Analytical Development and Characterization of the Metal-binding *cla*MP Tag for Generation of Inline Bioconjugates to Enable Targeted Metal Delivery: Evaluation of Tag Placement, Metal Insertion and Retention, and Stability of Ni(II)-*cla*MP-Tagged Protein Variants

By

Copyright 2014

Brittney J. Mills

Submitted to the graduate degree program in Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Chairperson/Jennifer S. Laurence, Ph.D.

_____

Heather Desaire, Ph.D.

_____

Mario Rivera, Ph.D.

_____

Timothy A. Jackson, Ph.D.

_____

Teruna Siahaan, Ph.D.

Date Defended: May 12, 2014

The Dissertation Committee for Brittney J. Mills certifies that this is the
approved version of the following dissertation:

Analytical Development and Characterization of the Metal-binding *cla*MP Tag for Generation of
Inline Bioconjugates to Enable Targeted Metal Delivery: Evaluation of Tag Placement, Metal
Insertion and Retention, and Stability of Ni(II)-*cla*MP-Tagged Protein Variants

_____
Chairperson/Jennifer S. Laurence, Ph.D.

Date Approved: May 12, 2014

THE CONTENTS OF THIS DISSERTATION ARE CONFIDENTIAL AND ARE NOT TO BE PUBLICLY DISCLOSED. THIS DISSERTATION WILL BE EMBARGOED FOR AT LEAST TWO YEARS.

CONFLICT OF INTEREST STATEMENT
JENNIFER S. LAURENCE IS CO-OWNER OF ECHOGEN INC., A LIMITED LIABILITY COMPANY THAT HAS LICENSED THE PATENT-PROTECTED METAL ABSTRACTION PEPTIDE (MAP) TECHNOLOGY, ON WHICH THE *cla*MP TAG IS BASED, FROM THE UNIVERSITY OF KANSAS.

ABSTRACT

Metal-based protein conjugates have become the primary means of achieving site-specific delivery of metal ions for use in clinical applications.   Small, organic chelators are the primary entities used for metal delivery, but they require chemical conjugation to be attached to the protein surface. Conjugation using this method requires extensive optimization and generates a highly heterogeneous product. Therefore, metal-binding peptide tags capable of being engineered into a predetermined position, inline within a protein sequence are highly advantageous. The metal-binding *cla*MP Tag is capable of being inserted inline within a protein sequence to generate a "linker-less" bioconjugate. The *cla*MP Tag module is based upon the metal abstraction peptide (MAP) sequence, NCC (Asn-Cys-Cys), which has been shown to bind transition metals extremely tightly and to retain metal binding in the presence of longer polypeptide sequences, thus allowing for use of this tag as an inline metal-binding agent for the generation of bioconjugates. The primary focus of this work was to characterize the effects of *cla*MP Tag addition on various protein systems to establish its applicability as a metal-delivery platform. Ni(II) was inserted into the tag in these studies because the Ni-*cla*MP complex had previously been extensively characterized using small peptides and it was determined to be highly stable and to have a number of advantageous properties to enable quantitative analysis. The properties of this unique metal-peptide module suggested that it would be amenable to use in a larger protein construct, where retention of the metal would be enabling to targeted delivery. Herein, the unique charge, spectroscopic, and catalytic properties of the Ni-*cla*MP complex were used as the basis for establishing methods to evaluate the structure and stability of *cla*MP Tag-based inline metal-protein conjugates.

Two model protein systems were chosen for analytical development that are different in size, composition, and structure to assess the applicability of the *cla*MP Tag within unrelated proteins and using different methods of purification. The placement of the *cla*MP Tag within the protein sequence was varied to determine if tag position differentially affects metal insertion. Because the *cla*MP Tag itself contains two cysteine residues, which may interfere with native disulfide bond formation, the first protein investigated for expression, correct folding and higher order structure, and stability was epidermal growth factor (EGF). EGF is a small, difficult to express, disulfide-containing protein, which is capable of being produced recombinantly in *E. coli* when fused to thioredoxin, a thiol-modulating protein. The study described herein demonstrates addition of the *cla*MP Tag to either terminus of EGF does not negatively affect the expression or function of the protein. NMR analysis was used to confirm that the disulfide bonds are formed correctly and that the tertiary structure of the protein is well maintained in the metal-*cla*MP Tag conjugate. The unique absorption spectrum and characteristic 2^- charge of the properly formed metal-*cla*MP structure were utilized to quantitatively assess both metal incorporation and stability of the conjugate. The data show that addition of the cysteine-containing *cla*MP Tag to a thiol- and disulfide-containing protein does not negatively impact the critical attributes or protein properties, making it amenable to use as an inline metal carrier.

A polyhistidine tag was engineered into EGF to enable facile purification using immobilized metal affinity chromatography (IMAC), but this approach leads to concomitant uptake of Ni(II) into the *cla*MP Tag. To evaluate the ability to purify the tagged protein prior to metal insertion as well as different approaches to metal incorporation, maltose-binding protein (MBP) was examined. This second protein system allows for decoupling of the metal insertion and purification steps because amylose resin, which does not rely on metal in any way, enables

affinity purification of MBP. This allowed for metal insertion efficiencies to be determined

before and after purification. It also demonstrated that the *cla*MP Tag may be successfully

applied to larger protein systems, because MBP is approximately six-times larger than EFG.

Finally, because of its larger size, retention of metal in the *cla*MP Tag within the MBP variant

could be evaluated more easily in the presence of a large excess of a potentially competitive

high-affinity chelator. This study was used to establish that the unique chemistry that the *cla*MP

tag undergoes to bind metal generates a highly stable product.

ACKNOWLEDGEMENTS

The research in this dissertation was supported by funds from the Wallace H. Coulter Foundation (CTRA) and KU Cancer Center Pilot Award. This work was also supported by the NIGMS Biotechnology Predoctoral Training Grant (T32 GM-08359) and the Chemistry Department at the University of Kansas.

This work would not have been possible without assistance from others within the KU community. I would like to thank past members of the Laurence Lab and other collaborators who have contributed to this work. Early experiments that contributed to the formation of this project were completed by Drs. A. Andrew Vartia, Andria Skinner, Kevin Han, and Mary Krause, under the direction of Dr. Jennifer Laurence. I would also like to thank Dr. Qingxin Mu who completed the cell culture studies presented in this work. This work made use of the KU Biomolecular NMR Core Facility, and I would like to thank Dr. Asokan Anabanadam for his guidance and training on the NMR spectrometers. I would also like to thank the KU Microscopy and Imaging Laboratory for use of their equipment and Heather Shinogle for her assistance. This work also made use of the KU Mass Spectrometry Lab, and I would like to thank their staff for collecting and analyzing the mass spectrometry data. Finally, I would like to thank Drs. C. Russ Middaugh and David Volkin at the KU Macromolecule and Vaccine Stabilization Center for allowing me to use their equipment and Ozan Kumru for assistance in running the experiments and data analysis. I would also like to thank current and former members of the Laurence lab for their encouragement and guidance during my graduate career. I would especially like to thank Dr. Mary Krause for her support and training during my early years as a graduate student. I would also like to thank the members of my committee (Drs. Jennifer Laurence, Heather Desaire, Mario Rivera, Timothy Jackson, and Teruna Siahaan) for serving on my committee and

offering their constructive feedback to improve my research. I would also like to thank Hospira, Inc. and Dr. Melissa Perkins for providing me the opportunity to complete a three-month internship. During this time, I learned a great deal about pharmaceutical product development and gained a great mentor who has helped to shape my career as a scientist.

I would also like to thank my husband, James, and my parents and sisters. Their support and encouragement during my graduate career were very important factors in my success. Even though there were many times I simply wanted to unwind, James always kept me thinking about research and was a great resource for helping me to think about my research in a different way and find solutions to any problems. Also having gone through graduate school, James showed me the work ethic that was required to succeed and understood the demands of the job, even though it sometimes meant coming back up to MRB with me late at night. My parents and sisters have also been constant sources of encouragement and ears to listen, although they barely understood what my research entailed. Having my husband, family, and friends by my side helped push me through to get my degree.

Finally, I would like to thank my advisor, Jen, who not only directed my research, but also is great mentor. I really appreciate that she took me on, even though I was a student outside of her department. She provided support, but also pushed me to find my own solutions, which helped me grow as a scientist. Her feedback helped me to improve my writing skills and critically evaluate my own work. Jen's enthusiasm for science makes it exciting to come in to lab everyday, and I know that she will always be there if I need someone to talk to or advice. Jen has significantly contributed to my growth as a scientist, and I am very appreciative of all that she has done to support me on both the professional and personal level.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

In alphabetical order:

ADC	Antibody drug conjugate
AEC	Anion exchange chromatography
ANS	8-Anilino-1-naphthalenesulfonic acid
βME	β-mercaptoethanol
CD	Circular dichroism
CV	Column volumes
Da	Daltons
DAR	Drug-to-antibody ratio
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
E. coli	*Escherichia coli*
Epo	Erythropoietin
FcRn	Neonatal Fc receptor
HC	Heavy chain
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
IMAC	Immobilized metal affinity chromatography
IFN	Interferon
IL-5	Interleukin 5

IPTG	Isopropyl β-D-1-thiogalactopyranoside
KPi	Potassium phosphate
LB	Luria broth
LBT	Lanthanide-binding peptide tags
LC	Light chain
LIC	Ligation independent cloning
MAP	Metal abstraction peptide
MBP	Maltose binding protein
MRI	Magnetic resonance imaging
MWCO	Molecular weight cut-off
NaPi	Sodium phosphate
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
OD	Optical density
PCR	Poylmerase chain reaction
PK	Pharmacokinetics
PEG	Polyethylene glycol
PRL-1	Phosphatase of regenerating liver
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SLS	Static light scatter
SOD	Superoxide dismutase
T-DM1	Trastuzumab DM1

T_m Melting temperature

CHAPTER 1.
EFFECTS OF LOCALIZED INTERACTIONS AND SURFACE
PROPERTIES ON STABILITY OF PROTEIN-BASED THERAPEUTICS

1.1 INTRODUCTION

Antibody drug conjugates (ADCs) are the primary means of achieving site-specific delivery of cytotoxic agents, but the impact of conjugation on antibody structure and stability may limit the molecule from making it to market. ADCs utilize conjugation through surface-accessible lysine residues,[1-2] which alters the surface charge of the molecule, or selective reduction or amino acid mutation to conjugate through lysine residues.[3-5] Conjugation using either strategy can have adverse effects on the physical stability of the molecule. Altering the surface charge through lysine conjugation directly affects the stability of the molecule.[1,6] Amino acid mutation to achieve site-specific conjugation also causes changes in surface properties, but its effects on physical stability may be more drastic due to local structural changes that perturb the stability of the entire molecule.[3-4] Because high-resolution data are difficult to obtain due to the large size of antibodies, mutations within smaller, therapeutically relevant proteins may be useful in assessing the effect of conjugation on the physical stability of the molecule. These mutations are akin to large modifications required in ADCs because they both alter the surface properties of the molecule, which in turn affects physical stability. Although the mechanism of affecting stability may not be the same in all cases, these smaller proteins can serve as model systems for assessing the impact of certain mutations on protein structure and stability within ADCs.

Protein structure is modulated by different types of interactions among residues, with some interactions exhibiting greater contribution to protein structure than others. Secondary structural elements, such as α-helices and β-sheets, are formed by a diverse network of hydrogen

bonding interactions among atoms located within the polypeptide backbone. The hydrogen bonds are essential in maintaining secondary structure and a lack of hydrogen bonds is associated with, or may lead to, disordered structures.[7] Formation of the three-dimensional fold of the protein requires additional interactions among amino acid side chains. Hydrophobic association is critical to achieving the condensed state of a folded protein, but ionic interactions between charged residues on the surface of the molecule are important contributing factors in protecting the core to stabilize the protein fold. Covalent bond formation between Cys side chains in the form of disulfide bonds cross-bridge the structure, which also can contribute significantly to the overall protein fold and its stability. The core structure, packing interactions and surface composition impact the physical and structural stability of proteins. The surface properties determine solvent accessibility to key structural elements, which contribute to the overall three-dimensional conformation and aggregation propensity of the molecule.[7] Secondary structural features and disulfide bonds are the main structural elements that are monitored when assessing changes in protein structure. These features and changes in them are easily observed using high-throughput approaches, which enable rapid comparison among analogs. When coupled to thermal titration, this approach has been particularly useful because, as the largely hydrophobic tertiary contacts are broken, individual secondary structure elements are disrupted.[7] Amino acid mutation can be used to modulate these elements to investigate contributions to stability and in protein engineering to generate a more stable structure. The mutation may cause only local structural perturbations that lead to structural stabilization, or it may cause far-reaching alterations that alter the global conformation of the molecule. Structural stabilization may be caused by the addition of stabilizing interactions, such as salt bridges, hydrogen bonds, or

hydrophobic associations; but amino acid mutation may also be destabilizing to protein structure due to elimination of a disulfide bond or glycosylation site.

Addition of different entities to the protein may also be required to develop a protein therapeutic that exhibits the desirable pharmacological properties or therapeutic use, such as a drug molecule in ADC development. A larger molecular weight species such as a polyethylene glycol (PEG) or lipid molecule can be attached to the protein to increase the molecular weight of the entire molecule, thus increasing the circulating half-life and improving the pharmacokinetics of low molecular weight proteins that are too quickly cleared from the bloodstream to complete their desired function.[8-12] Addition of a PEG molecule is also desirable as it can shield regions of hydrophobic surface area[10] or affect the ability of molecules to come in contact with each other, both of which increase the physical stability. Glycans have been shown to assist in stabilizing protein structure[13-14] by interacting with the protein directly through hydrogen bonds with the polypeptide backbone or through interactions with surface accessible amino acids,[15-17] although the extent of stabilization is highly protein dependent. Glycan and PEG molecules can also modulate the physical stability of the protein by shielding hydrophobic regions of the surface that may otherwise be prone to self-association with other protein molecules.[10, 18-20]

Even though amino acid mutation and addition of larger molecules to a protein may seem only to have effects on nearby residues, alterations in the local structure of the protein can modify the global structure and physical stability of the entire molecule. These changes can have detrimental effects on protein function, which is undesirable for use of a protein as a therapeutic entity. As such, assessing the impact of local alterations on protein structure is critical during the development of protein therapeutics. This review provides an overview of how the structural

stability of current protein therapeutics is affected by amino acid mutation and addition of larger molecules to aid in the design of ADCs with desirable structural stability.

1.1.1 Insulin

Insulin is a small hormone whose main function is to regulate blood glucose levels. As a therapeutic entity, it is primarily used for the treatment of diabetes mellitus, which is caused by decreased signaling that can lead to a lack of insulin secretion. Human insulin is a small fifty-one amino acid globular protein that is composed of an A-chain (amino acids A1-21) and B-chain (amino acids B1-30). The secondary and tertiary structure of the insulin molecule is referred to as the insulin fold and is shared by members of the insulin fold super-family. This insulin fold is composed of three α-helices, two located in the A-chain (A2-A8, A12-A19) and one located in the B-chain (B9-B19), and a hydrophobic core composed of nonpolar amino acids. Insulin is folded as a precursor protein and then cleaved to achieve the two-chain secreted form. Apart from the typical insulin fold, the two chains are also connected through interchain disulfide bonds between residues A7-B7 and A20-B19. An intrachain disulfide bond is present between residues A6-A11, as well. In addition to the three disulfide bonds, the conformation of the B-chain also influences the structure and bioactivity of the insulin molecule. Insulin is known to have two structures characterized by two distinct B-chain conformations that determine the bioactivity of the molecule: the non-active R-state and the bioactive T-state. The main difference between the two states is the conformation of the B1-B8 segment. In the non-active R-state, the B1-B8 segment combines with the B9-B19 segment so the entire region from B1-B19 exists in α-helical form. The B1-B8 segment is more flexible in the active T-state as it adopts an extended conformation, which allows for interaction of this region with the receptor.

Insulin is active as a monomer, but many therapeutic versions of insulin are formulated as hexameric species to improve in vivo stability and generate a slower release profile. The association state of the molecule is dependent on a number of factors such as temperature, ionic strength, protein concentration, and the presence of molecular stabilizing agents.[21-24] Therefore, in order to evaluate the structure and solution dynamics of the monomeric form of insulin, two different insulin mutants have been generated that contain mutations at the dimer and hexamer interface: DKP insulin and 4E insulin. The first mutant analyzed, DKP insulin, contains three mutations to destabilize the interface responsible for dimer and hexamer formation: HisB10Asp, ProB28Lys, and LysB29Pro (Figure 1.1).[25-27] DKP insulin exhibits similar secondary structure to native insulin as determined by far-UV CD[26] and NMR analysis.[25, 28-29] Formation of the dimeric and hexameric forms of insulin occurs through coordination of Zn^{2+} ions by the HisB10 residues so mutation of this residue eliminates the coordination site,[30] destabilizing hexamer formation. The dimer interface is also destabilized by Pro28Lys mutation because interaction of this residue with TyrB26 on the opposite molecule contributes to dimer stabilization, and LysB29 participates in a network of ionic contacts to stabilize the hexamer interface.[30] Computational modeling also shows that ProB28 has the potential to make nonpolar contacts with the backbone of GlyB23 and ArgB22.[31] Because the combined effect of these mutations allow for reduced self-association propensity and these mutations do not alter the global structure of the insulin molecule, DKP insulin can be used to determine the effect of single amino acid substitutions on the local and global structure of monomeric insulin.

In 4E insulin, PheB1, HisB10, TyrB16, and ThrB27 are mutated to glutamic acid and ThrB30 is removed to yield a mutant that is predominantly monomeric at millimolar concentrations.[32] The near-UV CD and NMR spectra of this mutant exhibited similar spectral
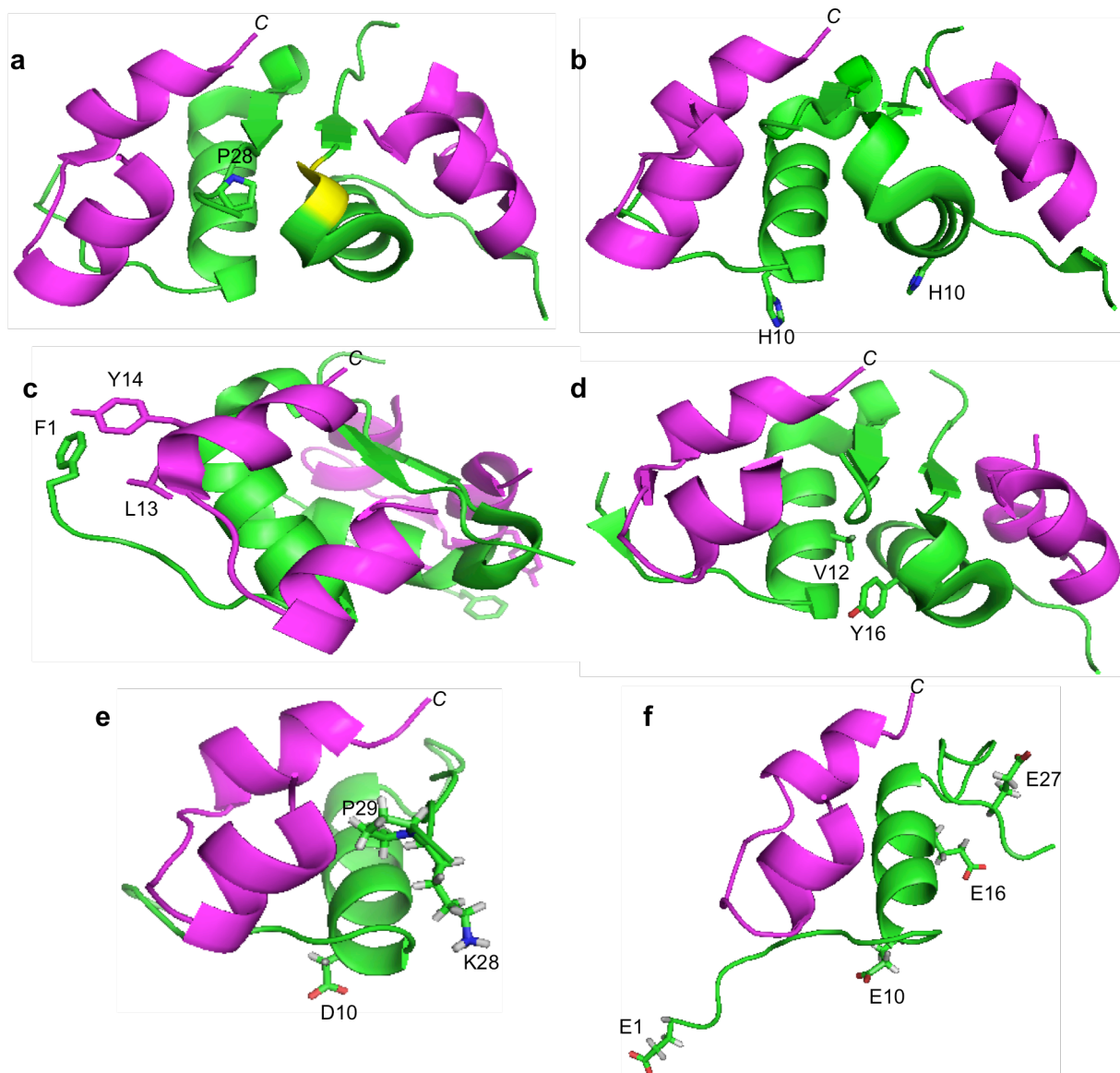
Figure 1.1 Mutations within dimer/hexamer interface of insulin allow for stabilization of monomeric form. a) P28 stabilizes dimer formation through an interaction with the peptide backbone of residues Arg22 and Gly23 (yellow). b) His10 from each monomer contributes to chelating the metal ion required for hexamer formation. c) F1 makes contacts with the side chains of L13 and Y14. d) Y16 also stabilizes the dimer interface by interacting with Val12. e) DKP insulin contains mutations at H10, P28, and K29, which in the native molecule lead to self-association. f) 4E insulin contains four different mutations at F1, H10, Y16, and T27 that are contact interfaces for self-association within the native molecule. A-strand shown in magenta and B-strand shown in green, and the C-terminus of the A-strand labeled to aid in visualization. PDB files 4EWZ, 1HUI, 2JMN were used to generate this figure.

features as native insulin,[32-33] confirming that the protein is mainly well-ordered, except for the B-chain termini.[32] Mutation of TyrB16 to either Glu[34] or His[33] results in the formation of a

mutant that maintains native secondary and tertiary structure as determined by CD and NMR, respectively. Substitution of Tyr with His results in similar denaturation stability as native insulin, yet this mutation is commonly used because it decreases self-association and allows for generation of a monomeric product at millimolar concentrations.[33] Within the insulin hexamer, PheB1 contacts the side chains of LeuA13 and TyrA14 within the same molecule,[35] and ThrB27 side chain hydrogen bonds with the AsnA18 amide on the adjacent hexamer.[30] Computational modeling indicates that TyrB16 has the potential to make van der Waals interaction with ValB12 or GlyB8,[31] further stabilizing the interface. Therefore, PheB1Glu, TyrB10His/Glu and ThrB27Glu mutation eliminates these interactions, which destabilizes the self-association interface.

1.1.2 Erythropoietin

Erythropoietin (Epo) is a glycoprotein hormone and cytokine that regulates red blood cell production. It is used to treat anemia that results predominantly from chronic kidney failure, Crohn's disease, or chemotherapy.[36] This glycoprotein consists of 166 amino acids and four carbohydrate motifs, three N-linked (Asn24, Asn38, Asn83) and one O-linked (Ser126), that account for nearly 40% of the molecule by weight. The primary role of the carbohydrate chains is to stabilize protein structure to limit self-association, but they also have been shown to affect the *in vivo* activity of the protein.[18] The solution structure of Epo shows that it contains the left-handed four-helix bundle commonly observed in cytokines.[37] The four helices are denoted as Helix A, Helix B, Helix C, and Helix D and exhibit an up-up-down-down direction.[37] The helical components of Epo are very important in maintaining the global conformation and bioactivity,[38] whereas modifications in the regions between the helices, known as interconnecting loops, do not impact the structure of the protein as drastically.[39-42] Three

different loops of varying sizes are present in Epo: the AB loop consists of 36 amino acids, two N-glycosylation sites, and a disulfide bond between Cys29-Cys33, the BC loop is smaller in size at only five residues in length, and it contains the third N-glycosylation site, and the CD loop contains 23 amino acids and the O-glycosylation site. An additional disulfide bond is present between Cys7-Cys161 that is responsible for maintaining the global structure as it connects Helix A and Helix D. The implications of modifying these structural elements on molecule stability will be discussed later in this review.

1.1.3 Interferon

Interferons are glycoprotein cytokines that are produced by the body in response to the introduction of pathogens such as bacteria or viruses. The Type I interferons, IFN-α and IFN-β, exhibit antiviral, antiproliferative, and immunomodulatory functions, which has led to their development into protein therapeutics.[43-46] IFN-α-2 is primarily used for the treatment of different viral diseases[47] and IFN-β-1 is used in the treatment of multiple sclerosis,[19, 48-52] whereas both have been useful in the treatment of cancer. Similarly to Epo, IFN-α-2 and IFN-β-1 exhibit the four-helix bundle with up-up-down-down topology that is commonly found in cytokines.[15, 53-54] Both also include a fifth helix that is not involved in the four-helix bundle, but still contributes to the overall structure of the protein. Apart from containing a similar helical topology, the primary sequence of the two is quite different with only approximately 50% amino acid homology between the two.[55-56] This difference in amino acid composition leads to differences in the disulfide network and glycosylation pattern. IFN-α-2 is a 165 amino acid protein that contains two disulfide bonds between Cys1-Cys98 and Cys29-Cys138[57] and an O-glycosylation site at Thr106.[58] On the other hand, IFN-β-1 contains 166 amino acids, one disulfide bond between Cys31-Cys141, and a N-glycosylation site at N80.

Two IFN-β variants exist, IFN-β-1a and IFN-β-1b, with the primarily difference between the two variants reliant upon the expression host used to produce the two molecules. IFN-β-1a is the form found in the human body so it is glycosylated at Asn80 and as a therapeutic entity, must be produced in a host capable of performing glycosylation. The IFN-β-1b form is a recombinant version produced in an *E. coli*-based system, thus it is not glycosylated. It also lacks the N-terminal Met residue and has a Cys17Ser mutation to assist in obtaining the correctly folded form. Even with these two alterations, the structure of the two IFN-β variants are expected to be very similar because the mutated Cys residue is not involved in the formation of any disulfide bonds.[19, 59] On the other hand, the stability of the two variants differs due to the presence of the glycan group in the IFN-β-1a variant and its absence in IFN-β-1b.[60-61] The glycan shields an uncharged region located on the surface of the molecule and directly interacts with nearby residues to increase the solubility and structural stability of IFN-β-1a over the non-glycosylated variant, IFN-β-1b.[19, 61]

1.2 MODULATION OF STRUCTURAL ELEMENTS THROUGH MUTATION OR DISULFIDE BOND REARRANGEMENT

1.2.1 Mutations that Alter Hydrophobic Interactions

The likelihood of amino acid side chains to associate with water is a primary factor governing protein folding and physical stability. Nonpolar amino acids are preferentially excluded from the protein surface,[62-65] resulting in hydrophobic association among these residues to form the core of the molecule and drive the formation of the tertiary structure. Therefore, mutation of residues that contribute to the hydrophobic core would be expected to alter the structure and stability of the molecule. The degree of alteration, though, depends on the amount of energy contributed to maintaining the core of the molecule, which varies with the extent of

hydrophobic surface area and aromatic character. In 4E insulin, ValA3Gly and ThrA8His/Arg mutation results only in local structural alterations that have opposite effects on the hydrophobic content of the region.[66] ValA3Gly substitution results in complete loss of the α-helical character in the A2-A8 region (Figure 1.2), which leads to decreased hydrophobic surface area in this region due to elimination of the ValA3 side chain and shifting of the GluA4 side chain to encompass a region of the protein that was predominantly hydrophobic in nature in the native structure.[66] Mutation of ThrA8 to His or Arg, though, leads to tightening of the A2-A8 α-helix, and increased exposure of the hydrophobic residues IleA2 and ValA3,[66] without adversely affecting the global structure. The A3 mutations within DKP insulin also suggest that nonpolar
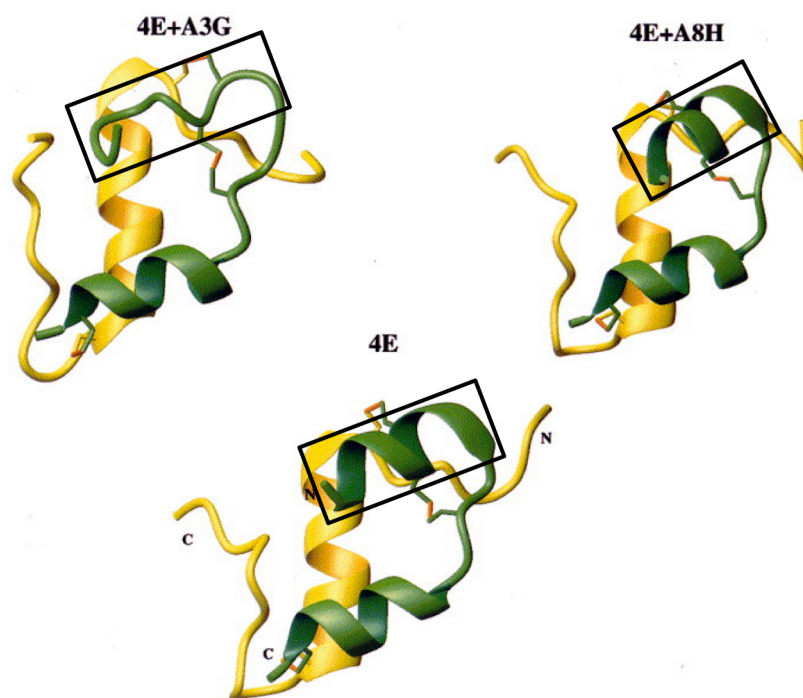
Figure 1.2 A3 and A8 mutations affect helicity of local segment. Mutation of ValA3 or ThrA8 only cause local structural alterations in the A2-A8 helix of the A-chain (boxed). Whereas ValA3Gly mutation eliminates helical feature, ThrA8His/Arg mutation leads to tightening of the helix. Reproduced with permission from Olsen, H. B., Ludvigsen, S., Kaarsholm, N. C. (1998) The relationship between insulin bioactivity and structure in the NH2-terminal A-chain helix. *J. Mol. Biol.* *284*, 477-488, © 1998 Elsevier, Philadelphia, PA.

interactions are essential for maintaining structural stability. Although ValA3Thr mutation did not affect protein conformation, it did alter the denaturation stability of the molecule due to the introduction of a polar moiety in a region of high hydrophobicity,[67] which alters the packing efficiency and solvent accessibility in the region.

Even if a residue seems to contribute minimally to protein structure and the hydrophobic core of the molecule, it may exert a large effect on protein structure by protecting buried, stabilizing interactions. Elimination of the protection provided by specific residues leads to greater solvent accessibility and disruption of the stabilizing interaction, which adversely affects protein stability.[68] Within insulin, PheB24 is important for maintaining the hydrophobic core of the molecule as it packs against other nonpolar residues such as ValB12, LeuB15, and the A20-B19 disulfide bond in both crystal structures and in solution.[21, 25, 28-30, 69-70] Substitution of this residue with non-aromatic L-amino acids, such as Ser or Leu, and aromatic residues, such as His, alters the solution structure of the molecule, which leads to decreases in bioactivity.[26, 71-72] In the PheB24His mutant, the His side chain is more tightly packed into the B24 hydrophobic pocket than Phe due to the ability to form additional hydrogen bonds.[72] The PheB24 to Gly mutation also drastically decreases the folding stability of the molecule. NMR illustrates that the core structure and α-helical content of the PheB24Gly mutant remain, but the mutant is more flexible, which leads to instability and unfolding of the B20-B30 region.[34, 73] Within Epo, mutation of Gly151, which contributes substantially to core formation by bringing the side chain of Lys152 into contact with Val63, Trp51, and Phe148,[37] results in global rearrangements within the hydrophobic core of the molecule.[74] In addition to maintaining the core of the molecule, hydrophobic residues are also important for forming interactions that stabilize different domains of Epo. Mutation of Helix B residues Leu69, Leu70, and Ala73, which make important

hydrophobic contacts with Ile39 and Val41 of the AB loop,[37] alters protein conformation. Leu69Ala resulted in only slight conformational changes, whereas Leu70Ala, but not Leu70Ile, substitution resulted in drastic conformational changes,[74] illustrating the necessity of a larger nonpolar side chain at this position. Substantial conformational changes are also observed upon substitution of Val63, Leu67, and Val74 with Ser.[74] Val63, Leu67, and Val 74 are buried within the helices and appear to make hydrophobic contacts holding the helices together thus confirming their importance.[37]

The presence of aromatic side chains, as opposed to solely hydrophobic side chains, may be necessary to form the most stable structure. Within insulin, removal of the aromatic side chain of PheB1 eliminates the packing interaction with LeuA13 and causes destabilization of the B1-B4 region.[33] Mutation of TyrB26 to Thr also leads to a decrease in folding stability due to the elimination of stabilizing aromatic interactions.[75] The importance of aromatic interactions in maintaining global structure is also observed within Epo and the interferons. The aromatic side chains of Phe142/Tyr145 and Phe148/Tyr156 interact with hydrophobic residues located on the AC and BC helix, to form the hydrophobic core of the Epo molecule (Figure 1.3).[37] Substitution of these residues leads to disruption of the core, leading to protein unfolding.[74] Mutation of Phe142 to Ile causes slight perturbations in global conformation, whereas mutation of Tyr145 to either Ile or Phe causes substantial conformational rearrangements.[74] Therefore, Phe142 must primarily make hydrophobic interactions so Ile will not alter the structure, but placement of Tyr at position 145 is necessary for both aromatic interactions, as determined by the Ile substitution, as well as a type of dipole interaction or stabilizing hydrogen bond because substitution with Phe at this site also resulted in conformational rearrangement. Phe148Val or Tyr156Ile/Phe mutation confirms that a larger hydrophobic entity is required at this site because Val/Ile substitution
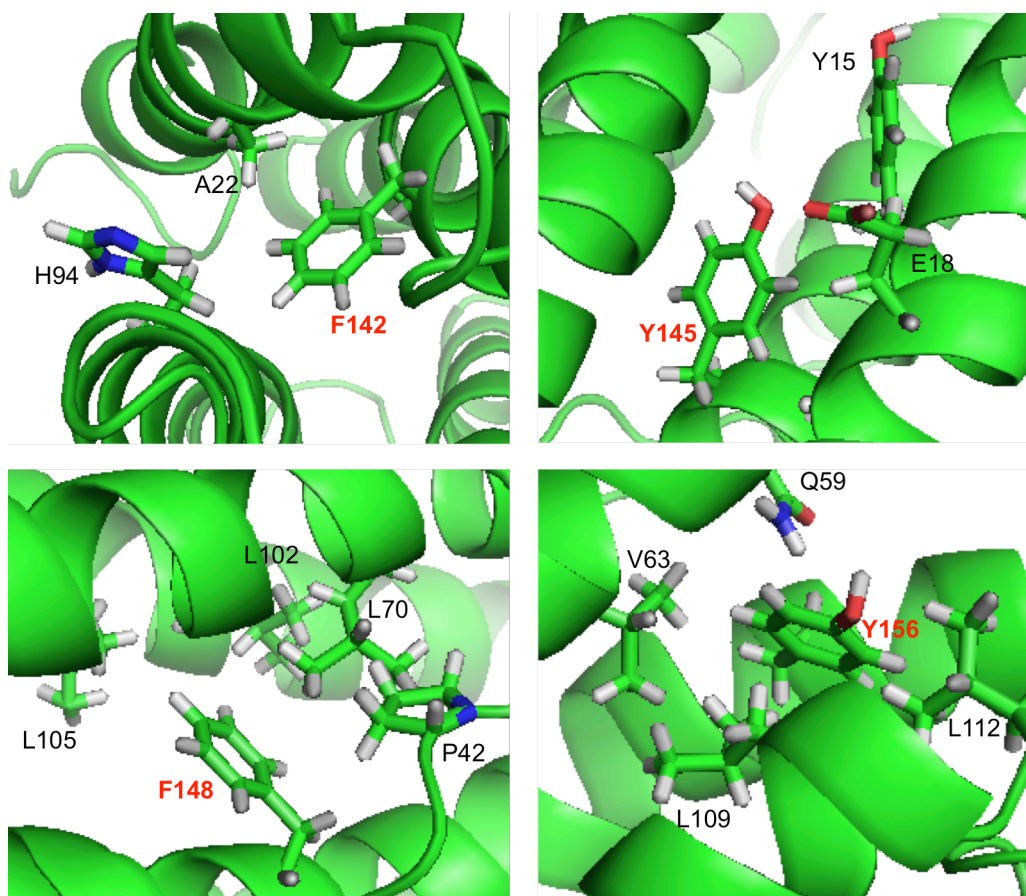
Figure 1.3 Four aromatic residues within Epo make large contributions to protein core. Substitution of the four primary residues responsible for maintaining the hydrophobic core leads to substantial conformational rearrangements. PDB file 1BUY was used to generate this figure.

results in substantial alterations in global conformation, but Phe substitution does not and is still successfully secreted.[74, 76-77] Substitution of Phe148 to Tyr results in a slight conformational change,[74] which could be caused by the addition of the hydroxide moiety into the hydrophobic pocket, whereas elimination of the moiety in the Tyr156Phe substitution would not be expected to alter the structure of the core.

As the exposure of nonpolar side chains to a water-based environment is unfavorable and destabilizing,[25, 29] substitution of a solvent-exposed nonpolar side chain for a more polar residue decreases the exposed hydrophobic surface area and increases the physical stability. Introduction of Asp, His, or Ser into the B25 position of insulin leads to an increase in the

folding stability due to subtle rearrangements in the global structure.[26, 75] The side chain of PheB25 is pointed toward the solvent, suggesting that it does not contribute significantly to maintaining protein structure, thus implying that the increases in folding stability are due to the elimination of hydrophobic surface area. The introduction of a polar side chain into the hydrophobic core will have the opposite effect and destabilize protein structure, as is illustrated by the conformation changes observed within Epo upon substitution of the core residues Ala98 or Leu105 with Ser.[74] The Leu105Asp mutation causes even more drastic conformational changes than the Leu105Ser mutation,[74] suggesting that, as expected, the core of Epo is more destabilized by introduction of a charged residue than simply a polar residue. Charged residues are more water-soluble than polar residues and in order for these to be stably contained within the hydrophobic core, salt bridges and/or hydrogen-bonding requirements must be satisfied.[78-81] Structural rearrangements in the four-disulfide insulin mutant (4SS-insulin) lead to a drastic decrease in the solvent-exposed hydrophobic surface area due to mutation of the hydrophobic residue Ile to Cys and additional shielding of hydrophobic residues,[82] thus increasing the physical stability of the monomer. The T_m of the four-disulfide mutant is 34.6 °C greater than that of native insulin, and extensive shaking for forty-five hours shows complete fibrillation of the native molecule, whereas the four-disulfide mutant exhibits none,[82] illustrating that improvements in physical stability can be made by decreasing exposed hydrophobic surface area.

1.2.2 Mutations that Affect Net Surface Charge

Charged residues help stabilize protein structure through electrostatic interactions that pair residues of opposite charge. This can be accomplished on the surface of the protein[79, 83-84] or in the form of internally buried salt bridges.[79, 85] Because these interactions contribute to local structural elements within domains or assist in structure formation by connecting different

domains, mutation of these residues can cause substantial conformational changes. Large changes in protein conformation have been observed due to electrostatic repulsion when an oppositely charged residue is introduced in close proximity to a similarly charged residue. The extreme case of this effect is apparent in many intrinsically disordered proteins (IDPs). Within Epo, the alterations in the three-dimensional fold observed upon Gln78Glu, His94Glu, or Lys97Glu/Asp substitution[74] are due in part to charge repulsion between Glu78 and Asp96, Glu94 and Glu18 or Glu21, and Glu/Asp97 and Glu18 or Glu21 (Figure 1.4). Ala substitution at position 78 or 97 results in slight structural changes,[74] suggesting that the substantial alterations in protein conformation observed upon introduction of Glu in these regions are not solely due to charge repulsion, but may also be caused by the elimination of stabilizing interactions such as a
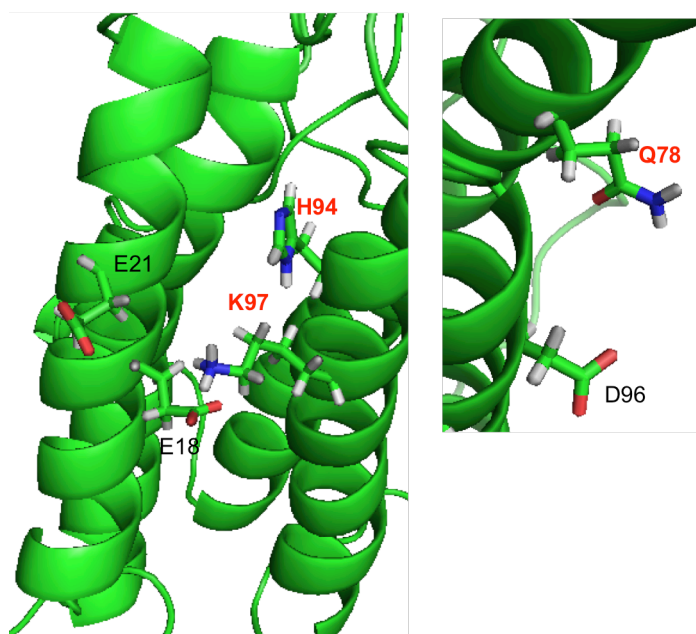
Figure 1.4   Alterations in the three-dimensional fold of Epo are caused by charge repulsion. Mutation of residues shown in red to Glu causes conformational changes because it leads to similarly charged residues within close proximity, which will cause charge repulsion and structural alterations. PDB file 1BUY was used to generate this figure.

hydrogen bond or salt bridge. Within IFN-α-2, mutation of Lys31, which works cooperatively with Lys133 to sandwich the side chain of Asp32,[53-54] to Glu causes conformational changes that lead to large changes in the bioactivity.[86] Introduction of a charged side chain in close proximity to similarly charged amino acid would be expected to alter the global conformation of the molecule due to repulsion of the similar two side chains.

Mutations that lead to elimination or addition of ionic interactions are destabilizing and stabilizing for protein structure, respectively. Charged residues within the B20-B23 loop of insulin stabilize the tertiary conformation by forming electrostatic interactions with C-terminal A-chain residues to correctly orient the two chains of the molecule. Removal of these interactions through mutation of GluB21 or ArgB22 to Ala results in a 25% decrease in protein expression levels relative to native insulin,[87] confirming that interactions within this loop are important in maintaining protein structure. The global structure of Epo is also substantially altered upon elimination of a charged residue (Glu23Ser).[74] As Glu23 is buried between the A and D helix interface,[37] and appears to form a salt bridge with a residue located on the D helix, Ser substitution eliminates this potential interaction. The results observed with insulin and Epo confirm the importance of salt bridges in maintaining protein structure. In contrast to the destabilizing effects observed by salt bridge elimination, addition of a non-native salt bridge through amino acid mutation is stabilizing to protein structure. Deletion of ThrB27 within insulin causes rearrangements that position the side chains of LysB29/GluA4 and the N-terminus of the A-chain and C-terminus of the B-chain in close proximity of one another. Therefore, the increase in folding stability observed with this mutant may be caused by the addition of two non-native sites for ionic interactions to occur.[75] Modulation of the thermal stability of Epo has been investigated by mutation of Arg103. Ala substitution causes minimal impact on protein

structure,[74, 88] whereas introduction of a negatively charged side chain through Glu/Asp mutation results in a substantial increase in thermal stability.[88-89] Multiple mutations were tested at this site to determine their effect on thermal stability with Asn, Asp, and Glu substitution increasing the stability, Ala and His substitution having similar stability as the native molecule, Gln substitution slightly decreasing the stability, and Lys substitution substantially reducing the stability.[89] Even though a few positively charged residues were located nearby, the exact interaction(s) responsible for the observed increase in thermal stability is undetermined.[89] Charged residues can exert a field effect in which one charged moiety interacts with several oppositely charged residues within the bonding radius to further stabilize the protein conformation.

The physical stability of protein therapeutics can also be improved by increasing the number of charged residues on the protein surface. This principle has been applied to non-glycosylated Epo, which has a lower stability than its glycosylated counterpart.[18] Mutation of the N-glycosylation sites to Lys results in the formation of a mutant with the same structural properties as native Epo produced in *E. coli*,[12, 34, 35] but improved thermal structural and aggregation stability and solubility in comparison to native, non-glycosylated Epo.[18, 37, 90] The denaturation stability of the two variants is very similar, and the observed melting temperature was also determined to be scan-rate dependent.[90] The data suggest the increase in net charge does not necessarily rigidify the protein fold to impart stability; rather, it alters the temperature-dependent aggregation kinetics of the unfolded state by decreasing the aggregation rate for the Lys mutant in comparison to the native protein.[90]

Akin to mutation, chemical modification alters the surface properties of a protein. Although the compounds attached induces a significantly larger change in size than mutation,

chemical principles still apply. The most noted property is decreased solubility, due to the relatively more hydrophobic character of the linkers and drugs. Self-selection occurs, where higher drug-to-antibody ratio (DAR) species precipitate during the reaction, such that an average DAR is typically below 4. The net surface charge commonly becomes more negative upon formation of conjugates because the most commonly used method relies on coupling to surface exposed lysine residues, incorporating the basic amine into a neutral amide bond. In addition to generating a heterogeneous mixture of molecules, multiple charge states have been observed with modification of a single lysine, depending on the position of the residue within the protein structure,[91] further complicating structural analysis. In some cases, this may be due to the formation of different chemical isomers upon linking at the same residue. Conjugation to lysine residues may also impart structural modifications that negatively impact the physiochemical properties of the molecule. Addition of the drug entity to lysine residues in T-DM1 caused destabilization of the structural stability of the CH2 domain, which is the least stable domain in antibodies and is where the majority of conjugation occurs.[2] Alterations in the structure of this single domain led to decreased thermal stability of the entire conjugate in comparison to the parent antibody.[1] Although lysine conjugation was shown to alter the thermal stability of T-DM1, it has a more limited effect on thermal stability than thiol conjugation.[6] This difference may be due to the more severe structural alterations caused by modification of cysteine residues, which is discussed later in the manuscript. In both cases, the number of drug molecules attached is low, particularly with respect to the large size of an antibody and the high number of basic residues.

1.2.3 Mutations that Alter Hydrogen Bond Formation

Amino acid mutation may also result in the introduction of non-native hydrogen bonds that stabilize protein structure. In native insulin, PheB25His/Ser mutation leads to an increase in folding stability due to subtle structural rearrangements that allow for the formation of an additional hydrogen bond between the substituted residue and TyrA19,[26, 75] thus stabilizing protein structure. ThrA8His/Arg mutation also introduces an additional hydrogen bond within insulin between A8 and GluA4. A hydrogen bond at this location is not present in native insulin because it would result in exposure of the nonpolar portion of the Thr side chain to water, which is highly unfavorable.[75] This additional hydrogen bond present in the mutant tightens the helix, not altering the global conformation, but increasing the folding stability of the molecule.[66, 75] This mutation also enhances insulin activity 3-fold,[66, 75, 92] as changes in the conformation of the A2-A8 helical region has been shown to greatly impact insulin potency.[66, 93] The importance of hydrogen bonds in maintaining the active conformation has also been illustrated within IFN-β-1; Arg147Ala substitution, which eliminates a crucial hydrogen bond between Arg147 and Leu24 responsible for stabilization of the AB loop,[15] causes changes in conformational leading to decreased function.[94] Therefore, modulation of the hydrogen bonding patterns within crucial structural elements, such as the A2-A8 α-helix within insulin, can be used to develop molecules that exhibit both high structural stability and bioactivity (Figure 1.2).

Addition of hydrogen bonds can also lead to increases in structural stability of therapeutic proteins. HisB10 is mutated to form the 4E and DKP insulin molecules discussed earlier as mutation of this residue leads to reduced self-association and generation of monomeric species. Mutation of HisB10 to either a Glu or Asp introduces the potential for another hydrogen bond within the molecule between the carbonyl oxygen of the side chain of the substituted residue and

the backbone amide NH of CysB7 (Figure 1.5). The additional interaction leads to large increases in the chaotropic denaturation stability of the molecule[75] and stabilization of structural motifs required for high affinity receptor binding.[26, 75, 95] Even though HisB10Thr mutation also provides the potential for a similar stabilizing hydrogen bond, Thr substitution decreases the denaturation stability because formation of the hydrogen bond
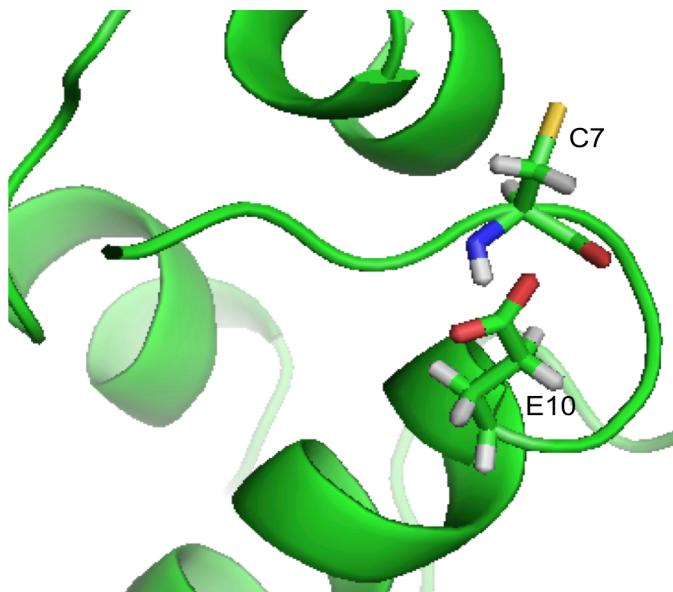
Figure 1.5 HisB10Glu mutation within insulin introduces additional hydrogen bond. Mutation of HisB10 to either Glu or Asp introduces the possibility of an additional hydrogen bond within the molecule between the side chain of Glu/Asp and the backbone amide NH of CysB7. PDB file 1HUI was used to generate this figure.

leaves the methyl side chain of Thr exposed to solvent, which is highly unfavorable and deleterious to protein stability.[75] The thermal stability of insulin can also be increased through the addition of non-native hydrogen bonds. Introduction of a fourth disulfide bond within insulin (A10/B4) results in the addition of at least four non-native hydrogen bonds and increases the T_m by 34.6 °C,[82] confirming the importance of hydrogen bonds and disulfides in modulating the structural flexibility that can directly impact the stability profile of protein therapeutics.

1.2.4 Mutations that Eliminate Evolutionarily Conserved Amino Acids

Evolutionarily conserved residues are commonly mutated to determine the exact role of these residues in protein structure. Within insulin, mutation of LeuB6 with Gly, Ala, Met, Val, or Phe does not alter the secondary structure, but does change the bioactivity of the molecule

with the extent of alteration dependent on the size and hydrophobicity of the substituted residue.[93] As discussed earlier, the B1-B8 region of the insulin molecule determines whether the molecule exists in the R- or T-state. The increased flexibility of the T-state is necessary for high affinity interactions of residues within this segment with the receptor. Therefore, the decreased bioactivity observed with mutation of LeuB6 suggests that local alterations in structure occur without affecting the secondary structural elements. For example, substitution with Met increases the bioactivity more than Gly substitution, suggesting that an increase in the size and/or decrease in flexibility of the B6 residue positively contributes to maintaining the bioactive conformation.[93] From these mutations, several different roles of LeuB6, such as correctly orienting the amino-terminal portion of the B-chain relative to the rest of the molecule, were hypothesized.[87, 93] Results of investigations into the effects of these mutations on the local structure of the insulin molecule are not yet available but could provide needed insight into the structural alterations that occur to accommodate the substituted residues and how these changes impact the stability of the molecule.

In addition to modulating bioactivity, mutation of evolutionarily conserved residues can also impart structural changes that alter the folding and stability of the molecule. Mutation of GlyB8 to Ala would not seem to impart much difference with regard to structure because it simply introduces a methyl group, but this mutation causes 2-3 fold decreases in refolding yield.[96] Similar decreases in refolding yields are observed mutation of GlyB8 to Ser, Thr, or Leu.[97] The refolded mutants also exhibit reduced α-helical character, with the largest modifications caused by the Ser and Thr mutations. Denaturation-induced unfolding confirms that substitution with Ser, Thr, or Ala leads to the generation of the most structurally unstable molecule as the denaturant concentration needed for unfolding was decreased by 15%, whereas

Leu had a very similar midpoint as native insulin.[96-97] The presence of a Gly residue at the B8 position is essential for achieving the desired structural stability due to the nature of the dihedral angle required at this position to maintain the B7-B10 turn within the T-state of insulin.[98] The mutants are also much more easily reduced than the native form,[96-97] which would be expected because GlyB8 is directly adjacent to the A7-B7 disulfide bond and structural alterations caused by mutation of GlyB8 would be expected to cause differences in disulfide stability.

1.2.5 Deletion of Segments of Protein Sequence

Deletion of multiple sequential amino acid residues can have varying effects, which are dependent on the role of these residues in maintaining protein structure, as well as the size of the deleted segment. Although it is difficult to elucidate the exact residue responsible for changes in protein structure or stability when multiple residues are eliminated, general trends can be established. For example, NMR confirms that deletion of the B25-B30 region in insulin does not alter the tertiary structure of the molecule.[21, 100-101] Deletion of this region does lead to a decrease in denaturation folding stability[75] and alterations in the molecular dynamics of the mutant,[25] suggesting that the structural stability of the molecule is altered by deletion of this segment. The results from Epo mutants are more difficult to interpret.   Although deletion of segments within the Epo sequence have been shown to affect protein secretion with the extent dependent on the deleted segment,[40] the differences in secretion efficiency may be reflective of cellular processing issues, as opposed to structural instability. Therefore, results based on deletion of multiple residues should not be used to draw conclusions about the molecule's structural stability because these mutations cause many different opportunities for structural change that are difficult to interpret.

1.2.6 Modulation of Native Disulfide Network

1.2.6.1 Elimination of Disulfide Bonds

Each disulfide bond within a protein exhibits differential effects on the folding efficiency and structural stability of the molecule, depending on extent of contribution of the bond to protein structure. The interchain disulfide bonds of insulin have a greater effect on the folding efficiency than the intrachain bond. Deletion of the intrachain disulfide bond through Cys to Ser mutation, referred to as the A6-11 mutant, results in a 25% decrease in folding efficiency, whereas deletion of the interchain disulfide bonds A7-B7 and A20-B19 through Cys to Ser mutation, referred to as the A7-B7 and A20-B19 mutants, substantially decrease the folding efficiency by 46% and 72%, respectively.[27, 99-100] Refolding of each mutant containing two of the three native disulfide bonds yields a mixture of misfolded monomer and cross-linked multimers, but it is possible for correct formation of the remaining two disulfide bonds to occur with minimal impacts to secondary structure.[99, 101-103] Removal of any of the three disulfide bonds causes increased structural flexibility and unfolding with the A20-B19 mutant exhibiting the greatest extent of unfolding.[99-100, 102] The A6-A11 and A7-B7 mutants contain 25% helical content with respect to native insulin with structural rearrangements only observed locally in the A-chain, but elimination of the A20-B19 disulfide bond results in complete abolition of any helical character.[27, 99-100, 104-105] Similarly to the A20-B19 disulfide bond within insulin, one disulfide bond within Epo and IFN-α-1 is more critical for preserving protein structure than the other. Because the Cys7-Cys161 disulfide bond of Epo connects Helix A and Helix D, which is required for formation of the up-up-down-down topology, deletion or mutation of either Cys7 or Cys161 causes drastic alterations in the tertiary structure.[39, 74] On the other hand, the Cys29-Cys33 disulfide bond is not a strong contributing factor in determining the topology of Epo as

replacement of either Cys29 or Cys33 with Tyr results in the successful secretion.[39] Within IFN-α-1, the Cys29-Cys138 disulfide bond contributes greater to maintaining the global conformation, as the presence of this disulfide bond alone, but not Cys1-Cys98, is sufficient for maintaining the biological activity of IFN-α-2.[106] Because the AB loop is important in maintaining the biologically active conformation,[94, 107] deletion of this disulfide bond, which connects loop AB and Helix E,[53-54] will result in global structural rearrangements that are highly detrimental to bioactivity. IFN-β-1 contains only one disulfide bond that is responsible for connecting the AB loop and Helix E,[15, 108] which is essential for biological activity.[46, 109] Elimination of this disulfide bond through Cys141Tyr mutation results in complete loss of biological function,[108] suggesting that large structural rearrangements occur in the absence of this bond. Further mutagenesis studies could provide insight into whether it is possible to introduce stabilizing interactions within the proteins discussed to stabilize the native conformation in the absence of a disulfide bond.

Because disulfide bond removal eliminates key components of protein structure, the stability of these mutants would be expected to be less than that of the native molecule. Current work suggests that the A7-B7 insulin mutant exhibits a severe loss in ordered structure within only the A-chain and is more easily unfolded by denaturant and temperature in comparison to the A6-A11 mutant.[27] The decreased folding stability could be due to the increase in exposed hydrophobic surface area on the A7-B7 mutant,[99-100] which makes it more prone to undergo unfolding or an association-induced unfolding event. The A20-B19 mutant was also more easily unfolded by denaturant and exhibited an increase in molecular hydrodynamic volume,[99] suggesting that elimination of this bond causes structural instability. The thermal stability of IFN-α-2 is also altered in the absence of the disulfides. It was reported that incubation of a

chemically reduce, disulfide-free mutant at 25 °C does not infer conformational changes and the disulfides can be reformed, but incubation of the mutant at 37 °C eliminates the native structure, does not allow reformation of the disulfides using thiol exchange, and results in the formation of HMWS and visible precipitation.[106] The elevated temperature could promote conformational changes, followed by aggregation, or it could increase the aggregation rate and drive the conformation change.

Elimination of disulfide bonds has also been shown to increase protein stability. In phosphatase of regenerating liver (PRL-1), the reduced form of the protein exhibits higher physical stability than the oxidized form.[68] Elimination of the disulfide leads to local structural alterations near these Cys residues, which allows for the formation of an additional, stabilizing hydrogen bond. Disulfide bond elimination within Hsp16.3 has also been shown to increase the stability of the molecule.[110] In this case, formation of the disulfide causes undesirable conformational rigidity and increased hydrophobic surface area, which leads to protein self-association and aggregation.[110] Therefore, the effect of disulfide bond elimination on molecular stability is dependent on both local and global structural changes that occur in the absence of the bond.

Modulation of disulfide networks is also performed in the generation of ADCs. Drug conjugation to native cysteine residues can be performed after breaking intermolecular disulfide bonds made accessible through selective reduction.[3-4] Non-covalent interactions maintain the structure near the hinge region in the absence of the disulfide bonds,[111] but the greater fragmentation rates observed at higher ionic strength confirm that these interactions are not sufficient to maintain protein structure.[3] Even though the quaternary structure is well-maintained, elimination of these structural features and introduction of a hydrophobic entity at

the site could cause local structural changes that cause the molecule to be more susceptible to unfolding when thermally stressed.[4] The rapid aggregation of cysteine-conjugated ADCs and lower T_m of the CH2 domain indicate that structural alterations have occurred within the molecule,[3-4] although they may be not be observable with low resolution techniques such as CD. It is suggested that the CH2 domain is the cause of aggregate formation due to its instability in the absence of the disulfides and its presence as a partially unfolded domain within the aggregate.[4] The number of conjugated entities or DAR has also been shown to influence the physical stability of the molecule, confirming that the decreased physical stability of ADCs is due to both elimination of disulfides and introduction of a hydrophobic entity on the surface of the protein structure. Conjugation at Cys occurs preferentially within the Fab region for the lower DAR species (2-4), whereas the higher DAR species are conjugated at both the Fab and hinge region. The decrease in T_m observed within higher DAR conjugates suggests that the number of drug entities, as opposed to disulfide elimination, causes the largest impact on protein structure. Unfolding of the CH2 domain has also been observed at low DAR,[3] where conjugation primarily occurs in the Fab region, illustrating the far-reaching effects that surface modification can have on the structural stability of ADCs.

1.2.6.2 Addition of Cysteine Pairs

In addition to disulfide bond elimination, non-native disulfide bonds have also been introduced to determine their effect on protein structure and stability. Using the high-resolution structure, cysteine residues can be introduced into precise locations within the protein structure that are favorable and don't affect the native disulfide bonds. Addition of a third interchain disulfide bond in insulin between residues A10 and B4 (4SS-insulin) minimally alters expression levels,[82] the structure of the four-disulfide mutant is altered. The additional disulfide bond

constrains the N-terminal portion of the B-chain in a conformation that does not allow it to achieve the R-state and leads to the formation of a non-native β-sheet between A10-A12 and B2-B4 (Figure 1.6).[82] Structural rearrangements in the four-disulfide insulin mutant (4SS-insulin) lead to a drastic decrease in the hydrophobic surface area as discussed earlier, and formation of addition hydrogen bonds between the A- and B-chain, which cause the 4SS-insulin mutant to exhibit a T_m 34.6 °C greater than that of native insulin.[82]
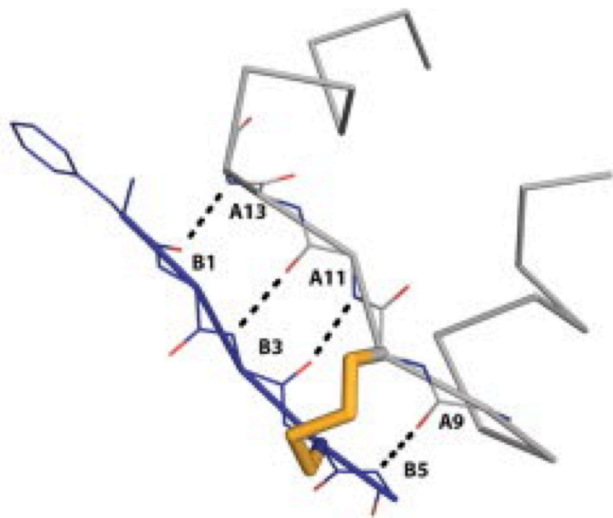
Figure 1.6   Addition of a non-native disulfide within insulin increases thermal stability. With the mutation of A10 and B4 to Cys, an additional disulfide bond is formed between the two strands leading to additional hydrogen bonding interactions. These additional stabilizing interactions drastically increase the T_m of the mutant. Modified and reproduced with permission from Vinther, T. N., Norrman, M., Ribel, U., Huus, K., Schlein, M., Steensgaard, D. B., *et*. *al*. (2013) Insulin analog with additional disulfide bond has increased stability and preserved activity. *Protein Sci*. *22*, 296-305, © 2013 John Wiley & Sons, Inc., Hoboken, NJ.

Rearrangement of the non-essential disulfide bond within Epo also leads to stabilizing modifications in protein structure. Epo(NDS) contains four mutations (His32Gly, Cys33Pro, Trp88Cys, and Pro90Ala) that rearrange the disulfide bond from Cys29-Cys33 to Cys29-Cys88, thus connecting the AB and BC loops.[112] The Fc-Epo(NDS) fusion protein exhibited better folding with 93% monomer observed in the NDS mutant and only 65% monomer content in the native counterpart.[112] In the solution structure, Cys29 is very near in space to Trp88 so formation of a disulfide bond at this location stabilizes Helices A and C in the four-helix bundle. The His32Gly and Pro90Ala

mutations alleviate structural strain near the BC loop caused by the new disulfide bond, with the Pro90Ala mutation contributing greater to expression of the protein in a non-aggregated state. The Epo(NDS) mutant also exhibited a 20-fold greater resistance to enzymatic cleavage than native Epo in the absence of glycosylation.[112] Although the reason for the increased stability is not fully understood, it is suggested that the NDS mutant is packed more tightly than the native protein which decreases water accessibility into the core.[112]

Engineered disulfide bonds have also served useful in increasing the thermal stability of antibody fragments[113-116] and whole antibodies. For example, modulation of the disulfide network within the Fab domain of IgG4 leads to increased thermal stability of the entire antibody.[117] Although the Fab domain of IgG4 has a drastically lower thermal stability than IgG1,[118] altering the interchain disulfide pattern of IgG4 (DSB between LC Cys214 and HC Cys229 as opposed to HC Cys214) to be more similar to that of IgG1 increases the thermal stability of IgG4 to similar values observed for IgG1. Although CD analysis illustrates that minimal structural changes accompany these substitutions, this low-resolution method is not sufficient to probe how local structural perturbations cause increased thermal stability. Additional investigation into the physical stability of the disulfide-mutants could provide details regarding the key structural interactions necessary for stabilization, and if the potential exists to further modulate the disulfide network of other therapeutic proteins to increase their physical stability.

Disulfide bonds are commonly engineered into protein structures to increase the structural stability of the molecule, but introduction of a non-native disulfide bond can also have a destabilizing effect on protein structure. Although molecular modeling can provide a good assessment of the optimal location of an engineered disulfide within the protein structure, the

theoretical results do not always align with what is observed experimentally, leading to structural destabilization.[119] Introduction of a non-native disulfide bond can lead to elimination of crucial stabilizing interactions, such as hydrogen bonds, or cause atypical dihedral angles, both of which will have a destabilizing effect on protein structure.[120-121] Engineered disulfide bonds have also been shown to stabilize local stability, whereas the global stability of the molecule was decreased.[122] In this case, the non-native disulfide bond stabilized the denatured state, which decreased the stability of the entire molecule. The stabilization or destabilization effect observed upon introduction of non-native disulfide bonds is dependent upon the structural elements altered by introduction of the disulfide. Many of the engineered disulfide bonds that induce structural stabilization link regions between 25 and 75 residues apart that exhibit high conformational mobility.[123] Improvements in thermal stability have also been observed through engineering of disulfide bonds into regions of high flexibility.[124-125] The stabilizing effect of disulfide bonds is caused by a decrease in the conformation entropy of the unfolded state.[123, 126] Therefore, engineered disulfide bonds lead to structural stabilization by increasing the structural rigidity of the molecule, but can also eliminate crucial stabilizing interactions and destabilize protein structure.

1.3 MODULATION OF STABILITY THROUGH CHEMICAL MODIFICATION

Chemical modifications such as glycosylation and PEGylation impact protein stability, but the mechanism of stabilization can be more complex and difficult to elucidate than mutation. The carbohydrate entity or PEG molecule may increase protein stability through direct interactions with the protein or indirectly by shielding the protein from interaction with other protomers, often by occluding regions of hydrophobic surface area.[15, 19, 127] Therefore, modulating the glycosylation pattern of proteins, such as Epo and interferon, may afford

increased stability, although the means by which this occurs is highly protein dependent.[18-19, 128-130] PEGylation provides an additional means of increasing physical stability due to the large size of the PEG molecule, which sterically hinders association among protein molecules.[131-133]

1.3.1 Glycosylation

Glycosylation, or the covalent attachment of carbohydrate moieties to the protein surface, can occur within a native protein sequence or non-native sites can be engineered into the native sequence. The two main types of glycosylation are N- and O-glycosylation, which occur at Asn and Ser/Thr residues, respectively. The presence of these residues alone does not always lead to glycosylation at this site. N-glycosylation generally requires a consensus sequence,[134] and the location of these residues within the three-dimensional structure of the protein can also dictate whether or not glycosylation occurs. In Epo, full glycosylation is reliant upon preservation of the four-helix bundle, as mutants missing helical regions were not fully glycosylated.[39] Also, the introduction of a bulky carbohydrate chain near the hydrophobic core would be detrimental in achieving the correct fold, which leads to lack of glycosylation near these regions.

The addition of the glycan occurs during protein folding, suggesting it may contribute to folding *in vivo* with some glycosylation sites contributing more than others to the folding process.[129, 135-141] Native O- and N-glycosylation in Epo has been shown to have minimal effects on protein structure and secretion.[142, 143] Removal of the O-glycosylation site in Epo through substitution of Ser126 has no impact of the global conformation of the molecule.[74] Even though elimination of all three N-glycosylation sites causes Epo to be produced mainly in the inclusion bodies in the *E. coli* system,[18, 39, 144] the refolded molecule contains the same secondary and tertiary structure of the native protein,[18, 144] suggesting that the structure of the protein is not modulated by the presence of the glycan moieties, but its stability may be. The addition of non-

native carbohydrate groups has also been shown to have minimal influence on the global structure of insulin[145-146] and IFN-α-2b.[130] On the other hand, native N-glycosylation of IFN-β-1 assists in the formation of the stable three-dimensional structure.[19] IFN-β-1b, which folds in the absence of the glycan, exhibits a T_m 7 °C lower than that of enzymatically deglycosylated IFN-β-1a, which is folded in the presence of the glycan and then the glycan is removed.[19] Removal of the N-glycan in IFN-β-1b also does not affect protein secondary structure once folding is achieved, [147] suggesting that the glycan assists in achieving, but not maintaining, the desired fold.

The addition of glycan entities on the protein surface increases structural stability by limiting the structural mobility.[129, 148] The size of the carbohydrate chains also contributes to folding stability by increasing the free energy associated with the unfolded state of the protein.[13-14] This increased folding stability afforded by glycosylation leads to measurable difference in thermal stability, as the glycosylated forms of IFN-β,[19] IL-5,[128] Epo[18], yeast external invertase, bovine serum fetuin, and glucoamylase[129] all exhibit increased thermal stability in comparison to the deglycosylated forms. The N-glycan motifs drastically influence Epo structure because their removal caused a decrease in molecular stability, whereas O-glycosylation does not contribute an appreciable amount to the stability of the molecule.[149] Non-glycosylated Epo is also more susceptible to unfolding in the presence of denaturant, and it is less stable in acidic conditions.[18, 150] Acid-induced denaturation results in complete elimination of the tertiary structure of the mutant, but not of the glycosylated form, which suggests that the glycan moieties assist in the stabilization of the global structure of the protein at acidic pH.[18] Glycan molecules can also modulate the thermal stability of IFN-α-2. The IFN-α-2 variant containing four non-native glycans (Pro4Asn, Arg23Asn, Lys70Asn, Asp77Asn) exhibited a T_m approximately thirty degrees higher than the native O-glycosylated and *E. coli* generated non-glycosylated variant.[130]

The T_m values of the enzymatically deglycosylated 4N variant were the same as the native protein, confirming that it is the addition of the glycan moieties and not simply the amino acid mutagenesis that is responsible for the drastic increase in thermal stability, although a detailed description of the interaction between the protein surface and glycan has not yet been reported. The glycan within IgG1-Fc also directly interacts with the protein surface to stabilize protein conformation. The outer and inner regions of the N-glycan at Asn297 within IgG1-Fc have the largest effect on protein conformation[151] and are hypothesized to stabilize protein conformation through non-covalent interactions with nearby residues on the protein surface.[16-17] Removal of the terminal *N*-acetylglucosamine resulted in the largest reduction in T_m (3.1 °C) with an additional decrease in T_m by 1.4 °C upon complete deglycosylation.[152] In the solid-state structure, the loop containing the N-glycosylation site and the oligosaccharide are most perturbed by removal of the terminal N-acetylglucosamine or mannose sugar residues, suggesting an important role of these motifs in maintaining protein conformation.

Even though glycosylation generally increases the folding stability, deglycosylation can also increase stability due to alterations in the protein conformation resulting in decreased hydrophobic surface area.[153] The effect of the O-glycan on the thermal stability of IFN-α-2 is debated. Johnston *et al*. reported T_m values of 65.7 °C and 63.8 °C for the non-glycosylated and O-glycosylated variants of IFN-α-2b,[143] whereas Ceaglio *et al*. reported similar T_m values near 66 °C for both forms.[130] The O-glycosylation site is located within the flexible CD loop, and addition of GalNAc or Gal(β1,3)GalNAc, which are two of the three major glycan moieties found in naturally produced IFN-α-2, decreases the conformational dynamics of the loop.[154] This suggests that flexibility of the loop is required for structural stability, as the rigidity imparted by the glycan molecule leads to decreased stability. Because the observed difference in

thermal stability between the two variants is so small, additional experiments concerning the interactions between the glycan motifs and protein surface would prove useful in determining if the O-glycan is indeed destabilizing in IFN-α-2.

In addition to contributing to protein folding, glycosylation may also increase the physical stability of the molecule. At the high concentrations necessary for protein therapeutics, protein molecules may associate to form aggregates.[155] Attachment of the glycan to the protein surface increases the solubility of the molecule due to the hydrophilic nature and steric effects of the carbohydrate,[13] which leads to decreased aggregation propensity.[134, 156-157] Introduction of non-native glycosylation sites increases the physical stability of IFN-α-2, as has been shown with other therapeutic proteins.[18, 145-146, 150] The solubility of the monomeric form of insulin can also be increased through the chemical attachment of glycan molecules at the amino groups of the N- terminus of the A- or B-chain, and LysB29. The tri-substituted molecule is the most monomeric derivative, but the stabilizing effects of glycan addition to the three locations are not completely additive.[145-146] Shaking tests illustrate that the N-terminal A-chain modification increases the physical stability five-fold, N-terminal B-chain 20-fold and N-terminal B-chain-LysB29 and trisubstituted molecule 30-fold.[145-146] The N-terminal B-chain site and LysB29, are known to be dimer-forming surfaces (Figure 1.1),[25-26, 32] and fibril formation has been shown to occur at the N-terminal portion of the B-strand,[158] so glycan addition at this site in particular will lead to increased stability of the monomeric form by limiting self-association and fibril formation.

The carbohydrate moiety may also increase the physical stability by shielding regions of hydrophobic surface area that are prone to interacting. Severe precipitation of non-glycosylated Epo occurs after incubation at elevated temperatures and the helical content of the remaining

soluble portion is nonexistent; but, the glycosylated protein retains the native helical structure and remains soluble.[18, 144] ANS fluorescence suggests that non-glycosylated Epo contains more solvent-exposed hydrophobic residues in comparison to the fully glycosylated form.[18] Simply adding free N-glycans to non-glycosylated Epo results in decreased ANS fluorescence suggesting that the glycans do not need to be chemically attached to have the protective effect.[150] In Epo, the inner regions of the N-glycans were determined to be the part of the glycan moiety responsible for interacting with the protein surface.[150] Removal of the galactose groups resulted in a decrease in denaturation stability and an increase in ANS florescence, suggesting that these groups are responsible for making the primary contacts with the hydrophobic surface area of Epo. The glycan is also responsible for modulating the physical stability of IFN-β-1a.[19, 60, 159] Non-glycosylated IFN-β-1a exhibits a T_m 5 °C lower than glycosylated IFN-β-1a.[19-20] The glycan forms hydrogen bonds with surface residues Gln23 and Asn86 of IFN-β-1a, further increasing the structural integrity of the molecule.[15, 19, 61] It also shields a region of uncharged surface area so the observed increases in thermal stability could be due to either of these factors. In native IFN-α-2, glycosylation at this position is not required for maintaining physical stability, as this region contains charged residues that are unlikely to self-associate (Figure 1.7).[53-54] In addition, the presence of the negatively charged sialic acid residues on the glycan may further decrease the self-associate propensity through electrostatic repulsion between neighboring molecules, as is observed with Epo.[160-161]

The type of aggregate formed in the absence of the glycan also varies between IFN-β-1a and IFN-β-1b. In IFN-β-1b, mainly soluble aggregates are formed that do not exhibit covalent properties,[19] which suggests that the aggregation in this variant is primarily due to self-association of non-charged regions within the protein. On the other hand, the aggregates that
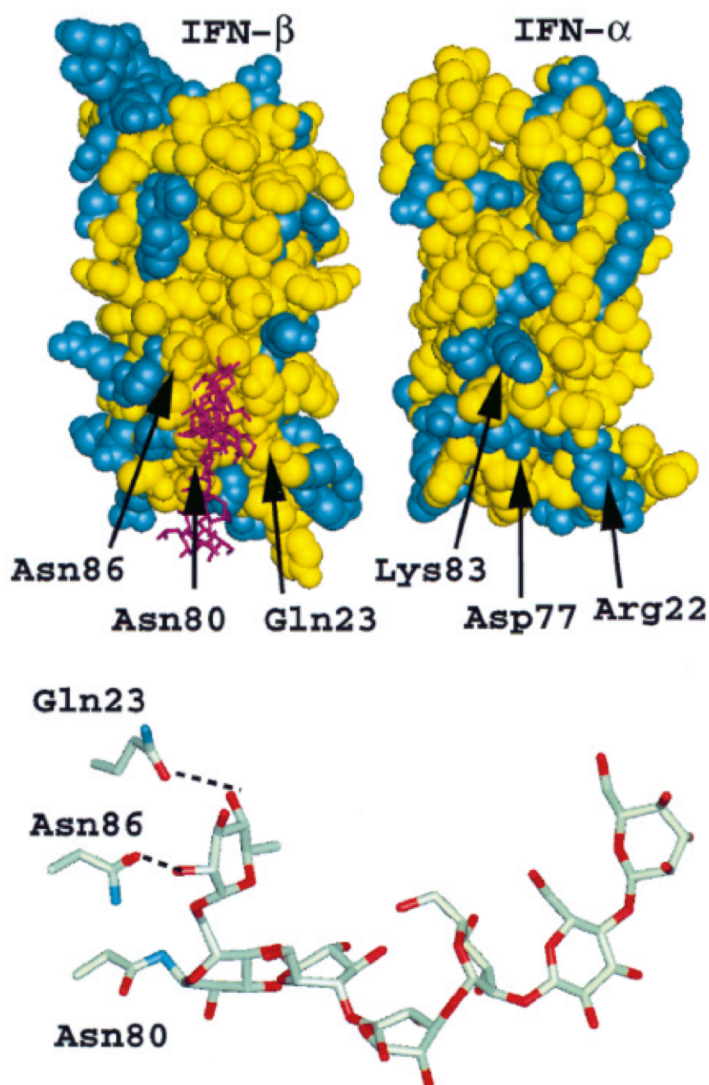
Figure 1.7 The glycan moiety within IFN-β-1 increases the structural and physical stability of the molecule. The glycan forms stabilizing hydrogen bonds with residues Gln23 and Asn86 to maintain the structural integrity of the protein, and it also shields a surface-exposed region of uncharged residues. IFN-α does not require a glycan at this location because it contains charged residues in the analogous region, thus eliminating the region prone to self-association. Reproduced with permission from Karpusas, M., Whitty, A., Runkel, L., Hochman, P. (1998) The structure of human interferon-β. Implications for activity. *Cell. Mol. Life Sci. 54*, 1203-1216, © 1998 Springer, New York.

form in enzymatically deglycosylated IFN-β-1a are primarily composed of disulfide-linked complexes.[19] Cys17Ser mutation results in increased physical stability so the decreased stability

upon deglycosylation of IFN-β-1a suggests that the glycan may modulate the reactivity of Cys17. Partial unfolding may result in the ability of Cys17 to form intermolecular disulfide bonds due to increased exposure, which leads to decreased stability.

1.3.2 PEGylation

Covalent attachment of PEG to protein therapeutics has been shown to increase serum half-life and solubility, while also decreasing immunogenicity.[8-12] PEGylation also affects the physical stability of proteins, and PEGylation of smaller protein therapeutics can be used to investigate the effect on a protein when site-specific chemical conjugation at different positions within the protein is accomplished. Conjugation of a PEG molecule or drug entity usually occurs through the amino groups located on the N-terminus of the protein or surface-exposed lysine side chains but can also occur through non-native, engineered cysteine residues. Conjugation through lysine or histidine residues is usually nonspecific and results in a highly heterogeneous product,[2, 162-165] whereas conjugation to non-native engineered cysteine residues allows for some degree of site specificity.[5, 166] The conjugation site is mainly determined by properties of the protein, the number of possible conjugation sites within the protein, and the type of conjugation chemistry used.

Site-specific PEGylation using native residues has been accomplished using knowledge of the reactivity of the potential conjugation sites in the protein of interest to develop the reaction conditions necessary for the desired modification site. Insulin contains three potential PEGylation sites (N-terminus of the A-chain, LysB29, N-terminus of the B-chain), but site-specific PEGylation at LysB29 is possible at high pH conditions (pH>9.5) because this amino group exhibits the highest reactivity under these reaction conditions.[167] PEGylation at either site within insulin results in slight structural differences that depend on the conjugation site but not

the size of the PEG molecule,[131-133, 168] suggesting that the residue modification and not size of conjugated entity causes structural perturbations. The disubstituted A-chain N-terminus/LysB29 molecule exhibits a more intense α-helical feature due to the PEG molecule reducing the overall flexibility of native monomeric insulin.[168] Substitution at LysB29 also causes restriction in the rotation of B-chain aromatic residues,[131-132] suggesting that conjugation at LysB29 impacts the flexibility of the monomer more so than conjugation at the N-terminus of the A-chain. On the other hand, PEGylation of IFN-α-2 at native lysine or histidine residues has minimal impacts on protein structure. The attachment of a 12 kDa PEG molecule at His34 of IFN-α-2b does not cause substantial changes in the secondary or tertiary structure,[164-165] and PEGylation at lysine residues is expected to have minimal effects on protein structure, as was the case with IFN-β-1b.[163] Although it allows for site-specificity, conjugation to engineered cysteine residues is performed less often because introduction of additional cysteine residues, particularly in the absence of a high-resolution structure, may result in glutathione adducts and product heterogeneity[169-170] or have negative implications in protein folding.[166] Selective reduction of disulfides within antibodies is used in ADC conjugation,[3-4] but insertion of a much larger PEG molecule at this location would be highly detrimental to structural stability.

Although PEGylation has been shown to minimally impact the structure of protein therapeutics, it does modulate the stability of the molecule. PEGylation of Epo is hypothesized to occur either at His32 or His 94 and drastically increases its unfolding and aggregation thermal stability in comparison to the non-glycosylated form.[144] A two-fold decrease in temperature-induced aggregation was also observed with random lysine PEGylation of IFN-α-2b in comparison to the non-PEGylated counterpart.[171] Because the global structure is not affected by addition of the PEG molecule, additional investigations are needed to determine if the PEG

molecule increases the thermal stability through modifications in local structure due to direct interactions with nearby residues. On the other hand, the thermal unfolding of IFN-β-1b is not influenced by random PEGylation at surface exposed lysine residues, but the aggregation propensity of the molecule is decreased.[163] This suggests that in the case of IFN-β-1b, the PEG molecule does not interact with the protein surface; rather it increases the stability of the molecule by sterically interfering with protein self-association.

Because PEGylation can alter the exposed hydrophobic surface of the protein, it has been shown to increase protein stability in a similar manner as non-native glycosylation, although these effects are not as dramatic as the increases in PK activity provided by PEGylation.[10] PEGylation of insulin can increase the physical stability of the monomeric form by sterically hindering self-association at known contact points. Conjugation at the N-terminus of the B-chain, which is known to be a self-association contact point (Figure 1.1), increases the physical stability of monomeric insulin 36-40 fold in comparison to the 4-8 fold increase observed by conjugation at the LysB29 position.[131-132] Addition of a PEG molecule at the PheB1 location prevents self-association by limiting the hydrophobic interaction site known to contribute to aggregation of monomeric insulin[158, 172] and sterically restricting self-association. The N-terminal portion of the A-chain must also contribute to fibril formation as disubstitution of the N-terminus of the A-chain and LysB29 resulted in a 8-fold increase in physical stability over monosubstitution of LysB29.[133] As with glycosylation, the self-association propensity of the trisubstituted molecule was the lowest, suggesting monomer stability increases with increasing PEG substitution and size of the PEG molecule.[133] PEGylation has also been shown to increase the physical stability of antibody fragments, allowing for the generation of high-concentration formulations > 200 mg/mL without protein aggregation.[173] Because PEGylation allows for

increased physical stability with minimal structural implications, it is a very useful tool for generating protein therapeutics with the desired properties.

1.4 CONCLUSION

The stability of protein therapeutics is modulated by many different factors including structural elements and surface properties, and investigating the effects of changing these factors is of critical importance, particularly for advancing beyond empirical evaluation to implement design of more stable therapeutics. The generation of ADCs requires modulation in both the structure and surface of the antibody so the stability of the conjugate relative to the native molecule must be determined. Because high-resolution data on antibody structure is limited, it is difficult to connect observed changes in molecule's stability to the specific structural elements responsible for altering the stability. As such, smaller protein therapeutics can serve as model systems for assessing the molecular mechanisms behind changes in stability. In this review, the impact of modifying key elements responsible for stabilizing protein structure, such as hydrogen bonds, hydrophobic associations, and ionic interactions, within smaller protein therapeutics was analyzed and the trends observed were considered with respect to larger antibody systems. Mutations within smaller protein therapeutics are akin to chemical conjugation to antibodies because both modulate structural and surface properties, which directly impact stability. Until a significant body of site-specific data are reported regarding the impact of conjugation on the structural stability of ADCs, the trends observed with mutation and modification of smaller protein therapeutics provide foundational principles for understanding structural stability, which can be applied when attempting to design site-specific conjugates and ADCs that exhibit the best possible stability profiles.

1.5 REFERENCES

1.   Wakankar, A. A., Feeney, M. B., Rivera, J., Chen, Y., Kim, M., Sharma, V. K., Wang, Y. J. (2010) Physicochemical Stability of the Antibody-Drug Conjugate Trastuzumab-DM1: Changes due to Modification and Conjugation Processes. *Bioconjugate Chem.* *21*, 1588-1595.

2.   Wang, L., Amphlett, G., Blättler, W. A., Lambert, J. M., Zhang, W. (2005) Structural characterization of the maytansinoid-monoclonal antibody immunoconjugate, huN901-DM1, by mass spectrometry. *Protein Sci.* *14*, 2436-2446.

3.   Adem, Y. T., Schwarz, K. A., Duenas, E., Patapoff, T. W., Galush, W. J., Esue, O. (2014) Auristatin Antibody Drug Conjugate Physical Instability and the Role of Drug Payload. *Bioconjugate Chem.* *25*, 656-664.

4.   Beckley, N. S., Lazzareschi, K. P., Chih, H.-W., Sharma, V. K., Flores, H. L. (2013) Investigation into Temperature-Induced Aggregation of an Antibody Drug Conjugate. *Bioconjugate Chem.* *24*, 1674-1683.

5.   Junutula, J. R., Raab, H., Clark, S., Bhakta, S., Leipold, D. D., Weir, S., Chen, Y., Simpson, M., Tsai, S. P., Dennis, M. S., Lu, Y., Meng, Y. G., Ng, C., Yang, J., Lee, C. C., Duenas, E., Gorrell, J., Katta, V., Kim, A., McDorman, K., Flagella, K., Venook, R., Ross, S., Spencer, S. D., Wong, W. L., Lowman, H. B., Vandlen, R., Sliwkowski, M. X., Scheller, R. H., Polakis, P., Mallet, W. (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* *26*, 925-932.

6.   Acchione, M., Kwon, H., Jochheim, C. M., Atkins, W. M. (2012) Impact of linker and conjugation chemistry on antigen binding, Fc receptor binding and thermal stability of model antibody-drug conjugates. *MAbs 4*, 362-72.

7.   Laurence, J. S., Middaugh, C. R. In *Fundamental structures and behaviors of proteins*, John Wiley & Sons, Inc.: 2010; pp 1-61.

8.   Greenwald, R. B. (2001) PEG drugs: an overview. *J. Controlled Release 74*, 159-171.

9.   Pepinsky, R. B., Lepage, D. J., Gill, A., Chakraborty, A., Vaidyanathan, S., Green, M., Baker, D. P., Whalley, E., Hochman, P. S., Martin, P. (2001) Improved pharmacokinetic properties of a polyethylene glycol-modified form of interferon-β-1a with preserved in vitro bioactivity. *J. Pharmacol. Exp. Ther.* *297*, 1059-1066.

10. Reuss, R. (2013) PEGylated interferon beta-1a in the treatment of multiple sclerosis - an update. *Biol.: Targets Ther. 7*, 131-138.

11. Wylie, D. C., Voloch, M., Lee, S., Liu, Y.-H., Cannon-Carlson, S., Cutler, C., Pramanik, B. (2001) Carboxyalkylated histidine is a pH-dependent product of pegylation with SC-PEG. *Pharm. Res. 18*, 1354-1360.

12. Zhang, C., Yang, X.-l., Yuan, Y.-h., Pu, J., Liao, F. (2012) Site-specific PEGylation of therapeutic proteins via optimization of both accessible reactive amino acid residues and PEG derivatives. *BioDrugs 26*, 209-215.

13. Shental-Bechor, D., Levy, Y. (2008) Effect of glycosylation on protein folding: a close look at thermodynamic stabilization. *Proc. Natl. Acad. Sci. U. S. A. 105*, 8256-8261.

14. Banks, D. D. (2011) The Effect of Glycosylation on the Folding Kinetics of Erythropoietin. *J. Mol. Biol. 412*, 536-550.

15. Karpusas, M., Nolte, M., Benton, C. B., Meier, W., Lipscomb, W. N., Goelz, S. (1997) The crystal structure of human interferon β at 2.2-Å resolution. *Proc. Natl. Acad. Sci. U. S. A. 94*, 11813-11818.

16. Deisenhofer, J. (1981) Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-Å resolution. *Biochemistry 20*, 2361-70.

17. Padlan, E. A. In *X-ray diffraction studies of antibody constant regions*, Am. Soc. Microbiol.: 1990; pp 12-30.

18. Narhi, L. O., Arakawa, T., Aoki, K. H., Elmore, R., Rohde, M. F., Boone, T., Strickland, T. W. (1991) The effect of carbohydrate on the structure and stability of erythropoieitin. *J. Biol. Chem. 266*, 23022-6.

19. Runkel, L., Meier, W., Pepinsky, R. B., Karpusas, M., Whitty, A., Kimball, K., Brickelmaier, M., Muldowney, C., Jones, W., Goelz, S. E. (1998) Structural and functional differences between glycosylated and non-glycosylated forms of human interferon-β (IFN-β). *Pharm. Res. 15*, 641-649.

20. Watanabe, Y., Kawade, Y. (1983) Properties of non-glycosylated human interferon-beta from MG63 cells. *J Gen Virol 64 (Pt 6)*, 1391-5.

21. Blundell, T., Dodson, G., Hodgkin, D., Mercola, D. (1972) Insulin. Structure in the crystal and its reflection in chemistry and biology. *Adv. Protein Chem. 26*, 279-402.

22. Kang, S., Brange, J., Burch, A., Vølund, A., Owens, D. R. (1991) Subcutaneous insulin absorption explained by insulin's physicochemical properties. Evidence from absorption studies of soluble human insulin and insulin analogues in humans. *Diabetes Care 14*, 942-8.

23. Derewenda, U., Derewenda, Z., Dodson, G. G., Hubbard, R. E., Korber, F. (1989) Molecular structure of insulin: the insulin monomer and its assembly. *Br. Med. Bull. 45*, 4-18.

24. Mark, A. E., Jeffrey, P. D. (1990) The self-association of zinc-free bovine insulin: four model patterns and their significance. *Biol. Chem. Hoppe-Seyler 371*, 1165-74.

25. Weiss, M. A., Hua, Q.-X., Lynch, C. S., Frank, B. H., Shoelson, S. E. (1991) Heteronuclear 2D NMR studies of an engineered insulin monomer: assignment and characterization of the receptor-binding surface by selective deuterium and carbon-13 labeling with application to protein design. *Biochemistry 30*, 7373-89.

26. Shoelson, S. E., Lu, Z.-X., Parlautan, L., Lynch, C. S., Weiss, M. A. (1992) Mutations at the dimer, hexamer, and receptor-binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions. *Biochemistry 31*, 1757-67.

27. Hua, Q.-X., Nakagawa, S. H., Jia, W., Hu, S.-Q., Chu, Y.-C., Katsoyannis, P. G., Weiss, M. A. (2001) Hierarchical Protein Folding: Asymmetric Unfolding of an Insulin Analogue Lacking the A7-B7 Interchain Disulfide Bridge. *Biochemistry 40*, 12299-12311.

28. Kline, A. D., Justice, R. M., Jr. (1990) Complete sequence-specific proton NMR assignments for human insulin. *Biochemistry 29*, 2906-13.

29. Weiss, M. A., Nguyen, D. T., Khait, I., Inouye, K., Frank, B. H., Beckage, M., O'Shea, E., Shoelson, S. E., Karplus, M., Neuringer, L. J. (1989) Two-dimensional NMR and photo-CIDNP studies of the insulin monomer: assignment of aromatic resonances with application to protein folding, structure, and dynamics. *Biochemistry 28*, 9855-73.

30. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin Crowfoot, D. M., Hubbard, R. E., Isaacs, N. W., et, a. (1988) The structure of 2-zinc pig insulin crystals at 1.5 Å resolution. *Philos. Trans. R. Soc. London, B 319*, 369-456.

31.   Zoete, V., Meuwly, M., Karplus, M. (2005) Study of the insulin dimerization: Binding free energy calculations and per-residue free-energy decomposition. *Proteins: Struct., Funct., Bioinf. 61*, 79-93.

32.   Olsen, H. B., Ludvigsen, S., Kaarsholm, N. C. (1996) Solution Structure of an Engineered Insulin Monomer at Neutral pH. *Biochemistry 35*, 8836-8845.

33.   Ludvigsen, S., Roy, M., Thøgersen, H., Kaarsholm, N. C. (1994) High-Resolution Structure of an Engineered Biologically Potent Insulin Monomer, B16 Tyr → His, As Determined by Nuclear Magnetic Resonance Spectroscopy. *Biochemistry 33*, 7998-8006.

34.   Ludvigsen, S., Olsen, H. B., Kaarsholm, N. C. (1998) A structural switch in a mutant insulin exposes key residues for receptor binding. *J. Mol. Biol. 279*, 1-7.

35.   Smith, G. D., Swenson, D. C., Dodson, E. J., Dodson, G. G., Reynolds, C. D. (1984) Structural stability in the 4-zinc human insulin hexamer. *Proc. Natl. Acad. Sci. U. S. A. 81*, 7093-7.

36.   Markham, A., Bryson, H. M. (1995) Epoetin alfa: a review of its pharmacodynamic and pharmacokinetic properties and therapeutic use in nonrenal applications. *Drugs 49*, 232-54.

37.   Cheetham, J. C., Smith, D. M., Aoki, K. H., Stevenson, J. L., Hoeffel, T. J., Syed, R. S., Egrie, J., Harvey, T. S. (1998) NMR structure of human erythropoietin and a comparison with its receptor bound conformation. *Nat. Struct. Biol. 5*, 861-866.

38.   Bittorf, T., Jaster, R., Brock, J. (1993) Structural and functional characterisation of recombinant human erythropoietin analogues. *FEBS Lett 336*, 133-6.

39.   Boissel, J.-P., Lee, W.-R., Presnell, S. R., Cohen, F. E., Bunn, H. F. (1993) Erythropoietin structure-function relationships. Mutant proteins that test a model of tertiary structure. *J. Biol. Chem. 268*, 15983-93.

40.   Boissel, J.-P., Bunn, H. F. (1990) Erythropoietin structure-function relationships. *Prog. Clin. Biol. Res. 352*, 227-32.

41.   Sytkowski, A. J., Feldman, L., Zurbuch, D. J. (1991) Biological activity and structural stability of N-deglycosylated recombinant human erythropoietin. *Biochem. Biophys. Res. Commun. 176*, 698-704.

42.   Chern, Y., Chung, T., Sytkowski, A. J. (1991) Structural role of amino acids 99-110 in recombinant human erythropoietin. *Eur. J. Biochem. 202*, 225-9.

43.   Arduini, R. M., Strauch, K. L., Runkel, L. A., Carlson, M. M., Hronowski, X., Foley, S. F., Young, C. N., Cheng, W., Hochman, P. S., Baker, D. P. (1999) Characterization of a soluble ternary complex formed between human interferon-beta-1a and its receptor chains. *Protein Sci 8*, 1867-77.

44.   Mager, D. E., Neuteboom, B., Efthymiopoulos, C., Munafo, A., Jusko, W. J. (2003) Receptor-mediated pharmacokinetics and pharmacodynamics of interferon-β1a in monkeys. *J. Pharmacol. Exp. Ther. 306*, 262-270.

45.   Runkel, L., deDios, C., Karpusas, M., Baker, D., Li, Z., Zafari, M., Betzenhauser, M., Muldowney, C., Miller, S., Redlich, P. N., Grossberg, S. E., Whitty, A., Hochman, P. S. (2001) Mapping of IFN-beta epitopes important for receptor binding and biologic activation: comparison of results achieved using antibody-based methods and alanine substitution mutagenesis. *J Interferon Cytokine Res 21*, 931-41.

46.   Runkel, L., deDios, C., Karpusas, M., Betzenhauser, M., Muldowney, C., Zafari, M., Benjamin, C. D., Miller, S., Hochman, P. S., Whitty, A. (2000) Systematic mutational mapping of sites on human interferon-beta-1a that are important for receptor binding and functional activity. *Biochemistry 39*, 2538-51.

47.   Gutterman, J. U. (1994) Cytokine therapeutics: lessons from interferon α. *Proc. Natl. Acad. Sci. U. S. A. 91*, 1198-205.

48.   Antonetti, F., Finocchiaro, O., Mascia, M., Terlizzese, M. G., Jaber, A. (2002) A comparison of the biologic activity of two recombinant IFN-β preparations used in the treatment of relapsing-remitting multiple sclerosis. *J. Interferon Cytokine Res. 22*, 1181-1184.

49.   Bertolotto, A., Deisenhammer, F., Gallo, P., Sørensen, P. S. (2004) Immunogenicity of interferon beta: differences among products. *J. Neurol. 251*, II/15-II/24.

50.   Durelli, L., Verdun, E., Barbero, P., Bergui, M., Versino, E., Ghezzi, A., Montanari, E., Zaffaroni, M. (2002) Every-other-day interferon beta-1b versus once-weekly interferon beta-1a for multiple sclerosis: results of a 2-year prospective randomised multicentre study (INCOMIN). *Lancet 359*, 1453-1460.

51.   Horowski, R. (2002) Multiple sclerosis and interferon beta-1b, past, present and future. *Clin Neurol Neurosurg 104*, 259-64.

52. Williams, G. J., Witt, P. L. (1998) Comparative study of the pharmacodynamic and pharmacologic effects of Betaseron and AvonexTM. *J. Interferon Cytokine Res. 18*, 967-975.

53. Klaus, W., Gsell, B., Labhardt, A. M., Wipf, B., Senn, H. (1997) The three-dimensional high resolution structure of human interferon α-2a determined by heteronuclear NMR spectroscopy in solution. *J. Mol. Biol. 274*, 661-675.

54. Radhakrishnan, R., Walter, L. J., Hruza, A., Reichert, P., Trotta, P. P., Nagabhushan, T. L., Walter, M. R. (1996) Zinc mediated dimer of human interferon-α2b revealed by x-ray crystallography. *Structure (London) 4*, 1453-1463.

55. Pestka, S., Langer, J. A., Zoon, K. C., Samuel, C. E. (1987) Interferons and their actions. *Annu. Rev. Biochem. 56*, 727-77.

56. Weissmann, C., Weber, H. (1986) The interferon genes. *Prog. Nucleic Acid Res. Mol. Biol. 33*, 251-300.

57. Wetzel, R. (1981) Assignment of the disulfide bonds of leukocyte interferon. *Nature (London) 289*, 606-7.

58. Adolf, G. R., Kalsner, I., Ahorn, H., Maurer-Fogy, I., Cantell, K. (1991) Natural human interferon-α2 is O-glycosylated. *Biochem. J. 276*, 511-18.

59. Mark, D. F., Lu, S. D., Creasey, A. A., Yamamoto, R., Lin, L. S. (1984) Site-specific mutagenesis of the human fibroblast interferon gene. *Proc. Natl. Acad. Sci. U. S. A. 81*, 5662-6.

60. Conradt, H. S., Egge, H., Peter-Katalinic, J., Reiser, W., Siklosi, T., Schaper, K. (1987) Structure of the carbohydrate moiety of human interferon-β secreted by a recombinant Chinese hamster ovary cell line. *J. Biol. Chem. 262*, 14600-5.

61. Karpusas, M., Whitty, A., Runkel, L., Hochman, P. (1998) The structure of human interferon-β. Implications for activity. *Cell. Mol. Life Sci. 54*, 1203-1216.

62. Dill, K. A. (1985) Theory for the folding and stability of globular proteins. *Biochemistry 24*, 1501-9.

63. Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry 29*, 7133-55.

64. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation. *Advances in Protein Chem. (C. B. Anfinsen, M. L. Anson, Kenneth Bailey, and John T. Edsall, editors, Academic Press Inc.) 14*, 1-63.

65. Privalov, P. L. (1979) Stability of proteins. Small globular proteins. *Adv. Protein Chem. 33*, 167-241.

66. Olsen, H. B., Ludvigsen, S., Kaarsholm, N. C. (1998) The relationship between insulin bioactivity and structure in the NH2-terminal A-chain helix. *J. Mol. Biol. 284*, 477-488.

67. Huang, K., Chan, S. J., Hua, Q.-x., Chu, Y.-C., Wang, R.-y., Klaproth, B., Jia, W., Whittaker, J., De Meyts, P., Nakagawa, S. H., Steiner, D. F., Katsoyannis, P. G., Weiss, M. A. (2007) The A-chain of Insulin Contacts the Insert Domain of the Insulin Receptor: Photo-cross-linking and mutagenesis of a diabetes-related crevice. *J. Biol. Chem. 282*, 35337-35349.

68. Skinner, A. L., Laurence, J. S. (2010) Probing residue-specific interactions in the stabilization of proteins using high-resolution NMR: a study of disulfide bond compensation. *J. Pharm. Sci. 99*, 2643-2654.

69. Blundell, T. L., Cutfield, J. F., Dodson, G. G., Dodson, E., Hodgkin, D. C., Mercola, D. (1971) Structure and biology of insulin. *Biochem. J. 125*, 50p-51p.

70. Blundell, T. L., Dodson, E., Dodson, G., Vijayan, M. (1971) Structure of a protein hormone, insulin. *Contemp. Phys. 12*, 209-28.

71. Shoelson, S., Fickova, M., Haneda, M., Nahum, A., Musso, G., Kaiser, E. T., Rubenstein, A. H., Tager, H. (1983) Identification of a mutant human insulin predicted to contain a serine-for-phenylalanine substitution. *Proc. Natl. Acad. Sci. U. S. A. 80*, 7390-4.

72. Záková, L., Kletvíková, E., Veverka, V., Lepsík, M., Watson, C. J., Turkenburg, J. P., Jirácek, J., Brzozowski, A. M. (2013) Structural Integrity of the B24 Site in Human Insulin Is Important for Hormone Functionality. *J. Biol. Chem. 288*, 10230-10240.

73. Hua, Q. X., Shoelson, S. E., Kochoyan, M., Weiss, M. A. (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature (London) 354*, 238-41.

74. Elliott, S., Lorenzini, T., Chang, D., Barzilay, J., Delorme, E., Giffin, J., Hesterberg, L. (1996) Fine-structure epitope mapping of antierythropoietin monoclonal antibodies reveals a model of recombinant human erythropoietin structure. *Blood 87*, 2702-13.

75. Kaarsholm, N. C., Norris, K., Jørgensen, R. J., Mikkelsen, J., Ludvigsen, S., Olsen, O. H., Sørensen, A. R., Havelund, S. (1993) Engineering stability of the insulin monomer fold with application to structure-activity relationships. *Biochemistry 32*, 10773-8.

76. Elliott, S., Lorenzini, T., Chang, D., Barsilay, J., Delorme, E. (1997) Mapping of the active site of recombinant human erythropoietin. *Blood 89*, 493-502.

77. Son, H., Lee, J. H., Chung, T. (1995) None of the four tyrosine residues is essential for the biological activity of erythropoietin. *Arch. Pharmacal Res. 18*, 371-5.

78. Forsyth, W. R., Antosiewicz, J. M., Robertson, A. D. (2002) Empirical relationships between protein structure and carboxyl pKa values in proteins. *Proteins: Struct., Funct., Genet. 48*, 388-403.

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

80. Marti, D. N., Rudolf Bosshard, H. (2003) Electrostatic interactions in leucine zippers: thermodynamic analysis of the contributions of Glu and His residues and the effect of mutating salt bridges. *J. Mol. Biol. 330*, 621-637.

81. Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M., Pace, C. N. (2006) Hydrogen Bonding Markedly Reduces the pK of Buried Carboxyl Groups in Proteins. *J. Mol. Biol. 362*, 594-604.

82. Vinther, T. N., Norrman, M., Ribel, U., Huus, K., Schlein, M., Steensgaard, D. B., Pedersen, T. A., Pettersson, I., Ludvigsen, S., Kjeldsen, T., Jensen, K. J., Hubálek, F. (2013) Insulin analog with additional disulfide bond has increased stability and preserved activity. *Protein Sci. 22*, 296-305.

83. Spek, E. J., Bui, A. H., Lu, M., Kallenbach, N. R. (1998) Surface salt bridges stabilize the GCN4 leucine zipper. *Protein Sci. 7*, 2431-2437.

84. Strop, P., Mayo, S. L. (2000) Contribution of Surface Salt Bridges to Protein Stability. *Biochemistry 39*, 1251-1255.

85. Legg-E'Silva, D., Achilonu, I., Fanucchi, S., Stoychev, S., Fernandes, M., Dirr, H. W. (2012) Role of Arginine 29 and Glutamic Acid 81 Interactions in the Conformational Stability of Human Chloride Intracellular Channel 1. *Biochemistry 51*, 7854-7862.

86. Mitsui, Y., Senda, T., Shimazu, T., Matsuda, S., Utsumi, J. (1993) Structural, functional and evolutionary implications of the three-dimensional crystal structure of murine interferon-β. *Pharmacol. Ther. 58*, 93-132.

87. Kristensen, C., Kjeldsen, T., Wiberg, F. C., Schäffer, L., Hach, M., Havelund, S., Bass, J., Steiner, D. F., Andersen, A. S. (1997) Alanine scanning mutagenesis of insulin. *J. Biol. Chem. 272*, 12978-12983.

88. Grodberg, J., Davis, K. L., Sytkowski, A. J. (1993) Alanine scanning mutagenesis of human erythropoietin identifies four amino acids which are critical for biological activity. *Eur. J. Biochem. 218*, 597-601.

89. Grodberg, J., Davis, K. L., Sytkowski, A. J. (1996) Functional and structural role of arginine 103 in human erythropoietin. *Arch. Biochem. Biophys. 333*, 427-431.

90. Narhi, L. O., Arakawa, T., Aoki, K., Wen, J., Elliott, S., Boone, T., Cheetham, J. (2001) Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of Escherichia coli-derived erythropoietin. *Protein Eng. 14*, 135-140.

91. Boylan, N. J., Zhou, W., Proos, R. J., Tolbert, T. J., Wolfe, J. L., Laurence, J. S. (2013) Conjugation Site Heterogeneity Causes Variable Electrostatic Properties in Fc Conjugates. *Bioconjugate Chem. 24*, 1008-1016.

92. Marki, F., de, G. M., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Sieber, P. (1979) Synthesis and biological activity of seventeen analogues of human insulin. *Hoppe Seylers Z Physiol Chem 360*, 1619-32.

93. Nakagawa, S. H., Tager, H. S. (1991) Implications of invariant residue LeuB6 in insulin-receptor interactions. *J. Biol. Chem. 266*, 11502-9.

94. Senda, T., Saitoh, S.-i., Mitsui, Y. (1995) Refined crystal structure of recombinant murine interferon-β at 2.15 Å resolution. *J. Mol. Biol. 253*, 187-207.

95. Schwartz, G. P., Burke, G. T., Katsoyannis, P. G. (1987) A superactive insulin: [B10-aspartic acid]insulin(human). *Proc. Natl. Acad. Sci. U. S. A. 84*, 6408-11.

96. Guo, Z.-Y., Tang, Y.-H., Wang, S., Feng, Y.-M. (2003) Contribution of the absolutely conserved B8Gly to the foldability of insulin. *Biol. Chem. 384*, 805-809.

97. Guo, Z.-Y., Zhang, Z., Jia, X.-Y., Tang, Y.-H., Feng, Y.-M. (2005) Mutational analysis of the absolutely conserved B8Gly: Consequence on foldability and activity of insulin. *Acta Biochim. Biophys. Sin. 37*, 673-679.

98. Weiss, M. A. (2009) The structure and function of insulin: decoding the TR transition. *Vitam. Horm. (San Diego, CA, U. S.) 80*, 33-49.

99. Chang, S.-G., Choi, K.-D., Jang, S.-H., Shin, H.-C. (2003) Role of disulfide bonds in the structure and activity of human insulin. *Mol. Cells 16*, 323-330.

100. Guo, Z.-Y., Feng, Y.-M. (2001) Effects of cysteine to serine substitutions in the two inter-chain disulfide bonds of insulin. *Biol. Chem. 382*, 443-448.

101. Dai, Y., Tang, J.-G. (1994) Intra-A chain disulfide bond (A6-11) of insulin is essential for displaying its activity. *Biochem. Mol. Biol. Int. 33*, 1049-53.

102. Dai, Y., Tang, J.-G. (1996) Characteristic, activity and conformational studies of [A6-Ser, A11-Ser]-insulin. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol. 1296*, 63-68.

103. Hua, Q.-X., Hu, S.-Q., Frank, B. H., Jia, W., Chu, Y.-C., Wang, S.-H., Burke, G. T., Katsoyannis, P. G., Weiss, M. A. (1996) Mapping the functional surface of insulin by design: structure and function of a novel A-chain analogue. *J Mol Biol 264*, 390-403.

104. Hua, Q.-X., Gozani, S. N., Chance, R. E., Hoffmann, J. A., Frank, B. H., Weiss, M. A. (1995) Structure of a protein in a kinetic trap. *Nat. Struct. Biol. 2*, 129-38.

105. Weiss, M. A., Hua, Q.-X., Jia, W., Chu, Y.-C., Wang, R.-Y., Katsoyannis, P. G. (2000) Hierarchical Protein "Un-Design": Insulin's Intrachain Disulfide Bridge Tethers a Recognition α-Helix. *Biochemistry 39*, 15429-15440.

106. Morehead, H., Johnston, P. D., Wetzel, R. (1984) Roles of the 29-138 disulfide bond of subtype A of human α interferon in its antiviral activity and conformational stability. *Biochemistry 23*, 2500-7.

107. Uzé, G., Di Marco, S., Mouchel-Vielh, E., Monneron, D., Bandu, M.-T., Horisberger, M. A., Dorques, A., Lutfalla, G., Mogensen, K. E. (1994) Domains of interaction between alpha interferon and its receptor components. *J. Mol. Biol. 243*, 245-57.

108. Shepard, H. M., Leung, D., Stebbing, N., Goeddel, D. V. (1981) A single amino acid change in interferon (IFN)-β1 abolishes its antiviral activity. *Nature (London) 294*, 563-5.

109. Runkel, L., Pfeffer, L., Lewerenz, M., Monneron, D., Yang, C. H., Murti, A., Pellegrini, S., Goelz, S., Uzé, G., Mogensen, K. (1998) Differences in activity between α and β type I interferons explored by mutational analysis. *J. Biol. Chem. 273*, 8003-8008.

110. Fu, X., Li, W., Mao, Q., Chang, Z. (2003) Disulfide bonds convert small heat shock protein Hsp16.3 from a chaperone to a non-chaperone: implications for the evolution of cysteine in molecular chaperones. *Biochem. Biophys. Res. Commun. 308*, 627-635.

111. Wang, X., Kumar, S., Singh, S. K. (2011) Disulfide Scrambling in IgG2 Monoclonal Antibodies: Insights from Molecular Dynamics Simulations. *Pharm. Res. 28*, 3128-3144.

112. Way, J. C., Lauder, S., Brunkhorst, B., Kong, S.-M., Qi, A., Webster, G., Campbell, I., McKenzie, S., Lan, Y., Marelli, B., Nguyen, L. A., Degon, S., Lo, K.-M., Gillies, S. D. (2005) Improvement of Fc-erythropoietin structure and pharmacokinetics by modification at a disulfide bond. *Protein Eng., Des. Sel. 18*, 111-118.

113. Gong, R., Vu, B. K., Feng, Y., Prieto, D. A., Dyba, M. A., Walsh, J. D., Prabakaran, P., Veenstra, T. D., Tarasov, S. G., Ishima, R., Dimitrov, D. S. (2009) Engineered Human Antibody Constant Domains with Increased Stability. *J. Biol. Chem. 284*, 14203-14210.

114. Hagihara, Y., Mine, S., Uegaki, K. (2007) Stabilization of an Immunoglobulin Fold Domain by an Engineered Disulfide Bond at the Buried Hydrophobic Region. *J. Biol. Chem. 282*, 36489-36495.

115. Hussack, G., Hirama, T., Ding, W., MacKenzie, R., Tanha, J. (2011) Engineered single-domain antibodies with high protease resistance and thermal stability. *PLoS One 6*, e28218.

116. Saerens, D., Conrath, K., Govaert, J., Muyldermans, S. (2008) Disulfide Bond Introduction for General Stabilization of Immunoglobulin Heavy-Chain Variable Domains. *J. Mol. Biol. 377*, 478-488.

117. Peters, S. J., Smales, C. M., Henry, A. J., Stephens, P. E., West, S., Humphreys, D. P. (2012) Engineering an Improved IgG4 Molecule with Reduced Disulfide Bond Heterogeneity and Increased Fab Domain Thermal Stability. *J. Biol. Chem. 287*, 24525-24533.

118. Schur, P. H. (1972) Human gamma-G subclasses. *Progr. Clin. Immunol. 1*, 71-104.

119. Betz, S. F. (1993) Disulfide bonds and the stability of globular proteins. *Protein Sci. 2*, 1551-8.

120. Wells, J. A., Powers, D. B. (1986) In vivo formation and stability of engineered disulfide bonds in subtilisin. *J. Biol. Chem. 261*, 6564-70.

121. Katz, B. A., Kossiakoff, A. (1986) The crystallographically determined structures of atypical strained disulfides engineered into subtilisin. *J. Biol. Chem. 261*, 15480-5.

122. Betz, S. F., Marmorino, J. L., Saunders, A. J., Doyle, D. F., Young, G. B., Pielak, G. J. (1996) Unusual effects of an engineered disulfide on global and local protein stability. *Biochemistry 35*, 7422-7428.

123. Dani, V. S., Ramakrishnan, C., Varadarajan, R. (2003) MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins. *Protein Eng. 16*, 187-193.

124. Melnik, B. S., Povarnitsyna, T. V., Glukhov, A. S., Melnik, T. N., Uversky, V. N. (2012) SS-stabilizing proteins rationally: intrinsic disorder-based design of stabilizing disulphide bridges in GFP. *J. Biomol. Struct. Dyn. 29*, 815-824.

125. Yu, X.-W., Tan, N.-J., Xiao, R., Xu, Y. (2012) Engineering a disulfide bond in the lid hinge region of Rhizopus chinensis lipase: Increased thermostability and altered acyl chain length specificity. *PLoS One 7*, e46388.

126. Craig, D. B., Dombkowski, A. A. (2013) Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics 14*, 346.

127. Toyoda, T., Arakawa, T., Yamaguchi, H. (2002) N-glycans stabilize human erythropoietin through hydrophobic interactions with the hydrophobic protein surface: studies by surface plasmon resonance analysis. *J. Biochem. 131*, 511-515.

128. Kodama, S., Tsujimoto, M., Tsuruoka, N., Sugo, T., Endo, T., Kobata, A. (1993) Role of sugar chains in the in-vitro activity of recombinant human interleukin 5. *Eur. J. Biochem. 211*, 903-8.

129. Wang, C., Eufemi, M., Turano, C., Giartosio, A. (1996) Influence of the Carbohydrate Moiety on the Stability of Glycoproteins. *Biochemistry 35*, 7299-7307.

130. Ceaglio, N., Etcheverrigaray, M., Kratje, R., Oggero, M. (2010) Influence of carbohydrates on the stability and structure of a hyperglycosylated human interferon alpha mutein. *Biochimie 92*, 971-978.

131. Hinds, K., Koh, J. J., Joss, L., Liu, F., Baudyš, M., Kim, S. W. (2000) Synthesis and Characterization of Poly(ethylene glycol)-Insulin Conjugates. *Bioconjugate Chem. 11*, 195-201.

132. Hinds, K. D., Kim, S. W. (2002) Effects of PEG conjugation on insulin properties. *Adv. Drug Delivery Rev. 54*, 505-530.

133. Zhang, M., Dou, H., Yin, L., Zhang, Y., Zhu, S., Yin, C. (2009) Preparation and characterization of monomethoxypoly(ethylene glycol)-insulin conjugates. *Pharmazie 64*, 190-196.

134. Sinclair, A. M., Elliott, S. (2005) Glycoengineering: The effect of glycosylation on the properties of therapeutic proteins. *J. Pharm. Sci. 94*, 1626-1635.

135. Elliott, S., Chang, D., Delorme, E., Eris, T., Lorenzini, T. (2004) Structural requirements for additional N-linked carbohydrate on recombinant human erythropoietin. *J. Biol. Chem. 279*, 16854-16862.

136. Elliott, S., Lorenzini, T., Asher, S., Aoki, K., Brankow, D., Buck, L., Busse, L., Chang, D., Fuller, J., Grant, J., Hernday, N., Hokum, M., Hu, S., Knudten, A., Levin, N., Komorowski, R., Martin, F., Navarro, R., Osslund, T., Rogers, G., Rogers, N., Trail, G., Egrie, J. (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat. Biotechnol. 21*, 414-421.

137. Kalies, K.-U., Hartmann, E. (1998) Protein translocation into the endoplasmic reticulum (ER) Two similar routes with different modes. *Eur. J. Biochem. 254*, 1-5.

138. Kiely, M. L., McKnight, G. S., Schimke, R. T. (1976) Studies on the attachment of carbohydrate to ovalbumin nascent chains in hen oviduct. *J. Biol. Chem. 251*, 5490-5.

139. Nilsson, I., von Heijne, G. (1993) Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem. 268*, 5798-801.

140. Rothman, S. S. (1975) Protein transport by the pancreas. *Science 190*, 747-53.

141. Helenius, A., Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem. 73*, 1019-1049.

142. Wasley, L. C., Timony, G., Murtha, P., Stoudemire, J., Dorner, A. J., Caro, J., Krieger, M., Kaufman, R. J. (1991) The importance of N- and O-linked oligosaccharides for the biosynthesis and in vitro and in vivo biologic activities of erythropoietin. *Blood 77*, 2624-32.

143. Wilson Johnston, M. J., Frahm, G., Li, X., Durocher, Y., Hefford, M. A. (2011) O-linked glycosylation leads to decreased thermal stability of interferon alpha 2b as measured by two orthogonal techniques. *Pharm. Res. 28*, 1661-1667.

144. Wang, Y.-J., Liu, Y.-D., Chen, J., Hao, S.-J., Hu, T., Ma, G.-H., Su, Z.-G. (2010) Efficient preparation and PEGylation of recombinant human non-glycosylated erythropoietin expressed as inclusion body in E. coli. *Int. J. Pharm. 386*, 156-164.

145. Baudyš, M., Uchio, T., Hovgaard, L., Zhu, E. F., Avramoglou, T., Jozefowicz, M., Říhová, B., Park, J. Y., Lee, H. K., et, a. (1995) Glycosylated insulins. *J. Controlled Release 36*, 151-7.

146. Baudyš, M., Uchio, T., Mix, D., Wilson, D., Kim, S. W. (1995) Physical Stabilization of Insulin by Glycosylation. *J. Pharm. Sci. 84*, 28-33.

147. Utsumi, J., Yamazaki, S., Hosoi, K., Shimizu, H., Kawaguchi, K., Inagaki, F. (1986) Conformations of fibroblast and E. coli-derived recombinant human interferon-βs as studied by nuclear magnetic resonance and circular dichroism. *J. Biochem. 99*, 1533-5.

148. Solá, R. J., Rodríguez-Martínez, J. A., Griebenow, K. (2007) Modulation of protein biophysical properties by chemical glycosylation: biochemical insights and biomedical implications. *Cell. Mol. Life Sci. 64*, 2133-2152.

149. Smith Dordal, M., F. F., Goldwasser, E. (1985) The role of carbohydrate in erythropoietin action. *Endocrinology (Baltimore) 116*, 2293-9.

150. Toyoda, T., Itai, T., Arakawa, T., Aoki, K. H., Yamaguchi, H. (2000) Stabilization of human recombinant erythropoietin through interactions with the highly branched N-glycans. *J. Biochem. 128*, 731-737.

151. Krapp, S., Mimura, Y., Jefferis, R., Huber, R., Sondermann, P. (2003) Structural Analysis of Human IgG-Fc Glycoforms Reveals a Correlation Between Glycosylation and Structural Integrity. *J. Mol. Biol. 325*, 979-989.

152. Mimura, Y., Church, S., Ghirlando, R., Ashton, P. R., Dong, S., Goodall, M., Lund, J., Jefferis, R. (2000) The influence of glycosylation on the thermal stability and effector function expression of human IgG1-Fc: properties of a series of truncated glycoforms. *Mol. Immunol. 37*, 697-706.

153. Spiriti, J., Bogani, F., van der Vaart, A., Ghirlanda, G. (2008) Modulation of protein stability by O-glycosylation in a designed Gc-MAF analog. *Biophys. Chem. 134*, 157-167.

154. Ghasriani, H., Belcourt, P. J. F., Sauvé, S., Hodgson, D. J., Brochu, D., Gilbert, M., Aubin, Y. (2013) A Single N-Acetylgalactosamine Residue at Threonine 106 Modifies the Dynamics and Structure of Interferon α2a around the Glycosylation Site. *J. Biol. Chem. 288*, 247-254.

155. Shire, S. J., Shahrokh, Z., Liu, J. (2004) Challenges in the development of high protein concentration formulations. *J. Pharm. Sci. 93*, 1390-1402.

156. Kern, G., Kern, D., Jaenicke, R., Seckler, R. (1993) Kinetics of folding and association of differently glycosylated variants of invertase from Saccharomyces cerevisiae. *Protein Sci. 2*, 1862-8.

157. Paul, G., Lottspeich, F., Wieland, F. (1986) Asparaginyl-N-acetylgalactosamine. Linkage unit of halobacterial glycosaminoglycan. *J. Biol. Chem. 261*, 1020-4.

158. Brange, J., Dodson, G. G., Edwards, D. J., Holden, P. H., Whittingham, J. L. (1997) A model of insulin fibrils derived from the X-ray crystal structure of a monomeric insulin (despentapeptide insulin). *Proteins: Struct., Funct., Genet. 27*, 507-516.

159. Fazeli, A., Shojaosadati, S. A., Fazeli, M. R., Khalifeh, K., Ariaeenejad, S., Moosavi-Movahedi, A. A. (2013) The role of trehalose for metastable state and functional form of recombinant interferon beta-1b. *J. Biotechnol. 163*, 318-324.

160. Tsuda, E., Kawanishi, G., Ueda, M., Masuda, S., Sasaki, R. (1990) The role of carbohydrate in recombinant human erythropoietin. *Eur. J. Biochem. 188*, 405-11.

161. Dissing-Olesen, L., Thaysen-Andersen, M., Meldgaard, M., Højrup, P., Finsen, B. (2008) The function of the human interferon-β1a glycan determined in vivo. *J. Pharmacol. Exp. Ther. 326*, 338-347.

162. Bailon, P., Palleroni, A., Schaffer, C. A., Spence, C. L., Fung, W.-J., Porter, J. E., Ehrlich, G. K., Pan, W., Xu, Z.-X., Modi, M. W., Farid, A., Berthold, W., Graves, M. (2001) Rational Design of a Potent, Long-Lasting Form of Interferon: A 40 kDa Branched Polyethylene Glycol-Conjugated Interferon α-2a for the Treatment of Hepatitis C. *Bioconjugate Chem. 12*, 195-202.

163. Basu, A., Yang, K., Wang, M., Liu, S., Chintala, R., Palm, T., Zhao, H., Peng, P., Wu, D., Zhang, Z., Hua, J., Hsieh, M.-C., Zhou, J., Petti, G., Li, X., Janjua, A., Mendez, M., Liu, J., Longley, C., Zhang, Z., Mehlig, M., Borowski, V., Viswanathan, M., Filpula, D. (2006) Structure-Function Engineering of Interferon-β-1b for Improving Stability, Solubility, Potency, Immunogenicity, and Pharmacokinetic Properties by Site-Selective Mono-PEGylation. *Bioconjugate Chem.* *17*, 618-630.

164. Grace, M., Youngster, S., Gitlin, G., Sydor, W., Xie, L., Westreich, L., Jacobs, S., Brassard, D., Bausch, J., Bordens, R. (2001) Structural and biologic characterization of pegylated recombinant IFN-α2b. *J. Interferon Cytokine Res.* *21*, 1103-1115.

165. Wang, Y.-S., Youngster, S., Grace, M., Bausch, J., Bordens, R., Wyss, D. F. (2002) Structural and biological characterization of pegylated recombinant interferon alpha-2b and its therapeutic implications. *Adv. Drug Delivery Rev.* *54*, 547-570.

166. Rosendahl, M. S., Doherty, D. H., Smith, D. J., Carlson, S. J., Chlipala, E. A., Cox, G. N. (2005) A Long-Acting, Highly Potent Interferon α-2 Conjugate Created Using Site-Specific PEGylation. *Bioconjugate Chem.* *16*, 200-207.

167. Uchio, T., Baudyš, M., Liu, F., Song, S. C., Kim, S. W. (1999) Site-specific insulin conjugates with enhanced stability and extended action profile. *Adv. Drug Delivery Rev.* *35*, 289-306.

168. Caliceti, P., Veronese, F. M. (1999) Improvement of the physicochemical and biopharmaceutical properties of insulin by polyethylene glycol conjugation. *S.T.P. Pharma Sci.* *9*, 107-113.

169. Lim, A., Wally, J., Walsh, M. T., Skinner, M., Costello, C. E. (2001) Identification and location of a cysteinyl posttranslational modification in an amyloidogenic ϰ1 light chain protein by electrospray ionization and matrix-assisted laser desorption/ionization mass spectrometry. *Anal. Biochem.* *295*, 45-56.

170. Mawatari, S., Murakami, K. (2004) Different types of glutathionylation of hemoglobin can exist in intact erythrocytes. *Arch. Biochem. Biophys.* *421*, 108-114.

171. Ramon, J., Saez, V., Baez, R., Aldana, R., Hardy, E. (2005) PEGylated Interferon-α2b: A Branched 40K Polyethylene Glycol Derivative. *Pharm. Res.* *22*, 1374-1386.

172. Brange, J., Andersen, L., Laursen, E. D., Meyn, G., Rasmussen, E. (1997) Toward Understanding Insulin Fibrillation. *J. Pharm. Sci.* *86*, 517-525.

173. Chapman, A. P. (2002) PEGylated antibodies and antibody fragments for improved therapy: a review. *Adv. Drug Delivery Rev. 54*, 531-545.

CHAPTER 2.
*cla*MP TAG: A VERSATILE INLINE METAL-BINDING PLATFORM BASED ON THE METAL ABSTRACTION PEPTIDE

2.1 INTRODUCTION

Metals are extremely useful tools for clinical and biotechnology applications. Metal-based diagnostic imaging is common and its applications include molecular targeting to identify specific cell populations.[1-2] Addition of metal-binding modalities to proteins has been accomplished primarily through chemical conjugation.[3] Attaching a chelating agent requires use of a secondary chemistry, which leads to a heterogeneous mixture of products and requires optimization of the reaction conditions and additional purification steps to recover the desired product. [3-6] These reactions are inefficient, utilizing significant excesses of materials, and significantly increase the time needed to obtain a purified final product.

Incorporation of a peptide-based metal-binding tag into a protein during expression to generate an inline carrier eliminates the need for subsequent modification, greatly simplifying the production of metal-based conjugates. Inline lanthanide-binding peptide tags (LBTs) have been developed for binding to larger metal ions, such as gadolinium, demonstrating the validity of this principle.[7-10] Lanthanide binding tags have found utility in a variety of applications, including imaging and biotechnology research. They are large (~14-20 amino acids) because they must contain six residues positioned appropriately to enable binding lanthanide ions (Tb(III), Gd(III), Dy(III))[7] using eight oxygen ligands. The presence of a ninth exchangeable water ligand allows for the application of LBTs as MRI contrast agents.[7, 9] In the laboratory, LBTs are used in luminescence energy transfer studies[10] and for luminescence-based detection on gels.[7] In addition, paramagnetic lanthanide ions are incorporated into LBTs for use as magnetic-field relaxation and alignment agents in NMR spectroscopy for evaluating protein

structure.[8, 11]   Lanthanide binding tags have found large utility as imaging agents, however, lanthanides are reported to induce significant adverse effects in some patient populations,[12-14] indicating additional approaches are needed.   Utilization of more biologically compatible transition metals is desirable, but the inability to achieve appropriate binding to these smaller metal ions by traditional chelators has hindered their use in the clinic.[15-17]   Nature typically accomplishes tight binding and control of transition metals using sophisticated protein scaffolds in which the metal is well protected from facile exchange. These highly structured environments are absent in small peptides, which typically results in weaker binding and easier release of the metal to high affinity chelators, such as EDTA. Thus a tag capable of sequestering smaller metals would enable new tools and provide opportunity to expand therapeutic and diagnostic applications.

Our lab discovered the metal abstraction peptide (MAP) and its novel chemistry that binds transition metals and develops a uniquely structured product with compatible properties for use in clinical applications.[18] The structure is extremely stable under basic conditions, resisting metal release and loss to high affinity chelators in competition experiments, but undergoes rearrangement upon acidification, permitting rapid exchange or release of the metal (manuscript in preparation). MAP is composed of a three amino acid sequence (Asn-Cys-Cys; NCC), and the metal is coordinated by two thiolate ligands and two deprotonated amide nitrogens from the peptide backbone in a square planar organization when inserted inline with a protein sequence.[19] Because the sequence is composed of natural amino acids, it can be engineered as a tag into a genetic construct. The desired protein product encodes the inline metal carrier, which we refer to as the *cla*MP Tag. In order to generate the specific metal-bound peptide structure having these unique properties, the *cla*MP Tag must be reacted with a metal ion that is partially coordinated

by a chelator (Fig. 2.1), such as those used in immobilized metal affinity chromatography (IMAC), e.g. nitrilotriacetic acid (NTA). The peptide abstracts the metal ion from the high-affinity chelator; this reaction is extremely efficient and yields a highly stable, single product.[19-21]
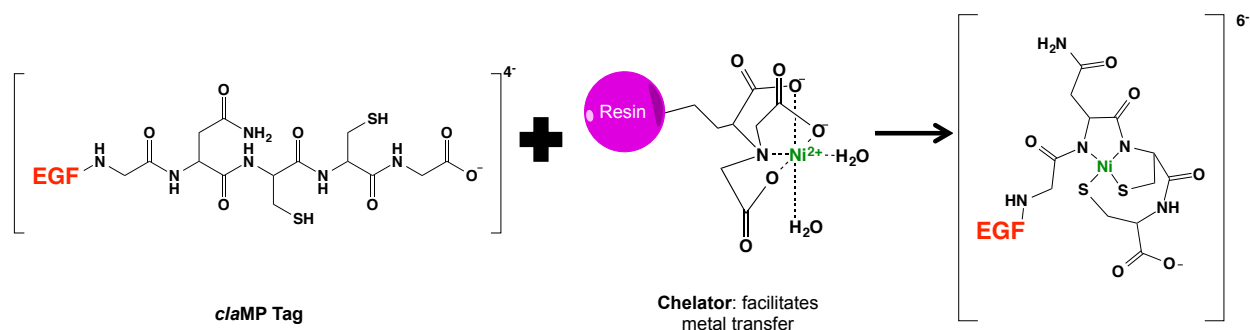
Figure 2.1 Schematic that illustrates the unique method by which metal is inserted into the *cla*MP Tag. The *cla*MP Tag is reacted with a chelated metal ion (i.e. Ni-NTA), which leads to metal insertion into the tag. A metal ion with a 2^+ charge, such as Ni(II), reacted with the net neutral *cla*MP Tag sequence generates a single product with a charge of 2^-. EGF has a 4^- charge, resulting in a net charge of 6^- for the inline *cla*MP Tag conjugate.

Here, epidermal growth factor (EGF) was used as a model system to demonstrate that the *cla*MP sequence can be used effectively as a tag. EGF contains six cysteine residues, which comprise three disulfide bonds, and its expression is known to require the assistance of disulfide modulating enzymes to fold correctly.[22-25] Because the *cla*MP Tag itself contains two cysteine residues, these additional thiols could complicate the folding process by increasing the number of possible folded states. EGF was chosen because it exemplifies the extremes of protein systems to which the *cla*MP Tag may be applied. Production of EGF as a thioredoxin-fusion protein was performed to assess the ability to include the *cla*MP Tag in a thiol- and disulfide-containing fusion protein. Because the Ni-MAP complex has catalytic superoxide dismutase (SOD) activity,[19] we confirmed that the SOD activity was retained when the Ni-*cla*MP complex was incorporated inline with the EGF sequence and used this activity to quantitatively confirm formation of the desired product. Here, the effects of the tag and its placement within EGF on

expression, structure and native function of EGF are presented to demonstrate the compatibility and versatility of the *cla*MP Tag for use as an inline metal carrier.

2.2 MATERIALS AND METHODS

2.2.1 Cloning and Construction of the Expression Plasmid

Three DNA sequences were prepared to generate EGF variants: control EGF and two individually tagged variants containing the *cla*MP Tag at either the N-terminus (GNCCG-EGF) or C-terminus (EGF-GNCCG) of the native sequence. A plasmid containing human EGF was obtained (Origene, #RC210817), and the EGF sequence was amplified using PCR and subcloned into the pET-32Xa/LIC vector. GNCCG was added to either the N-terminus or C-terminus of EGF and amplified using PCR for insertion into a LIC plasmid using the primers below (IDT). In each primer, the LIC sequence is underlined and the portion corresponding to EGF and the claMP Tag are not. 5'-<u>GGTATTGAGGGTCGCG</u>GAAACTGCTGCGGCAATAGTGACT CTGAATGTCCC-3' (forward primer with tag on the N-terminus), 5'<u>AGAGGAGAGTTAGA GCCG</u>TCAGCGCAGTTCCCACCACTTCAG-3' (reverse primer for N-terminally tagged construct),   5'-<u>GGTATTGAGGGTCGC</u>AATAGTGACTCTGAATGTCCCCTGTCCCACGAT GGG (forward primer for C-terminal tagged construct) and 5'-<u>AGAGGAGAGTTAGAGCCG</u> TCAGCCGCAGCAGTTTCCGCGCAGTTCCCACCACTTCAGGTC-3' (reverse primer with tag on the C-terminus). The amplification reactions were purified using the QIAquick PCR Purification Kit (Qiagen). The fragments were inserted into the vector using the manufacturer's protocol from the Xa/LIC cloning kit, according to the manufacturer's instructions (Novagen). These plasmids contain a Thioredoxin tag, a His6 tag, thrombin cleavage site, S-tag, and Factor Xa site, which are all positioned on the N-terminal side of each EGF variant (Fig. 2.2). The ligation product was transformed into the DH5α *Escherichia coli* cell strain using the standard

heat shock method, and the DNA was harvested after selection and growth on LB containing 100 µg/mL ampicillin. Plasmid DNA was purified using a miniprep kit (Qiagen), and the intended product was verified by DNA sequencing (UC Berkeley DNA Sequencing Facility).

2.2.2 Protein Expression and Purification

Expression and purification of the protein from these plasmids was completed in a similar manner as previously reported for EGF.[26] Each plasmid was transformed into the Origami B (DE3) *E. coli* strain (Novagen) using the standard heat shock protocol. Colonies were grown for ~24 hours at 37 °C on LB agar plates with 100 µg/mL ampicillin, 30 µg/mL kanamycin, and 12.5 µg/mL tetracycline. To prepare ^{15}N-labeled cultures, individual colonies were selected to inoculate 175 mL M9ZB media supplemented with 0.4% glucose, 1 mM $MgSO_4$, and 100 µg/mL ampicillin. These starter cultures were grown for 16 hours at 37 °C, 250 rpm. Fifty milliliters of starter culture was used to inoculate 1 L minimal media containing minimal salts and trace minerals.[27] To prepare non-labeled cultures, individual colonies were selected to inoculate 50 mL LB media containing 100 µg/mL ampicillin. Again, starter cultures were grown for 16 hours at 37 °C, 250 rpm. Twenty milliliters starter culture was used to inoculate 1 L LB supplemented with 100 µg/mL ampicillin. Cultures of ^{15}N-labeled media and non-labeled media were allowed to incubate at 200 rpm and 37 °C until an OD_{600} of 0.7 was met. At this point, protein expression was induced through the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The cultures were allowed to incubate for an additional 16 hours at 25 °C, 200 rpm, and then they were harvested by ultracentrifugation. Cell pellets were stored at -80 °C until use.

Each pellet (1 L of culture) was resuspended in 25 mL 50 mM Tris-Cl, 20 mM imidazole, pH 7.9 and lysed using a French Press at 21,000 psi. The cellular debris was removed by centrifugation at 21,000 x g for 1 hour at 4 °C. The supernatant, which contained the protein,

was filtered through a 0.45 μm filter and applied to a 5 mL Hi-Trap Chelating HP column charged with nickel (GE Lifesciences). EGF constructs containing the *cla*MP Tag were allowed to incubate on the column for one hour to facilitate metal transfer to the tag. Cellular proteins were removed by washing the column with six column volumes (CV) of 50 mM Tris-Cl, 40 mM imidazole, 500 mM NaCl, pH 7.9. The protein was eluted from the column using a linear gradient from 0-100% 20 mM sodium phosphate (NaPi), 500 mM NaCl, 500 mM imidazole, pH 9.5 over 12 CV. The thioredoxin tag was cleaved using 10 units thrombin (FisherScientific) per mg of protein at 16 °C for 9.5 hours, and the two cleavage fragments were separated on a Superdex75 column (GE Lifesciences). The remaining tags were cleaved using 9 units Factor Xa (Novagen) per mg of protein at 25 °C for 16 hours, and the two cleavage fragments were separated on a Superdex75 column. Samples were concentrated using an Amicon Ultra 3 kDa MWCO concentrator (Millipore) to a final concentration of approximately 0.15 mM. The purity of the protein was confirmed with SDS-PAGE and size exclusion chromatography (TSKgel G3000SW, 7.8 mm x 30 cm, 5 μm particle size, Fisher Scientific) and the molecular weight was validated with ESI-MS. Samples were quantified using a Bradford assay.

2.2.3 SDS-PAGE Analysis

Tris-tricine gels were prepared as reported.[28] EGF constructs were separated using a discontinuous system consisting of a 4 % (v/v) stacking gel and a 18 % (v/v) resolving gel. The samples were prepared in non-reducing Laemmli buffer and heated for 10 minutes at 90 °C before being loaded onto the gel. A prestained, dual-color molecular weight marker was used for reference (BioRad, #161-0374). Gels were stained using Coomassie (R-250).

Densitometry analysis was performed on Coomassie-stained tris-tricine gels using the Typhoon TRIO Variable Mode Imager (Amersham Biosciences). Relative quantitation was

performed using the ImageQuantTL software (Amersham Biosciences). The amount of *cla*MP-Tagged EGF was determined based upon the intensity observed for native EGF.

2.2.4 Absorption Analysis

After purification, samples were diluted and analyzed using electronic absorption spectroscopy to confirm Ni(II) incorporation. Samples were placed in a cuvette with a 1-cm path length and spectra were acquired from 200-800 nm using a Carey 100 Bio UV-visible spectrophotometer (Varian).

2.2.5 HPLC Analysis

Anion exchange chromatography (AEC) was performed on a 4 x 250 mm BioLC ProPac Wax10 column (Dionex). The column was equilibrated with 20 mM Tris-Cl, 10 mM KCl, pH 7.5 before injection of the sample. A linear gradient from 0-100% 20 mM Tris-Cl, 500 mM KCl, pH 7.5 over 70 mL with a constant flow rate of 1 mL/min was used to elute the protein from the column and detected by absorbance at 220 nm. The sample volume used for all injections was 20 μL, and each sample was run in duplicate.

2.2.6 ^{1}H-^{15}N HSQC Analysis

Two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra were obtained at 25 °C using a 600 MHz Bruker Avance NMR spectrometer with a cryogenic, triple resonance probe. ^{15}N-labeled protein was in prepared 50 mM KPi, 10 mM NaCl, pH 7.3, with 6% D_2O. A Bradford assay was used to determine the concentration of the samples. Data were obtained with 177 and 217 scans for EGF and EGF-Ni-*cla*MP, respectively. A different number of scans were used to account for variations in sample concentrations and achieve a similar signal to noise ratio for each 2D spectrum. Analysis of the data was performed using NMRPipe[29] and the NMR assignment program Sparky.[30]

2.2.7 Xanthine/Xanthine Oxidase SOD Activity Assay

This assay was performed as previously described.[19, 31-32] Briefly, 600 μM cytochrome c from bovine heart and 300 μM xanthine were added to 50 mM potassium phosphate with 100 μM EDTA, pH 7.8 to yield final concentrations of 10 μM cytochrome c and 50 μM xanthine. Enough xanthine oxidase was added to cause a change in absorbance at 550 nm of 0.02 to 0.04 AU/min. Several concentrations were analyzed to develop an IC_{50} curve.

2.2.8 Cell Culture

Human epidermoid carcinoma cells (A431, ATCC CRL-1555) were purchased from American Type Culture Collection (ATCC) and grown in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin. Cells were cultured in an incubator at a constant temperature of 37 °C and 5 % CO_2.

2.2.9 EGF Cell Viability Assay

A431 cells were seeded in a 96-well clear bottom black plate at a density of 10,000 cells per well and incubated overnight in the growth conditions described above. The following day, the media was replaced with media containing the test samples and controls. Test samples were diluted into media at the appropriate concentrations and then transferred onto the cells. Nine different concentrations were used, and each sample was run in triplicate. The cells were incubated with EGF variants for 72 hours, when the control cells reached confluency. Cell viability was measured using the CellTiter-Blue assay (Promega) following manufacturer's protocol. Briefly, the media was replaced with cell culture medium containing reagent and allowed to incubate for 30 minutes. Following this incubation period, a microplate reader (SpectraMax GeminiXS) was used to determine the fluorescence intensity of the dye

$(560_{ex}/590_{em})$. The fluorescence intensity was compared to control wells to determine percent viability.

2.3 RESULTS

2.3.1 *cla*MP Tag Addition Does Not Impair EGF Expression

Because native EGF contains three disulfide bonds, it was important to determine the extent to which addition of the two cysteine residues in the *cla*MP Tag might affect protein expression or lead to non-native disulfide bond formation and misfolding in an *E. coli* system. Native EGF, like many disulfide-containing proteins, accumulates in inclusion bodies when expressed into the reducing cytosolic environment of *E. coli* [24, 33] and addition of the *cla*MP Tag did not affect this outcome (unpublished data). By expressing EGF in an engineered strain that contains a more oxidizing cytoplasmic environment, proper folding is achieved (Fig. 2.2).[26, 34] Using this approach, EGF and *cla*MP-Tagged EGF variants were produced in the soluble fraction of the cell lysate.
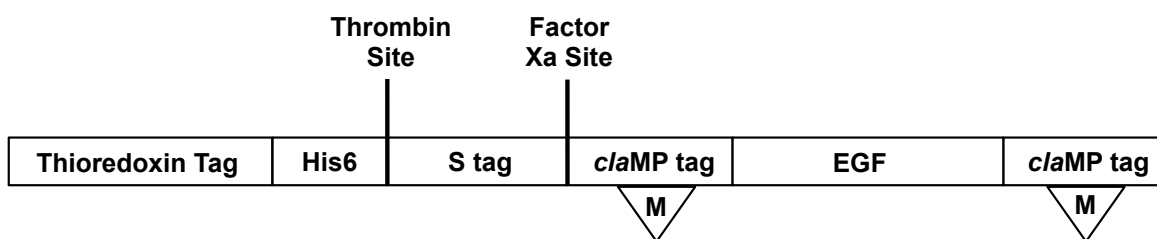
Figure 2.2 Cartoon of pET-32 expression construct of recombinant *cla*MP-Tagged EGF.

SDS-PAGE was used to verify the *cla*MP Tag did not adversely affect expression of *cla*MP-Tagged EGF; all EGF variants showed excellent expression and reasonably similar expression levels (Fig. 2.3). The molecular weight of each fusion protein was expected to be approximately 24 kDa. As shown in Figure 2.3a, no band is present at this size in the pre-induction samples (lanes 2, 4, and 6), but a band appears in the post-induction sample for each of
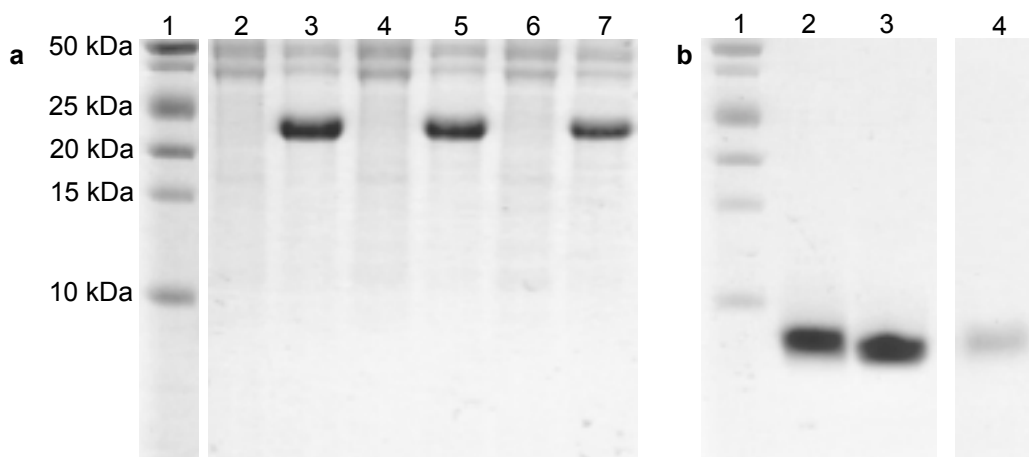
Figure 2.3 SDS-PAGE analysis of *cla*MP-Tagged EGF expression and final product. Addition of *cla*MP Tag to EGF did not significantly hinder expression or purification of desired product. Coomassie stained 18 % tris-tricine gels illustrating the expression of pET-32-EGF, pET-32-*cla*MP-EGF, and pET-32-EGF-*cla*MP and final purified EGF products. (a) (Lane 1) molecular weight standard, (lanes 2, 4, 6) *E. coli* lysate of pET-32-EGF, pET-32-*cla*MP-EGF, and pET-32-EGF-*cla*MP before IPTG induction, (lanes 3, 5, 7) *E. coli* lysate of pET-32-EGF, pET-32-*cla*MP-EGF, pET-32-EGF-*cla*MP 16 hours after IPTG induction. (b) (Lane 1) molecular weight standard, (lane 2) purified EGF-Ni-*cla*MP, (lane 3) purified EGF, and (lane 4) purified Ni-*cla*MP-EGF.

the various constructs (lanes 3, 5, and 7), confirming successful expression. Native EGF was used as a standard to establish relative expression of the *cla*MP-Tagged forms. N-terminal and C-terminal placement led to a 14% and 29% decrease in EGF expression yield, respectively. Differences in expression among the variants were on the same order as batch-to-batch variation among replicates of the same protein. Therefore, insertion of the *cla*MP Tag into the protein sequence has a negligible effect on protein expression, as determined by densitometric analysis of whole cell lysates.

2.3.2 *cla*MP-Tagged EGF Is Soluble

The EGF variants accumulated in the soluble fraction of the cell lysate, as determined by SDS-PAGE analysis of the supernatant and pellet from each sample following centrifugation of

the lysate (data not shown). Following individual processing and purification steps, the yield of each protein variant was determined (Table 2.1). SDS-PAGE was performed to examine the purity and amount of protein present at each step, and the pure final product is shown in Figure 2.3b. Cleavage of the fusion tags from EGF-*cla*MP worked as expected, completely cleaving the fusion protein and yielding the expected amount of each product. During Factor Xa cleavage, a 30% lower yield was observed with the N-terminally tagged construct in comparison to EGF-Ni-*cla*MP (Table 2.1). The *cla*MP Tag abuts the Factor Xa recognition sequence, and therefore, it is hypothesized that cleavage efficiency is reduced because the Ni-*cla*MP complex kinks the

protein conformation near the site, limiting access by the protease to the cleavage site. Sample purity of the final EGF proteins also was assessed using size exclusion chromatography (Fig. 2.4). As a control, native EGF was examined, and it elutes at 12 minutes. EGF-Ni-*cla*MP elutes earlier than EGF at 10 minutes, which is expected because it is slightly larger than EGF.
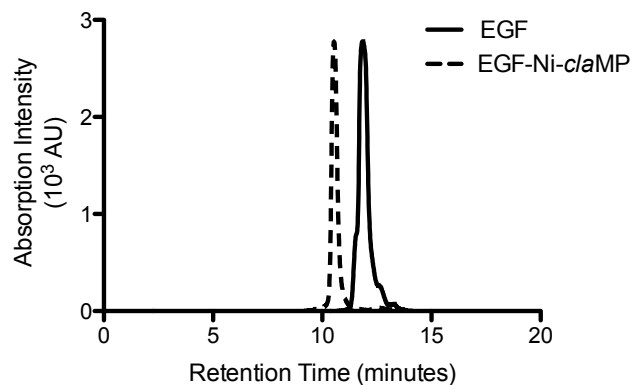

Figure 2.4 Size exclusion chromatogram of EGF and *cla*MP-Tagged EGF. Size-exclusion chromatography verifies that one main species is present in the sample. EGF-Ni-*cla*MP elutes slightly earlier than EGF.

Table 2.1. Relative Yields of EGF Constructs

Conjugate	Trx-EGF (mg)	EGF Mass Equivalents (mg)[a]	EGF Mass Before Cleavage Final (mg)[b]	Final Product (mg)	% Yield[c]	Average Percent Yield
Ni-*cla*MP-EGF	8.2	2.3	3.2	0.5	14%	
	7.2	2.0	1.7	0.1	9%	12 ± 3%
EGF-Ni-*cla*MP	6.9	1.9	2.3	1.1	45%	
	15.1	4.2	4.8	1.6	34%	
	12	3.3	3.0	1.5	50%	43 ± 7%
EGF	11.8	2.3	3.2	0.5	14%	
	7.6	2.0	1.3	0.7	60%	53 ± 7%

[a] The mass of EGF within the fusion protein was determined by taking into account the molecular weight of the fusion protein and EGF in the construct.
[b] The mass of EGF within the fusion protein was determined by taking into account the molecular weight of the remaining portion of the fusion tags and EGF in the construct.
[c] Percent yield was determined by comparing the final yield to the theoretical amount of EGF before the final cleavage step. The second cleavage step was used to determine the percent yield.

2.3.3 *cla*MP-Tagged EGF Is Correctly Folded

The three disulfide bonds present in EGF are the primary contributing factors to the protein structure. As such, conservation of the native disulfide bonds is essential to preserve protein structure and maintain biological function. Formation of the native disulfide bonds in EGF occurs through multiple iterations of incorrect cysteine pairs until the correct disulfide network is achieved,[35] and addition of the *cla*MP Tag introduces two additional non-native cysteine residues into EGF, which could interfere with the native disulfide network. To validate that addition of the *cla*MP Tag to EGF did not affect protein structure, two-dimensional heteronuclear NMR was used to compare the variants to native EGF (Fig. 2.5).[26] A ^{1}H-^{15}N HSQC spectrum provides a fingerprint of the protein; each amide from the backbone generates one peak, which reflects the unique conformation of the corresponding residue within the structured protein. ^{1}H-^{15}N HSQC spectra obtained for EGF and EGF-Ni-*cla*MP are highly similar and have specific, localized differences. The spectra show that the native fold of the
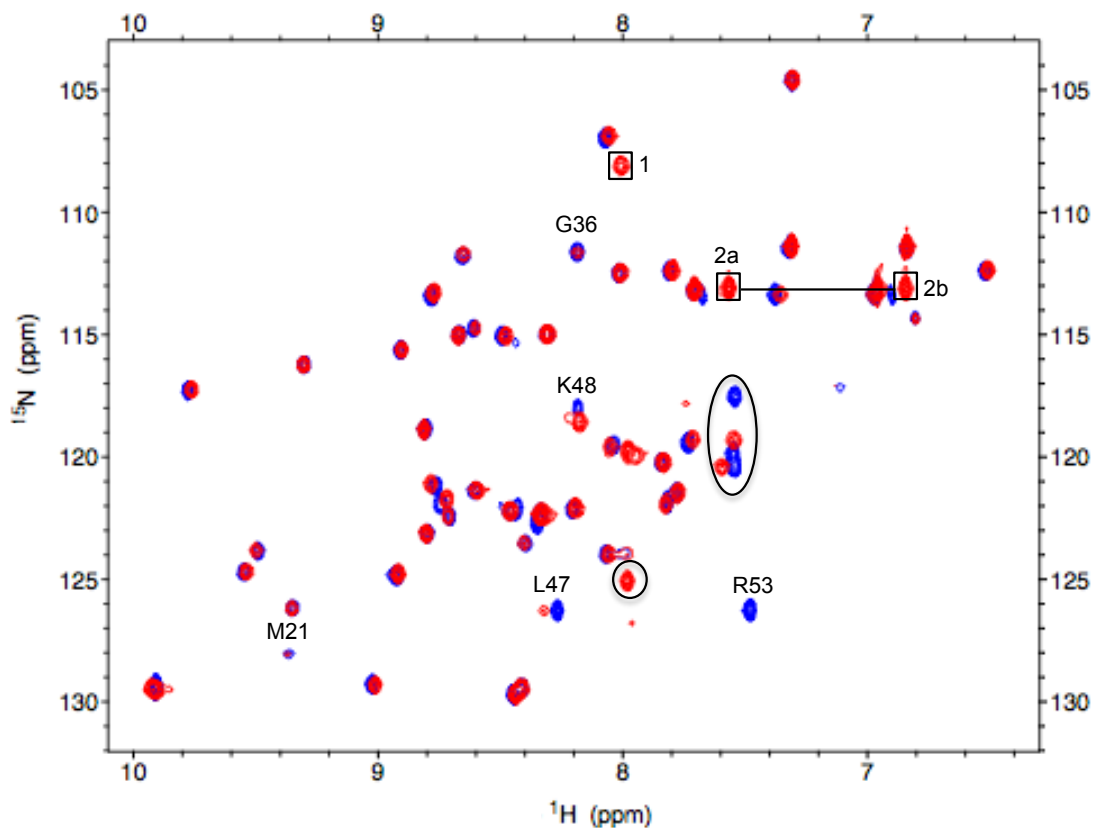
Figure 2.5 **^{1}H-^{15}N HSQC spectrum of EGF and EGF-Ni-*cla*MP.** Higher order protein structure is maintained in the presence of *cla*MP Tag. ^{1}H-^{15}N HSQC spectra of EGF (blue) and EGF-Ni-*cla*MP (red) at 0.1 mM and 0.08 mM, respectively. Peaks from native EGF that decrease in intensity when the tag is present are labeled with their corresponding residues. Shifted peaks corresponding to C-terminal residues are circled. New peaks 1, 2a, and 2b appear due to the addition of the *cla*MP Tag. Peak 1 corresponds to a glycine residue and Peaks 2a and 2b correspond to the asparagine side chain amide.

protein is maintained in the presence of the *cla*MP Tag, as the majority of the residues in both spectra overlay. The disulfide network is also maintained; the chemical shift positions of the five assigned cysteine residues remain unperturbed. The data show that the tertiary structure is not altered by the addition of the *cla*MP Tag. In the EGF-Ni-*cla*MP spectrum, a few notable differences from native EGF are observed; residues near the C-terminus of the protein (K48, W49/50, L52, and R53) are shifted. There also are differences in peak intensity observed between the variants for M21, G36, and L47. Because NMR is exquisitely sensitive to small

changes in local environment, the presence of the metal-bound *cla*MP Tag would be expected to influence the chemical shift of neighboring residues. The chemical shift positions of EGF with and without the tag were compared and the differences mapped onto the structure of EGF to show the residues affected by the tag. The structure reveals that the negatively charged Ni-*cla*MP module likely associates with a neighboring patch of positive charge (Fig. 2.6).
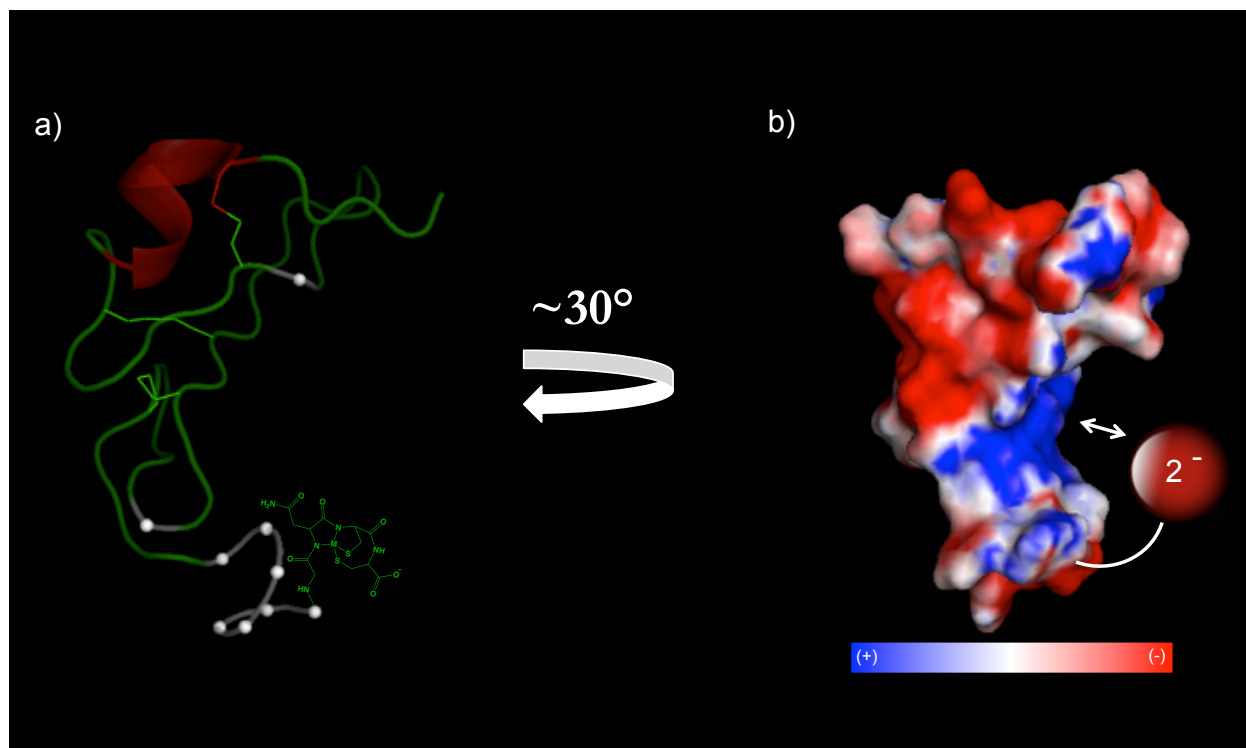
Figure 2.6 Illustration of the potential location of Ni-*cla*MP within the EGF molecule. (a) Schematic of EGF molecule; residues affected by addition of the *cla*MP Tag on the C-terminus are shown in grey. (b) Electrostatic surface map of EGF with the Ni-*cla*MP complex added onto the C-terminus. The negatively charged Ni-*cla*MP complex is hypothesized to associate with the positively charged pocket shown in blue.

In addition, three peaks emerge in the EGF-Ni-*cla*MP spectrum, which may be due to the addition of residues composing the tag. We expect Peak 1 corresponds to the amide backbone of one of the glycine residues because it appears in the glycine region. The backbone of the asparagine and first cysteine residues do not generate peaks because the *cla*MP Tag binds to the

metal via these nitrogens, deprotonating them. The other two residues do not generate observable peaks in the HSQC. The asparagine side chain from the tag is visible in the spectrum, and it appears as a new pair of peaks with unique ^{1}H values and the same ^{15}N chemical shift (Peaks 2a, 2b).

2.3.4 Ni(II) Is Selectively Inserted into *cla*MP Tag in the Presence of EGF

For the initial analytical development, Ni(II) was used to occupy the *cla*MP Tag, because this is a well understood system in our lab.[19-21] Insertion of Ni(II) into the *cla*MP Tag can be accomplished using Ni(II)-charged IMAC resin or via solution transfer using aqueous Ni-NTA,[19-21] both resulting in formation of the desired Ni-*cla*MP complex. The *cla*MP-Tagged EGF was produced as a His-tagged fusion protein, which accomplished affinity purification and facilitated complete Ni(II) transfer into the *cla*MP Tag during the purification process. Use of IMAC resin for transfer has been shown to lead to full incorporation of Ni(II) into *cla*MP-Tagged proteins as determined by ICP-OES and very efficiently limits nonspecific binding of Ni(II) to proteins (unpublished data). EGF has previously been purified using Ni(II)-based affinity chromatography, and Ni(II) was not reported to nonspecifically bind to EGF.[26] Control reactions performed with native protein in this study confirm this result.

The Ni-*cla*MP complex of the correct structure is rusty orange in color, allowing incorporation in the context of the protein to be confirmed with absorption spectroscopy, as established with the peptide system.[21] In the EGF-Ni-*cla*MP spectrum, there are broad features present in the visible region that reflect the structure of the unique metal complex, as seen in Figure 2.7a. These features are absent from the control spectrum of native EGF, which confirms they are due to correct Ni-*cla*MP formation, and not nonspecific binding of Ni(II) to EGF. The Ni-*cla*MP peptide module also exhibits additional features at longer wavelengths. These features

are not apparent in the absorption spectrum of EGF-Ni-*cla*MP shown in Figure 2.7a because a much higher protein concentration would be necessary, and this would cause the absorption from the protein to be off-scale.

AEC also was used to verify that proper formation of the Ni-*cla*MP complex was achieved. Correct incorporation of Ni(II) into the *cla*MP Tag creates a net charge of 2⁻. Native and metal-free *cla*MP-Tagged EGF each have an overall charge of 4⁻, but proper nickel insertion into *cla*MP-Tagged EGF develops an overall charge of 6⁻. This difference in charge results in unique AEC elution profiles for the two species. As shown in Figure 2.7b, EGF elutes at 35 minutes, whereas EGF-Ni-*cla*MP elutes at 75 minutes. The greater negative charge of EGF-Ni-*cla*MP causes it to bind more strongly to the cationic resin allowing AEC to be utilized to confirm metal binding and to separate the two distinct charge species.

a)

b)
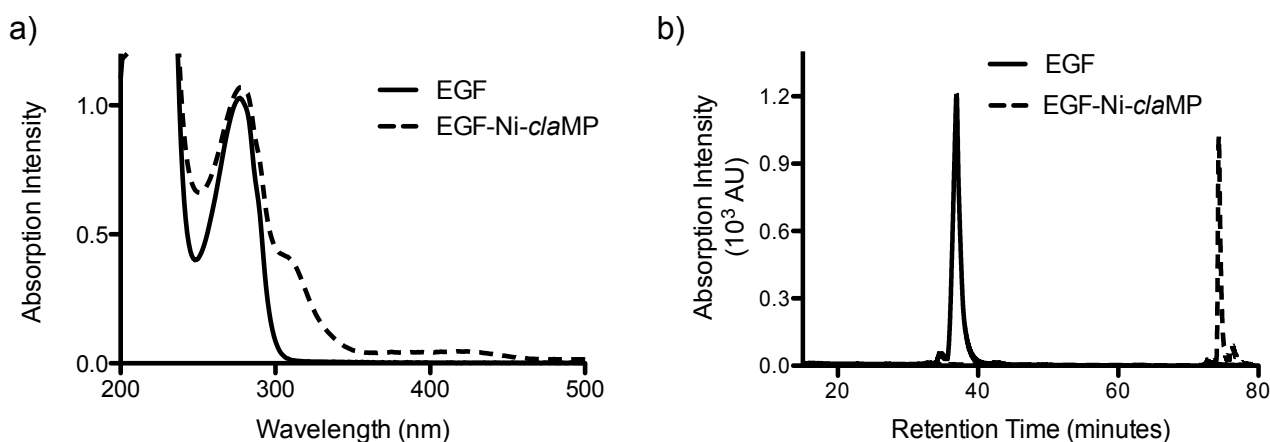
Figure 2.7 Absorption spectra and AEX chromatogram of EGF and EGF-Ni-*cla*MP. Ni(II) is successfully incorporated into the *cla*MP Tag in the presence of EGF. (a) Absorption spectroscopy validates Ni(II) incorporation into the *cla*MP Tag. In comparison to EGF, EGF-Ni-*cla*MP contains distinct features in the visible region, which confirm Ni(II) incorporation. (b) Anion exchange chromatography demonstrates the difference in net charge between EGF and EGF-Ni-*cla*MP due to the addition of Ni-*cla*MP.

As it has been shown that the Ni-*cla*MP complex acts as a superoxide mimic,[19] verification of complex formation in the presence of EGF can be confirmed using a standard SOD activity assay. The IC_{50} values for the Ni-MAP module and EGF-Ni-*cla*MP are 35 ± 8 μM

and 44 ± 4 μM, respectively, which are comparable within experimental error for the xanthine/xanthine oxidase assay. This confirms the Ni-*cla*MP Tag specifically and efficiently forms the unique structure determined for the original Ni-MAP module and this module maintains comparable SOD-like activity in the context of the protein.

2.3.5 EGF Remains Functional with Addition of *cla*MP Tag

In order to be used successfully as an inline metal carrier, the *cla*MP Tag cannot be deleterious to the function of the protein. A standard cell viability assay using the A431 cell strain was completed to assess the impact of the *cla*MP Tag on EGF function. This strain is known for its overexpression of the epidermal growth factor receptor (EGFR) on the cell surface.[36] Overexpression of EGFR leads to the
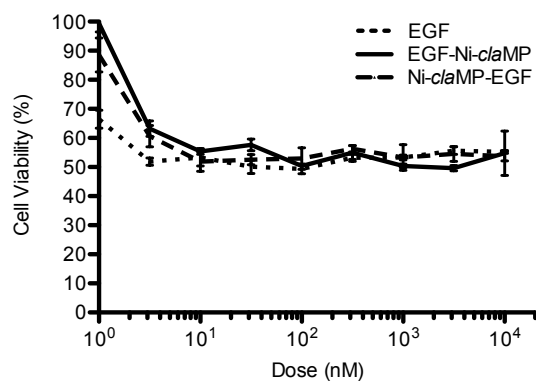
Figure 2.8 EGF Activity Assay. Addition of the *cla*MP Tag to either the N- or C-terminus of EGF has little to no effect on the function of the protein. The decrease in cell viability of A431 cells caused by addition of EGF is comparable to that of both Ni-*cla*MP-EGF and EGF-Ni-*cla*MP.

unexpected outcome of growth inhibition upon stimulation with EGF.[36] The N- and C-terminus of EGF are not responsible for high-affinity binding to the receptor[37] so placement of the *cla*MP Tag on either terminus would not be expected to diminish EGF activity and indeed it does not. EGF or either *cla*MP-Tagged variant of EGF elicits a similar cytostatic response (Fig. 2.8). At nanomolar concentrations, addition of EGF or either Ni-*cla*MP-EGF protein results in a 50% loss in cell viability in the A431 cell line, confirming EGF remains active in the presence of the tag at either position.

2.4 DISCUSSION

Metal-binding agents enable a wide range of biotechnology, healthcare, and research applications. Peptide tags and organic molecules capable of binding metal ions can be combined with the appropriate protein to allow for targeted delivery of metal ions for imaging and therapeutic applications and for use as analytical research tools.[3, 8, 10, 38-40] Incorporation of a metal-binding tag into a protein can be accomplished using either chemical conjugation or inline addition. Conjugating metal-binding organic molecules to proteins is typically performed using chemical means, primarily conjugation to lysine or cysteine residues using moiety specific chemistry.[3-5, 41-42] Chemical conjugation using these approaches typically generates heterogeneous mixtures and the reactions often have poor yield of the desired product.[5-6, 43-44] Protein conjugates used in imaging and therapeutic applications must retain native protein structure and function in the presence of the peptide carrier containing the metal ion. The *cla*MP tag is similar in size to small chelating agents, but it bears the added advantage that placement within the protein can be controlled precisely. These features of the metal *cla*MP complex make it favorable for use in many applications.

Typically, proteins are modified by attachment to lysine residues. With chemical conjugation, protein function may be altered because the exact attachment site cannot be controlled,[45] and lysine residues are often present in receptor recognition sites of many proteins, including antibodies.[46-47] In addition, in vivo PK may be affected.[48] For example, modification of the Fc region can alter binding to FcRn, resulting in changes in biodistribution, endosomal recycling, and PK.[49] Incorporation of an inline metal-binding peptide eliminates the need for conjugation chemistry and allows for the formation of engineered site-selective conjugates designed to preserve function of the native protein and optimize performance characteristics.

The tripeptide *cla*MP Tag can be easily inserted into the gene of interest using standard cloning methodologies. Here, it is demonstrated that addition of the *cla*MP Tag into the EGF sequence allows for the generation of an inline metal carrier that can be positioned to avoid disrupting the protein's function.

While chelators have been developed that bind transition metals, their properties are often incompatible with clinical use. Tight binding chelators to copper have been developed, but those with sufficiently slow off rates for in vivo use require very high temperature and long incubation times to accomplish binding. SarAr technology is an exception, which achieves tight binding under reasonable conditions for use with proteins,[50] but this chelator is selective for and only used with copper.[51] The *cla*MP Tag is capable of binding a wide range of transition metals, including many species that are used to enable imaging and therapeutic applications, under conditions amenable to retention of higher order protein structure. Cytotoxic metal ions, such as Pt or Pd, can be loaded into *cla*MP and have potential to be used in targeted oncology applications (manuscripts in preparation). Radioactive tracer metals, such as ^{99m}Tc, ^{55}Co, $^{64/67}$Cu, or ^{212}Pb also can be incorporated into the tag (manuscripts in preparation), and these isotopes have proven utility in diagnostic imaging.[52-58] While not currently in wide usage, ^{57}Ni could be inserted in the tag and monitored based on its positron or x-ray emission.[59-60] Ni-*cla*MP also allows for the generation of a catalytic antioxidant,[19-21] which may be useful as a targeted anti-inflammatory agent or biotechnology reagent for selective detection of tagged proteins.

In order to investigate the potential of using the *cla*MP Tag as an inline metal carrier, its effects on the expression, structure, and function of a thiol- and disulfide-containing protein and metal insertion into the tag were analyzed. The spectroscopic and chromatographic data verify that incorporation of Ni(II) into the *cla*MP Tag is not affected when the tag is placed inline with

a protein sequence. The overall structure of *cla*MP-Tagged EGF is also maintained, as the chemical shift positions of the native cysteine residues remain unaltered. In addition, native function is maintained with *cla*MP Tag inclusion at either terminus. The study described here demonstrates successful inline incorporation of the *cla*MP Tag into a protein, illustrating this tag provides opportunity to expand the range of applications in which metals may be used in biotechnology and healthcare.

2.5 REFERENCES

1.   Aoki, I., Takahashi, Y., Chuang, K.-H., Silva, A. C., Igarashi, T., Tanaka, C., Childs, R. W., Koretsky, A. P. (2006) Cell labeling for magnetic resonance imaging with the T1 agent manganese chloride. *NMR Biomed*. *19*, 50-59.

2.   Zheng, Q., Dai, H., Merritt, M. E., Malloy, C., Pan, C. Y., Li, W.-H. (2005) A New Class of Macrocyclic Lanthanide Complexes for Cell Labeling and Magnetic Resonance Imaging Applications. *J. Am. Chem. Soc. 127*, 16178-16188.

3.   De León-Rodríguez, L. M., Kovacs, Z. (2008) The Synthesis and Chelation Chemistry of DOTA-Peptide Conjugates. *Bioconjugate Chem. 19*, 391-402.

4.   Lewis, M. R., Kao, J. Y., Anderson, A.-L. J., Shively, J. E., Raubitschek, A. (2001) An improved method for conjugating monoclonal antibodies with N-Hydroxysulfosuccinimidyl DOTA. *Bioconjugate Chem. 12*, 320-324.

5.   Lewis, M. R., Raubitschek, A., Shively, J. E. (1994) A Facile, Water-Soluble Method for Modification of Proteins with DOTA. Use of Elevated Temperature and Optimized pH To Achieve High Specific Activity and High Chelate Stability in Radiolabeled Immunoconjugates. *Bioconjugate Chem. 5*, 565-76.

6.   Knör, S., Modlinger, A., Poethko, T., Schottelius, M., Wester, H.-J., Kessler, H. (2007) Synthesis of novel 1,4,7,10-tetraazacyclodecane-1,4,7,10-tetraacetic acid (DOTA) derivatives for chemoselective attachment to unprotected polyfunctionalized compounds. *Chem. - Eur. J. 13*, 6082-6090, S6082/1-S6082/20.

7.   Franz, K. J., Nitz, M., Imperiali, B. (2003) Lanthanide-binding tags as versatile protein coexpression probes. *ChemBioChem 4*, 265-271.

8.   Martin, L. J., Hähnke, M. J., Nitz, M., Wöhnert, J., Silvaggi, N. R., Allen, K. N., Schwalbe, H., Imperiali, B. (2007) Double-Lanthanide-Binding Tags: Design, Photophysical Properties, and NMR Applications. *J. Am. Chem. Soc. 129*, 7106-7113.

9.   Nitz, M., Sherawat, M., Franz, K. J., Peisach, E., Allen, K. N., Imperiali, B. (2004) Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide. *Angew. Chem., Int. Ed. 43*, 3682-3685.

10.   Sculimbrene, B. R., Imperiali, B. (2006) Lanthanide-Binding Tags as Luminescent Probes for Studying Protein Interactions. *J. Am. Chem. Soc. 128*, 7346-7352.

11.   Wöhnert, J., Franz, K. J., Nitz, M., Imperiali, B., Schwalbe, H. (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. *J. Am. Chem. Soc. 125*, 13338-13339.

12.   Yang, J. J., Yang, J., Wei, L., Zurkiya, O., Yang, W., Li, S., Zou, J., Zhou, Y., Wilkins Maniccia, A. L., Mao, H., Zhao, F., Malchow, R., Zhao, S., Johnson, J., Hu, X., Krogstad, E., Liu, Z.-R. (2008) Rational Design of Protein-Based MRI Contrast Agents. *J. Am. Chem. Soc. 130*, 9260-9267.

13.   Perazella, M. A. (2007) Nephrogenic systemic fibrosis, kidney disease, and gadolinium: is there a link? *Clin. J. Am. Soc. Nephrol. 2*, 200-202.

14.   Wadas, T. J., Wong, E. H., Weisman, G. R., Anderson, C. J. (2010) Coordinating Radiometals of Copper, Gallium, Indium, Yttrium, and Zirconium for PET and SPECT Imaging of Disease. *Chem. Rev. (Washington, DC, U. S.) 110*, 2858-2902.

15.   McQuade, P., Miao, Y., Yoo, J., Quinn, T. P., Welch, M. J., Lewis, J. S. (2005) Imaging of Melanoma Using 64Cu- and 86Y-DOTA-ReCCMSH(Arg11), a Cyclized Peptide Analogue of α-MSH. *J. Med. Chem. 48*, 2985-2992.

16.   Chen, X., Hou, Y., Tohme, M., Park, R., Khankaldyyan, V., Gonzales-Gomez, I., Bading, J. R., Laug, W. E., Conti, P. S. (2004) Pegylated Arg-Gly-Asp peptide: 64Cu labeling and PET imaging of brain tumor αvβ3-integrin expression. *J. Nucl. Med. 45*, 1776-1783.

17.   Ping Li, W., Lewis, J. S., Kim, J., Bugaj, J. E., Johnson, M. A., Erion, J. L., Anderson, C. J. (2002) DOTA-D-Tyr1-Octreotate: A Somatostatin Analogue for Labeling with Metal and Halogen Radionuclides for Cancer Imaging and Therapy. *Bioconjugate Chem. 13*, 721-728.

18.   Laurence, J. A. S., Vartia, A. A., Krause, M. E. Metal abstraction peptide (MAP) tag and associated methods. Singapore Patent 166345, June 29, 2012.

19.   Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2010) Novel Tripeptide Model of Nickel Superoxide Dismutase. *Inorg. Chem. 49*, 362-364.

20.   Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2011) MAPping the Chiral Inversion and Structural Transformation of a Metal-Tripeptide Complex Having Ni-Superoxide Dismutase Activity. *Inorg. Chem. 50*, 2479-2487.

21. Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2013) Embedding the Ni-SOD Mimetic Ni-NCC within a Polypeptide Sequence Alters the Specificity of the Reaction Pathway. *Inorg. Chem. 52*, 77-83.

22. Chang, J.-Y., Schindler, P., Ramseier, U., Lai, P.-H. (1995) The disulfide folding pathway of human epidermal growth factor. *J. Biol. Chem. 270*, 9207-9216.

23. Lee, J. Y., Yoon, C. S., Chung, I. Y., Lee, Y. S., Lee, E. K. (2000) Scale-up process for expression and renaturation of recombinant human epidermal growth factor from Escherichia coli inclusion bodies. *Biotechnol. Appl. Biochem. 31*, 245-248.

24. Le, P. U., Lenferink, A. E. G., Pinard, M., Baardsnes, J., Massie, B., O'Connor-McCourt, M. D. (2009) Escherichia coli expression and refolding of E/K-coil-tagged EGF generates fully bioactive EGF for diverse applications. *Protein Expression Purif. 64*, 108-117.

25. Wu, J., Yang, Y., Watson, J. T. (1998) Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci. 7*, 1017-1028.

26. Huang, H.-W., Mohan, S. K., Yu, C. (2010) The NMR solution structure of human epidermal growth factor (hEGF) at physiological pH and its interactions with suramin. *Biochem. Biophys. Res. Commun. 402*, 705-710.

27. Ciaccio, N. A., Moreno, M. L., Bauer, R. L., Laurence, J. S. (2008) High-yield expression in E. coli and refolding of the bZIP domain of activating transcription factor 5. *Protein Expression Purif. 62*, 235-243.

28. Schägger, H. (2006) Tricine-SDS-PAGE. *Nat. Protoc. 1*, 16-22.

29. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277-93.

30. Goddard, T. D., Kneller, D. G. (2004) SPARKY. San Francisco: University of California.

31. Crapo, J. D., McCord, J. M., Fridovich, I. (1978) Preparation and assay of superoxide dismutases. *Methods Enzymol. 53*, 382-393.

32. Tabbi, G., Driessen, W. L., Reedijk, J., Bonomo, R. P., Veldman, N., Spek, A. L. (1997) High Superoxide Dismutase Activity of a Novel, Intramolecularly Imidazolato-Bridged Asymmetric Dicopper(II) Species. Design, Synthesis, Structure, and Magnetism of

Copper(II) Complexes with a Mixed Pyrazole-Imidazole Donor Set. *Inorg. Chem. 36*, 1168-1175.

33. Sharma, K., Cherish Babu, P. V., Sasidhar, P., Srinivas, V. K., Krishna Mohan, V., Krishna, E. (2008) Recombinant human epidermal growth factor inclusion body solubilization and refolding at large scale using expanded-bed adsorption chromatography from Escherichia coli. *Protein Express. Purif. 60*, 7-14.

34. Hammarström, M., Hellgren, N., Van Den Berg, S., Berglund, H., Härd, T. (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. *Protein Sci. 11*, 313-321.

35. Chang, J.-Y., Li, L., Lai, P.-H. (2001) A major kinetic trap for the oxidative folding of human epidermal growth factor. *J. Biol. Chem. 276*, 4845-4852.

36. Gill, G. N., Lazar, C. S. (1981) Increased phosphotyrosine content and inhibition of proliferation in EGF-treated A431 cells. *Nature (London) 293*, 305-307.

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

38. Daughtry, K. D., Martin, L. J., Sarraju, A., Imperiali, B., Allen, K. N. (2012) Tailoring Encodable Lanthanide-Binding Tags as MRI Contrast Agents. *ChemBioChem 13*, 2567-2574.

39. Breeman, W. A. P., Kwekkeboom, D. J., de Blois, E., de Jong, M., Visser, T. J., Krenning, E. P. (2007) Radiolabelled regulatory peptides for imaging and therapy. *Anti-Cancer Agents Med. Chem. 7*, 345-357.

40. Wängler, C., Buchmann, I., Eisenhut, M., Haberkorn, U., Mier, W. (2007) Radiolabeled peptides and proteins in cancer therapy. *Protein Pept. Lett. 14*, 273-279.

41. Lewis, M. R., Shively, J. E. (1998) Maleimidocysteineamido-DOTA Derivatives: New Reagents for Radiometal Chelate Conjugation to Antibody Sulfhydryl Groups Undergo pH-Dependent Cleavage Reactions. *Bioconjugate Chem. 9*, 72-86.

42. Li, L., Tsai, S.-W., Anderson, A.-L., Keire, D. A., Raubitschek, A. A., Shively, J. E. (2002) Vinyl Sulfone Bifunctional Derivatives of DOTA Allow Sulfhydryl- or Amino-

Directed Coupling to Antibodies. Conjugates Retain Immunoreactivity and Have Similar Biodistributions. *Bioconjugate Chem.* *13*, 110-115.

43.   Lewis Phillips, G. D., Li, G., Dugger, D. L., Crocker, L. M., Parsons, K. L., Mai, E., Blättler, W. A., Lambert, J. M., Chari, R. V. J., Lutz, R. J., Wong, W. L. T., Jacobson, F. S., Koeppen, H., Schwall, R. H., Kenkare-Mitra, S. R., Spencer, S. D., Sliwkowski, M. X. (2008) Targeting HER2-Positive Breast Cancer with Trastuzumab-DM1, an Antibody-Cytotoxic Drug Conjugate. *Cancer Res.* *68*, 9280-9290.

44.   Wakankar, A. A., Feeney, M. B., Rivera, J., Chen, Y., Kim, M., Sharma, V. K., Wang, Y. J. (2010) Physicochemical Stability of the Antibody-Drug Conjugate Trastuzumab-DM1: Changes due to Modification and Conjugation Processes. *Bioconjugate Chem.* *21*, 1588-1595.

45.   Shen, B.-Q., Xu, K., Liu, L., Raab, H., Bhakta, S., Kenrick, M., Parsons-Reponte, K. L., Tien, J., Yu, S.-F., Mai, E., Li, D., Tibbitts, J., Baudys, J., Saad, O. M., Scales, S. J., McDonald, P. J., Hass, P. E., Eigenbrot, C., Nguyen, T., Solis, W. A., Fuji, R. N., Flagella, K. M., Patel, D., Spencer, S. D., Khawli, L. A., Ebens, A., Wong, W. L., Vandlen, R., Kaur, S., Sliwkowski, M. X., Scheller, R. H., Polakis, P., Junutula, J. R. (2012) Conjugation site modulates the in vivo stability and therapeutic activity of antibody-drug conjugates. *Nat. Biotechnol.* *30*, 184-189.

46.   Bachran, C., Schneider, S., Riese, S. B., Bachran, D., Urban, R., Schellmann, N., Zahn, C., Sutherland, M., Fuchs, H. (2011) A lysine-free mutant of epidermal growth factor as targeting moiety of a targeted toxin. *Life Sci.* *88*, 226-232.

47.   Wang, Y.-J., Liu, Y.-D., Chen, J., Hao, S.-J., Hu, T., Ma, G.-H., Su, Z.-G. (2010) Efficient preparation and PEGylation of recombinant human non-glycosylated erythropoietin expressed as inclusion body in E. coli. *Int. J. Pharm.* *386*, 156-164.

48.   Boswell, C. A., Mundo, E. E., Zhang, C., Bumbaca, D., Valle, N. R., Kozak, K. R., Fourie, A., Chuh, J., Koppada, N., Saad, O., Gill, H., Shen, B.-Q., Rubinfeld, B., Tibbitts, J., Kaur, S., Theil, F.-P., Fielder, P. J., Khawli, L. A., Lin, K. (2011) Impact of Drug Conjugation on Pharmacokinetics and Tissue Distribution of Anti-STEAP1 Antibody-Drug Conjugates in Rats. *Bioconjugate Chem.* *22*, 1994-2004.

49.   Junutula, J. R., Raab, H., Clark, S., Bhakta, S., Leipold, D. D., Weir, S., Chen, Y., Simpson, M., Tsai, S. P., Dennis, M. S., Lu, Y., Meng, Y. G., Ng, C., Yang, J., Lee, C.

C., Duenas, E., Gorrell, J., Katta, V., Kim, A., McDorman, K., Flagella, K., Venook, R., Ross, S., Spencer, S. D., Wong, W. L., Lowman, H. B., Vandlen, R., Sliwkowski, M. X., Scheller, R. H., Polakis, P., Mallet, W. (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol. 26*, 925-932.

50.   Di Bartolo, N., Sargeson, A. M., Smith, S. V. (2006) New 64Cu PET imaging agents for personalised medicine and drug development using the hexa-aza cage, SarAr. *Org. Biomol. Chem. 4*, 3350-3357.

51.   Voss, S. D., Smith, S. V., DiBartolo, N., McIntosh, L. J., Cyr, E. M., Bonab, A. A., Dearling, J. L. J., Carter, E. A., Fischman, A. J., Treves, S. T., Gillies, S. D., Sargeson, A. M., Huston, J. S., Packard, A. B. (2007) Positron emission tomography (PET) imaging of neuroblastoma and melanoma with 64Cu-SarAr immunoconjugates. *Proc. Natl. Acad. Sci. U. S. A. 104*, 17489-17493.

52.   Wu, Y., Zhang, X., Xiong, Z., Cheng, Z., Fisher, D. R., Liu, S., Gambhir, S. S., Chen, X. (2005) microPET imaging of glioma integrin alpha(V)beta(3) expression using Cu-64-labeled tetrameric RGD peptide. *J. Nucl. Med. 46*, 1707-1718.

53.   Waibel, R., Alberto, R., Willuda, J., Finnern, R., Schibli, R., Stichelberger, A., Egli, A., Abram, U., Mach, J. P., Plückthun, A., Schubiger, P. A. (1999) Stable one-step technetium-99m labeling of His-tagged recombinant proteins with a novel Tc(I)-carbonyl complex. *Nat. Biotechnol. 17*, 897-901.

54.   Cai, H., Conti, P. S. (2013) RGD-based PET tracers for imaging receptor integrin alpha(v)beta(3) expression. *J. Labelled Compd. Radiopharm. 56*, 264-279.

55.   Pandya, D. N., Bhatt, N., Dale, A. V., Young Kim, J., Lee, H., Su Ha, Y., Lee, J.-E., Il An, G., Yoo, J. (2013) New Bifunctional Chelator for Cu-64-Immuno-Positron Emission Tomography. *Bioconjugate Chem. 24*, 1356-1366.

56.   Goethals, P., Volkaert, A., Vandewielle, C., Dierckx, R., Lameire, N. (2000) 55Co-EDTA for renal imaging using positron emission tomography (PET): a feasibility study. *Nucl. Med. Biol. 27*, 77-81.

57.   Jansen, H. M., Decoo, D., Minderhoud, J. M. (1997) Co-registration of PET and MRI in different courses of MS using Cobalt-55 as a Calcium-tracer. *Acta Neurol. Belg. 97*, 178-182.

58. Meredith, R. F., Torgue, J., Azure, M. T., Shen, S., Saddekni, S., Banaga, E., Carlise, R., Bunch, P., Yoder, D., Alvarez, R. (2014) Pharmacokinetics and Imaging of (212)Pb-TCMC-Trastuzumab After Intraperitoneal Administration in Ovarian Cancer Patients. *Cancer Biother. Radiopharm 29*, 12-17.

59. Nielsen, G. D., Andersen, O., Jensen, M. (1993) Toxicokinetics of nickel in mice studied with the γ-emitting isotope nickel-57. *Fundam. Appl. Toxicol. 21*, 236-43.

60. Zweit, J., Carnochan, P., Goodall, R., Ott, R. (1994) Nickel-57-doxorubicin, a potential radiotracer for pharmacokinetic studies using PET: production and radiolabelling. *J. Nucl. Biol. Med. 38*, 18-21.

CHAPTER 3.
STABILITY ANALYSIS OF AN INLINE PEPTIDE-BASED CONJUGATE FOR METAL DELIVERY: NICKEL(II)-*cla*MP TAG EPIDERMAL GROWTH FACTOR AS A MODEL SYSTEM

3.1 INTRODUCTION

Coupling of chelators to proteins allows for the generation of conjugates that enable metal-based imaging or therapeutic applications.[1-3] Generally this requires optimization of linker chemistry and additional purification steps to obtain the desired product.[1, 4-6] Even though the conjugated entities are much smaller in size than the protein molecule, the impact of their addition may not be insignificant. These alterations often affect pharmaceutical properties, including solubility, chemical and physical stability, and colloidal behavior of the protein.[7] The magnitude and type of effect depends on the specific position modified and the structural attributes of the protein,[8] as well as the properties of the linker and attached compound.

The physical stability of proteins depends on the core structure, packing interactions and surface composition. The surface properties that develop in the folded state affect solubility and colloidal stability, but they also determine solvent access to core structural elements that contribute to the overall three-dimensional fold and aggregation propensity of the molecule.[9] As such, the composition of the molecule's surface significantly impacts its stability, which is altered by conjugation. This principle is apparent when comparing homologous proteins or mutants having the same three-dimensional fold because each variant can exhibit widely different stability profiles depending on their distinct surface properties.[10-11] Charged residues located on the surface of structured proteins often increase the physical stability by protecting against thermal unfolding[10-11] or by decreasing aggregation propensity due to charge-charge repulsion between protein molecules.[12] Conjugation to surface-exposed lysine residues is commonly used in formation of bioconjugates, and it alters the surface charge and causes large

product heterogeneity.[13] Even though conjugates are successfully formed using this chemical method, altering the surface charge can affect the stability of the molecule[7, 14] and also introduce unexpected complexity, as multiple charge states have been observed when modifying a single lysine, depending upon its position in the structure.[15] To be able to control and/or better characterize stability, conjugation to native or engineered cysteine residues and non-natural amino acids has been pursued to increase site specificity.[16] Modification of native cysteine residues in antibodies has been performed by reducing inter-strand disulfide bonds and conjugating at these positions. Heterogeneity is decreased greatly compared to lysine-based approaches, but complexity remains and physical stability can be adversely affected.[8, 17-18] Introduction of additional cysteine residues permits site-specific attachment, and the THIOMAB approach has demonstrated success in identifying suitable positions within antibodies for this purpose.[19] Site-specific attachment of any type permits detailed analysis of the structure and stability of the protein conjugate, allowing for selection of desirable attributes.

The *cla*MP Tag is based on the metal abstraction peptide (MAP) composed of the amino acid sequence Asn-Cys-Cys (NCC).[20-22] It has been shown to bind Ni(II), forming a single specific metal-peptide complex structure having an overall charge of 2^- [20] and unusual binding and release properties.[23] Incorporation of the NCC module into longer peptide sequences has been shown to have no discernable effect on complex formation.[22] Therefore, the *cla*MP Tag can be biosynthetically encoded into a protein to generate a linker-less bioconjugate,[24] which eliminates the need to chemically link to a chelating moiety. The specific metal abstraction chemistry performed by NCC occurs with a variety of transition metal ions.[23] The *cla*MP Tag also binds cytotoxic metals, such as Pt and Pd, and radioisotopes, such as ^{99m}Tc, $^{64/67}$Cu, and ^{60}Co

(unpublished data); thus, it has the potential to serve as a metal-binding agent for use in a variety of imaging and therapeutic applications.

Our previous study demonstrated that the *cla*MP Tag can be successfully incorporated at different positions within a thiol- and disulfide-containing fusion protein composed of thioredoxin and epidermal growth factor (EGF) without any negative effects on the expression yield, structure, or function of the tagged protein.[24] Herein is presented the solution stability of this model system, the Ni(II)-*cla*MP-EGF conjugate and the Ni-*cla*MP Tag module in the context of the EGF system, to demonstrate the potential of the *cla*MP Tag in serving as an inline metal carrier. The inline Ni-conjugates were stored in aqueous solution using two different buffers, and UV-vis absorption spectroscopy, liquid chromatography, mass spectrometry, circular dichroism, and NMR were used to characterize stability and changes in the conjugate structure over a 24-week period.

3.2 EXPERIMENTAL METHODS

3.2.1 Protein Expression and Purification

Wild-type and *cla*MP Tag constructs were expressed and purified as previously described.[24] Briefly, the DNA sequence was inserted into the pET-32Xa vector using a pET-32Xa/LIC cloning kit (Novagen). The plasmid was transformed into the Origami B (DE3) cell strain (Novagen) followed by expression in either *Luria Broth* or MMTM (for NMR samples). All of the samples for NMR analysis were grown on ^{15}N-labeled ammonium chloride (Isotech) as the nitrogen source. After expression, a 1 L pellet was resuspended, lysed using a French Press, and purified using a Hi-Trap Chelating HP column (GE Healthcare) charged with Ni(II). Cleavage with thrombin and Factor Xa (Novagen) were completed to obtain the final product. Protein concentration was determined using the Bradford assay, and the samples were

concentrated to approximately 0.2 mM. EGF and EGF-Ni-*cla*MP were each prepared in either 50 mM Tris-Cl, 10 mM NaCl, pH 7.3 or 50 mM potassium phosphate (KPi), 10 mM NaCl, pH 7.3 and stored at 4 °C.

3.2.2 Absorption Analysis

UV-vis spectroscopy was used to monitor Ni(II) incorporation into the *cla*MP Tag. Samples were placed in a 1-cm path length cuvette. Spectra were acquired from 200-800 nm using a Cary 100 Bio UV-visible spectrophotometer (Varian).

3.2.3 Size Exclusion Chromatography (SEC)

Size exclusion chromatography was used to assess the purity of the sample and determine if larger molecular weight species formed over time. Analysis was performed using a 4.6 x 300 mm Yarra SEC-2000, 3 μm pore size column (Phenomenex). The column was equilibrated in "Buffer A" (20 mM Tris-Cl, 10 mM KCl, pH 7.5 before injection of the sample. A constant flow rate of 0.35 mL/min "Buffer A" over 40 minutes was used to elute the protein from the column. UV absorption at 220 nm was used to monitor protein elution from the column. The area of the peak at 8 minutes was used to quantify the amount of monomer at each incubation time point.

3.2.4 Anion Exchange Chromatography (AEC)

AEC was performed using a 4.6 x 250 mM BioLC ProPac WAX-10 column (Dionex). Before sample injection, the column was equilibrated using "Buffer A". Protein was eluted using a linear gradient from 0-100% "Buffer B" ("Buffer A" + 500 mM KCl) over 70 minutes. A constant flow rate of 1 mL/min was used, and UV absorption at 220 nm was used to monitor protein elution from the column. The area of the main peak at 75 minutes was used to estimate the amount of protein present. Fluctuations in peak area of approximately 20% were observed

for duplicate samples analyzed on the same day. Because large run-to-run instrument variation was observed, the elution profile was used to establish trends and not for absolute quantitation.

3.2.5 ^{1}H-^{15}N HSQC NMR

To examine the structural stability of the constructs using 2D NMR, EGF and EGF-Ni-*cla*MP were prepared in two buffer systems at pH 7.3, 50 mM Tris-Cl and 50 mM KPi, each containing 10 mM NaCl and 6% D_2O. The samples were concentrated to approximately 0.1 mM for analysis. ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) spectra were acquired on a 600 MHz Bruker Avance NMR spectrometer with a triple resonance cryoprobe. The spectra for EGF-Ni-*cla*MP were acquired with 120 and 217 scans for the samples prepared in Tris-Cl and KPi, respectively, with 2048 points in ^{1}H and 128* increments in ^{15}N. The numbers of scans used were selected to account for slight variations in sample concentration and achieve a similar signal to noise ratio. All spectra were obtained at 25 °C. Spectra were obtained at multiple time intervals to evaluate time-dependent changes to protein structure. NMRpipe[25] and SPARKY[26] were used to process the spectra and for data analysis.

3.2.6 Static Light Scattering

Static light scatter (SLS) data was collected with a Photon Technology International (PTI) spectrafluorometer (Edison, NJ) equipped with a four-position sample holder and a Peltier temperature controller. Samples were prepared at a protein concentration of 0.5 mg/mL in 50 mM KPi, 10 mM NaCl, pH 7.3 and loaded into a 0.2 cm x 1 cm cuvette. The intensity of scattered light was detected by a photomultiplier tube at an angle 90° to the arc lamp white light source. The excitation wavelength was set at 295 nm, and the emission wavelength range was from 250 nm to 350 nm. Spectra were collected every 2.5 °C from a temperature range of 5 to 80 °C, with a two-minute equilibration at each temperature. The intensity of the scattered light

was measured as a function of temperature. The background was subtracted from each data point based upon a sample containing only buffer with no protein present. Two samples were analyzed for each variant.

3.2.7 Circular Dichroism

Circular dichroism (CD) spectra were collected using Chirascan™ Plus CD Spectrophotometer (Applied Photophysics, Surrey, UK) equipped with a Peltier temperature controller. Samples were prepared at a protein concentration of 0.035 mg/mL in 50 mM KPi, 10 mM NaCl, pH 7.3 and loaded into a 0.1-cm path length cuvette. Spectra were acquired every 2.5 °C from 5 to 80 °C, with a one-minute equilibration at each temperature. An averaging time of 0.5 seconds/nm and a bandwidth of 1 nm were used. The molar ellipticity at 200 nm was plotted as a function of temperature. The background was subtracted from each data point based upon a sample containing only buffer with no protein present. Two samples were analyzed for each variant.

3.3 RESULTS AND DISCUSSION

The emergence of bioconjugates in imaging and therapeutic applications has led to the need to develop assays suitable for investigating the stability of the entire conjugate, which encompasses the stability of the modified protein and the conjugated entity. The conjugation approach can affect the analysis methods used to analyze the bioconjugate and monitor stability. Typically, lysine conjugation is used, but it can result in a large degree of product heterogeneity that complicates analysis,[10, 13] and it also can lead to differences in stability among different preparations. The *cla*MP Tag is advantageous in that it does not require chemical conjugation because inline insertion of this tag within the protein sequence allows for the generation of a homogeneous bioconjugate, which simplifies analysis of the molecule and its stability.

The *cla*MP Tag has been shown not to adversely affect EGF expression, folding and function.[24] Here, *cla*MP-Tagged EGF was used as a model system to analyze the stability of the *cla*MP Tag in the context of a protein system and determine how the *cla*MP Tag modulates protein stability. The tag rapidly becomes fully occupied in the presence of the metal transfer agent, efficiently yielding a highly stable single species in solution.[20, 22, 24] Nonetheless, in the event of incomplete loading the metal-occupied *cla*MP-Tagged protein is easily separated from any residual unreacted proteins using ion exchange chromatography because the correctly formed Ni(II) complex develops a 2⁻ charge, differentiating it from the unoccupied form. This approach efficiently generates a single, homogeneous conjugate species, permitting the stability of the resulting product to be well characterized.

3.3.1 Stability of EGF and Ni-*cla*MP Tag within the Conjugate

The stability of the metal-bound *cla*MP Tag is important to investigate in order to successfully implement the tag as an inline conjugate. Most biologics and conjugates are stored as lyophilized formulations, but they must be stable in solution for 1-2 months to retain activity during manufacture and administration. Accelerated stability studies of ADCs have demonstrated that linking drugs molecules to antibodies leads to formation of high molecular weight species (HMWS) and diminished structural stability.[8, 17] Here, dimers or HMWS could result from either covalent crosslinking via disulfide bond formation or Ni-based oxidation or bridging via the metal upon liberation from the tag. SEC was used to investigate formation of multimers over time. In the initial chromatogram, EGF-Ni-*cla*MP elutes as a fully monomeric species exhibited by one main peak at 7.5 minutes (Figure 3.1). This is in contrast to native EGF, which elutes as two peaks at approximately 8.5 and 9.5 minutes. The presence of two peaks is dependent on buffer concentration, such that analysis of EGF on the SEC column at
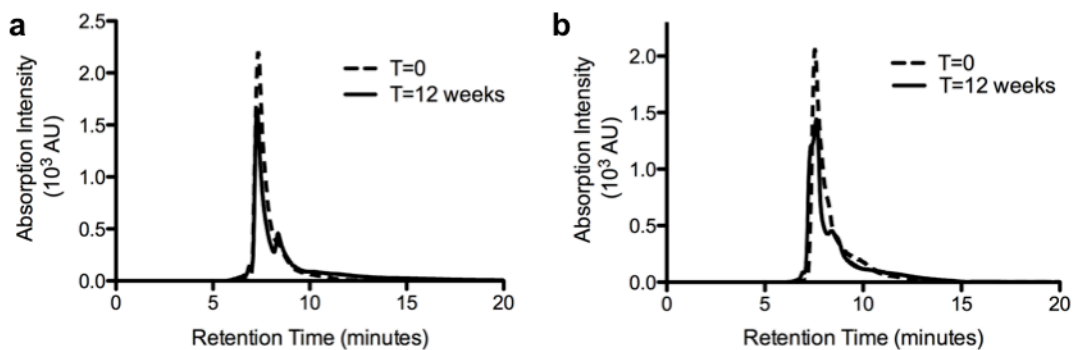
Figure 3.1 Size exclusion chromatography assessment of monomeric content. EGF-Ni-*cla*MP prepared in either Tris-Cl (a) or KPi (b). At the initial time point, only one species is present in both samples. After three months, mostly monomeric protein remains with the emergence of a shoulder on the backside of the peak suggests the formation of a smaller molecular weight species.

slightly higher buffer concentration (30 mM Tris, 15 mM KCl) yields a broad single peak at 9.0 min (data not shown). Because the behavior is reversible, these peaks likely represent two conformers. The presence of only one peak in the *cla*MP-Tagged EGF chromatogram suggests the protein's structural flexibility may be restricted to a single conformation in the presence of the negatively charged tag, which is adjacent to a patch of basic residues. At the 12-week time point, the EGF-Ni-*cla*MP peak shape is slightly altered, gaining a shoulder peak centered at 8.4 minutes, corresponding to a smaller molecular weight species or the emergence of a second conformer, as observed with EGF. These SEC data show no evidence of dimers or HMWS even at 24 weeks. Earlier eluting species indicative of protein crosslinking or aggregation are not present, implying the global conformation and the Cys residues within EGF-Ni-*cla*MP are unaltered. The presence of Ni(II) has been observed to cause oxidative crosslinking of proteins.[27] As such, the SEC data further implies that nickel is not released readily from the *cla*MP Tag because dimers and HMWS do not emerge.

To investigate changes in the Ni-*cla*MP complex, such as loss of metal or alteration of the ligands around the metal center, absorption spectroscopy and charge-based separation was
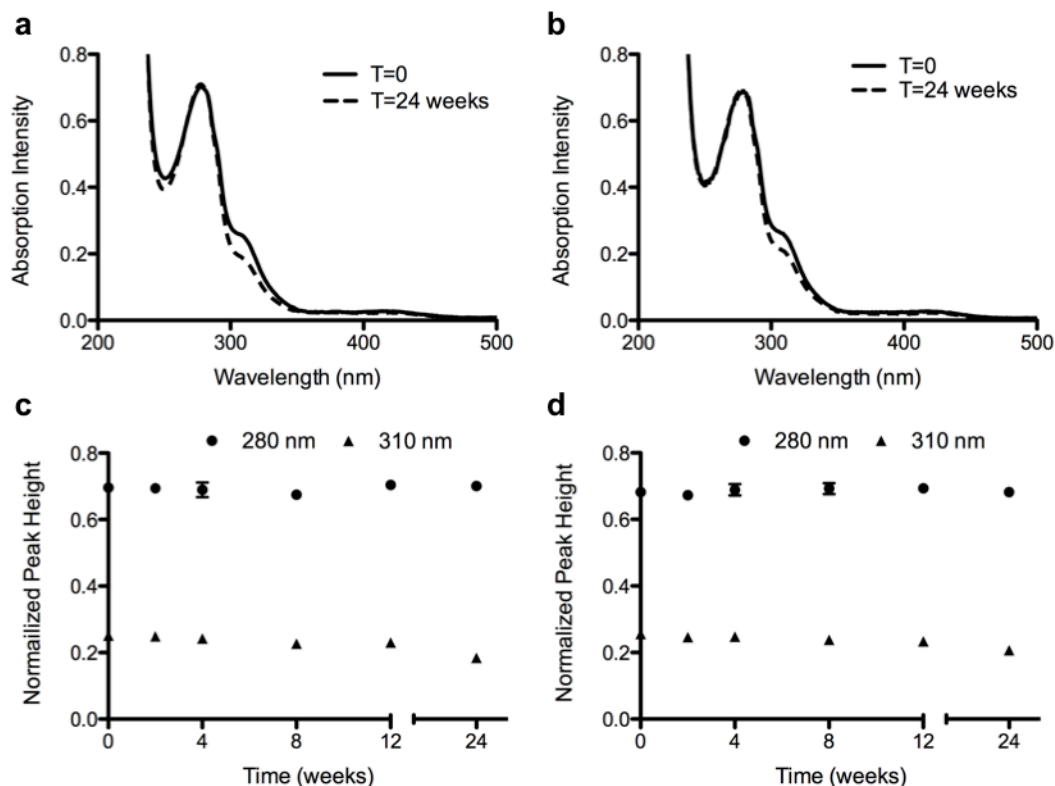
Figure 3.2 **UV-vis spectroscopy confirms Ni(II) is incorporated into the *cla*MP Tag**
Absorption spectra of samples prepared in both Tris-Cl (a) and KPi (b). Ni(II) remains
incorporated in the *cla*MP Tag regardless of buffer system chosen, as the intensity of the feature
at 310 nm remains constant over the 12-week period with Tris-Cl (c) and KPi (d). At the 24-
week time point, the intensity of the feature at 310 nm decreases, suggesting alterations occur in
the Ni-*cla*MP complex. The intensity of the feature at 280 nm remains constant, indicating the
protein concentration remains the same over the time period investigated.

used. The nickel(II) complex was used to perform this initial stability study because its

absorption spectrum is rich in structural detail,[22] and any chemical or structural transformation of

the Ni-*cla*MP complex would alter the features in the UV-vis spectrum and the unique 2⁻ charge

of the complex. The UV-vis absorption data shows that EGF-Ni-*cla*MP is stable over a 12-week

period, as no decrease in peak height was observed for the protein feature and only

approximately a 10% decrease observed for the Ni-*cla*MP feature (Figure 3.2). After 24 weeks,

a decrease in intensity of the feature at 310 nm, which corresponds to the Ni-*cla*MP complex,

begins to emerge. Because correct formation of the *cla*MP complex introduces a change in charge of 2⁻, Ni-*cla*MP-Tagged EGF exhibits a significantly different elution profile than EGF when separated on an anion exchange column. AEC was used to investigate the stability of Ni-*cla*MP by evaluating the elution profile of soluble *cla*MP-Tagged EGF over time. EGF elutes at approximately 35 minutes, whereas EGF-Ni-*cla*MP elutes at 75 minutes. Samples prepared in either 50 mM Tris-Cl, pH 7.3 or 50 mM KPi, pH 7.3, each containing 10 mM NaCl, exhibited the same chromatographic feature at the initial time point, confirming the conjugate has the same surface charge regardless of the buffer system in which it is prepared (Figure 3.3). This supports the SEC data, confirming the buffer choice does not noticeably impact the behavior of the molecule.

Like native EGF, the AEC peak area of EGF-Ni-*cla*MP prepared in either buffer is constant through eight weeks within instrument variability (+/- 20%), indicating the protein remains intact and soluble. Peak area for these preparations decreased below the variability threshold to approximately 25% and 35% at the 12-week time point for samples prepared in Tris-Cl and KPi, respectively (Figure 3.3). Multiple samples were analyzed, and peak areas acquired from different batches were observed to deviate substantially at the same time point. A 15-35% decrease in peak area was observed at the 12-week time point depending on the sample analyzed. Because the AEC peak area values were extremely noisy, these data were not useful for absolute quantitation and instead were used solely to identify general trends and changes in the profile. EGF-Ni-*cla*MP prepared in KPi exhibits a downward trend in peak area starting at the 12-week time point, while the sample prepared in Tris-Cl also decreases in peak area, although to a lesser extent. There is also the emergence of an additional peak around 50 minutes in the AEC chromatogram of the sample prepared in KPi at the 12-week time point, and this feature grows in

intensity at later time points. Because earlier elution indicates a specie is less negatively charged than the conjugate, this data suggests a cleavage event, likely resulting in removal of the Ni-*cla*MP complex, and/or alterations in the protein structure near this region has occurred.
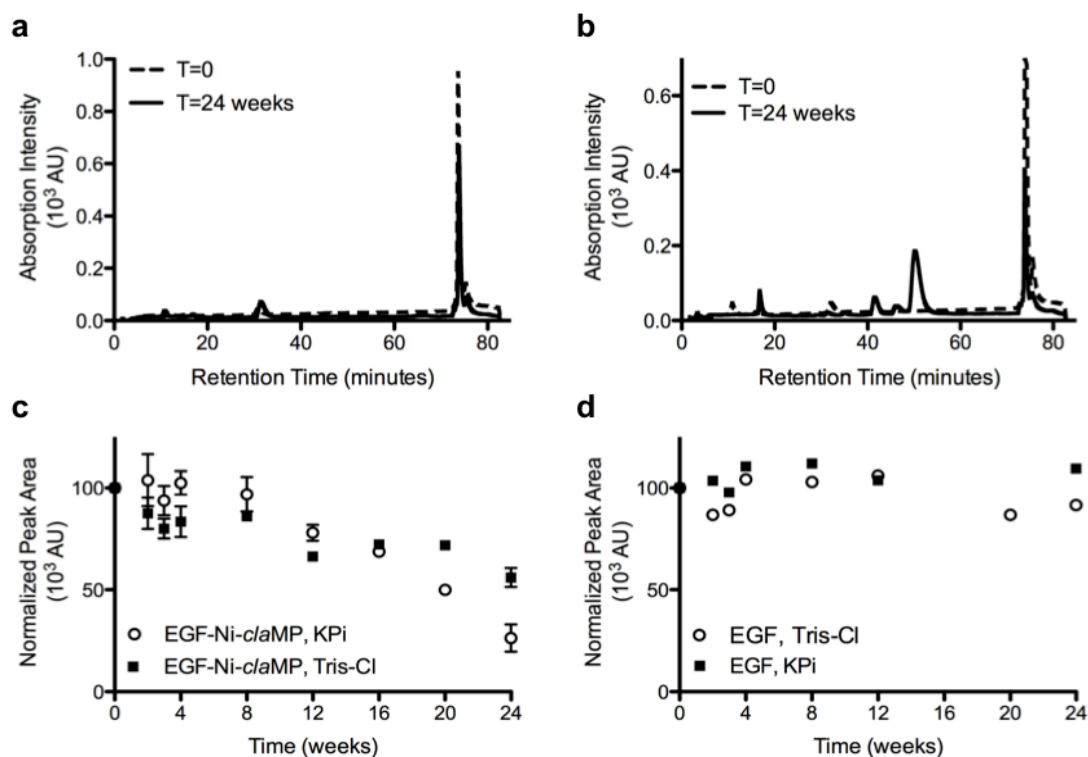
Figure 3.3 Trend analysis by anion exchange chromatography to assess retention of intact EGF-Ni-*cla*MP. The initial chromatograms of the conjugate prepared in either Tris-Cl (a) or KPi (b) are very similar, indicating the buffer choice does not alter formation of conjugate. In the sample prepared in KPi buffer, a peak emerges at 50 minutes, indicating formation of a species with a less negative charge. c) The area of the peak at 74 minutes was plotted to reflect the amount of intact EGF-Ni-*cla*MP remaining. At 12 weeks, the area of this peak decreases by 25% and 35% for the samples prepared in Tris-Cl and KPi, respectively. At 24 weeks, the area of this peak in samples prepared in Tris-Cl and KPi decreases by approximately 40% and 70%, respectively. d) The area of the peak at 35 minutes was plotted to reflect the amount of intact EGF in solution, and it remains constant over the 24-week time period. 20% variability was observed for both EGF and EGF-Ni-*cla*MP.

3.3.2 Identification of Cleavage Fragment

To further investigate the possible cleavage event and determine if it leads to removal of the Ni-*cla*MP complex from EGF, mass spectrometry (MS) was performed. MS analysis of the

sample after 24 weeks reveals hydrolysis of the peptide backbone occurs following L52 and preceding R53, the C-terminal residue of the native EGF sequence. This validates the SEC data, which shows emergence of a smaller molecular weight species. It also aligns with the AEC results, which indicate the degradant is less negative than the conjugate but more negative than native EGF. EGF without its terminal Arg and the Ni-claMP complex would exhibit a charge of 5⁻, causing it to elute later than native EGF but earlier than EGF-Ni-*cla*MP. The MS data support hydrolytic cleavage of the peptide backbone. Because of the large variability observed between different sample preparations in the rate at which this species emerges and that cleavage occurs between Leu52 and Arg53, we hypothesize the primary degradant observed is caused by a minor, undetectable protease contaminant. Both chymotrypsin and pepsin have been shown to cleave on the C-terminal side of leucine,[28] which is consistent with the location of cleavage in this EGF conjugate. Our system undergoes a long lab phase before cleavage occurs, suggesting that the protease activity may be dependent on sufficient accumulation of free Ni(II). Over time, Ni(II) may be released from the *cla*MP Tag, which then catalyzes the proteolytic cleavage of the protein. However, evidence of metal-induced hydrolysis has been reported to occur in the presence of various ions, including Ni(II).[29] This is specific to the sequence SHHK and depends on the presence of one of two embedded His. Complete cleavage occurs in a substantially shorter time frame, suggesting that it is not the same mechanism since our system undergoes a long lag period before degradation begins. In addition, it is clear that despite the presence of molecular oxygen in the samples throughout the incubation, no evidence of oxidation (e.g. methionine sulfoxide) was detected in the MS data, indicating the presence of the Ni-*cla*MP Tag complex does not lead to chemical incorporation of oxygen into the targeting protein even after 24 weeks.

The emergence of the cleaved fragment is accompanied by visible precipitation. The absorption at 280 nm remains constant at 24 weeks, but the absorption of the Ni-*cla*MP feature at 310 nm decreases, suggesting selective precipitation of the Arg53-Ni-*cla*MP fragment. Analysis of the NMR peak intensities also confirms precipitation of neither EGF nor EGF-Ni-*cla*MP occurs. It was determined previously that the cysteine residues are unaltered by addition of the tag and as such these resonances can be used to

Figure 3.4 Plot of NMR peak heights. The peak height corresponding to Cys residues confirms the protein structure is maintained and no loss of protein from solution occurs. The peak heights corresponding to the Gly from the *cla*MP tag and C-terminal residues from EGF (L52, R53) decrease, indicating a change in structure in this region.

quantify the total amount of soluble monomeric protein in the sample. The peak heights of the cysteine residues in the HSQC spectra remain constant over the 24-week period, confirming the intact EGF-Ni-*cla*MP and cleaved EGF components remain in solution. The peak heights corresponding to Arg53 and Gly from the *cla*MP Tag decrease in intensity over the 24-week time period, further indicating cleavage and precipitation of the Arg53-Ni-*cla*MP Tag fragment likely occurs (Figure 3.4). Because the net charge of the Arg53-Ni-*cla*MP fragment is 1⁻, presuming no other alterations, its aqueous solubility may be limited, particularly in the presence of the highly soluble, negatively charged protein, as might be predicted based on general principles and other peptides.[11]
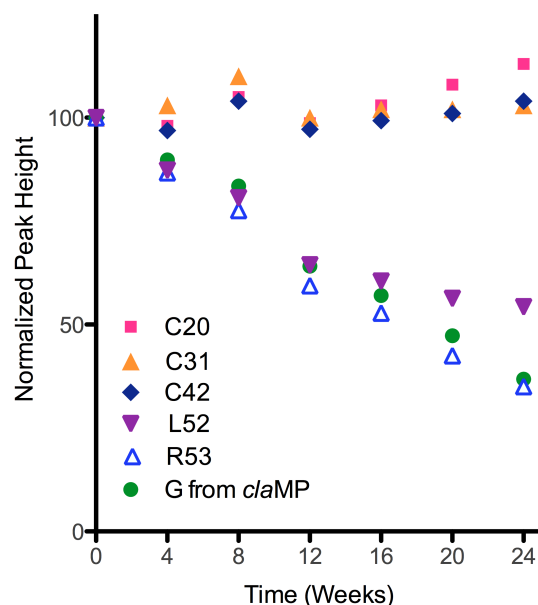
The likelihood of this cleavage event occurring with a *cla*MP-Tagged bioconjugate can be reduced by various standard measures, including frozen storage, lyophilization, and improving purification procedures to better eliminate proteolytic contaminants derived from the host, in this case *E. coli*. In this study, a simple two-step purification was used, and no precautions were taken to inhibit proteases, chelate residual or released metal, or eliminaate oxygen and light exposure during storage. Follow-up studies to examine each parameter individually are being pursued to characterize their influence on conjugate stability. In the construct examined here, a glycine spacer was included between EGF and the tag. Examination of a matrix of synthetic peptides and tagged protein variants has shown the spacer is unnecessary for metal complex formation (unpublished data), indicating the spacer may be removed or substituted. As such, if a spacer is desired, residues less prone to rapid cleavage than Gly can be utilized to reduce potential proteolysis. Optimization of these external factors will allow for the formation of highly stable *cla*MP-Tagged bioconjugates. The basal conditions analyzed provide a good starting point for identifying stable formulation conditions, which is an important consideration in the development of bioconjugates utilizing the *cla*MP Tag technology.

3.3.3 Impact of the claMP Tag on the Structural Stability of EGF

Because SEC showed the EGF proteins remained entirely monomeric and soluble at 24 weeks, 2D heteronuclear NMR was used to more closely investigate the local structural stability of *cla*MP-Tagged EGF. It has previously been shown that addition of the thiol-containing sequence to the C-terminus of EGF does not negatively impact native disulfide bond formation or the higher order structure of the molecule.[24] Here, the initial ^{1}H-^{15}N HSQC spectra of EGF-Ni-*cla*MP prepared in each buffer overlapped nicely, confirming the tertiary structure of EGF-Ni-*cla*MP is not altered by buffer selection (Figure 3.5), consistent with SEC and AEC results.
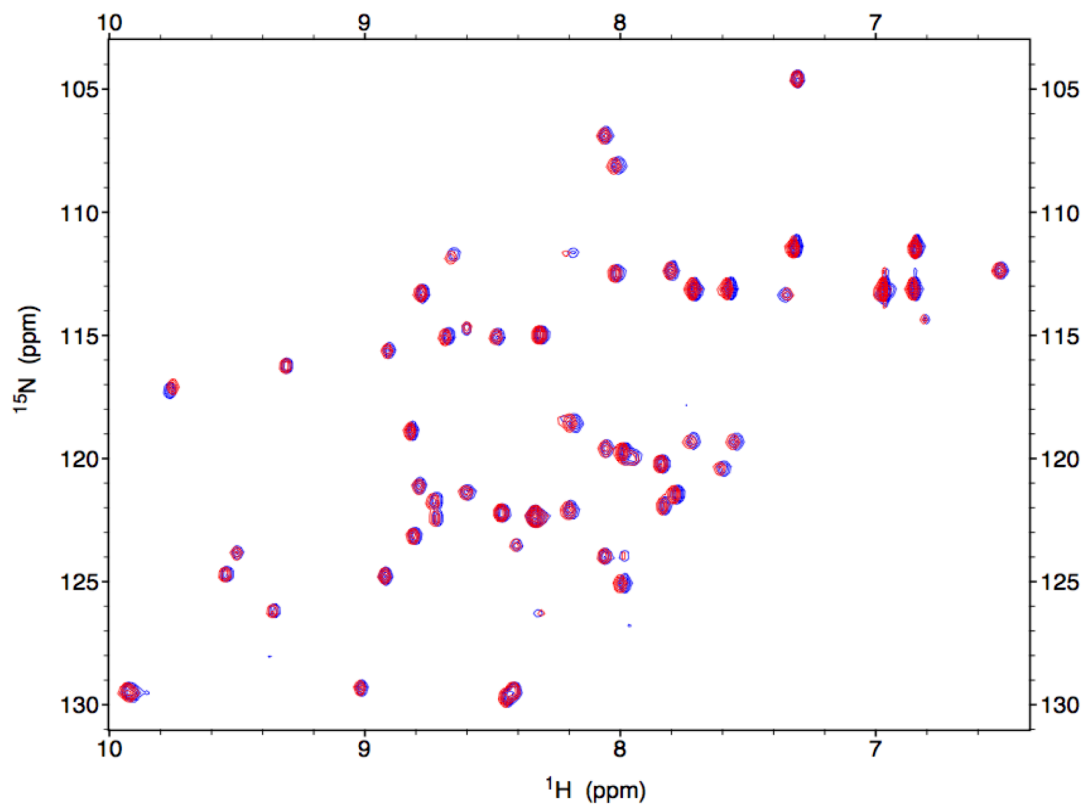
Figure 3.5 ^{1}**H-**^{15}**N HSQC spectra of EGF-Ni-*cla*MP in two buffer systems.** It confirms that buffer selection has no effect on the structure of EGF-Ni-*cla*MP, as the peaks of the samples prepared in Tris-Cl (red) and KPi (blue) overlap completely.

At the 12-week time point, additional peaks emerge near the initial resonance positions corresponding to the backbone amides from C-terminal residues W49, W50, E51 and L52. The indole NH from the side chains of W49 and W50 begin to split from one overlapped peak into two. These limited site-specific changes become more distinct at the 24-week time point, yet the global structure and disulfide network of the protein remain the same (Figure 3.6).
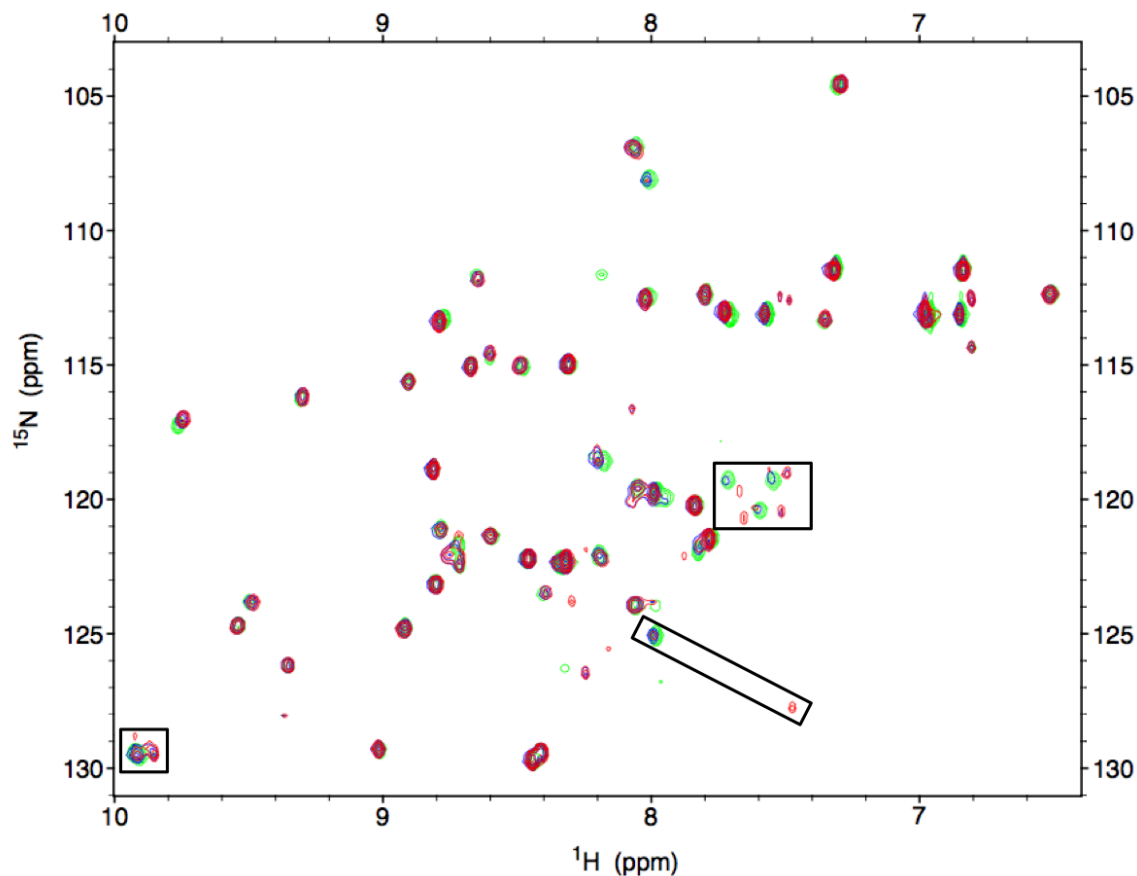
Figure 3.6 ^{1}H-^{15}N HSQC spectra of aged EGF-Ni-*cla*MP. EGF-Ni-*cla*MP spectra over a 24-week period with T=0 (green), T=12 weeks (blue), and T=24 weeks (red) shown. The tertiary structure of the protein is maintained, but specific differences are observed in peaks corresponding to the C-terminal region (boxed).

Altering the surface properties of the molecule through *cla*MP Tag addition did not significantly affect the structure or stability of EGF. Thermal titration of EGF and EGF-Ni-*cla*MP were performed to determine the effect of adding the tag on the unfolding and aggregation of the protein. The unusual fold of EGF produces a CD spectrum with atypical features. The CD features of Ni-*cla*MP overlap extensively with EGF above 220 nm and are opposite in sign, leading to the absence of an informative signal in this portion of the spectrum. As such, unfolding of EGF was monitored at 200 nm to detect changes in random coil. With and without the tag, EGF behaved identically (Figure 3.7). The onset temperature for unfolding
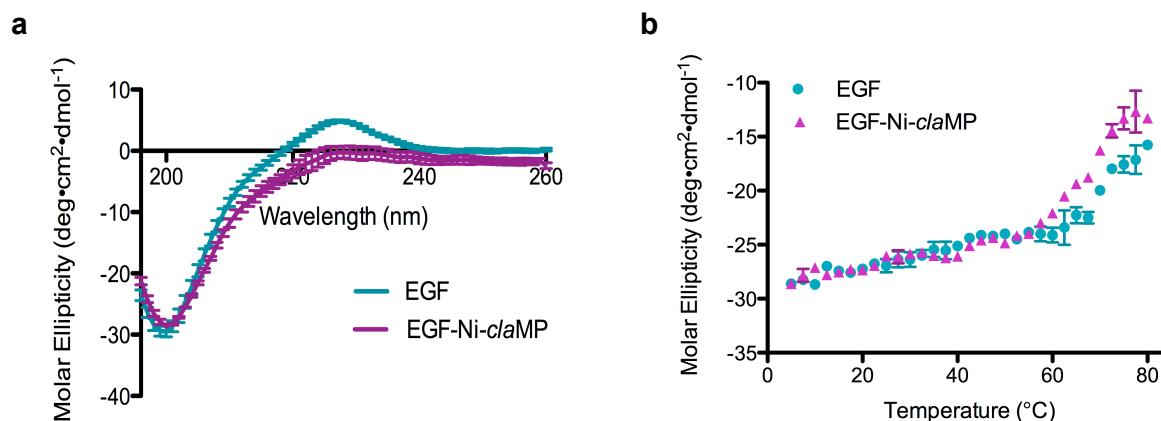
Figure 3.7 CD Thermal Analysis of EGF Variants. a) Initial CD spectra of EGF variants taken at 5 °C. b) Plot of CD intensity at 200 nm of EGF and EGF-Ni-*cla*MP as a function of temperature. The tagged and non-tagged variants exhibited similar changes in random coil, confirming that the *cla*MP Tag does not diminish protein stability.

($T_{m,onset}$) is indicative of the physical instability of the conjugate, as reported for ADCs.[8] As such, the fact that the $T_{m,onset}$ values of *cla*MP-Tagged EGF and native EGF are indistinguishable, indicates that incorporation of the *cla*MP Tag is innocuous. Visible precipitation was not observed and unfolding was reversible for both species. SLS was used to monitor formation of aggregates as a function of temperature, but no scattering could be detected for either protein up to the maximum temperature of 80 °C. Because native EGF is highly negatively charged, it is not unexpected that making it more negative by addition of the Ni-*cla*MP complex does not diminish stability. Charge repulsion and retention of the disulfide bonds likely protects EGF from aggregation. In the context of this system, alterations in surface charge due to *cla*MP Tag addition were shown to have minimal effects on protein stability, which is promising for application of this tag in biotechnology applications. Additional studies are ongoing to evaluate the effects on stability of incorporating the *cla*MP Tag into a diverse set of proteins that vary in isoelectric point, size, charge, and architecture.

3.4 CONCLUSION

In this study, the Ni(II)-*cla*MP Tag complex was shown to retain metal binding and not negatively alter the higher order structural stability of the protein molecule to which it is attached. Although cleavage of the Arg53-Ni-*cla*MP fragment occurs, this is likely due to protease contamination, which can be mitigated by increasing sample and/or reagent purity as well as storing it frozen and/or dried. Therefore, this stability analysis of the EGF-Ni-*cla*MP conjugate confirms that the *cla*MP Tag module is structurally and chemically stable and does not adversely impact the physicochemical stability of the protein to which it is attached, here EGF.

3.5 REFERENCES

1. De León-Rodríguez, L. M., Kovacs, Z. (2008) The Synthesis and Chelation Chemistry of DOTA-Peptide Conjugates. *Bioconjugate Chem.* *19*, 391-402.

2. Breeman, W. A. P., Kwekkeboom, D. J., de Blois, E., de Jong, M., Visser, T. J., Krenning, E. P. (2007) Radiolabelled regulatory peptides for imaging and therapy. *Anti-Cancer Agents Med. Chem.* *7*, 345-357.

3. Wängler, C., Buchmann, I., Eisenhut, M., Haberkorn, U., Mier, W. (2007) Radiolabeled peptides and proteins in cancer therapy. *Protein Pept. Lett.* *14*, 273-279.

4. Knör, S., Modlinger, A., Poethko, T., Schottelius, M., Wester, H.-J., Kessler, H. (2007) Synthesis of novel 1,4,7,10-tetraazacyclodecane-1,4,7,10-tetraacetic acid (DOTA) derivatives for chemoselective attachment to unprotected polyfunctionalized compounds. *Chem. - Eur. J.* *13*, 6082-6090, S6082/1-S6082/20.

5. Lewis, M. R., Kao, J. Y., Anderson, A.-L. J., Shively, J. E., Raubitschek, A. (2001) An improved method for conjugating monoclonal antibodies with N-Hydroxysulfosuccinimidyl DOTA. *Bioconjugate Chem.* *12*, 320-324.

6. Lewis, M. R., Raubitschek, A., Shively, J. E. (1994) A Facile, Water-Soluble Method for Modification of Proteins with DOTA. Use of Elevated Temperature and Optimized pH To Achieve High Specific Activity and High Chelate Stability in Radiolabeled Immunoconjugates. *Bioconjugate Chem.* *5*, 565-76.

7. Wakankar, A. A., Feeney, M. B., Rivera, J., Chen, Y., Kim, M., Sharma, V. K., Wang, Y. J. (2010) Physicochemical Stability of the Antibody-Drug Conjugate Trastuzumab-DM1: Changes due to Modification and Conjugation Processes. *Bioconjugate Chem.* *21*, 1588-1595.

8. Beckley, N. S., Lazzareschi, K. P., Chih, H.-W., Sharma, V. K., Flores, H. L. (2013) Investigation into Temperature-Induced Aggregation of an Antibody Drug Conjugate. *Bioconjugate Chem.* *24*, 1674-1683.

9. Laurence, J. S., Middaugh, C. R., Fundamental structures and behaviors of proteins. In *Aggregation of Therapeutic Proteins*, Wang, W.; Roberts, C. J., Eds. John Wiley & Sons, Inc.: Hoboken, NJ, 2010; pp 1-61.

10. Dominy, B. N., Minoux, H., Brooks, C. L., III (2004) An electrostatic basis for the stability of thermophilic proteins. *Proteins: Struct., Funct., Bioinf.* *57*, 128-141.

11. Glyakina, A. V., Garbuzynskiy, S. O., Lobanov, M. Y., Galzitskaya, O. V. (2007) Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms. *Bioinformatics 23*, 2231-2238.

12. Narhi, L. O., Arakawa, T., Aoki, K., Wen, J., Elliott, S., Boone, T., Cheetham, J. (2001) Asn to Lys mutations at three sites which are N-glycosylated in the mammalian protein decrease the aggregation of Escherichia coli-derived erythropoietin. *Protein Eng. 14*, 135-140.

13. Wang, L., Amphlett, G., Blättler, W. A., Lambert, J. M., Zhang, W. (2005) Structural characterization of the maytansinoid-monoclonal antibody immunoconjugate, huN901-DM1, by mass spectrometry. *Protein Sci. 14*, 2436-2446.

14. Acchione, M., Kwon, H., Jochheim, C. M., Atkins, W. M. (2012) Impact of linker and conjugation chemistry on antigen binding, Fc receptor binding and thermal stability of model antibody-drug conjugates. *MAbs 4*, 362-72.

15. Boylan, N. J., Zhou, W., Proos, R. J., Tolbert, T. J., Wolfe, J. L., Laurence, J. S. (2013) Conjugation Site Heterogeneity Causes Variable Electrostatic Properties in Fc Conjugates. *Bioconjugate Chem. 24*, 1008-1016.

16. Strop, P., Liu, S.-H., Dorywalska, M., Delaria, K., Dushin, R. G., Tran, T.-T., Ho, W.-H., Farias, S., Galindo Casas, M., Abdiche, Y., Zhou, D., Chandrasekaran, R., Samain, C., Loo, C., Rossi, A., Rickert, M., Krimm, S., Wong, T., Michael Chin, S., Yu, J., Dilley, J., Chaparro-Riggers, J., Filzen, G. F., O'Donnell, C. J., Wang, F., Myers, J. S., Pons, J., Shelton, D. L., Rajpal, A. (2013) Location Matters: Site of Conjugation Modulates Stability and Pharmacokinetics of Antibody Drug Conjugates. *Chem. Biol. (Oxford, U. K.) 20*, 161-167.

17. Adem, Y. T., Schwarz, K. A., Duenas, E., Patapoff, T. W., Galush, W. J., Esue, O. (2014) Auristatin Antibody Drug Conjugate Physical Instability and the Role of Drug Payload. *Bioconjugate Chem.*, Ahead of Print.

18. Jackson, D., Atkinson, J., Guevara, C. I., Zhang, C., Kery, V., Moon, S.-J., Virata, C., Yang, P., Lowe, C., Pinkstaff, J., Cho, H., Knudsen, N., Manibusan, A., Tian, F., Sun, Y., Lu, Y., Sellers, A., Jia, X.-C., Joseph, I., Anand, B., Morrison, K., Pereira, D. S., Stover, D. (2014) In vitro and in vivo evaluation of cysteine and site specific conjugated herceptin antibody-drug conjugates. *PLoS One 9*, e83865/1-e83865/14, 14 pp.

19. Junutula, J. R., Raab, H., Clark, S., Bhakta, S., Leipold, D. D., Weir, S., Chen, Y., Simpson, M., Tsai, S. P., Dennis, M. S., Lu, Y., Meng, Y. G., Ng, C., Yang, J., Lee, C. C., Duenas, E., Gorrell, J., Katta, V., Kim, A., McDorman, K., Flagella, K., Venook, R., Ross, S., Spencer, S. D., Wong, W. L., Lowman, H. B., Vandlen, R., Sliwkowski, M. X., Scheller, R. H., Polakis, P., Mallet, W. (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol. 26*, 925-932.

20. Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2010) Novel Tripeptide Model of Nickel Superoxide Dismutase. *Inorg. Chem. 49*, 362-364.

21. Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2011) MAPping the Chiral Inversion and Structural Transformation of a Metal-Tripeptide Complex Having Ni-Superoxide Dismutase Activity. *Inorg. Chem. 50*, 2479-2487.

22. Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2013) Embedding the Ni-SOD Mimetic Ni-NCC within a Polypeptide Sequence Alters the Specificity of the Reaction Pathway. *Inorg. Chem. 52*, 77-83.

23. Laurence, J. A. S., Vartia, A. A., Krause, M. E. Metal abstraction peptide (MAP) tag and associated methods. U.S. Patent 8,110,402, February 7, 2012.

24. Mills, B. J., Mu, Q., Krause, M. E., Laurence, J. S. (2014) *cla*MP Tag: A versatile inline metal-binding platform based on the metal abstraction peptide. *Bioconjugate Chem. accepted.*

25. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277-93.

26. Goddard, T. D., Kneller, D. G. (2004) SPARKY. San Francisco: University of California.

27. Gill, G., Richter-Rusli, A. A., Ghosh, M., Burrows, C. J., Rokita, S. E. (1997) Nickel-dependent oxidative cross-linking of a protein. *Chem Res Toxicol 10*, 302-9.

28. Keil, B., *Specificity of Proteolysis*. Springer-Verlag: Berlin; New York, 1992.

29. Bal, W., Liang, R., Lukszo, J., Lee, S.-H., Dizdaroglu, M., Kasprzak, K. S. (2000) Ni(II) Specifically Cleaves the C-Terminal Tail of the Major Variant of Histone H2A and Forms an Oxidative Damage-Mediating Complex with the Cleaved-Off Octapeptide. *Chem. Res. Toxicol. 13*, 616-624.

CHAPTER 4.
DECOUPLING PROTEIN PURIFICATION AND METAL INSERTION AND QUANTIFYING RETENTION OF NI(II) BY *cla*MP TAG IN PRESENCE OF HIGH AFFINITY CHELATOR: ANALYSIS OF MALTOSE-BINDING PROTEIN (MBP) VARIANTS

4.1 INTRODUCTION

Incorporation of metal-binding modalities within a protein sequence allows for the generation of a bioconjugate capable of achieving targeted delivery of metal ions for use in imaging and therapeutic applications. Most metal-binding modalities are small, organic chelators that require conjugation to the protein surface,[1-5] which is typically done after complete purification of the protein. Inline metal-binding tags are engineered into the gene of the desired protein to generate "linker-less" conjugates that do not require further chemical modification.[6-7] Inline tags do introduce additional complexities because they are attached to the protein before the protein is fully purified. Purification procedures are time-consuming, which provides the opportunity for undesirable reactivity of the metal-binding tag during this process. Therefore, when using peptide-based tags, the point of greatest metal insertion efficiency during the purification process must be determined to ensure that the highest yield of the desired product is achieved. The effects of tag placement on the metal insertion efficiency must also be considered, as the fold of the protein could alter tag accessibility for metal insertion. Chemical conjugation occurs when the protein is fully folded so surface accessible residues are the most common sites of conjugation, and the specific position of attachment cannot be controlled entirely,[8-10] although methods of site-specific conjugation are being developed.[11] Peptide tags may be engineered at any position within the protein sequence but are best positioned in places where no deleterious effects on the structure or stability occur. The fold of the protein near the peptide tag can alter

metal insertion efficiency by limiting accessibility of the peptide tag if it is folded into the protein structure so the optimal site of tag placement must be determined.

Here, we present how metal insertion efficiency into the *cla*MP Tag is altered by point of metal insertion during the purification process and placement of the *cla*MP tag within the protein. The *cla*MP Tag is a novel metal-binding tripeptide based upon the metal abstraction peptide (MAP) sequence Asn-Cys-Cys (NCC) that is capable of being inserted inline within a protein sequence. The tag has been extensively investigated in terms of its Ni(II) binding potential and generation of Ni-NCC produces an overall charge of 2^- for the complex.[12-14] This tag also has the ability to bind a variety of transition metal ions other than Ni(II) including cytotoxic metals, such as Pt and Pd, and radioisotopes, such as ^{99m}Tc, $^{64/67}$Cu, and ^{60}Co,[15] making it amenable to use as a tag in imaging and therapeutic applications. Incorporation of the tripeptide into longer peptide sequences does not negatively impact metal binding,[13] allowing for use of the *cla*MP Tag as a "linker-less" conjugate.[6] Because the *cla*MP Tag contains two cysteine residues, the effects of these free thiols on the purification process must be assessed in order to successfully apply the *cla*MP Tag as an inline metal-binding module. Purification of proteins is very time-consuming, which may lead to undesirable reactivity of free thiols and elimination of a metal-coordination site. To determine if this occurs within the *cla*MP Tag, we evaluated the metal insertion efficiency when metal is introduced immediately after cellular lysis prior to purification or alternatively after purification is complete.

Maltose-binding protein (MBP) was chosen for these analyses because it allows for separation of the metal insertion and purification steps. Clean decoupling of these steps is not possible with His-tagged proteins because the polyhisitidine sequence also is capable of binding metal making it more difficult to selectively identify binding to the two tags. The MBP-based

system permits analysis of different approaches to incorporate metal and the point of highest metal insertion efficiency during the purification process. Antibodies are purified most often based on their high Protein A/G affinity,[16-18] which, like MBP, does not involve use of metal.[19-21] The MBP system provided a more accessible approach for evaluating the importance of the timing of metal insertion into the *cla*MP tag in order to better understand application of the *cla*MP Tag technology to antibody systems, which are the primary scaffolds used for generating targeted therapeutic and diagnostic agents.[2, 9, 11, 22-24]

In order to ascertain the usefulness of the *cla*MP Tag as a metal-binding platform, the effects of the tag on different types of protein systems must be established. The *cla*MP Tag has been previously been shown to have no discernable affect on the expression, structure, or stability of the disulfide-containing protein epidermal growth factor (EGF).[6, 25] Because this tag would need to be used with much larger proteins than the 53 amino acid protein EGF for successful use as a bioconjugate, insertion of the *cla*MP Tag within the MBP sequence was used to assesses its affects on a larger protein with different structural properties. MBP is a 40 kDa single-domain protein, which has a more typical structure[26] and exhibits a greater negative charge than EGF. EGF is highly cross-bridged by disulfide bonds and contains few secondary structural elements, whereas MBP has an extensive amount of α-helical character defining its overall fold and no disulfide bonds.[26] The *cla*MP Tag was added to either the N- or C-terminus of MBP to determine if the site of tag addition would differentially modulate protein properties based on the structure of the region to which it is added.

The success of metal-based bioconjugates is also highly dependent on their ability to retain metal binding in the presence of very high affinity chelators, such as EDTA. It has been observed with peptide models that the *cla*MP Tag retains metal in the presence of competitive

species, but this had not been quantitatively determined for the tag within a protein system. Here, we investigate the effect of excess EDTA on retention of Ni(II) by the *cla*MP complex within *cla*MP-Tagged MBP.

4.2 MATERIALS AND METHODS

4.2.1 Cloning and Construction of the Expression Plasmid

Four different DNA sequences were prepared to generate MBP variants: control MBP and three individually tagged variants containing the *cla*MP Tag at either the N-terminus (NCC-MBP, GNCCG-MBP) or C-terminus (MBP-GNCCG) of the protein sequence. A plasmid containing the MBP sequence from the MBP fusion protein system was obtained (NEB, #N8108S), and the MBP sequence was amplified using PCR and subcloned into the pET-22b(+) vector. Custom primers (IDT) were used to generate N-terminal (NCC-MBP, GNCCG-MBP) and C-terminal variants (MBP-GNCCG) and amplified using PCR for insertion into pET-22b(+) vector using the NdeI and BamHI restrictions sites. In each primer, the restriction site sequence is bold, the portion corresponding to the *cla*MP Tag is italicized, and the remaining portion corresponds to MBP. The primers used are as follows: 5'- **GGGAATTCCATATG***AACTGCTGC*AAAATCGAAGAAGGTAAACTGGTAATCTGG -3' (forward primer with NCC on the N-terminus), 5'-**GGGAATTCCATATG***GGAAACTGCTGCGGC*AAAATCGAAGAAGGTAAACTGG-3' (forward primer with GNCCG on the N-terminus), 5'-**CGCGGATCC**TCAAGTCTGCGCGTCTTTCAGGGCTTCATCG-3' (reverse primer for N-terminally tagged variants), 5'-**GGGAATTCCATATG**AAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCG-3' (forward primer for C-terminally tagged variants), and 5'-**CGCGGATCC**TCAGCCGCAGCAGTTTCCAGTCTGCGCGTCTTTCAGGGC-3' (reverse primer with GNCCG on the C-terminus).

The amplification reactions were purified using the QIAquick PCR Purification Kit (Qiagen). The inserts and pET-22b(+) vector were digested using the NdeI and BamHIHF restriction enzymes for two hours at 37 °C. A four-fold molar excess of digested insert was added to digested vector and incubated with T4 ligase at 16 °C for 16 hours. The ligation product was transformed into the DH5α *Escherichia coli* (*E. coli*) cell strain using the standard heat shock method, and the DNA was harvested after selection and growth on LB medium containing 100 μg/mL ampicillin (AMP100). Plasmid DNA was purified using a miniprep kit (Qiagen), and the intended product was verified by DNA sequencing (UC Berkeley DNA Sequencing Facility).

4.2.2 Protein Expression

Each plasmid was transformed into the BL21 (DE3) *E. coli* strain (Novagen) using the standard heat shock protocol and transformants selected on LB agar plates with AMP100. Individual colonies were cultured in LB/AMP100 at 37 °C and 250 rpm for 16 hours and then 6 mL was transferred to 1 L in LB/AMP100 plus 0.2% glucose. Protein expression was induced at OD_{600} of 0.7 by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested after 4 hours using centrifugation, and pellets were stored at -80 °C until use.

To quantitatively determine the expression levels of the MBP variants, a sample was taken from each culture after the four-hour incubation. The total cell mass collected for each variant was determined based on the sample volume collected and OD_{600} value (volume sample=$(0.8*200)/OD_{600}$ at 4 hours). The sample was centrifuged and the supernatant removed to yield only the cell pellet. The pellet was solubilized in 8 M urea, followed by an equivalent amount of SDS loading buffer.

4.2.3 SDS-PAGE Analysis

MBP variants were separated using a discontinuous system consisting of a 5% (v/v) stacking gel and a 12% (v/v) resolving gel. The samples were prepared in non-reducing or reducing SDS loading buffer and heated for 10 minutes at 90 °C before being loaded onto the gel. A prestained, dual-color molecular weight marker was used for reference (BioRad, #161-0374). Gels were stained using Coomassie (R-250) to visualize protein.

Densitometry analysis was performed on Coomassie-stained SDS-PAGE gels using the Typhoon TRIO Variable Mode Imager (Amersham Biosciences). Relative quantitation was performed using the ImageQuantTL software (Amersham Biosciences).

4.2.4 Protein Purification

Each pellet (1 L of culture) was resuspended in 25 mL Lysis Buffer (50 mM Tris-Cl, 200 mM NaCl, pH 8.4 containing 2.5 mM β-mercaptoethanol (βME)) and lysed using a French Press at 21,000 psi. A control sample was also prepared in the absence of βME in the lysis buffer to verify that its addition does not alter metal incorporation. The cellular debris was removed by centrifugation at 21,000 rcf for 1 hour at 4 °C. The supernatant, which contained the protein, was filtered through a 1.2 μm followed by a 0.45 μm filter. Metal insertion before and after amylose resin purification was tested, using a protein concentration of approximately 0.1 mM during each metal-incubation step. For insertion of metal before amylose resin purification, the filtered supernatant was added to 15 mL IMAC Sepharose 6 Fast Flow resin charged with nickel (GE Lifesciences) and allowed to incubate for one hour to facilitate metal transfer to the *cla*MP Tag. Because the protein lacks the polyhisitidine tag, it is not expected to bind to the IMAC resin. As such, the flowthrough from the column was collected, washed with an additional three column volumes (CV) of Loading Buffer (50 mM Tris, 200 mM NaCl pH 8.0) to completely

elute the protein from the column. The protein fraction (75 mL) was diluted with an additional 25 mL Loading Buffer and passed over a 15/100 flex column containing 15 mL amylose high flow resin (NEB), using a peristaltic pump. After loading the protein, the amylose resin was washed with 2 CV Loading Buffer. The protein was eluted from the amylose resin using Loading Buffer containing 10 mM maltose.

The effects of inserting metal after complete purification, following elution from the amylose resin, also were tested. The filtered lysate was diluted five-fold, added to the amylose resin, washed, and eluted as above. The eluted protein was then incubated with 15 mL IMAC Sepharose 6 Fast Flow resin charged with nickel for one hour and eluted as above. The efficiency of metal insertion using either IMAC Sepharose 6 Fast Flow resin charged with nickel or the preformed, soluble complex of nickel incorporated nitrilotriacetic acid (Ni-NTA) was investigated. During these studies, the metal was incubated either with Ni-IMAC resin or 1:1 molar ratio of a solution of Ni-NTA for one hour before purification with the amylose resin. Because the *cla*MP Tag binds one Ni(II) per *cla*MP sequence, a 1:1 Ni(II): *cla*MP-Tagged protein ratio was chosen for the Ni-NTA incubation. Peptide studies have shown that incubation with higher ratios of Ni(II):*cla*MP sequence is unnecessary because full *cla*MP Tag saturation is achieved at a 1:1 ratio.

Following metal insertion and amylose purification, the protein was purified further using a self-packed Superdex75 C16/100 column (GE Lifesciences). Samples were concentrated using an Amicon Ultra 30 kDa MWCO concentrator (Millipore) to a final concentration of approximately 0.75 mM. Samples were quantified using a Bradford assay. The purity of the protein was confirmed with SDS-PAGE and size exclusion chromatography, and the molecular weight was confirmed using ESI-MS.

4.2.5 Absorption Analysis

After purification, samples were analyzed using UV-vis spectroscopy to confirm Ni(II) incorporation. Samples were placed in a cuvette with a 1-cm path length and spectra were acquired from 200-800 nm using a Cary 100 Bio UV-visible spectrophotometer (Varian).

4.2.6 HPLC Analysis

Anion exchange chromatography (AEC) was performed on a 4 x 250 mm BioLC ProPac Wax10 column (Dionex). The column was equilibrated in 30 mL "Buffer A" (20 mM Tris-Cl, 10 mM KCl, pH 7.5) before injection of the sample. A linear gradient from 0-100% "Buffer B" (Buffer A + 500 mM KCl) over 70 mL with a constant flow rate of 1 mL/min was used, and UV absorption at 220 nm was used to monitor protein elution from the column.

4.2.7 Determination of Ni(II) Insertion Efficiency Using SEC

The two peaks present in the SEC chromatogram were determined to correspond to Ni(II)-incorporated and Ni(II)-free versions of *cla*MP-Tagged MBP using AEC. To determine the insertion efficiency (%), the area of the peak corresponding to Ni(II)-incorporated *cla*MP-Tagged MBP was divided by the total area of the two peaks in the chromatogram and multiplied by 100%.

4.2.8 Circular Dichroism

Circular dichroism (CD) spectra were collected using Chirascan™ Plus CD Spectrophotometer (Applied Photophysics, Surrey, UK) equipped with a Peltier temperature controller. Samples were prepared at a protein concentration of 0.5 mg/mL in 50 mM KPi, pH 7.3, with 10 mM KCl, 120 mM KCl, or 140 mM KCl, which corresponds to an ionic strength of 142 mM, 252 mM, and 272 mM, respectively, and loaded into a 0.1-cm path length cuvette. Spectra were acquired every 2.5 °C from 5 to 85 °C, with a one-minute equilibration at each

temperature. An averaging time of 0.5 seconds/nm and a bandwidth of 1 nm were used. The molar ellipticity at 222 nm was plotted as a function of temperature. The background was subtracted from each data point based upon a sample containing only buffer with no protein present. Two samples were analyzed for each variant. The T_m was determined by applying a sigmoidal curve that best fit the data between 20 °C and 85 °C. The thermal unfolding curve of MBP-Ni-*cla*MP contained two transitions so two different T_m values were determined for each by applying a sigmoidal curve that best fit the data between 20 °C and 62.5 °C and between 57.5 °C and 85.0 °C. The R^2 values of these curves were all greater than 0.985.

4.2.9 EDTA Competition Analysis

EDTA was added to the samples to obtain a final concentration of 10 and 100 mM EDTA, which equates to a 10- and 100-fold molar excess over Ni-*cla*MP Tag protein in the competition reaction. The EDTA-spiked samples were allowed to incubate at 4 °C and aliquots were pulled at the 2 hour, 24 hour, and one week time points to be analyzed using UV-vis spectroscopy and HPLC as described above. Control samples without EDTA were also analyzed to confirm changes in intensity were due to EDTA addition and not protein degradation. All samples were tested in duplicate.

4.3 RESULTS AND DISCUSSION

4.3.1 Addition of *cla*MP Tag Does Not Hinder MBP Expression

It has previously been shown that *cla*MP Tag addition does not negatively impact the expression of a thiol- and disulfide-containing fusion protein composed of thioredoxin and epidermal growth factor (EGF).[6] Here, an unrelated protein having a different composition, architecture, and size was examined to further confirm *cla*MP Tag placement into recombinant proteins provides a robust, flexible platform. The molecular weight of the final Ni-*cla*MP
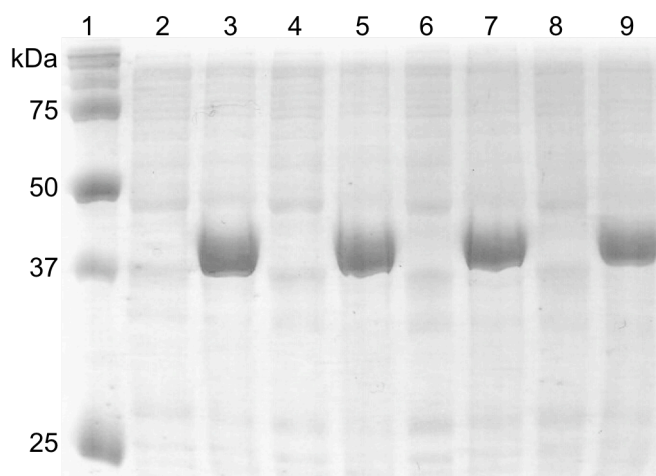
Figure 4.1 SDS-PAGE analysis of expression of *cla*MP Tag MBP variants. SDS-PAGE confirms biosynthetic inclusion of *cla*MP Tag into various positions does not appreciably impact expression of MBP. Coomassie-stained 12% SDS-PAGE gels show the expression of a 40 kDa protein from the pET-22b(+)-MBP, pET-22b(+)-NCC-MBP, pET-22b(+)-GNCCG-MBP, and pET-22b(+)-MBP-GNCCG constructs. (Lane 1) molecular weight standard, (lanes 2, 4, 6, 8) *E. coli* lysate of pET-22b(+)-MBP, pET-22b(+)-NCC-MBP, pET-22b(+)-GNCCGMBP, pET-22b(+)-MBP-GNCCG before IPTG induction, (lanes 3, 5, 7, 9) *E. coli* lysate of pET-22b(+)-MBP, pET-22b(+)-NCC-MBP, pET-22b(+)-GNCCGMBP, pET-22b(+)-MBP-GNCCG 4 hours after IPTG induction.

Tagged-MBP products is approximately 40 kDa, which is six times the size of the previously characterized EGF system. Regardless of tag location or sequence (GNCCG versus NCC), addition of the *cla*MP Tag did not negatively affect protein expression. SDS-PAGE showed that all MBP variants expressed very well with similar expression levels (Figure 4.1). No band is present at the expected molecular weight in the pre-induction samples (lanes 2, 4, 6, and 8), but a band is present at this size in the post-induction sample for each of the variants (lanes 3, 5, 7, and 9), indicating that the protein is successfully expressed. Native MBP was used as a control to determine the relative expression levels of the *cla*MP-Tagged variants. Densitometry analysis of whole cell lysates indicated that NCC-MBP expressed at similar levels as MBP, whereas GNCCG-MBP and MBP-GNCCG exhibited a 10% and 20% decrease in expression, respectively. These observed differences are within the normal batch-to-batch variation in

expression yields for a single variant and are not appreciably different. These results confirm that insertion of the *cla*MP Tag within a larger protein has a negligible effect on protein expression.

4.3.2 Effects of Metal Insertion Method on Insertion Efficiency

The generation of a metal-based bioconjugate requires high efficiency incorporation of the metal ion into the metal-binding modality. The optimal point and method of metal insertion are important parameters to investigate when generating a metal-based protein therapeutic and are highly dependent on the metal-binding entity. In this study, the metal-insertion efficiency into the *cla*MP Tag was investigated to determine the method of metal insertion and point during the purification process that leads to highest yield of the desired metal-bound bioconjugate. Whereas most metal-binding modalities are incubated with a salt form of the metal to be incorporated,[2, 7, 27] the *cla*MP Tag requires transfer of the metal ion from a chelator to achieve the desired structure. Insertion using a chelated metal ion also limits the amount of non-specific metal binding to the protein, which has been observed with incubation of MBP with $NiCl_2$.[28] It has previously been shown that Ni(II) can be successfully inserted into the *cla*MP Tag during purification of a His-tagged-protein using Ni-IMAC resin.[6] It was also demonstrated that small peptides containing the NCC module are rapidly formed using either a resin- or solution-based transfer.[12-13] In order to avoid metal insertion during the purification step, as occurs with His-tagged proteins isolated using Ni-IMAC, MBP was selected. The amylose affinity capture of MBP affords the use of a distinct separation scheme that does not involve metals. This approach, therefore, enables determination of the practical ability to separate purification of the *cla*MP Tagged protein from the metal incorporation step. The observed insertion efficiency when metal is inserted before or after purification had not been determined previously. Here, the MBP

system permits investigation into the effect of the free thiol moieties on metal incorporation efficiency, the potential for intermolecular cross-linking and the ability to include a thiol reducing agent in the solution during processing. Metal insertion efficiencies were investigated by incubating the protein with Ni-resin either in the cell lysate prior to purification or after elution of the purified protein from the amylose resin. All MBP variants were examined at approximately the same protein concentrations during the metal incubation steps for both insertion approaches, thus eliminating a concentration-based bias for the observed metal-insertion efficiencies.

Metal-bound and metal-free MBP variants were determined to elute differently on the size exclusion column (Figure 4.2). Because Ni(II) incorporation into the *cla*MP Tag causes a change in charge of 2⁻, metal-free variants will elute earlier than metal-bound variants. The

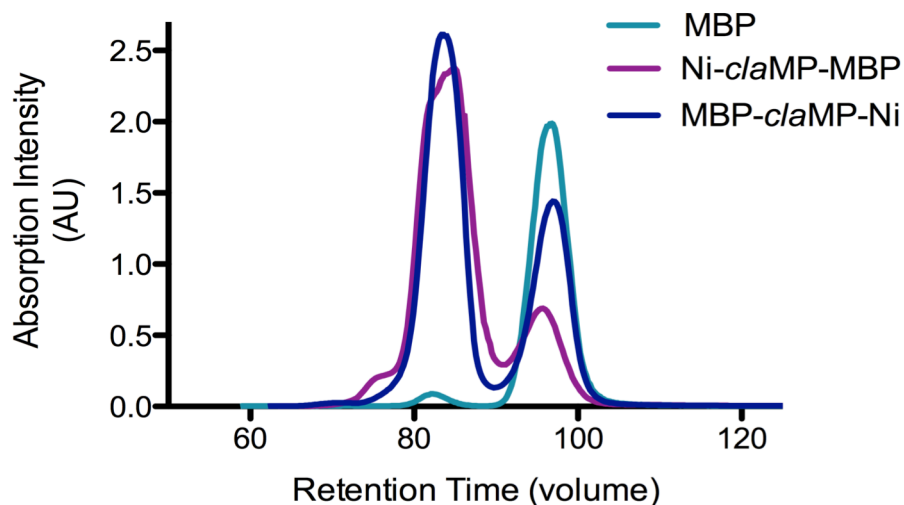**Figure 4.2   SEC separation of *cla*MP-Tagged variants of MBP.** Ni-incorporated *cla*MP-Tagged MBP variants elute earlier on the size exclusion column than unincorporated *cla*MP-Tagged variants, which allows for separation of the two species using this method.   Adequate resolution is achieved as Ni-incorporated and unincorporated variants elute approximately 15 mL apart at 80 and 95 mL, respectively.

charge state of each SEC peak was validated using anion exchange chromatography (AEC), as is shown in Figure 4.3. Metal-free *cla*MP-Tagged variants elute similarly to MBP, whereas Ni-*cla*MP-Tagged variants elute later in the gradient, indicating that metal is incorporated into the tag as they are more negatively charged. Using the AEC data, it was determined that metal-bound *cl*aMP-Tagged MBP elutes at 80 mL on the SEC, whereas metal-free *cla*MP-Tagged MBP elutes later at 95 mL (Figure 4.2). Because sufficient resolution is achieved, SEC can be used to separate metal-bound and metal-free forms of the MBP variants and quantitatively determine the yield of the two species. The metal insertion efficiency for each method of insertion is shown in Tables 4.1 and 4.2. Insertion of metal into the *cla*MP Tag prior to purification on the amylose resin leads to a 40% higher yield of metal-incorporated product than metal insertion after purification on the amylose resin. Therefore, metal insertion into the *cla*MP Tag is accomplished more effectively in the complex mixture of the cell lysate than when first fully purified, whereas the lanthanide binding tags (LBTs), which are also inserted inline within a protein sequence, are purified completely before metal insertion.[7,27] Addition of metal-binding entities to the antibody surface is also completed after antibody purification,[2] but these results suggest that metal insertion into the *cla*MP Tag inline within an antibody sequence may be most successful when completed before antibody purification using Protein A/G resin. The high incorporation efficiency observed in the cell lysate, where other metal scavenging proteins are present, indicates the claMP Tag effectively competes for binding. SDS-PAGE confirms the decrease in metal insertion efficiency observed when Ni(II) is presented after purification is not due to disulfide formation among cysteine residues within the tag. The free thiols within the tag are stable and do not become oxidized to form disulfides over the manufacturing time scale. The effect of reducing agents, such as β-mercaptoethanol (βME), on metal insertion efficiency was
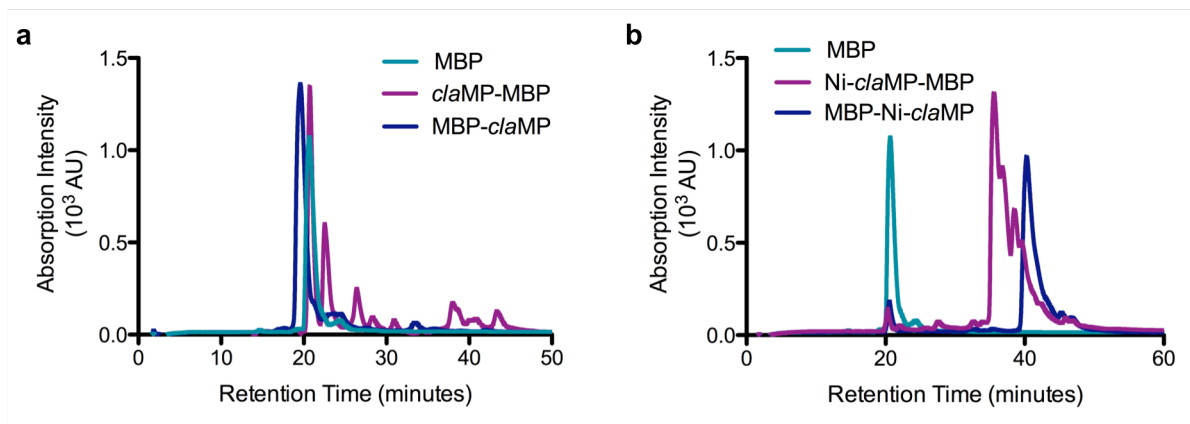
Figure 4.3 AEC analysis of *cla*MP-Tagged MBP variants. Anion exchange chromatography was used to confirm metal insertion into the tag. The MBP variants without metal bound into the *cla*MP Tag eluted similarly to native MBP at 20 minutes (a), whereas the Ni-occupied Ni-*cla*MP-MBP and MBP-Ni-*cla*MP proteins (b) eluted later.

also investigated. For these investigations, metal insertion before purification was performed in both the presence and absence of 2.5 mM βME. βME was shown to have a negligible effect on metal insertion into the tag, as the insertion efficiencies in the presence and absence of βME were similar (data not shown). Based on this data set, it is unclear why metal incorporation is less effective following purification.

Table 4.1 Metal Insertion Efficiency Before and After Purification

Variant	Pre-Amylose Purification	Post-Amylose Purification
GNCCG-MBP	80.01	36.96
NCC-MBP	70.33	X
MBP-GNCCG	66.06	30.87

*Values reported as the percent Ni-incorporated relative to the overall protein present. Locations marked with X were not investigated.

Successful incorporation of metal into the *cla*MP Tag requires transfer of the metal from a partial chelator to generate the desired complex. It also limits non-specific binding of the metal to the protein. Ni-charged IMAC resin or Ni-NTA was used to accomplish resin or solution transfer, respectively. The insertion efficiency can vary depending on which method is used.

Resin transfer resulted in approximately 75% yield of the desired metal-incorporated product. The one-hour metal incubation time used in these studies is much longer than what is required for metal transfer to the tag. Addition of $NiSO_4$ to the peptide module leads to immediate insertion of metal into the tag,[13-14] thus suggesting that the results observed with the one-hour incubation with the MBP variants are not due to the kinetics of metal-insertion into the tag. Because increasing the amount of Ni(II)-resin used during the incubation also had negligible impact on insertion efficiencies, the lack of full incorporation is not due to lack of Ni(II) ions for coordination. Rather, the lack of full incorporation may be due to restraints of diffusion within the solid resin matrix and/or the protein structure that limit accessibility of the tag to the surface of the resin where the nickel is coordinated. Resin transfer results in a 45% higher yield of the desired metal-incorporated product in comparison to solution transfer with Ni-NTA (Table 2). Because the chelator on the resin is proprietary and unknown, this result may suggest that metal transfer is more efficient when the metal is bound to a solid support or that the proprietary chelator better facilitates formation of the *cla*MP complex compared to NTA. A quantitative and kinetic comparison of these processes is not within the scope of this initial assessment, but rather the aim here was to establish that metal transfer into the *cla*MP Tag may be accomplished using either approach.

Table 4.2 Metal Insertion Efficiency using either Ni-NTA or Ni-IMAC Resin

Variant	Resin Transfer	Solution Transfer
GNCCG-MBP	75.78	30.07
NCC-MBP	70.33	X
MBP-GNCCG	66.06	X

*Values reported as the percent Ni-incorporated relative to the overall protein present. Locations marked with X were not investigated. Metal was inserted in the cell lysate before protein purification.

4.3.3 Impact of *cla*MP Tag Placement and Sequence on Insertion Efficiency

The tripeptide module responsible for directly binding Ni(II) consists of the Asn-Cys-Cys sequence, but residues surrounding this tripeptide can modulate the properties of the *cla*MP Tag (unpublished data). For this reason, the *cla*MP Tag was placed on either the N- or C-terminus of MBP to investigate if tag location impacted Ni(II) binding or protein properties. Because resin transfer resulted in a larger yield of the metal-bound product, metal insertion was completed using this method to generate a higher yield of the desired product. As shown in Table 4.1, the tag sequence or location does not significantly affect the insertion efficiency. GNCCG-MBP has the highest insertion efficiency of 80%, whereas NCC-MBP and MBP-GNCCG appear less efficient by 10% and 14%, respectively. These values are not appreciably different because insertion efficiencies varied up to 10% from batch-to-batch production of the same variant. Metal insertion during the lysis step would be expected to vary slightly as

Figure 4.4 Validation of inline Ni-*cla*MP Tag structure. UV-vis spectroscopy confirms Ni-insertion into the *cla*MP Tag generates the expected structure regardless of tag placement. Both N- and C-terminally tagged variants contain features in the visible region near 418 nm and 530 nm, which are absent in the native protein. The feature typically observed at 310 nm is obscured by the absorption of the aromatic side chains of the protein.

expression levels of each of the variants can vary up to 20% among batches. Regardless of tag placement, Ni(II) is selectively inserted into the tag and the product is primarily monomeric.
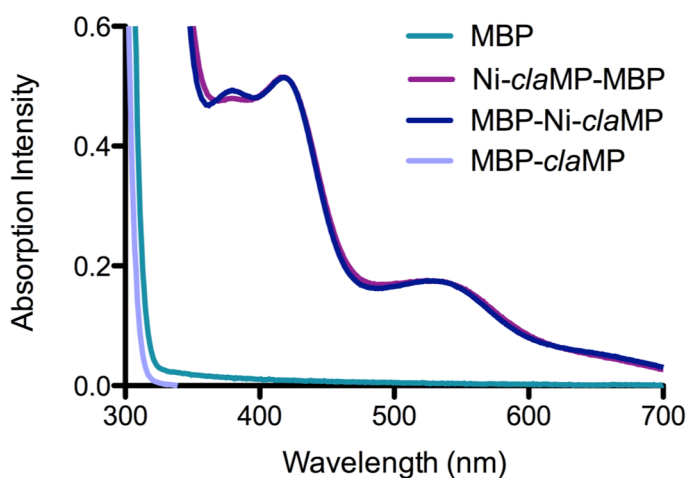
Ni(II) incorporation into the *cla*MP Tag results in two observable features in the UV-vis absorption spectrum at approximately 415 nm and 530 nm (Figure 4.4), due to the S-Ni(II) charge transfer and Ni(II) d-d transitions, respectively.[12] These features are present in both the N- and C-terminally tagged variants, indicating the Ni(II) is incorporated regardless of tag placement. A feature at approximately 315 nm also is indicative of Ni(II) incorporation, but because the absorbance of this band is substantially lower than the total absorbance of the 38 aromatic residues within MBP, this feature is obscured in the MBP variants. Non-reducing SDS-PAGE confirms that the product is primarily monomeric indicated by the large band at 37 kDa with approximately 3% and 6% dimeric species observed at 75 kDa for Ni-*cla*MP-MBP and MBP-Ni-*cla*MP, respectively (Figure 4.5). Because these bands are not present in the reducing SDS-PAGE gel, the larger molecular weight species must be due to aberrant disulfide bond formation among the cysteine residues of the *cla*MP Tag. Omission of βME from the lysis
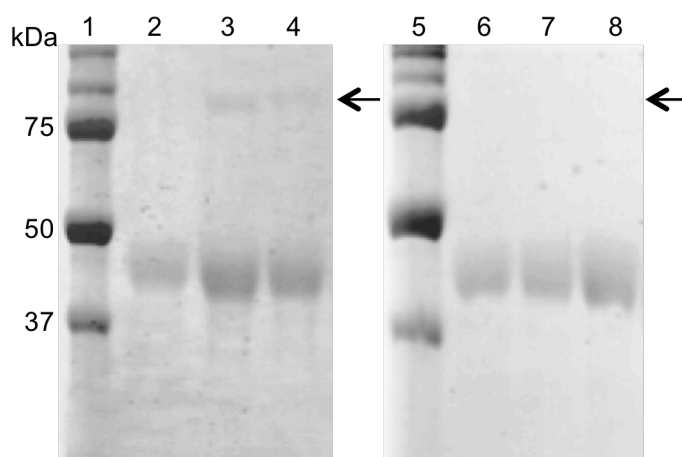
Figure 4.5 Analysis of intermolecular cross-linking. Non-reducing and reducing SDS-PAGE confirm the Ni-*cla*MP variants do not readily undergo intermolecular disulfide bond formation. Ni-*cla*MP-MBP and MBP-Ni-*cla*MP samples contain approximately 6% and 2% dimer, respectively. (Lanes 1 and 5) molecular weight markers (Bio-Rad Precision Plus Protein Dual Color Molecular Weight Standard), (lanes 2 and 6) purified MBP, (lanes 3 and 7) purified Ni-*cla*MP-MBP, and (lanes 4 and 8) purified MBP-Ni-*cla*MP. Lanes 2, 3, and 4 contain the non-reduced samples, and lanes 6, 7, and 8 contain the reduced samples. Position of dimer is indicated by the arrow.

buffer did not result in substantially higher amounts of dimer formation, suggesting that the reducing agent does not contribute appreciably to limiting dimer formation.

4.3.4 MBP Surface Charge Reliant on *cla*MP Tag Location

Even though the site of tag placement does not alter metal insertion efficiency, it does impact the surface charge profile of the molecule. Incorporation of Ni(II) into the *cla*MP Tag generates a complex with an overall charge of 2⁻ and addition of this entity to MBP, which has an overall charge of 10⁻, leads to a charge of 12⁻ for the Ni-*cla*MP-Tagged MBP variants. The effects of *cla*MP Tag location on molecule surface charge were investigated using AEC. Non-tagged MBP elutes 20 minutes into the gradient, whereas the Ni-*cla*MP-Tagged MBP variants elute later in the gradient, indicating their more negative surface charge. MBP-Ni-*cla*MP elutes as one clean peak at 40 minutes, whereas Ni-*cla*MP-MBP elutes up to five minutes earlier and contains multiple peaks (Figure 4.3b). The elution of both variants much later in the gradient than native MBP confirms Ni(II) is incorporated into the tag, but the presence of multiple peaks in the chromatogram of the N-terminally tagged variant suggests that the composition and/or structural features surrounding the *cla*MP tag in this position may impact the structure and/or conformational dynamics through local surface charge differently. The surface charge of the molecule, though, is altered by *cla*MP Tag location. The Ni-*cla*MP-Tagged variants elute later in the AEC chromatogram than MBP, which was expected because Ni(II) insertion into the *cla*MP Tag causes a change in charge of 2⁻. The peak profile is altered depending on *cla*MP Tag placement, which may reflect differences in the construct's ability to tolerate higher salt concentrations because both behave identically when analyzed by SEC. The N-terminus of MBP contains a single Lys residue embedded within a large region of negatively charged surface area (Figure 4.6) and is also highly flexible with no distinct structural elements.[26] Placement of the
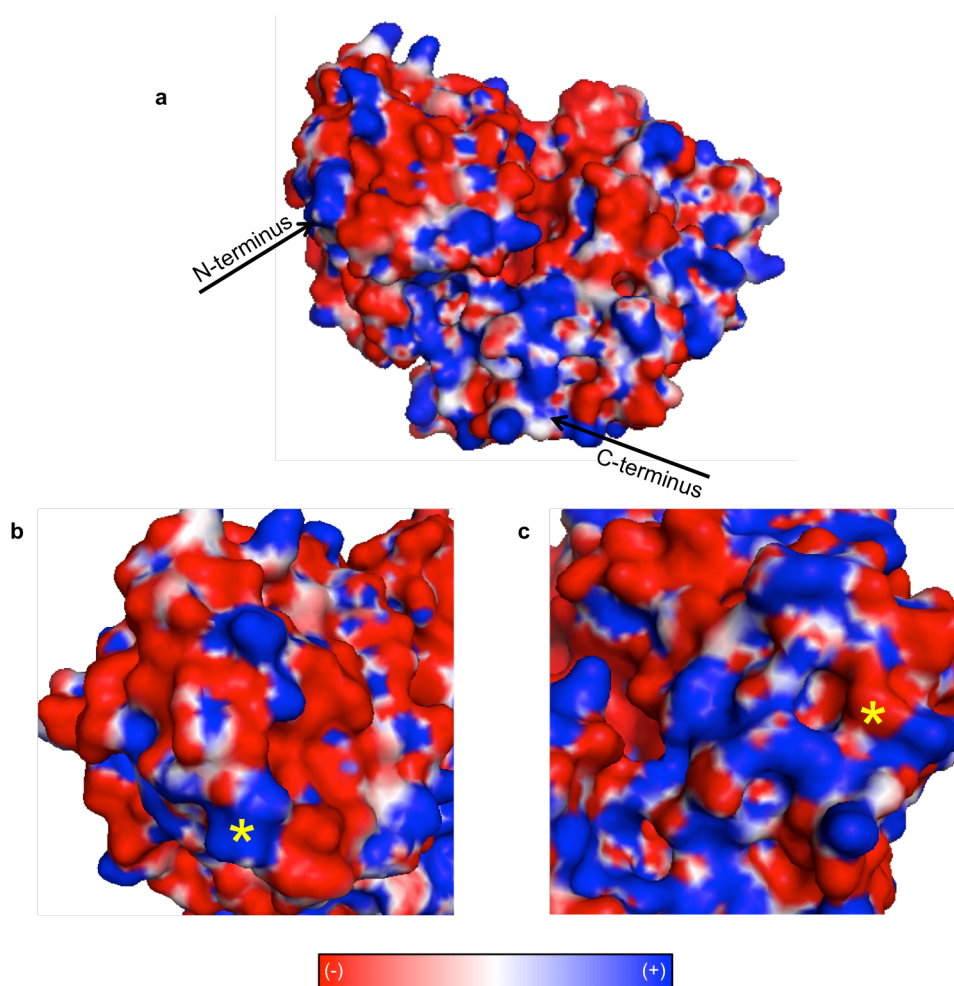
Figure 4.6 Electrostatic surface map of MBP. Placement of the Ni-*cla*MP Tag on the N- or C-terminus causes different effects on the surface charge of the molecule due to the local charge near these regions (a). The N-terminus contains a single lysine surrounded by a large region of negatively charged surface area (b), and the C-terminus has larger patches of positively charged surface area (c). The different surface charges observed in these regions will alter the interactions between the Ni-*cla*MP complex and the protein surface. N- and C-terminal residues are denoted by a yellow asterisk.

negatively charged Ni-*cla*MP complex within a region of high negative charge may cause charge repulsion between the complex and the surface of the protein or differential charge shielding of the complex by the protein surface, resulting in a less compact, more dynamic structure when placed in this position. Alterations in the salt concentration can alter the charge shielding effects, which would affect the surface charge of the molecule and alter the retention time on the anion

exchange column. The C-terminus of MBP is both more structured and more basic in character. As such, MBP may accommodate the additional negative charge well in this position and its overall structure be less affected by increased salt concentration. A high degree of flexibility within the N-terminal region may cause slight variations in surface charge due to the mobility of this segment, but the structural restrictions of the C-terminal region would lead to a more constant surface charge. The surface charge of antibodies is also differentially affected by protein structure, as multiple charge states are observed with modification of a single lysine within an antibody, depending on the position of the residue within the protein structure.[29] Here, the data shows placement of the *cla*MP Tag at the C-terminus of MBP to be a better choice than at the N-terminus.

Because the N- and C-terminally tagged MBP variants exhibited different AEC elution profiles, the effect of ionic strength on the structural stability of the variants was examined. Thermal titration of MBP, Ni-*cla*MP-MBP, and MBP-Ni-*cla*MP was performed at three different ionic strength values corresponding to the AEC retention times of the three variants, which are 142 mM, 252 mM, and 272 mM ionic strength, respectively. MBP contains a characteristic α-helical fold,[26] which allows for protein unfolding to be monitored at 222 nm. At all ionic strength values, the three variants exhibited similar onset temperatures for unfolding ($T_{m,onset}$). As the $T_{m,onset}$ has been reported to be indicative of the physical stability of ADCs,[9] similar values for all MBP variants suggests that addition of the Ni-*cla*MP Tag to MBP does not impact the physical stability of the molecule. MBP-Ni-*cla*MP does contain two separate unfolding transitions that are clearly evident at lower ionic strength (Figure 4.7a). At low ionic strength, MBP and Ni-*cla*MP-MBP do not unfold to the extent observed at higher ionic strength, but the second transition seen with MBP-Ni-*cla*MP at lower ionic strength does result in a similar degree
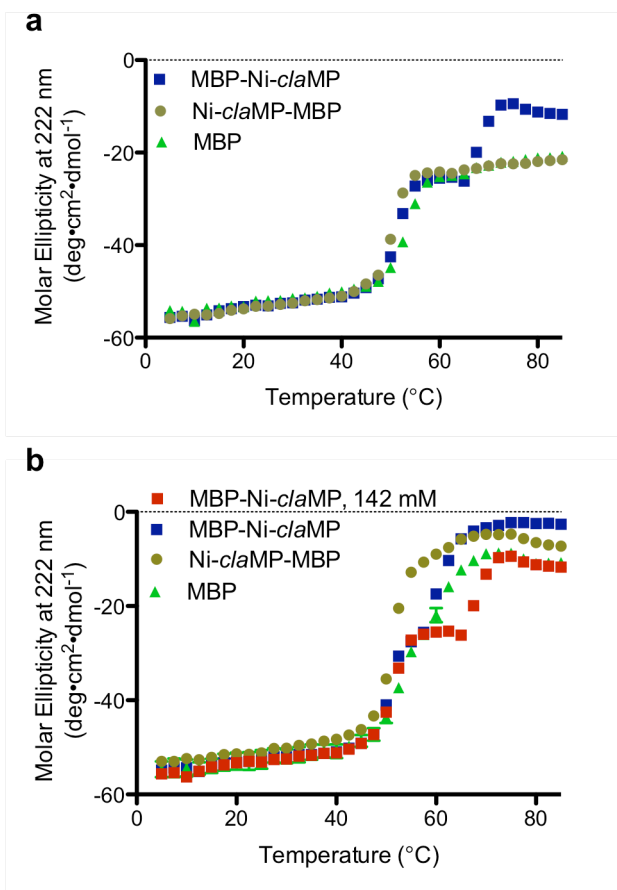
Figure 4.7 CD thermal plot of MBP variants. a) At low ionic strength (142 mM), the unfolding curve of MBP and N-terminally-tagged MBP are very similar and contain one transition, whereas the C-terminally-tagged variant contains two distinct transitions and a larger extent of unfolding. b) At higher ionic strength conditions (272 mM), a similar degree of unfolding occurs with MBP and the N-terminally-tagged variant as was observed for the second transition of the C-terminally-tagged variant at low ionic strength.

of unfolding as observed at higher ionic strength conditions (Figure 4.7b). Therefore, the different AEC profiles observed with the two tagged variants may be caused by greater interaction between the cationic resin and C-terminally tagged variant, as this variant may be more unfolded at higher ionic strength leading to exposure of additional charged residues. MBP-Ni-*cla*MP exhibited similar melting temperatures as native MBP (Table 4.3); however, the melting temperature of Ni-*cla*MP-MBP is 3-4 °C lower than the other variants. MBP contains a large degree of negatively charged surface area in the N-terminal region (Figure 4.6b), which may lead to charge repulsion between the protein surface and negatively charged Ni-*cla*MP

complex and structural destabilization. Increasing the ionic strength also minimally affects the T_m of *cla*MP-Tagged bioconjugates, but has been shown to adversely affect the structural stability of ADCs,[22] suggesting that an inline *cla*MP-Tagged bioconjugate may be more amenable to alterations in solution conditions.

Table 4.3 Effect of Ionic Strength on T_m of MBP Variants

Variant	Ionic Strength (mM)	T_m Value (°C)
MBP	142	53.2
	252	55.1
	272	55.3
Ni-*cla*MP-MBP	142	50.1
	252	51.8
	272	51.3
MBP-Ni-*cla*MP	142	50.9, 68.1[a]
	252	56
	272	55.5

[a]Two values reported due to the two transitions that occur with this variant at the specific ionic strength.

4.3.5 MBP-Ni-*cla*MP versus EDTA Competition Analysis

The effectiveness of a metal-based bioconjugate depends on the ability of the metal-binding modality to retain the metal in the presence of competitive agents. To investigate the relative binding strength of the *cla*MP Tag for Ni(II), MBP-Ni-*cla*MP was incubated with a ten-fold molar excess of EDTA (10 mM) and the percent retention of Ni-*cla*MP Tag over one week was determined. EDTA has extremely high affinity for Ni(II), which is on the order of 10^{18} under the conditions utilized.[30] This concentration of EDTA was chosen as a practical starting concentration because biological preparations generally do not contain concentrations higher than 10 mM EDTA. The proteins also were incubated with a 100-fold molar excess of EDTA. The objective of this initial experiment was not to establish a competitive binding constant, but rather to evaluate the loss of metal from the *cla*MP Tag when a high-affinity chelator is present.

Ni(II) incorporation into the *cla*MP Tag leads to the presence of multiple unique features within the UV-vis absorption spectrum, which allows for the percent retention to be assessed quantitatively (Figure 4.7a). If nickel is lost to EDTA, the intensity of Ni-*cla*MP Tag signal between 310-550 nm will decrease proportionally, and a much weaker peak will emerge near 700-800 nm, which reflects Ni-EDTA complex formation. The peak height of the features at 418 nm and 530 nm were used to quantitatively assess Ni(II) incorporation into the *cla*MP Tag. Analysis of both the N- and C-terminally tagged variants indicates that incubation with either concentration of EDTA leads to a linear decay in intensity of these features with differences in Ni(II) retention based upon *cla*MP Tag placement. Ni(II) retention by the *cla*MP Tag in the C-terminally-tagged variant decreases by approximately 9% after seven days (Figure 4.8a), and the N-terminally-tagged variant exhibits greater Ni(II) loss from the tag with a two fold larger decrease observed at 100 mM EDTA in comparison to 10 mM EDTA. AEC was also used to qualitatively assess metal retention in the *cla*MP Tag by monitoring the peak profile to determine if any differently-charged variants emerge due to Ni(II) loss from the tag. As shown in Figure 4.8b, the peak profile of the C-terminally-tagged variant remains unchanged after 7 days, further confirming that the metal is not easily chelated from the *cla*MP Tag by EDTA in this variant. The larger decreases in metal retention for the N-terminally-tagged variant indicate that the protein elements modulate the release rate, as repulsion between the negatively charged N-terminal region of MBP and the Ni-*cla*MP complex may cause the metal to be more easily chelated from the *cla*MP Tag than the C-terminally-tagged variant. The positively charged C-terminal region of MBP may stabilize the complex through electrostatic interactions, hence making metal chelation from the tag more difficult. EDTA competition experiments have been performed with the organic-based metal-binding agents in complex with lanthanide ions, and

EDTA is able to compete effectively for and chelate the metal from the metal-binding entity within minutes when added at an equivalent molar ratio or two-fold excess.[31-32] Metal loss from the *cla*MP Tag does not occur that quickly in the presence of ten- and 100-fold excess of EDTA, suggesting that the *cla*MP Tag is advantageous in comparison to organic-based metal-binding modalities that are commonly used in therapeutic and imaging applications.
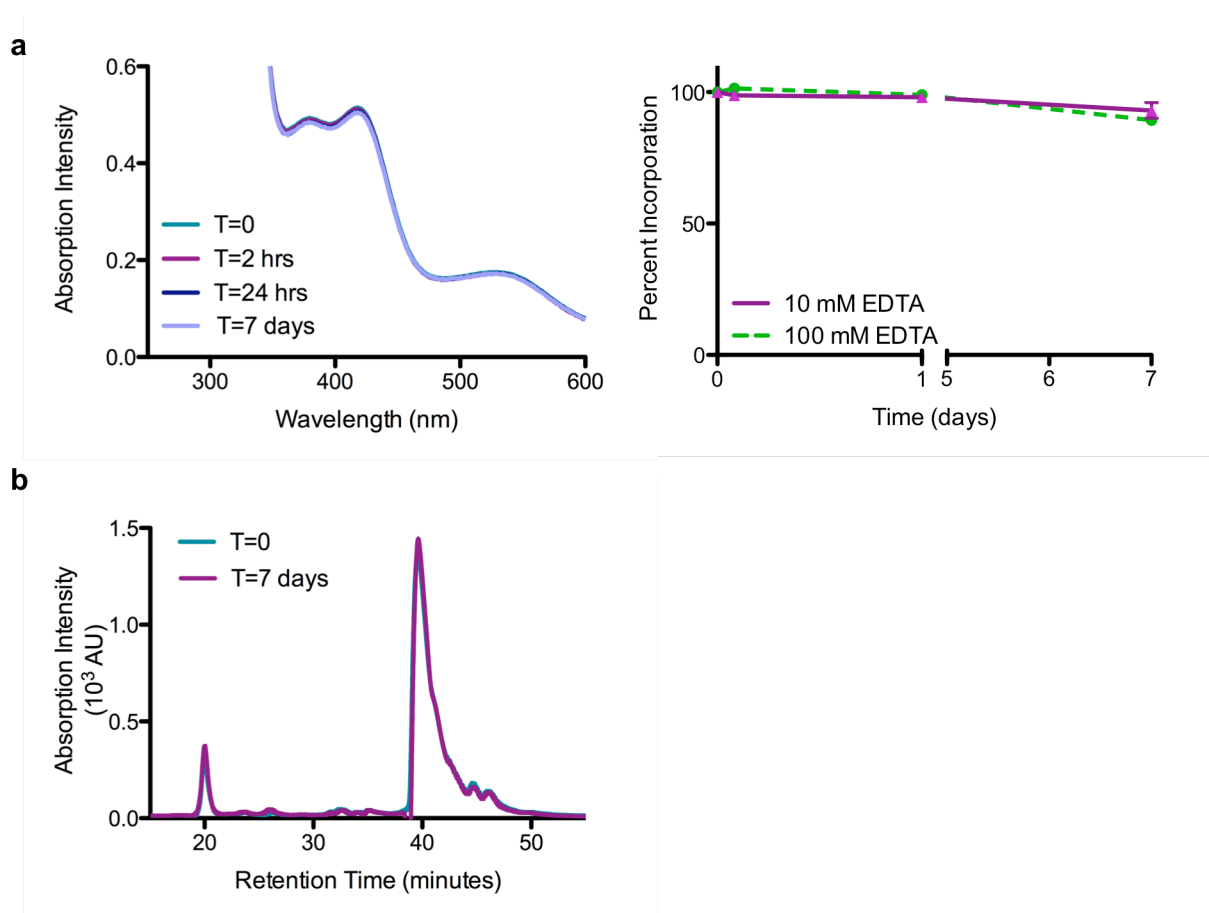

Figure 4.8 Evaluation of metal retention in Ni-*cla*MP Tag. EDTA competition analysis confirms that Ni(II) is not easily chelated from the *cla*MP Tag in the presence of 10 mM EDTA. a) The UV-vis spectrum of MBP-Ni-*cla*MP retains similar intensity and features as initially observed suggesting that the metal is retained by the *cla*MP Tag. The peak intensity at 418 and 530 nm in the UV-vis spectrum was used to quantitatively determine percent incorporation over the one-week time period, with approximately a 9% decrease in metal retention observed after one week with both concentrations of EDTA. b) AEC also indicates that Ni(II) remains incorporated within the *cla*MP Tag in the presence of EDTA. The peak at 40 minutes corresponds to MBP-Ni-*cla*MP and the peak profile does not change appreciably over the time period investigated.

4.4 CONCLUSION

Use of the *cla*MP Tag as an inline metal-binding modality requires optimization of different parameters to determine the method that results in the highest yield of the desired metal-incorporated product. In this investigation, it was determined that the highest metal insertion efficiency into the tag is achieved when metal is presented immediately after cellular lysis via a chelator attached to a solid support. The site of tag placement within the protein sequence and the size of the protein used have minimal effects on the insertion efficiency, but the charge profile of the region surrounding the *cla*MP Tag can lead to differential surface charge effects and influence the behavior of the conjugate. Placement of the Ni-*cla*MP complex on the N-terminus of MBP led to a slight decrease in structural stability in comparison to C-terminal placement. Finally, retention of Ni(II) by the *cla*MP Tag in the presence of large excesses of EDTA was shown to be influenced by site of tag placement, with the C-terminally-tagged variant retaining metal binding more effectively than the N-terminally-tagged variant. Therefore, the stability of the Ni-*cla*MP complex is modulated by nearby protein elements. The successful application of the *cla*MP Tag within a second, unrelated protein system further confirms the applicability of the tag as a platform to enable targeted metal delivery.

4.5 REFERENCES

1.   De León-Rodríguez, L. M., Kovacs, Z. (2008) The Synthesis and Chelation Chemistry of DOTA-Peptide Conjugates. *Bioconjugate Chem*. *19*, 391-402.

2.   Lewis, M. R., Kao, J. Y., Anderson, A.-L. J., Shively, J. E., Raubitschek, A. (2001) An improved method for conjugating monoclonal antibodies with N-Hydroxysulfosuccinimidyl DOTA. *Bioconjugate Chem*. *12*, 320-324.

3.   Lewis, M. R., Raubitschek, A., Shively, J. E. (1994) A Facile, Water-Soluble Method for Modification of Proteins with DOTA. Use of Elevated Temperature and Optimized pH To Achieve High Specific Activity and High Chelate Stability in Radiolabeled Immunoconjugates. *Bioconjugate Chem*. *5*, 565-576.

4.   Lewis, M. R., Shively, J. E. (1998) Maleimidocysteineamido-DOTA Derivatives: New Reagents for Radiometal Chelate Conjugation to Antibody Sulfhydryl Groups Undergo pH-Dependent Cleavage Reactions. *Bioconjugate Chem*. *9*, 72-86.

5.   Li, L., Tsai, S.-W., Anderson, A.-L., Keire, D. A., Raubitschek, A. A., Shively, J. E. (2002) Vinyl Sulfone Bifunctional Derivatives of DOTA Allow Sulfhydryl- or Amino-Directed Coupling to Antibodies. Conjugates Retain Immunoreactivity and Have Similar Biodistributions. *Bioconjugate Chem*. *13*, 110-115.

6.   Mills, B. J., Mu, Q., Krause, M. E., Laurence, J. S. (2014) *cla*MP Tag: A versatile inline metal-binding platform based on the metal abstraction peptide. *Bioconjugate Chem*. *accepted*.

7.   Franz, K. J., Nitz, M., Imperiali, B. (2003) Lanthanide-binding tags as versatile protein coexpression probes. *ChemBioChem 4*, 265-271.

8.   Wang, L., Amphlett, G., Blättler, W. A., Lambert, J. M., Zhang, W. (2005) Structural characterization of the maytansinoid-monoclonal antibody immunoconjugate, huN901-DM1, by mass spectrometry. *Protein Sci*. *14*, 2436-2446.

9.   Beckley, N. S., Lazzareschi, K. P., Chih, H.-W., Sharma, V. K., Flores, H. L. (2013) Investigation into Temperature-Induced Aggregation of an Antibody Drug Conjugate. *Bioconjugate Chem*. *24*, 1674-1683.

10.   Doronina, S. O., Toki, B. E., Torgov, M. Y., Mendelsohn, B. A., Cerveny, C. G., Chace, D. F., DeBlanc, R. L., Gearing, R. P., Bovee, T. D., Siegall, C. B., Francisco, J. A., Wahl,

A. F., Meyer, D. L., Senter, P. D. (2003) Development of potent monoclonal antibody auristatin conjugates for cancer therapy. *Nat. Biotechnol. 21*, 778-784.

11.   Junutula, J. R., Raab, H., Clark, S., Bhakta, S., Leipold, D. D., Weir, S., Chen, Y., Simpson, M., Tsai, S. P., Dennis, M. S., Lu, Y., Meng, Y. G., Ng, C., Yang, J., Lee, C. C., Duenas, E., Gorrell, J., Katta, V., Kim, A., McDorman, K., Flagella, K., Venook, R., Ross, S., Spencer, S. D., Wong, W. L., Lowman, H. B., Vandlen, R., Sliwkowski, M. X., Scheller, R. H., Polakis, P., Mallet, W. (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol. 26*, 925-932.

12.   Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2010) Novel Tripeptide Model of Nickel Superoxide Dismutase. *Inorg. Chem. 49*, 362-364.

13.   Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2013) Embedding the Ni-SOD Mimetic Ni-NCC within a Polypeptide Sequence Alters the Specificity of the Reaction Pathway. *Inorg. Chem. 52*, 77-83.

14.   Krause, M. E., Glass, A. M., Jackson, T. A., Laurence, J. S. (2011) MAPping the Chiral Inversion and Structural Transformation of a Metal-Tripeptide Complex Having Ni-Superoxide Dismutase Activity. *Inorg. Chem. 50*, 2479-2487.

15.   Laurence, J. A. S., Vartia, A. A., Krause, M. E. Metal abstraction peptide (MAP) tag and associated methods. U.S. Patent 8,110,402, February 7, 2012.

16.   Björck, L., Kronvall, G. (1984) Purification and some properties of streptococcal protein G, a novel IgG-binding reagent. *J. Immunol. 133*, 969-74.

17.   Hober, S., Nord, K., Linhult, M. (2007) Protein A chromatography for antibody purification. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 848*, 40-47.

18.   Low, D., O'Leary, R., Pujar, N. S. (2007) Future of antibody purification. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci. 848*, 48-63.

19.   Arbing, M. A., Chan, S., Harris, L., Kuo, E., Zhou, T. T., Ahn, C. J., Nguyen, L., He, Q., Lu, J., Menchavez, P. T., Shin, A., Holton, T., Sawaya, M. R., Cascio, D., Eisenberg, D. (2013) Heterologous expression of mycobacterial Esx complexes in Escherichia coli for structural studies is facilitated by the use of maltose binding protein fusions. *PLoS One 8*, e81753/1-e81753/15.

20.   Hewitt, S. N., Choi, R., Kelley, A., Crowther, G. J., Napuli, A. J., Van Voorhis, W. C. (2011) Expression of proteins in Escherichia coli as fusions with maltose-binding protein

to rescue non-expressed targets in a high-throughput protein-expression and purification pipeline. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun. 67*, 1006-1009.

21. Salema, V., Fernández, L. A. (2013) High yield purification of nanobodies from the periplasm of E. coli as fusions with the maltose binding protein. *Protein Expression Purif. 91*, 42-48.

22. Adem, Y. T., Schwarz, K. A., Duenas, E., Patapoff, T. W., Galush, W. J., Esue, O. (2014) Auristatin Antibody Drug Conjugate Physical Instability and the Role of Drug Payload. *Bioconjugate Chem. 25*, 656-664.

23. Wakankar, A. A., Feeney, M. B., Rivera, J., Chen, Y., Kim, M., Sharma, V. K., Wang, Y. J. (2010) Physicochemical Stability of the Antibody-Drug Conjugate Trastuzumab-DM1: Changes due to Modification and Conjugation Processes. *Bioconjugate Chem. 21*, 1588-1595.

24. Wängler, C., Buchmann, I., Eisenhut, M., Haberkorn, U., Mier, W. (2007) Radiolabeled peptides and proteins in cancer therapy. *Protein Pept. Lett. 14*, 273-279.

25. Mills, B. J., Laurence, J. S. (2014) Stability Analysis of an Inline Peptide-based Conjugate for Metal Delivery: Nickel(II)-*cla*MP Tag Epidermal Growth Factor as a Model System. *J. Pharm. Sci. submitted*.

26. Quiocho, F. A., Spurlino, J. C., Rodseth, L. E. (1997) Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure (London) 5*, 997-1015.

27. Daughtry, K. D., Martin, L. J., Sarraju, A., Imperiali, B., Allen, K. N. (2012) Tailoring Encodable Lanthanide-Binding Tags as MRI Contrast Agents. *ChemBioChem 13*, 2567-2574.

28. Carter, E. L., Hausinger, R. P. (2010) Characterization of the Klebsiella aerogenes urease accessory protein UreD in fusion with the maltose binding protein. *J. Bacteriol. 192*, 2294-2304.

29. Boylan, N. J., Zhou, W., Proos, R. J., Tolbert, T. J., Wolfe, J. L., Laurence, J. S. (2013) Conjugation Site Heterogeneity Causes Variable Electrostatic Properties in Fc Conjugates. *Bioconjugate Chem. 24*, 1008-1016.

30. Illemassene, M., Perrone, J., *Chemical Thermodynamics of Coupounds and Complexes of U, Np, Pu, Am, Se, Ni and Ar with Selected Organic Ligands*. Elsevier: Philadelphia, PA, 2005; Vol. 9.

31. Caravan, P., Comuzzi, C., Crooks, W., McMurry, T. J., Choppin, G. R., Woulfe, S. R. (2001) Thermodynamic Stability and Kinetic Inertness of MS-325, a New Blood Pool Agent for Magnetic Resonance Imaging. *Inorg. Chem. 40*, 2170-2176.

32. Sun, G., Feng, J., Jing, F., Pei, F., Liu, M. (2003) Synthesis and evaluation of novel polysaccharide-Gd-DTPA compounds as contrast agent for MRI. *J. Magn. Magn. Mater. 265*, 123-129.

CHAPTER 5.
CONCLUSIONS AND FUTURE DIRECTIONS

5.1 MAJOR CONCLUSIONS

This work details the first application of the *cla*MP Tag as a metal-binding tag in the context of larger protein systems. The *cla*MP Tag is able to be inserted inline within the gene of the desired protein, which allows for the formation of the bioconjugate as one molecule, eliminating the need for complex chemical conjugation chemistry. Insertion of the cysteine-containing *cla*MP Tag into the EGF sequence does not adversely affect expression or disulfide-bond formation within this protein system. When Ni(II) is selectively inserted into the *cla*MP Tag, the SOD-like activity of the Ni(II)-*cla*MP complex is maintained in the presence of the protein, and importantly, Ni(II) remains incorporated within the tag in the presence of a high – affinity competitive chelator. Finally, the stability analysis of the EGF-Ni-*cla*MP Tag conjugate confirms that the *cla*MP Tag module is structurally and chemically stable for months when stored in solution and the metal-occupied tag does not adversely impact the physicochemical stability of the protein to which it is attached. These studies illustrate the great potential for use of the *cla*MP Tag for targeted delivery of metal ions.

5.2 FUTURE DIRECTIONS

5.2.1 Confirming Cause of Conjugate Cleavage

During the stability analysis of EGF-Ni-*cla*MP, a cleavage event was observed to occur between residues Leu52 and Arg53 within EGF and results in release of Arg53-Ni-*cla*MP from the C-terminus of the protein. Over the time period investigated, the cleavage event could only be observed in KPi, but not Tris-Cl, suggesting that it may be buffer-mediated or due to the presence of a contaminant, such as residual protease and/or extraneous metal species, because even high-quality phosphate is known to contain trace metals. Although excess EDTA did not

diminish retention of nickel in the tag, it also remains possible that phosphate plays a role in alteration of the tag over long time periods. Therefore, all analyses should be completed in both buffer systems to control for the influence of buffer on the cleavage event. Although the exact cause of this cleavage event has not yet been determined, further experiments could be conducted to elucidate the reason for this occurrence. First, possible protease contamination could lead to hydrolysis of the peptide backbone preceding Arg53. Both chymotrypsin and pepsin have been shown to cleave C-terminal to Leu residues,[1] which is the amino acid immediately preceding the cleavage site. If cleavage were due to protease contamination, addition of an appropriate protease inhibitor could slow generation of these species and the possibility of cleavage.

Metal-catalyzed cleavage of peptide bonds has also been reported.[2-4] Metal-induced oxidative damage resulted in oxidation of the amine moieties, whereas the carboxyl end remained intact.[4] Our mass spectrometry results of the EGF fragment also are consistent with this type of mechanism, but because we could not detect the Arg53-Ni-*cla*MP fragment to confirm modification of the amine group, the current data set cannot distinguish between a hydrolytic and oxidative cleavage mechanism. Ni(II)-induced hydrolysis of the peptide backbone is also a possibility. This type of cleavage is sequence specific and, in all published reports of which we are aware, requires an embedded His residue.[2] Complete cleavage via this mechanism occurs at a constant linear rate and in a substantially shorter time frame, which is in contrast to our system that undergoes a long lag period before degradation begins. These studies were completed at 37-60 °C so it would be interesting to determine if incubation of EGF-Ni-*cla*MP at higher temperatures would expedite the cleavage event. Rates of Ni(II)-induced cleavage have been shown to increase as a function of temperature[5] so incubating EGF-Ni-

*cla*MP at elevated temperature would cause the cleavage event to occur at a greater rate. Insertion of varying polypeptide sequences containing the C-terminal region of EGF and the Ni-*cla*MP complex into other proteins, such as MBP, and adjacent sequences could be used to investigate whether this cleavage event is specific to the EGF formulation or if the Ni-*cla*MP complex is capable of performing peptide backbone cleavage. The specific spacing/residue requirements for cleavage could also be determined. If the Ni-*cla*MP complex is capable of performing metal-induced peptide backbone cleavage, the Ni-*cla*MP complex could be applied as a non-enzymatic method for removal of affinity tags, as is reported for the SHHK peptides.[6] Peptide controls could also be evaluated to determine the mechanism by which this cleavage reaction proceeds. Polypeptide sequences containing the last 8-10 residues of EGF and the Ni-*cla*MP complex on the C-terminus could be incubated at both 4 °C and 50 °C and then analyzed using AEC and mass spectrometry. Mass spectrometry analysis of the cleaved fragments could confirm whether oxidative cleavage or hydrolysis occurs because different masses would be observed for the C-terminal fragment in each mechanism. Performing the analysis in both buffer systems also will allow for evaluation of the buffer-dependency of the cleavage event. To determine if the reaction is dependent on released Ni(II), excess molar equivalents of $NiSO_4$ could be added to the Ni-*cla*MP-Tagged polypeptide. If the cleavage event is dependent on free Ni(II), the rate should increase with increasing $NiSO_4$ concentration.

We also hypothesize that the peptide fragment exhibits lower solubility upon cleavage, causing it to crash out of solution. Determining the time-dependent solubility of a control peptide could be used to support or disprove this hypothesis. Forming the Ni-*cla*MP complex within the RGNCCG peptide or N-terminally amidated RGNCCG in both Tris-Cl and KPi and monitoring its solubility over time using the distinctive Ni-*cla*MP features present within the

UV-vis absorption spectra could be used to evaluate the solubility. The effect of peptide concentration of solubility could also be examined by assessing the solubility of the R-Ni-*cla*MP complex at varying concentrations.

5.2.2 Probing Necessary Storage Conditions for *cla*MP-Tagged EGF

In the stability studies completed here, the EGF-Ni-*cla*MP conjugate was stored in aqueous conditions at 4 °C, but it would be desirable to store the conjugate in lyophilized form because this typically increases shelf life substantially. Many proteins are prone to physical and/or chemical degradation when formulated in aqueous solutions, which may be exacerbated by the stresses experienced during product handling, such as temperature excursions and excessive agitation.[7] The presence of a metal ion may also cause adverse effects on the stability of a *cla*MP-Tagged bioconjugate. Various metal ions such as Cu(II), Zn(II), and Ni(II) lead to cleavage within the hinge region of the IgG1 molecule,[3] and oxidative cleavage of the peptide backbone has also shown to occur in the presence of metal ions.[4] Therefore, formulation of a *cla*MP-Tagged bioconjugate as a lyophilized product may be required to limit chemical and physical degradation events and achieve the two-year average shelf-life necessary for protein therapeutic products.

Even though storage of therapeutics in a lyophilized form assists in achieving the required shelf life, there are many factors that must be considered when developing a lyophilized formulation, as opposed to an aqueous formulation. The lyophilization process contains three distinct steps –freezing, primary drying, and secondary drying – during which degradation of the protein product can occur. Suitable excipients, such as sucrose or trehalose, must be chosen to stabilize the protein during these extreme stresses, as many protein therapeutics are at least partially denatured during the lyophilization process in the absence of stabilizing agents.[7]

Because stabilization of protein therapeutics through lyophilization is widely reported, the generation of a lyophilized formulation of a *cla*MP-Tagged bioconjugate may help improve stability of the protein and/or inhibit possible metal-dependent degradation of the conjugate.

5.2.3 Production and Characterization of Inline *cla*MP-Tagged Antibody Conjugate

The "big picture" goal for application of the *cla*MP Tag within protein systems is the formation of a *cla*MP Tag-based antibody drug conjugate (ADCs) for use in targeted therapy and diagnostic imaging. Therapeutic ADC entities are capable of site-specific delivery of cytotoxic drugs, such as MMAE to treat cancer,[8-9] or because over 50% of all cancer patients receive a platinum-based drug, in the context of *cla*MP-Tagged conjugates, Pt(II). In this work, Ni(II) was inserted into the *cla*MP Tag, but other metals, including Pt(II), are capable of being inserted into the *cla*MP Tag (unpublished data). The first step in producing a *cla*MP Tag-based ADC is insertion of the *cla*MP Tag within the antibody sequence. Antibodies are very complex molecules with distinct structural elements required for protein function and stability, but acquiring high-resolution data is challenging due to the size of these molecules. Therefore, it is difficult to assess the most desirable site for *cla*MP Tag insertion within the protein sequence. Molecular modeling can be used to investigate the effects of *cla*MP Tag addition on structural elements, but the experimental methods for investigating local structural alterations that may occur due to *cla*MP Tag addition are limited. Evaluating the physical stability of the ADC in comparison to the native molecule is the most common method used to assess the effect of conjugation on antibody structure.[10-11]

Determining the drug-to-antibody ratio (DAR) is an important step in ADC characterization. This can be assessed using UV-vis absorption spectroscopy because the drug entity absorbs at a wavelength distinct from the antibody.[12] Quantitation of the DAR of Pt-

*cla*MP Tag-based ADCs is not possible using this method due to extensive overlap of the absorption from the antibody and Pt-*cla*MP complex. Mass spectrometry and hydrophobic interaction chromatography are also commonly used to determine the DAR. Characterization of *cla*MP Tag-based ADCs requires different methods than typical ADCs due to the unique properties of the system. The drug entity in typical ADCs is a hydrophobic molecule, whereas in our system, it is a negatively charged complex containing a Pt(II) ion. Incorporation of Pt(II) into the *cla*MP-Tagged antibody leads to a change in charge of the conjugate. Even though this may be very small, a slight shift in the AEC chromatogram would suggest successful metal incorporation. ICP-OES, which is much more quantitative, is the best method to use for the accurate determination of the DAR of a Pt-*cla*MP Tag-based ADC. Although Pt(II) has been widely used as a therapeutic entity in the form of *cis*platin, it exerts no specificity in the cell type that it attacks causing death of both healthy and diseased cells. Successful generation of a *cla*MP Tag-based ADC would be highly advantageous as a chemotherapeutic agent, as it would allow for site-specific delivery of Pt(II) with the desired target being modulated by the antibody selected.

There exists great potential for application of the *cla*MP Tag as a bioconjugate for use in both therapeutics and imaging, but the bioconjugate must be stable for at least the two-year shelf life. By determining the exact cause and mechanism of the cleavage event that occurred with the EGF-Ni-*cla*MP conjugate, the buffer conditions can be modulated to limit cleavage and/or a small amount of EDTA could be added to control metal-dependent cleavage, either chemical or enzymatic, deriving from either the buffer or release from the tag. Storage of the *cla*MP-Tagged bioconjugate as a lyophilized product may also assist achieving the desired product shelf life, but the parameters concerning successful lyophilization of the metal-*cla*MP complex while retaining

metal binding must also be assessed. Although not desirable for formation of a *cla*MP-Tagged

bioconjugate, if the Ni-*cla*MP complex is determined to cause peptide backbone hydrolysis, the

Ni-*cla*MP complex could be used in affinity tag removal from proteins. The results from these

additional studies will assist in determining the design space around the *cla*MP Tag system for

use as a biotechnology tool.

5.3 REFERENCES

1. Keil, B., *Specificity of Proteolysis*. Springer-Verlag: Berlin; New York, 1992.

2. Bal, W., Liang, R., Lukszo, J., Lee, S.-H., Dizdaroglu, M., Kasprzak, K. S. (2000) Ni(II) Specifically Cleaves the C-Terminal Tail of the Major Variant of Histone H2A and Forms an Oxidative Damage-Mediating Complex with the Cleaved-Off Octapeptide. *Chem. Res. Toxicol. 13*, 616-624.

3. Vlasak, J., Ionescu, R. (2011) Fragmentation of monoclonal antibodies. *MAbs 3*, 253-63.

4. Chei, W. S., Suh, J., Peptide- or Protein-Cleaving Agents Based on Metal Complexes In *Progress in Inorganic Chemistry*, Karlin, K. D., Ed. Wiley: Hoboken, NJ, 2007; Vol. 55.

5. Kopera, E., Krężel, A., Protas, A. M., Belczyk, A., Bonna, A., Wysłouch-Cieszyńska, A., Poznański, J., Bal, W. (2010) Sequence-Specific Ni(II)-Dependent Peptide Bond Hydrolysis for Protein Engineering: Reaction Conditions and Molecular Mechanism. *Inorg. Chem. 49*, 6636-6645.

6. Kopera, E., Belczyk-Ciesielska, A., Bal, W. (2012) Application of Ni(II)-assisted peptide bond hydrolysis to non-enzymatic affinity tag removal. *PLoS One 7*, e36350.

7. Carpenter, J. F., Pikal, M. J., Chang, B. S., Randolph, T. W. (1997) Rational design of stable lyophilized protein formulations: some practical advice. *Pharm. Res. 14*, 969-975.

8. Breij, E. C. W., de Goeij, B. E. C. G., Verploegen, S., Schuurhuis, D. H., Amirkhosravi, A., Francis, J., Miller, V. B., Houtkamp, M., Bleeker, W. K., Satijn, D., Parren, P. W. H. I. (2014) An Antibody-Drug Conjugate That Targets Tissue Factor Exhibits Potent Therapeutic Activity against a Broad Range of Solid Tumors. *Cancer Res. 74*, 1214-1226.

9. Li, D., Poon, K. A., Yu, S.-F., Dere, R., Go, M. A., Lau, J., Zheng, B., Elkins, K., Danilenko, D., Kozak, K. R., Chan, P., Chuh, J., Shi, X., Nazzal, D., Fuh, F., McBride, J., Ramakrishnan, V., de Tute, R., Rawstron, A., Jack, A. S., Deng, R., Chu, Y.-W., Dornan, D., Williams, M., Ho, W., Ebens, A., Prabhu, S., Polson, A. G. (2013) DCDT2980S, an Anti-CD22-Monomethyl Auristatin E Antibody-Drug Conjugate, Is a Potential Treatment for Non-Hodgkin Lymphoma. *Mol. Cancer Ther. 12*, 1255-1265.

10. Adem, Y. T., Schwarz, K. A., Duenas, E., Patapoff, T. W., Galush, W. J., Esue, O. (2014) Auristatin Antibody Drug Conjugate Physical Instability and the Role of Drug Payload. *Bioconjugate Chem. 25*, 656-664.

11.   Beckley, N. S., Lazzareschi, K. P., Chih, H.-W., Sharma, V. K., Flores, H. L. (2013) Investigation into Temperature-Induced Aggregation of an Antibody Drug Conjugate. *Bioconjugate Chem*. *24*, 1674-1683.

12.   Wakankar, A., Chen, Y., Gokarn, Y., Jacobson, F. S. (2011) Analytical methods for physicochemical characterization of antibody drug conjugates. *MAbs 3*, 161-72.