues were mutated to serines.

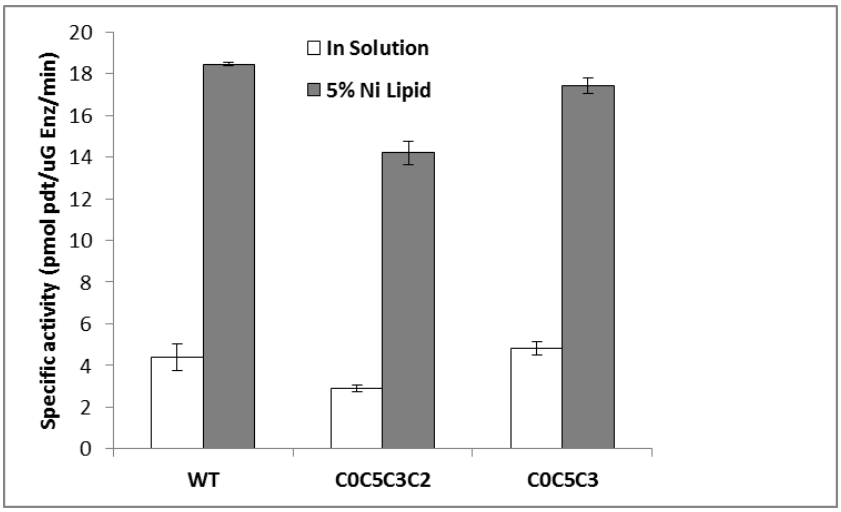


Figure 4.6. Effect of triple (C0C5C3) and quadruple (C0C5C3C2) cysteine mutations on kinase activity and activation on liposomes. Cysteines at respective residues were mutated to serines.

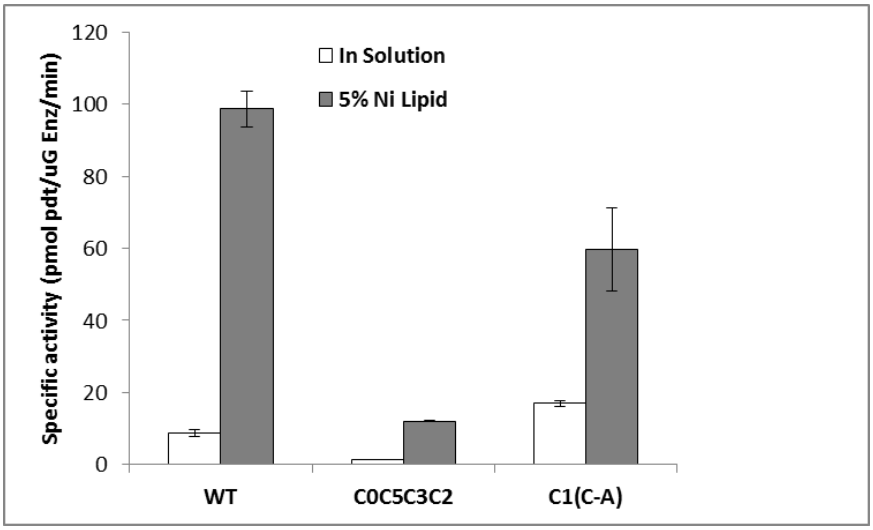


Figure 4.7. Effect of single (C1, C-A) and quadruple (C0C5C3C2) cysteine mutations on kinase activity and activation on liposomes. In the quadruple mutant, all cysteines have been mutated to serines, while in the single mutant, cysteine at position C1 has been mutated to alanine.

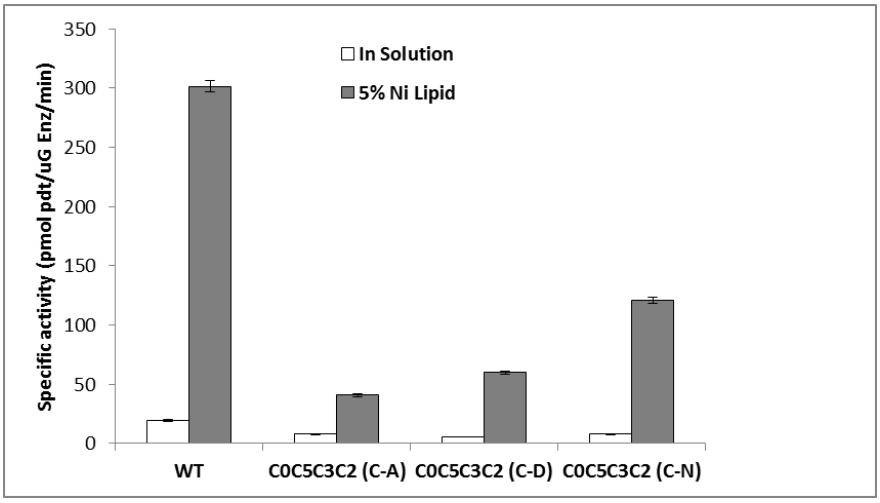


Figure 4.8. Effect of cysteine mutation at the C2 position in the quadruple mutant (C0C5C3C2) on kinase activity and activation on liposomes. Apart from the C2 positions that have been mutated to alanine, aspartate or asparagine, all other cysteines at the other positions have been mutated to serine.

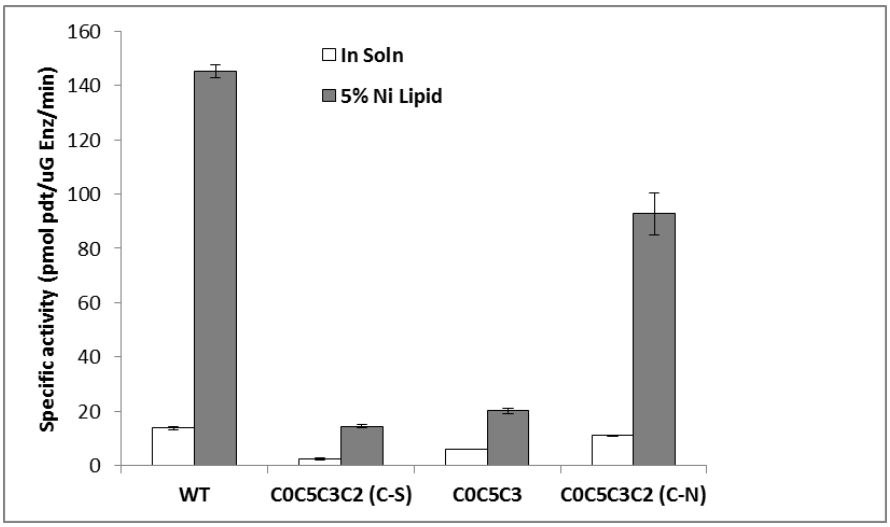


Figure 4.9. Effect of cysteine mutation at the C2 position in the quadruple mutant (C0C5C3C2) on kinase activity and activation on liposomes. Comparing cysteine to serine or cysteine to asparagine mutation at the C2 position in the quadruple mutant.

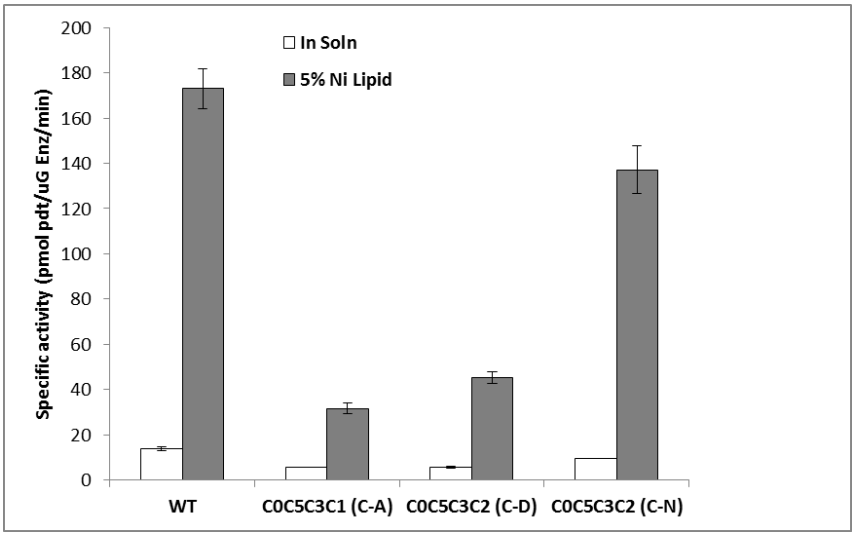


Figure 4.10. Effect of cysteine mutation at the C1 and C2 positions in two quadruple mutants (C0C5C3C2 and C0C5C3C1) on kinase activity and activation on liposomes. In the C0C5C3C1 mutant, the C1 has been mutated to alanine while all other cysteines are mutated to serine. In the C0C5C3C2 mutants, C2 is either mutated to aspartate or asparagine while the three other cysteines are mutated to serines.

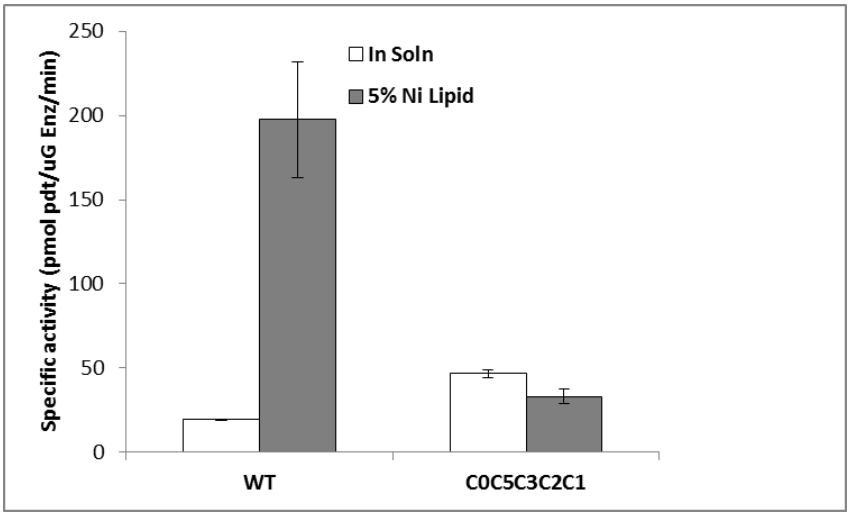


Figure 4.11. Effect of EGFR kinase domain pentamutant (C0C5C3C2C1) on kinase activity and activation on liposomes. While the other three cysteines (C0C5C3) have been mutated to serines, C1 is mutated to alanine while C2 is mutated to asparagine.

LplA ligase/Click chemistry conditions inhibit EGFR KD activity

We subjected the EGFR kinase domain to conditions that were required for click chemistry using LplA ligase. We also excluded some components of the reaction mixture to determine the effect of those components on the activity of EGFR KD. After 60 minutes of reaction, we measured the kinase activity of EGFR KD with and without the click chemistry components. Figure 4.12 shows that the addition of $CuSO_4$, sodium ascorbate and THPTA results in the loss of kinase activity.

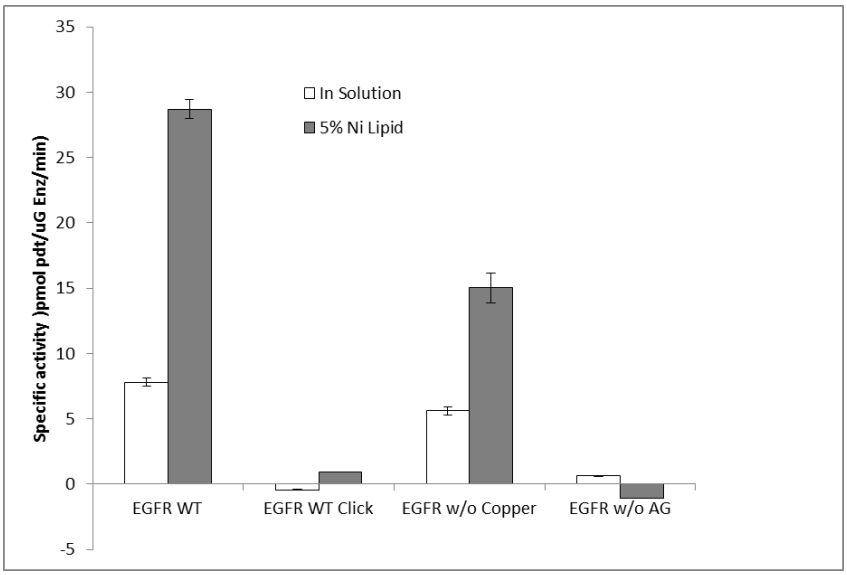


Figure 4.12. Kinase activity of EGFR kinase domain with LplA ligase and CuAAC conditions. EGFR WT, EGFR wildtype , EGFR WT click, EGFR wildtype mixed with CuSO4, Tris(3-hydroxypropyltriazolyl-methyl)amine, Sodium ascorbate, Aminoguanidine in buffer, EGFR w/o copper, EGFR wildtype mixed with, Tris(3-hydroxypropyltriazolyl-methyl)amine, Sodium ascorbate, Aminoguanidine in buffer, EGFR w/o AG, EGFR wildtype mixed with CuSO4, Tris(3-hydroxypropyltriazolyl-methyl)amine, Sodium ascorbate in buffer.

C-terminal fluorophore-labeled EGFR KD has low kinase activation

Following successful elution of the labeled EGFR-LPETG with fluorescein-labeled GGGCGGAWSHPQFEK peptide, our calculations resulted in >95% purity after the post-labeling purification. We obtained similar results when EGFR-LPETG was labeled with the QSY-labeled GGGCGGAWSHPQFEK peptide. We also compared the kinase activity and activation on nickel liposomes of EGFR KD WT to the labeled and unlabeled EGFR-LPETG.

The results show a diminished kinase activity and activation of the EGFR-LPETG (and fluorescein-labeled and QSY-labeled EGFR) as compared to wildtype (4.13)

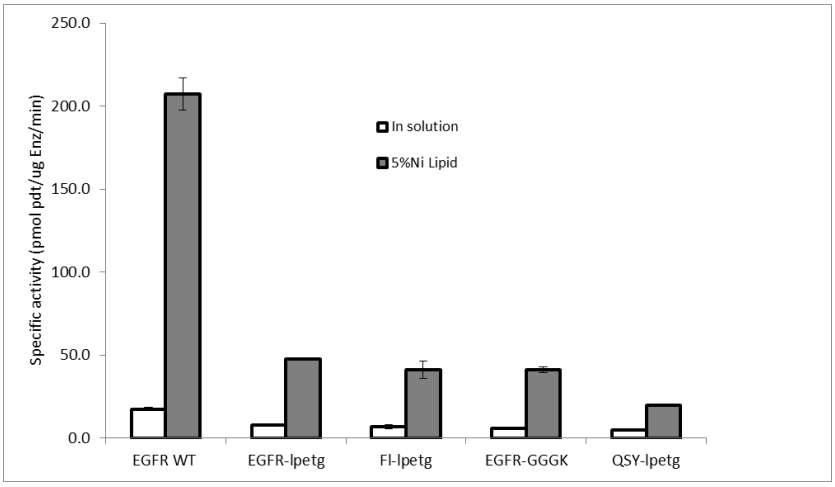


Figure 4.13. Effect of EGFR kinase domain C-terminal modifications on kinase activity and activation on liposomes. Samples; EGFR WT, EGFR wildtype, EGFR-lpetg, EGFR wildtype engineered with LPETG sortase recognition motif, FL-lpetg, EGFR wildtype engineered with LPETG sortase recognition motif and labeled with fluorescein at the C-terminus, EGFR-GGGK, EGFR wildtype engineered with LPETG sortase recognition motif and labeled with GGGK peptide, QSY-lpetg, EGFR wildtype engineered with LPETG sortase recognition motif and labeled with QSY quencher at the C-terminus.

N-terminal fluorophore-labeled EGFR KD has comparable kinase activity to wildtype

We designed and synthesized an affinity tag-containing peptide that was labeled with a fluorophore. This was used to label a TEV protease cleaved EGFR KD (g-EGFR) that generates a nucleophilic glycine necessary for the sortase reaction. For detailed results of the challenges, drawbacks and breakthroughs involved in successful labeling of EGFR KD at the N-terminus with a fluorophore, please refer to chapter 3. Initially, when we compared the kinase activities

and activation of fluorescein-labeled EGFR to wildtype EGFR, the data showed that the labeled EGFR kinase domains had comparatively lesser activity and lowered fold activation on nickel liposomes. We subjected the wildtype EGFR KD to a mock TEV protease digestion (containing no TEV protease), mock labeling (with no sortase) and subsequent purification with gel filtration and Ni-NTA columns. Here, we show that mock-labeled EGFR KD has similar activity and fold activation as labeled EGFR KD (Figure 4.14). We also show that fluorophore-labeled EGFR kinase domain retains activity, following the shortening of the time required for TEV protease cleavage and optimization of the purification steps (figure 4.15 and 4.16). These fluorophore-modified EGFR kinase domains had 95% purify after labeling and our calculations show that there is a 1:1 fluorophore: protein ratio.

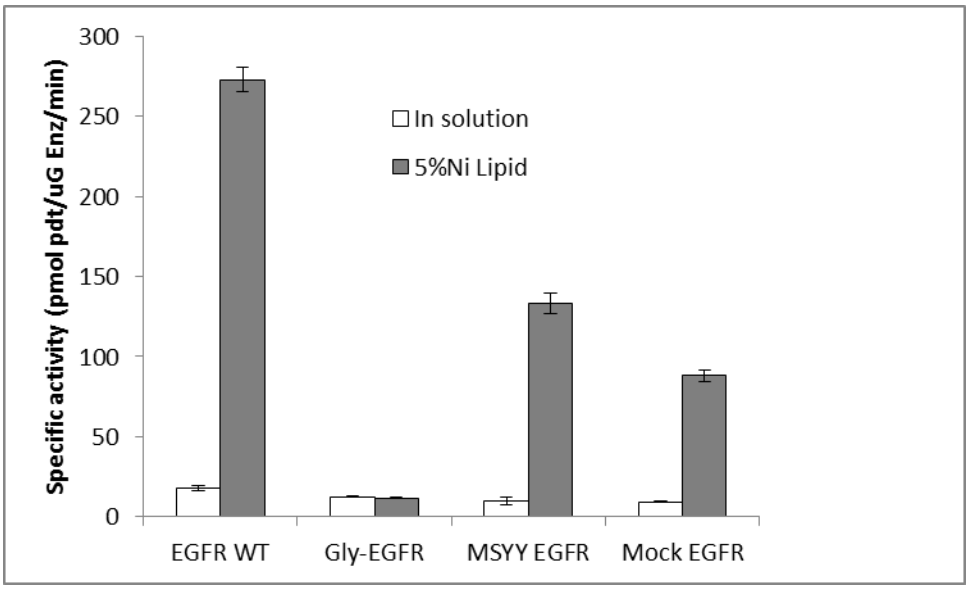


Figure 4.14. Effect of TEV protease cleavage, labeling and purification conditions on EGFR kinase domain activity

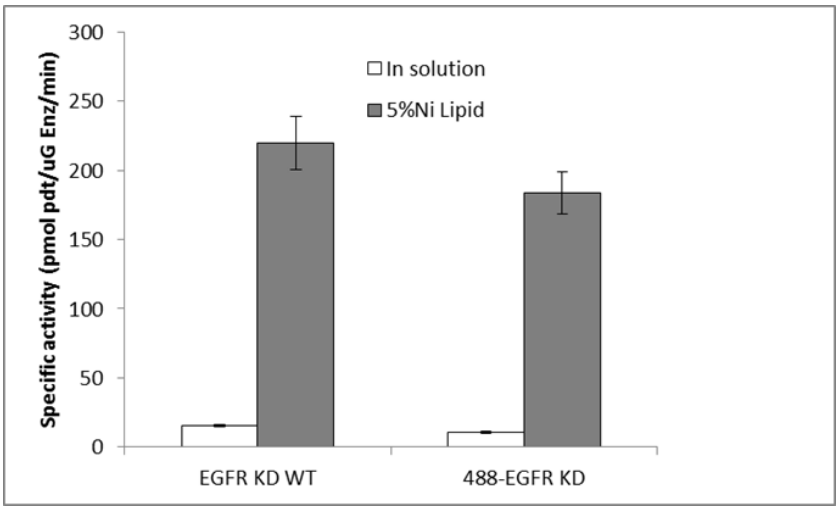


Figure 4.15. Effect of fluorophore labeling on EGFR kinase domain activity. EGFR kinase domain was labeled with Alexa488 maleimide.

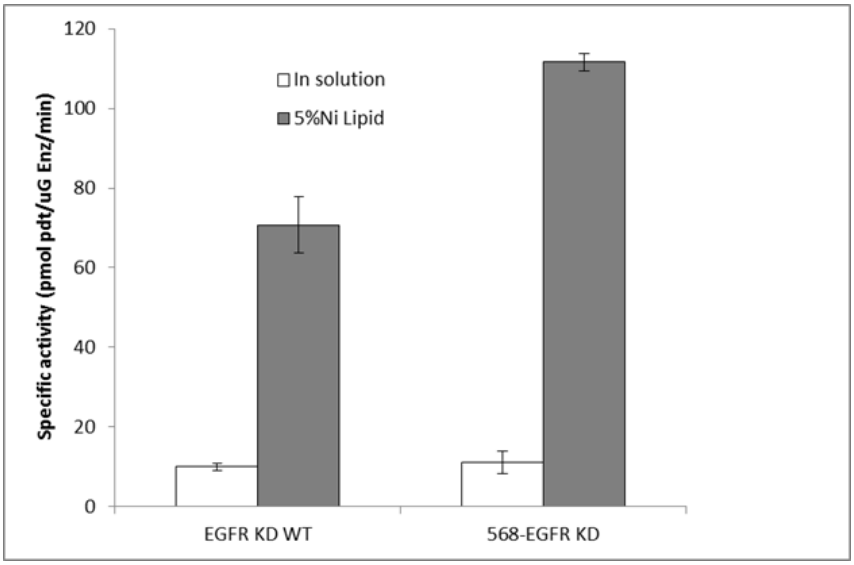


Figure 4.16. Effect of fluorophore labeling on EGFR kinase domain activity. EGFR kinase domain was labeled with Alexa568 maleimide.

Alexa488-EGFR interacts with Alexa568-EGFR when in close proximity

Following Nickel liposome incubation, the data shows a ~50% decrease in the donor (488-EGFR) fluorescence and a ~35% increase in acceptor (568-EGFR) fluorescence as shown in figure 4.17A. This fluorescence change was also reflected in the calculated FRET ratios (Table 4.2). Negative controls using dimerization-deficient EGFR KD (V924R mutant) labeled with the Alexa FRET pairs also showed increased FRET ratios on liposomes as observed with wildtype EGFR KD (Figure 4.17B). Lastly, peptide labeled with Alexa488 and Alexa568 produced data that suggested an increase in the FRET ratio on liposomes as compared to solution conditions. (figure 4.18)

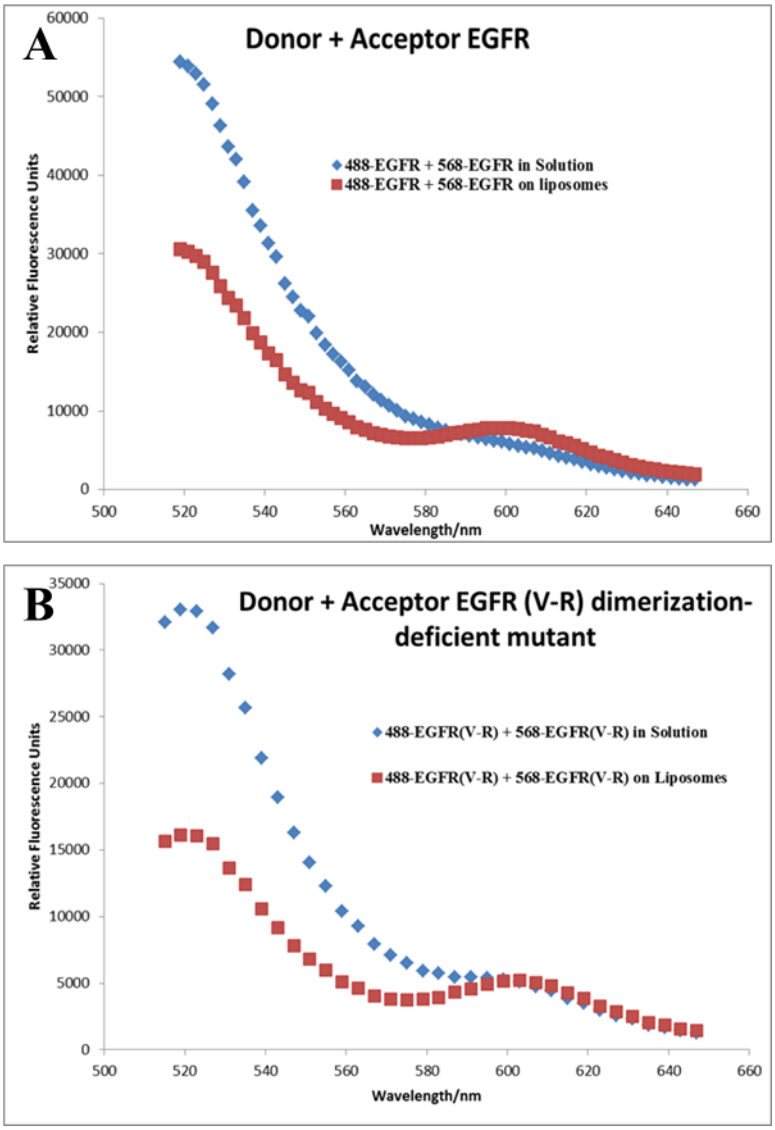


Figure 4.17. FRET measurements using Alexa488- and Alexa568-labeled EGFR kinase domains. A. Measuring FRET using Alexa488- and Alexa568-labeled wildtype EGFR kinase domain. B. Measuring FRET using Alexa488- and Alexa568-labeled dimerization-deficient EGFR kinase domain mutant (Val924Arg).

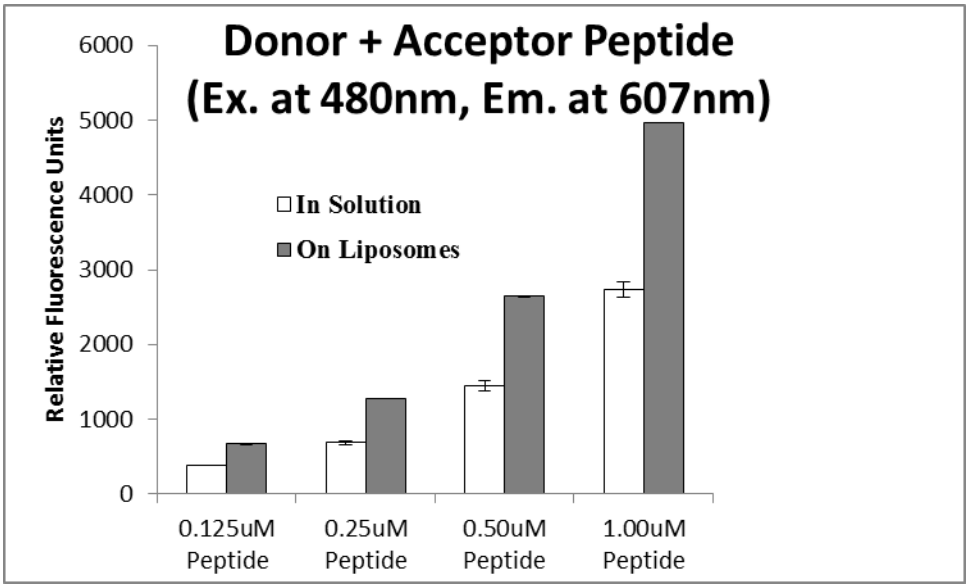


Figure 4.18. Effect of concentration of the fluorophore-labeled peptide alone on FRET measurements

	FRET Ratio (Emission at 607nm/Emission at 530nm)	
	Buffer	**Nickel Liposomes**
EGFR KD	0.09	0.24
EGFR KD (V924R)	0.1	0.20
MSYY peptide	0.05	0.1

Table 4.2. FRET ratios for EGFR kinase domain (and dimerization-deficient mutant) and MSYY peptide. FRET ratio between 0.1 – 10 is considered a range that allows for favorable efficiency calculations (127).

4.6 DISCUSSION

The formation of an asymmetric dimer by EGFR (HER2, HER3 and HER4) kinase domains is critical for activation and subsequent phosphorylation of their C-terminal tails. Notwithstanding, mechanistic details regarding how these kinase domains dimerize is under investigation. Most importantly, questions regarding binding affinities and kinetics of dimerization remain unanswered. In this work, we attempted to develop a biological probe that can be used to directly measure EGFR kinase domain dimerization on lipid surfaces. Our approach was to develop an EGFR kinase domain construct that is site-specifically labeled with two fluorophores, either a fluorescent donor-acceptor FRET pair or a fluorophore-quencher pair. Our enzymatic kinase activity results with EGFR kinase domain cysteine mutants suggested that two key cysteine residues (Cys781 and Cys797) were important for the structural and functional integrity of the kinase. These two cysteines have been shown to be solvent accessible (128) while Cys797 have been historically targeted by irreversible kinase inhibitors. The advent of chemoenzymatic approaches allowed us to take advantage of a sortase-mediated approach to site-specifically fluorophore-label the EGFR kinase domain at the C- and N-terminus. Unfortunately, C-terminal labeling compromised the activation of the kinase. X-ray crystal structures show little to no density of the distal region following the kinase domain core. From our analysis, we propose that this region is close to the dimer interface and its alteration would have deleterious effects on the structure and function of the kinase. The N-terminally labeled kinase domain provided an excellent opportunity for use in developing quantitative approaches to measure dimerization. We tested various donor-acceptor FRET pairs and a fluorophore-quenching system to develop this dimerization assay. Our findings suggest that the fluorophores interact differently with the nickel liposomes and this posed a huge challenge in establishing the requisite parameters for the assay.

During the development of this assay with Alexa488 and Alexa568, we obtained appreciable signal changes when EGFR kinase domains were incubated on liposomes. This result was encouraging and provided preliminary evidence that the kinase domains incubated on nickel liposomes were within a distance that allowed for transfer of energy. However, rigorous negative control experiments with the dimerization-deficient mutant (Val924Arg) and the fluorophore-labeled peptide produced similar FRET results. The ability of the fluorophore-labeled peptide to FRET (in the absence of EGFR kinase domain) shows that the assay might be measuring the proximity of the peptides or proteins and not the transfer of energy. In fact, incubation on nickel liposomes allows the concentration of the his-tagged EGFR KD, perhaps favoring conformational changes that result in dimerization and subsequent activation. Although my working concentration (protein) for my FRET assay is 1-2uM, this value translates to 1-2mM local concentration of EGFR KD on liposomes (4). At such high concentrations, the fluorophore pairs in the linker region are poised to FRET irrespective of the presence or absence of proteins.

In conclusion, this study presents a basis for further research to develop a biological probe to measure EGFR kinase domain dimerization on lipid surfaces. While the nickel liposome/his-tagged kinase domain system presents challenges for directly measuring dimerization affinities, our single-site fluorophore-labeled EGFR kinase domain can be useful in developing a different optical-based assay to determine how these proteins interact. For instance, we have successfully labeled the EGFR kinase domain containing the juxtamembrane segment with a FRET pair. Since studies have shown that the JM can dimerize without the need of lipid membranes (46), we can use this labeled JM construct and equilibrium studies (analytical ultracentrifugation) to monitor EGFR kinase domain dimerization in solution.

Chapter 5: Conclusion and Future Directions

5.1 Conclusion and future perspectives

The biology of the EGFR family of receptor tyrosine kinases is intimately connected to the structural organization of the different domains of these proteins. A deeper understanding of these structures has aided the development of therapeutics that target different structural states of both the extracellular and intracellular domains. However, the field has not made enough advancement in characterizing the C-terminal tails and dissecting their unique structural features. In this work, we provide experimental data that improves our basic understanding of the C-terminal tails in isolation. We also initiate experiments to build a platform that would enable the quantitative measurement of kinase domain dimerization on lipid surfaces using site-specifically fluorophore-labeled EGFR protein constructs.

We are the first to provide a comprehensive and detailed biophysical analysis of the C-terminal tail of a receptor tyrosine kinase. In chapter 2, we use multiple biophysical approaches to produce convincing experimental data that the EGFR and HER3 C-terminal tails are intrinsically disordered. Since the discovery of EGFR and its involvement in several human cancers, the field has made tremendous progress towards elucidating the structures of the various domains and enhancing the development of therapeutics against EGFR and HER2 using these structures. Our results provide stimulating impetus to investigate the dynamics of the C-terminal tails. Although the C-terminal tails have comparable molecular weights to globular proteins, their apparent size using the biophysical tools, are usually bigger than their globular counterparts. In fact, our data using SAXS shows that the EGFR C-terminal tail assumes greater shape and size compared to the compact, globular kinase domain. Such dynamic and flexible characteristics of the C-terminals are very important, given that they provide docking sites and interacting surfaces for multiple proteins to bind.

Several outstanding questions remain regarding the C-terminal tails. There is no structural or biophysical information regarding the interaction of kinase domain and the C-terminal tail, prior to activation. We have attempted to use Isothermal Calorimetry to determine the K_d of the tail interacting with the kinase domain and have been unsuccessful. Given that these intrinsically disordered regions have high on- and off rates, it is often a challenge to determine these thermodynamic parameters. This is also compounded by the high molar ratios needed for this interaction, and often requires the C-terminal tails to be in excess, a situation that can lead to aggregation and difficulty in interpreting the data. Secondly, it is unknown how phosphorylation affects the shape, size and conformation of the EGFR and HER3 C-terminals. In certain proteins, it has been shown that there is a disorder-to-order transition following phosphorylation (129). Since this is not a general phenomenon, it is important to determine how phosphorylation affects the intrinsic disorder characteristics of these proteins. The challenge that remains is the stoichiometric phosphorylation of these tails and effective separation of the fully phosphorylated C-terminal tail from the partially and unphosphorylated species. We circumvented this by expressing and purifying a phosphomimetic mutant (EGFR Y-E), where all the nine tyrosines have been replaced with a glutamate residue. Preliminary SAXS measurements show that wildtype EGFR C-terminal and the EGFR Y-E mutant have similar R_g values, and that there might be no difference in size between the phosphorylated and unphosphorylated constructs. Further biophysical studies need to be done to accurately measure the size and shape of the phosphorylated species. A major criticism for using this strategy is that the Y-E mutant does not fully represent the phosphorylated state of EGFR. Another important question that remains unanswered is the significance of intrinsic disorder to the function of EGFR family of receptor tyrosine kinases. In this realm, it might be necessary to combine these *in vitro* experiments with

cell-based approaches to fully understand their importance. Since intrinsically disordered regions are enriched in polar residues and have less hydrophobic residues, it will be important to create mutants that are either less disordered or more disordered, and determine their effects in cell-based assays on signaling outputs.

Our study in chapter 3 provides a strategy that can be used to N-terminally label any recombinant protein that contains a TEV cleavage site proximal to the protein core, using sortase. We successfully use this sortase-mediated strategy to label two proteins, the EGFR kinase domain and MSP1E3D1, a membrane scaffold protein. The genomic and post-genomic eras necessitate the development of strategies for protein bioengineering and incorporation of a variety of modalities to proteins for basic science research, biotherapeutics and biotechnological utility. Protein modification approaches must be simple and robust, requiring shorter reactions times for labeling. Such strategies should be amenable to functionalizing proteins with different handles at any desired location with minimum structural and functional consequences. Most importantly, protein labeling methods must be efficient, which enables their utility for quantitative and other biophysical studies. A major challenge faced by labeling approaches pertains to the purity of the labeled protein, and the related removal of unlabeled species. The N-terminal amino group offers unique functional properties for site-specific labeling of proteins. Efficient labeling with N-hydroxysuccinimide esters requires controlled pH experiments to selectively modify the N-terminal amino group over lysine residues which is always challenging as a result of the closeness in pKa of the α-amino group of the protein versus the ε-amino group of lysine. Sortase fulfilled most of the requirements for chemoenzymatic modification of proteins, but still faced the challenge of sub-stoichiometric labeling. However, post-labeling purification using an affinity tag in the peptide substrate resulted in >95% purity of the labeled protein. Our strategy

enables the use of labeled proteins for biophysical and biochemical studies as well as for diagnostic purposes.

In chapter 4, we employed the single-site fluorophore labeled EGFR kinase domain to develop a quantitative tool that can be used to measure dimerization on lipid surfaces. Our data suggests that the assay we developed was measuring the proximity of the kinase domains, rather than dimerization. This suggests that the nickel liposome system provides a platform that allows the population of histidine-tagged proteins and that dimerization might be occurring on these surfaces through a mechanism that is not completely understood, presumably altering the conformation of the kinase domain. While most studies have shed light on the structural features of the kinase domain and provided indirect measurement of the dimerization of kinase domains on lipid membranes, there needs to be further investigation regarding the direct measurement of dimerization and the determination of quantitative parameters that govern this. This would provide us with a molecular understanding of how EGFR kinase domain dimerizes. Such research provides signatures specific for each dimeric pair that would drive rational and structure-based design of therapies that target these dimeric interfaces. In fact, the crystal structure and chemical foot printing mass spectrometric analysis of the EGFR-HER3 and HER2-HER3 dimers respectively (130,131), provides stimulating evidence for the field to initiate investigations that would enhance the development of inhibitors that target the dimer interface of homo- and heterodimers. Our lab and other groups have shown that certain cancer-associated mutations are close to the dimer interface of the kinase domain. A robust dimerization assay would enable us determine the effect of these activating mutations on dimerization. Determination of these parameters with regard to how the kinase domain mutants interact with other family members will inform the field on the mechanism involved in the altered activation

and signaling differences exhibited by these mutants. It also provides a baseline of significant experimental data that would enable us make predictions about how addition of other regions (such as the juxtamembrane domain and C-terminal tail) may impact dimerization and guide us to test these regions.

REFERENCES

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

2. McKanna, J. A., Haigler, H. T., and Cohen, S. (1979) Hormone receptor topology and dynamics: morphological analysis using ferritin-labeled epidermal growth factor. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 5689-5693

3. Hynes, N.E., and MacDonald, G. (2009) ErbB receptors and signaling pathways in cancer. *Curr Opin Cell Biol.* **21,** 177-184

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

5. Jura, N., Shan, Y., Cao, X., Shaw, D.E., and Kuriyan, J. (2009) Structural analysis of the catalytically inactive kinase domain of the human EGF receptor 3. *Proc. Natl. Acad. Sci. U S A.* **106,** 1608-1613

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

7. Qiu, C., Tarrant, M.K., Choi, S.H., Sathyamurthy, A., Bose, R., Banjade, S., Pal, A., Bornmann, W.G., Lemmon, M.A., Cole, P.A., and Leahy, D.J. (2008) Mechanism of activation and inhibition of the Her4/ErB4 kinase. *Structure* **16,** 460-467.

8. Blobel, C.P. (2005) ADAMs: key components in EGFR signaling and development. *Nat. Rev. Mol. Cell Biol.* **6,** 32-43

9. Shi, F., Telesco, S.E., Liu, Y., Radhakrishnan, R., and Lemmon, M.A. (2010) ErbB3/HER3 intracellular domain is competent to bind ATP and catalyze autophosphorylation. *Proc. Natl. Acad. Sci. USA* **107,** 7692-7697

10. Zahnow, C.A. (2006) ErbB receptors and their ligands in the breast. *Expert reviews in molecular medicine* **8,** 1-21

11. Yarden, Y., and Silwkowski, M.X. (2001) Untangling the ErbB signaling network. *Nat. Rev. Mol. Cell Biol.* **2,** 127-137

12. Schlessinger J. (2000) Cell signaling by receptor tyrosine kinases. *Cell* **103,** 211-225

13. Monilola, A.O., Neve, R.M., Lane, H.A., and Hynes NE. (2000) The ErbB signaling

network: receptor heterodimerization in development and cancer. *EMBO J.* **19,** 3159-3167

14. Tzahar, E., Waterman, H., Chen, X., Levkowitz, G., Karunagaran, D., Lavi, S., Ratzin, B.J., and Yarden, Y. (1996) A hierarchical network of interreceptor interactions determines signal transduction by Neu differentiation factor/neuregulin and epidermal growth factor. *Mol. Cell Biol.* **6,** 5276-5287

15. Herbst, R.S., Heymach, J.V., and Lippman, S.M. (2008) Lung cancer. *New Engl. J. Med.* **359,** 1367-1380

16. Hatanpaa, K.J., Burma, S., Zhao, D., and Habib, A.A. (2010) Epidermal growth factor receptor in glioma: signal transduction, neuropathology, imaging, and radioresistance. *Neoplasia* **12,** 675-684

17. Bose, R., Kavuri, S.M., Searleman, A.C., Shen, W., Shen, D., Koboldt, D.C., Monsey, J., Goel, N., Aronson, A.B., Li, S., Ma, C., Ding, L., Mardis, E.R., and Ellis, M.J. (2013) Activating HER2 mutations in HER2 gene amplification negative breast cancer. *Cancer Discov.* **3,** 224-237

18. Jaiswal, B.S., Kljavin, N.M., Stawiski, E.W., Chan, E., Parikh, C., Durinck, S., Chaudhuri, S., Pujara, K., Guillory, J., Edgar, K.A., Janakiraman, V., Scholz, R-P., Bowman, K.K., Lorenzo, M., Li, H., Wu, J., Yuan, W., Peters, B.A., Kan, Z., Stinson, J., Mak, M., Modrusan, Z., Eigenbrot, C., Firestein, R., Stern, H.M., Rajalingam, K., Schaefer, G., Merchant, M.A., Sliwkowski, M.X., de Sauvage, F.J., and Seshagiri, S. (2013) Oncogenic ERBB3 mutations in human cancers. *Cancer Cell* **23,** 603-617

19. Kavuri, S.M., Jain, N., Galimi, F., Cottino, F., Leto, S.M., Migliardi, G., Searleman, A.C., Shen, W., Monsey, J., Trusolino, L., Jacobs, S.A., Bertotti, A., and Bose, R. (2015) HER2 activating mutations are targets for colorectal cancer treatment. *Cancer Discov.* **5,** 832-841

20. Alimandi, M., Romano, A., Curia, M.C., Muraro, R., Fedi, P., Aaronson, S.A., Di Fiore, P.P., and Kraus, M.H. (1995). Cooperative signaling of ErbB3 and ErbB3 in neoplastic transformation and human mammary carcinomas. *Oncogene* **10,**1813-1821

21. Noto, A., Vitis, C.D., Roscilli, G., Fattore, L., Malpicci, D., Marra, E., Luberto, L., D'Andrilli, A., Coluccia, P., Giovagnoli, M.R., Normanno, N., Ruco, L., Aurisicchio, L., Mancini, R., and Ciliberto, G. (2013) Combination therapy with anti-ErbB3 monoclonal antibodies and EGFR TKIs potently inhibits non-small cell lung cancer. *Oncotarget* **4,** 1253-1265

22. Sergina, N.V., Rausch, M., Wang, D., Blair, J., Hann, B., Shokat, K.M., and Moasser M.M. (2007) Escape from HER family tyrosine kinase inhibitor therapy by the kinase-inactive HER3. *Nature* **445,** 437-441

23. Baselga J, Swain S. (2009) Novel anticancer targets: revisiting ERBB2 and discovering ERBB3. *Nat. Rev. Cancer* **9,** 463-475

24. Xia, W., Petricoin, E.F. III., Zhao, S., Liu, L., Osada, T., Cheng, Q., Wulfkuhle, J.D., Gwin, W.R., Yang, X., and Gallagher R.I. (2013) An heregulin–EGFR–HER3 autocrine signaling axiscan mediate acquired lapatinib resistance in HER2+ breast cancer models. *Breast Cancer Res*. **15,** 1-15

25. Ritter, C.A., Perez-Torres, M., Rinehart, C., Guix, M., Dugger, T., Engelman, J.A., and Arteaga, C.L. (2007) Human breast cancer cells selected for resistance to trastuzumab in vivo overexpress epidermal growth factor receptor and ErbB ligands and remain dependent on the ErbB receptor network. *Clin. Cancer Res.* **13,** 4909-4919

26. Schoeberl, B., Pace, E.A., Fitzgerald, J.B., Harms, B.D., Xu, L., Nie, L., Linggi, B., Kalra, A., Paragas, V., and Bukhalid R. (2009) Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor–PI3K axis. *Sci. Signal.* **2,** 1-15

27. Nielsen, T.O., Poulsen, S.S., Journe, F., Ghanem, G., and Sorensen, B.S. (2014) HER4 and its cytoplasmic isoforms are associated with progression-free survival of malignant melanoma. *Melanoma Res.* **24,** 88-91

28. Awada, A., Bozovic-Spasojevic, I., and Chow, L. (2012) New therapies in HER2-positive breast cancer: a major step towards a cure of the disease? *Cancer Treat Rev.* **38,** 494-504

29. Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med.* **344,** 783-792

30. Hynes, N.E., and Lane, H.A. (2005) ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer.* **5,** 341-354

31. Burgess, A.W., Cho, H-S., Eigenbrot, C., Ferguson, K.M., Garrett, T.P.J., Leahy, D.J., Lemmon, M.A., Sliwkowski, M.X., Ward, C.W., and Yokoyama, S. (2003) An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. *Mol Cell.* **12,** 541-552

32. Citri A. (2003) The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp Cell Res.* **284,** 54-65

33. Endres, N.F., Das, R., Smith, A.W., Arkhipov, A., Kovacs, E., Huang, Y., Pelton, J.G., Shan, Y., Shaw, D.E., Wemmer, D.E., Groves, J.T., and Kuriyan, J. (2013) Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell* **152,** 543-556

34. Macdonald-Obermann JL1, Piwnica-Worms D, Pike LJ. (2012) Mechanics of EGF receptor/ErbB2 kinase activation revealed by luciferase fragment complementation imaging. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 137-142

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

36. Monsey, J., Shen, W., Schlesinger, P., and Bose, R. (2010) Her4 and Her2/neu tyrosine kinase domains dimerize and activate in a reconstituted in vitro system. *J. Biol. Chem.* **285,** 7035-7044

37. Arteaga, C.L., and Engelman, J.A. (2014) ERBB Receptors: From Oncogene Discovery to Basic Science to Mechanism-Based Cancer Therapeutics. *Cancer Cell* **25,** 282-303

38. Shah, D.R., Shah, R.R., and Morganroth, J. (2013) Tyrosine Kinase Inhibitors: Their On-Target Toxicities as Potential Indicators of Efficacy. *Drug Safety* **36,** 413-426

39. Zhang, X., Pickin, K.A., Bose, R., Jura, N., Cole, P.A., and Kuriyan, J. (2007) Inhibition of the EGFR receptor by binding of Mig6 to an activating kinase domain interface. *Nature* **450,** 741-744

40. Schlessinger, J., and Lemmon, M.A. (2003) SH2 and PTB domains in tyrosine kinase signaling. *Sci. STKE* **191,** RE12

41. Lucci, M.A., Orlandi, R., Triulzi, T., Tagliabue, E., Balsari, A., and Villa-Moruzzi, E. (2010) Expression profile of tyrosine phosphatases in HER2 breast cancer cells and tumors. *Cell Oncol.* **32,** 361-372

42. Tiganis, T. (2002) Protein tyrosine phosphatases: dephosphorylating the epidermal growth factor receptor. *IUBMB* **53,** 3-14

43. Lee, N. Y., Hazlett, T. L., and Koland, J. G. (2006) Structure and dynamics of the epidermal growth factor receptor C-terminal phosphorylation domain. *Protein Sci.* **15,** 1142-1152

44. Koland, J. G. (2014) Coarse-grained molecular simulation of epidermal growth factor receptor protein tyrosine kinase multi-site self-phosphorylation. *PLoS Comput. Biol.* **10,** e1003435

45. Keppel TR, Sarpong K, Murray EM, Monsey J, Zhu J, Bose R. (2017) Biophysical Evidence for Intrinsic Disorder in the C-terminal Tails of the Epidermal Growth Factor Receptor (EGFR) and HER3 Receptor Tyrosine Kinases. *J. Biol. Chem.* **292,** 597-610.

46. Jura, N., Endres, N.F., Engel, K., Deindl, S., Das, R., Lamers, M.H., Wemmer, D.E.,

Zhang, X., and Kuriyan, J. (2009) Mechanism for Activation of the EGF Receptor Catalytic Domain by the Juxtamembrane Segment. *Cell* **137,** 1293-1307

47. Kovacs, E., Das, R., Wang, Q., Collier, T. S., Cantor, A., Huang, Y., Wong, K., Mirza, A., Barros, T., Grob, P., Jura, N., Bose, R., and Kuriyan, J. (2015) Analysis of the role of the C-terminal tail in the regulation of the epidermal growth factor receptor. *Mol. Cell. Biol.* **35,** 3083-3102

48. Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18,** 756-764

49. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2008) Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37,** 215-246

50. Bae, S.-H., Dyson, H. J., and Wright, P. E. (2009) Prediction of the rotational tumbling time for proteins with disordered segments. *J. Am. Chem. Soc.* **131,** 6814-6821

51. Uversky, V. N., and Dunker, A. K. (2010) Understanding protein nonfolding. *Biochim. Biophys. Acta* **1804,** 1231-1264

52. Tantos, A., Han, K.-H., and Tompa, P. (2012) Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell. Endocrinol.* **348,** 457-465

53. Zhou, H.-X., Pang, X., and Lu, C. (2012) Rate constants and mechanisms of intrinsically disordered proteins binding to structured targets. *Phys. Chem. Chem. Phys.* **14,** 10466-10476

54. Uversky, V. N. (2013) The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini. *FEBS Lett.* **587,** 1891-1901

55. Habchi, J., Tompa, P., Longhi, S., and Uversky, V. N. (2014) Introducing protein intrinsic disorder. *Chem. Rev.* **114,** 6561-6588

56. Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N., and Dunker, A. K. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44,** 1989-2000

57. Iakoucheva, L. M., Brown, C. J., Lawson, J. D., Obradovic´, Z., and Dunker, A. K. (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **323,** 573-584

58. Kathiriya, J. J., Pathak, R. R., Clayman, E., Xue, B., Uversky, V. N., and Dave´, V. (2014) Presence and utility of intrinsically disordered regions in kinases. *Mol. Biosyst.* **10,** 2876-2888

59. Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem.* **269,** 2-12

60. Wyatt, P. J. (1993) Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* **272,** 1-40

61. Receveur-Bre´chot, V., Bourhis, J.-M., Uversky, V. N., Canard, B., and Longhi, S. (2006) Assessing protein disorder and induced folding. *Proteins* **62,** 24-45

62. Laue, T. M., Stafford, W. F., 3rd (1999) Modern applications of analytical ultracentrifugation. *Annu. Rev. Biophys. Biomol. Struct.* **28,** 75-100

63. Bernado´, P., and Svergun, D. I. (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8,** 151-167

64. Keppel, T. R., and Weis, D. D. (2013) Analysis of disordered proteins using a simple apparatus for millisecond quench-flow H/D exchange. *Anal. Chem.* **85,** 5161-5168

65. Keppel, T. R., and Weis, D. D. (2015) Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging. *J. Am. Soc. Mass. Spectrom.* **26,** 547-554

66. Neyroz, P., Zambelli, B., and Ciurli, S. (2006) Intrinsically disordered structure of *Bacillus pasteurii* UreG as revealed by steady-state and time-resolved fluorescence spectroscopy. *Biochemistry* **45,** 8918-8930

67. Uversky, V.N. (2009) Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes and macromolecular crowding. *Protein J.* **28,** 305-325

68. Sass L.E., Lanyi, C., Weninger, K., and Erie, D.A. (2010) Single-molecule FRET TACKLE reveals highly dynamic mismatched DNA-MutS complexes. *Biochemistry* **49** (14):3174-3190

69. Carrico, I.S. (2008) Chemoselective modification of proteins: hitting the target. *Chem. Soc. Rev.* **37,** 1423-1431

70. Sletten, E.M. and Bertozzi, C.R. (2009) Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48,** 6974-6998

71. Miseta, A., and Csutora, P. (2000) Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Mol. Biol. Evol.* **17,** 1232-1239

72. Rashidian, M., Dozier, J.K., and Distefano, M.D. (2013) Chemoenzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjug Chem.* **24,** 1277-1294

73. Mazmanian, S.K., Liu, G., Ton-That, H., and Schneewind, O. (1999*) Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285,**760-763

74. Ton-That, H., Liu, G., Mazmanian, S.K., Faull, K.F., and O. Schneewind, O. (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 12424-12429

75. Ton-That, H., Mazmanian, S.K., Faull, K.F., and O. Schneewind, O. (2000) Anchoring of surface proteins to the cell wall of *Staphylococcus aureus.* *J. Biol. Chem.* **275,** 9876-9881

76. Theile, C.S., Witte, M.D., Blom, A.E., Kundrat, L., Ploegh, H.L., and Guimaraes, C.P. (2013) Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8,** 1800-1807

77. Guimaraes, C.P., Witte, M.D., Theile, C.S., Bozkurt, G., Kundrat, L., Blom, A.E., and Ploegh, H.L. (2013) Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8,** 1787-1799

78. Lemmon, M. A., Schlessinger, J., and Ferguson, K. M. (2014) The EGFR family: not so prototypical receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* **6,** a020768

79. Lemmon, M. A., and Schlessinger, J. (2010) Cell signaling by receptor tyrosine kinases. *Cell* **141,** 1117-1134

80. Franklin, M. C., Carey, K. D., Vajdos, F. F., Leahy, D. J., de Vos, A. M., and Sliwkowski, M. X. (2004) Insights into ErbB signaling from the structure of the ErbB2-pertuzumab complex. *Cancer Cell* **5,** 317-328

81. Wood, E. R., Truesdale, A. T., McDonald, O. B., Yuan, D., Hassell, A., Dickerson, S. H., Ellis, B., Pennisi, C., Horne, E., Lackey, K., Alligood, K. J., Rusnak, D. W., Gilmer, T. M., and Shewchuk, L. (2004) A unique structure for epidermal growth factor receptor bound to GW572016 (lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **64,** 6652-6659

82. Dankort, D. L., Wang, Z., Blackmore, V., Moran, M. F., and Muller, W. J. (1997) Distinct tyrosine autophosphorylation sites negatively and positively modulate neu-mediated transformation. *Mol. Cell. Biol.* **17,** 5410-5425

83. Gajiwala, K. S. (2013) EGFR: tale of the C-terminal tail. *Protein Sci.* **22,** 995-999

84. Uversky, V. N., and Ptitsyn, O. B. (1994) “Partly folded” state, a new equilibrium state of

protein molecules: four-state guanidinium chloride-induced unfolding of -lactamase at low temperature. *Biochemistry* **33,** 2782-2791

85. Uversky, V. N., and Ptitsyn, O. B. (1996) Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-induced unfolding of carbonic anhydrase B at low temperature. *J. Mol. Biol.* **255,** 215-228

86. Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* **18,** 343-384

87. Liu, J., Faeder, J. R., and Camacho, C. J. (2009) Toward a quantitative theory of intrinsically disordered proteins and their function. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19819-19823

88. Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9,** Suppl. 1, S1

89. Xie, H., Vucetic, S., Iakoucheva, L. M., Oldfield, C. J., Dunker, A. K., Obradovic, Z., and Uversky, V. N. (2007) Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* **6,** 1917-1932

90. Redfield, C. (2004) Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. *Methods* **34,** 121-132

91. Ubersax, J. A., and Ferrell, J. E., Jr. (2007) Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol.* **8,** 530-541

92. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23,** 2947-2948

93. Whitmore, L., and Wallace, B. A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89,** 392-400

94. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32,** W668-W673

95. Compton, L. A., and Johnson, W. C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem.* **155,** 155-167

96. Manavalan, P., and Johnson, W. C., Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* **167,** 76-85

97. Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* **287,** 243-251

98. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287,** 252-260

99. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78,** 1606-1619

100. Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole, F. L., 2nd, Tsutakawa, S. E., Jenney, F. E., Jr., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S.-j., Scott, J. W., Dillard, B. D., Adams, M. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods* **6,** 606-612

101. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1—towards automated and web-supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40,** s223-s228

102. Svergun, D., Barberato, C., and Koch, M. H. J. (1995) CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28,** 768-773

103. Romero, P., Obradovic, Z., and Dunker, A. K. (1997) Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform.* **8,** 110-124

104. Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins* **42,** 38-48

105. Li, X., Romero, P., Rani, M., Dunker, A. K., and Obradovic, Z. (1999) Predicting protein disorder for N-, C-, and internal regions. *Genome Inform.* **10,** 30-40

106. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., So¨ding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539

107. Sapir, T., Eisenstein, M., Burgess, H. A., Horesh, D., Cahana, A., Aoki, J., Hattori, M., Arai, H., Inoue, K., and Reiner, O. (1999) Analysis of lissencephaly-causing LIS1 mutations. *Eur. J. Biochem.* **266,** 1011-1020

108. Cantor, C. R., and Schimmel, P. R. (1980) Biophysical Chemistry: Part II: Techniques for the Study of Biological Structure and Function, W. H. Freeman, Co, Oxford

109. Tcherkasskaya, O., and Uversky, V. N. (2001) Denatured collapsed states in protein folding: example of apomyoglobin. *Proteins* **44,** 244-254

110. Manon, F., and Ebel, C. (2010) Analytical ultracentrifugation, a useful tool to probe intrinsically disordered proteins, in Instrumental Analysis of Intrinsically Disordered Proteins, pp. 431-449, John Wiley & Sons, Inc., Hoboken, N. J.

111. Kikhney, A. G., and Svergun, D. I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589,** 2570-2577

112. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 8868-8873

113. Huang, Y., and Liu, Z. (2009) Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. *J. Mol. Biol.* **393,** 1143-1159

114. Lichty, J.J., Malecki, J.L., Agnew, H.D., Michelson-Horowitz, D.J., and Tan, S. (2006) Comparison of affinity tags for protein purification. *Protein Expr. Purif.* **41,** 98-105

115. Nilsson, J., Stahl, S., Lundeberg, J., Uhlen, M., and Nygren, P.A. (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Expr. Purif.* **11,** 1-16

116. Arnau, J., Lauritzen, C., Petersen, G.E., and Pedersen, J. (2006) Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expr. Purif.* **48,** 1-13

117. Warden-Rothman, R., Caturegli, I., Popik, V., and Tsourkas, A. (2013) Sortase-tag expressed protein ligation: combining protein purification and site-specific bioconjugation into a single step. *Anal. Chem.* **85,** 11090-11097

118. Kobashigawa, Y., Kumeta, H., Ogura, K., and Inagaki, F. (2009) Attachment of an NMR invisible solubility enhancement tag using a sortase-mediated protein ligation method. *J. Biomol. NMR* **43,** 145-150

119. Yamamura, Y., Hirakawa, H., Yamaguchi, S., and Nagamune, T. (2011) Enhancement of sortase A-mediated protein ligation by inducing a b-hairpin structure around the ligation site. *Chem, Commun* **47,** 4742-4744

120. Williamson, D.J., Fascione, M.A., Webb, M.E., and Turnbull, W.B. (2012) Efficient N-terminal labeling of proteins by use of sortase. *Angew. Chem. Int. Ed.* **51,** 9377-9380

121. Chen, I., Dorr, B.M., and Liu, D.R. (2011) A general strategy for the evolution of bond forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 11399-11404

122. Suree, N., Yi, S.W., Thieu, W., Marohn, M., Damoiseaux, R., Chan, A., Jung, M.E., and Clubb, R.T. (2009) Discovery and structure-activity relationship analysis of *Staphylococcus aureus* sortase A inhibitors. *Bioorg. Med. Chem.* **17,** 7174-7185

123. Petrache, A.I., Machin, D.C., Williamson, D.J., Webb, M.E., and Beales, P.A. (2016) Sortase-mediated labeling of lipid nanodiscs for cellular tracing. *Mol. Biosyst.* **12,** 1760-1763

124. Denisov, I.G., Grinkova, Y.V., Lazarides, A.A., and Sligar, S.G. (2004) Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size. *J. Am. Chem. Soc.* **126,** 3477-3487

125. Uttamapinant, C., Sanchez, M.I., Liu, D.S., Yao, J.Z., White, K.A., Grecian, S., Clark, S., Gee, K.R., and Ting, A.Y. (2013) Site-specific protein labeling using PRIME and chelation-assisted click chemistry. *Nat. Protoc.* **8,** 1620-1634

126. Presolski, S.I., Hong, V.P., and Finn, M.G. (2011) Copper-catalyzed azide-alkyne click chemistry for bioconjugation. *Curr. Protoc. Chem. Biol.* **3,** 153-162

127. Berney, C., and Danuser, G. (2003) FRET or No FRET; A quantitative approach. *Biophysical Journal* **84,** 3992-4010

128. Wani, R., Nagata, A., and Murray, B.W. (2014) Protein redox chemistry: post-translational cysteine modifications that regulate signal transduction and drug pharmacology. *Front. Pharmacol.* **5,** 1-8

129. Bah, A., Vernon, R.M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L.E., and Forman-Kay, J.D. (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **519,** 106-109

130. Littlefield, P., Liu, L., Mysore, V., Shan, Y., Shaw, D.E., and Jura, N. (2014) Structural analysis of the EGFR/HER3 heterodimer reveals the molecular basis for activating HER3 mutations. *Science Sig.* **7**(354):ra114

131. Collier, T.S., Diraviyam, K., Monsey, J., Shen, W., Sept, D., and Bose, R. (2013) Carboxyl group footprinting mass spectrometry and molecular dynamics identify key interactions in the HER2-HER3 receptor tyrosine kinase interface. *J. Biol. Chem.* **288,** 25254-25264