2013-12-01

Investigation into the Effects of PEGylation on the Thermodynamic Stability of the WW Domain

Sam S. Matthews
Brigham Young University - Provo

Investigation into the Effects of PEGylation on the Thermodynamic Stability of the WW Domain

Sam Scowcroft Matthews

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Joshua L. Price, Chair
Paul B. Savage
Steven W. Graves

Department of Chemistry and Biochemistry

Brigham Young University

December 2013

ABSTRACT

Investigation into the Effects of PEGylation on the Thermodynamic Stability of the WW Domain

Sam Scowcroft Matthews
Department of Chemistry and Biochemistry BYU
Master of Science

The covalent attachment of poly(ethylene glycol) (PEG) to a protein surface (known as PEGylation), has been demonstrated to increase the serum half-life of therapeutic proteins by reducing kidney clearance and immunogenicity and by protecting against proteolysis. Theses beneficial effects could be further enhanced if PEGylation consistently increased protein conformational stability (i.e. the difference in free energy between the folded and unfolded states). However, the effects of PEGylation on protein conformational stability are unpredictable; PEGylation has been reported to increase, decrease, or have no effect on the conformational stability of medicinal proteins.

This thesis details the results of two studies aimed at discovering the structural determinants which influence the thermodynamic impact of PEGylation on the WW domain, a small model protein. Chapter 1 is a brief introduction to protein therapeutics and protein PEGylation. Chapter 2 describes a study which demonstrates that the thermodynamic impact of PEGylation is strongly dependent on the site to which PEG is conjugated. The studies described in Chapter 3 elaborate on this site dependence, and demonstrate that PEG stabilizes the WW domain through interactions with the surface of the folded peptide, and that two factors – the orientation of the PEG chain (relative to the protein surface) and the identity of nearby side chains – play a critical role in determining the thermodynamic impact of PEGylation.

Keywords: PEGylation, Therapeutic Proteins, Thermodynamic Stability, Circular Dichroism, β-sheet, D-Amino Acids

ACKNOWLEDGEMENTS

Thanks to my advisor, Josh Price. From the beginning of my graduate program, he encouraged me to “own” my projects, and has allowed me to play an active role in the design and implementation of every project I was involved in during these past 2 years.

Thanks also to my lab-mates, with whom I collaborated on many of the experiments presented in this thesis. Thanks specifically to Paul Lawrence, with whom I shared the experience of being Dr. Price’s first graduate student. Whenever something went wrong, he was the first person I turned to. Many of the results in Chapter 1 were obtained with his help. Thanks also to Brijesh Pandey, our post-doc, who, among other things, agreed to watch over my shoulder and tell me what I was doing wrong when none of my peptide syntheses worked properly. Finally, thanks to all the undergraduate students who helped in the synthesis, purification, and characterization of many of the peptides used in my research. To Mindy, Chad, Mason, Ryan, Minnie, and Cameron, thank you.

Table of Contents

Chapter 1: Introduction to Protein Therapeutics and PEGylation

1.1 Protein Therapeutics

1.1.1 Insulin

On January 23, 1922, Leonard Thompson, a 14-year old boy with an acute case of diabetic ketoacidosis, was injected with extracts from an ox pancreas;[1] his symptoms rapidly improved, and insulin therapy soon became the standard treatment for diabetes. For this breakthrough, Frederick Banting and J.J.R. Macleod were awarded the 1923 Nobel Prize in physiology or medicine. In 1977, recombinant human insulin was produced in E. coli by Genentech, and marketed by Eli Lily. It was the first recombinant human protein therapeutic to receive FDA approval,[2,3] and ushered in a new era of protein-based drugs.

1.1.2 Advantages of Protein Therapeutics

There are currently over 120 FDA-approved biopharmaceuticals (comprising primarily protein therapeutics but also DNA- and RNA-based drugs) on the market today, and from 2006-2010, protein therapeutics comprised 21% of all new approved drugs in the US.[4] Some, like insulin, are administered to address a deficiency in a protein normally produced by the body, while others, such as L-asparaginase, provide a novel function. Protein drugs are a promising class of pharmaceuticals which do not suffer from many of the problems plaguing conventional "small molecule" drugs.[2]

The limited surface area of small molecule drugs means they must bind to concave clefts in the protein surface;[5] this means that less that less than 15% of the human proteome have the necessary structural features to be considered "druggable." [5] In contrast, protein-protein interactions tend to involve broad interfaces; this means that virtually any endogenous protein could theoretically be a target for a protein-based drug (although protein based drugs have their own unique challenges, which will be detailed below).

The large surface areas involved in protein-protein interactions tend to make protein therapeutics more specific than small molecule therapeutics. This high specificity means that biopharmaceuticals often

have fewer side effects than conventional therapies[2]. With the exception of immune-response mediated reactions to drug administration, the side effects of protein therapeutics are almost always derived from the action of the protein on its specific target, rather than from off-target effects.[6] Monoclonal antibodies, a major subclass of protein therapeutics, are noted for their lack of off-target toxicity.[7]

1.1.3 Challenges for Protein Therapeutics

Despite their many advantages, protein therapeutics face many unique challenges, which can limit their effectiveness. Because the digestive system tends to degrade proteins, and because intact proteins are not absorbed readily by the intestinal epithelium, very few protein drugs can be administered orally. Most are administered by intravenous, intramuscular, or subcutaneous injection.[8] Proteins can also be excreted by kidney filtration, degraded by proteases, neutralized by host antibodies, and can be prone to aggregation and denaturation.[9-11] These challenges can lead to a short serum half-life in vivo, necessitating more frequent dosing regimens, which leads to increased cost of treatment, lower patient compliance and reduced quality of life.

1.2 Protein Folding

These final four challenges (proteolysis, immunogenicity, denaturation, and aggregation) are, at a fundamental level, related to the conformational stability of a protein,[12-14] meaning the difference in free energy between the folded and unfolded states. The central dogma of molecular biology states that genetic information is stored in a linear sequence of DNA, transcribed to a linear sequence of RNA, and then translated into a linear sequence of amino acids, which is known as a protein. However, while proteins are sequenced in a linear fashion, and can therefore be visualized as 'beads on a string,' it is the three-dimensional structure of a protein that leads to its function; most proteins are entirely inactive when not in their folded state.

The folded state of the protein is determined almost exclusively by the sequence of amino acids - most cytosolic proteins fold to the correct state spontaneously (although some proteins require chaperone proteins to prevent aggregation while they fold). Protein folding is widely thought to be entropically

driven. In the unfolded state, hydrophobic amino-acid side-chains are surrounded by highly-ordered water molecules. When the protein folds, these molecules are released into the bulk solvent, increasing the entropy of the system; this more than compensates for the decreased conformational entropy of the peptide chain.[15]

All proteins exist in equilibrium between their final folded state(s), their unfolded states (random coil), and any partially-folded intermediates. The relative population of a state depends on the energy of that state - the lower the energy of the state, the more the protein samples that state. The thermodynamic stability of a protein is defined broadly as the difference in energy between the folded and the unfolded states; the more stable a protein is, the greater the population of the folded state relative to the unfolded state. In proteins with low conformational stability, significant populations of unfolded or partially unfolded intermediate conformations exist at equilibrium. These unfolded conformations are especially prone to proteolysis[16-20], aggregation[21-23], and recognition/neutralization by antibodies [13,23-29].

1.3 Strategies to Improve Protein Pharmacokinetics

The pharmaceutical industry employs a variety of tools to overcome the challenges intrinsic to protein therapeutics. Efforts to increase protein oral bioavailability will not be detailed here, but have been reviewed recently.[8] Increasing protein serum half-life must involve reducing excretion through kidney filtration, inhibiting proteolytic cleavage, preventing or reducing the immune response to the protein, and/or preventing denaturation and aggregation. Several strategies have been employed to accomplish these tasks, including modification of the amino acid sequence or chemical modification of individual side chains on the protein surface.[30,31]

1.3.1 Sequence Modification

The simplest way to alter the pharmacokinetics (meaning the absorption, distribution, and elimination) of a protein drug is to alter its sequence. This can be done to alter sites commonly targeted by proteases, as in the case of TX14[32]. After identifying the site of proteolytic cleavage, Taylor et al. replaced leucine with isoleucine in two sites, and arginine with valine in a third, to generate a new peptide,

which was resistant to the proteases found in the endothelial and epithelial cells of the blood-brain barrier. Sequence modification can also be used to affect the solubility and self-association properties of the protein, as in the case of several insulin analogues currently on the market.[30,33] Insulin lispro (brand name Humalog) is identical to human insulin, except at positions 28 and 29 in the B-chain, which are reversed; this reduces the ability of the protein to hexamerize, leading to faster absorbtion and faster elimination.[34] Conversely, Insulin glargine (brand name Lantus) contains several amino acid substitutions which reduce its solubility at neutral pH; this causes it to precipitate in subcutaneous tissue, leading to slow and sustained release.[35]

A more extreme example of sequence modification is a fusion protein, in which the genes encoding multiple different proteins are joined, to create proteins containing the domains of multiple different components. When proteins are fused to the Fc region of immunoglobulins, they bind to the salvage receptor FcRN; this causes the endothelial cells in the kidney to internalize the protein and recycle it into the bloodstream;[36] this reuptake causes the proteins, which would normally have half-lives of 2-3 hours, to remain in circulation for up to several weeks[37].

1.3.2 Glycosylation

In addition to the 20 amino acids coded for by DNA, the cell can also add diversity to its proteome by covalently modifying the amino acid side chains after the protein has been translated (posttranslational modification). Common modifications include glycosylation (installation of a sugar), phosphorylation (installation of a phosphate), and acetylation (installation of an acetate group). Among other effects, these modifications can stabilize the protein or even cause it to adopt a different folded shape. One common way to increase the serum half-life of a protein drug is to engineer additional glycosylation sites into the protein. The cellular machinery involved in N-glycosylation recognizes a consensus sequence (Asn-X-Ser/Thr, where X is any amino acid except proline)[38-40] and glycosylates the first Asn. Recombinant human erythropoietin (rHuEPO) naturally contains four glycosylation sites (three Asn and one Ser). Although EPO normally needs to be administered several times per week, by adding

two additional glycosylation sites, Elliot et al were able to prolong the in-vivo circulation of EPO to the point where once-weekly injections achieved the same result[41,42].

1.3.3 Non-Natural Modifications to Protein Surfaces

In addition to natural methods of modifying protein side chains, many protein therapeutics are enhanced by non-natural covalent modifications. This generally involves appending a large, amphipathic polymer to the side chain of one or more surface exposed amino acid residues. When anti-tumor protein neocarzinostatin (NCS) is conjugated with poly(styrene-co-maleic acid/anhydride) (SMA), the resultant protein (designated SMANCS) has a tenfold increase in serum half-life (19 min in mice vs 1.8 minutes), and interestingly, accumulated preferentially in tumor tissues[43], which have leaky vasculature and poor clearance mechanisms[44]. One interesting way that this latter feature is exploited in polymer-directed enzyme prodrug therapy (PDEPT). Satchi et al[45] conjugated poly(hydroxypropyl) methylacrylamide (HPMA) to both a cytotoxic prodrug (doxorubicin) and to an activating enzyme (cathepsin B). Since both compounds accumulated in tumors, it was only at the site of the tumor that the prodrug was converted to the active form, resulting in tumor-selective cytotoxicity.

1.4 PEGylation

By far the most common non-natural protein conjugate is poly(ethylene glycol), also known as PEGylation, which was pioneered by Davis and Abuchowsky[46,47] in 1977. The first PEGylated protein was approved for use by the FDA in 1990; the PEGylated form of adenosine deaminase (pedemase bovine) was developed for the treatment of severe combined immunodeficiency disease (SCID)[48] and marketed as Adagen®. There are currently ten FDA-approved PEGylated proteins on the market,[10] four of which are considered "blockbusters" (generating more than $1 billion per year in revenue).[49] Poly(ethylene glycol) is amphiphilic, non-toxic (although at very high doses it can induce apparently benign vacuole formation)[50] and is generally thought to be non-immunogenic (although some evidence exists that PEGylated proteins can elicit an anti-PEG immune response in some patients[44] and animal models[51])

1.5 Effects of PEGylation on Serum Half-Life

One interesting feature of PEG is its high degree of hydration (two water molecules per monomer unit); as a result, a PEG polymer has an effective hydrodynamic radius of 5- to 10-times as large as would be calculated based strictly on molecular weight (although this effect may also depend on the site of PEGylation[52]). The large effective size of conjugated PEG polymers are thought to be the primary means by which PEGylation prolongs the circulation time of therapeutic proteins[11] (although there are exceptions where alternate mechanisms must be in play, as shown below in section 1.5.3) With a PEGylated peptide, antibodies and proteolytic enzymes are physically blocked from interacting with the protein surface, and, since the glomerular filtration is largely a function of size, the larger PEG-protein conjugate is filtered from the blood much slower than its parent protein.

1.5.1 PEGylation Reduces Immunogenic Response

Early insulin therapeutics, derived from bovine pancreas extracts, were known to trigger an immune response in patients.[26] In most cases where the protein therapeutic is non-human in origin, foreign proteins are taken up into dendritic cells via endocytosis, and cleaved into smaller peptide fragments. These fragments bind to the Major Histocompatability Complex (MHC), which is then presented on the cell surface. Fragments that are identified as "non-self" (antigenic determinants, or epitopes) trigger the activation of T-helper Cells (TH-cells). These cells proliferate and activate B-Cells, which proliferate and secrete antibodies that bind specifically to the epitope presented.

In addition, antibodies against insulin that has been derived from recombinant human DNA have been reported, albeit at much lower levels.[53] Repeated presentation of self-antigens, such as recombinant insulin, cause the dendritic cells to begin recognizing "self" proteins as non-self. This problem is exacerbated if the injection contains impurities[54] or protein aggregates.[55] An immune response against protein therapeutics that supplement endogenous proteins can exacerbate the original condition, as the antibodies will neutralize the endogenous proteins secreted naturally by the body as well.

To one degree or another, all biopharmaceuticals are potentially immunogenic;[56] the results of an immune response to a protein therapeutic can range from partial or total inactivation of the drug (resulting in reduced or nullified efficacy) to anaphylaxis (although this effect has largely been eliminated by better protein purification techniques).[57]

Pegaspargase (marketed as Oncaspar ®) was one of the first PEGylated proteins to win FDA approval (1994). L-Aspariginase, a peptidase secreted by E. coli,[58] is used to treat acute lymphocytic leukemia (ALL). In this and similar cancers, the ability to synthesize L-asparagine is lost. By depleting the body's reserves of circulating asparagine, it is possible to selectively "starve" the cancer cells.[59] However, this treatment suffers from two primary problems: it is highly immunogenic and has a short serum half-life (~20 hours). Pegaspargase is significantly less immunogenic and has a significantly longer serum half-life (~350 hours)[60]. Although pegaspargase is almost twenty times as expensive per dose ($980.00 vs $52.38) as L-asparaginase, the reduced dosing schedule, with corresponding reductions in hospital fees and staffing needs, makes pegaspargase the more inexpensive treatment[61].

1.5.2 PEGylation Decreases Kidney Clearance

It is interesting to note that, although the majority of the beneficial effect of PEGylating L-asparaginase can be attributed to a reduction in immunogenicity, pegaspargase has a longer circulation time than L-asparaginase even when comparing the results of intial injections (when no antibodies would be present),[62] suggesting an additional mechanism of action. Small proteins are primarily excreted via kidney filtration, which is accomplished largely on the basis of size (although charge selection has also been observed). Molecules (including macromolecules) below 7 kDa pass into Bowman's capsule completely unhindered (the ratio of concentrations in the filtrate and in blood plasma, or F/P ratio, is 1) while species of mass between 7 and approximately 70 kDa show decreased filtration with increased mass[63]. Species above ~70 kDa are not filtered by the kidneys. PEGylation of small proteins (below the 70 kDa threshold for kidney clearance) usually decreases kidney filtration and thus increases serum half-life[11]. As recombinant human proteins replace animal- derived sources, the likelihood of immunological

reactions to protein therapeutics is expected to decrease, and decreased kidney clearance is anticipated to be the most valuable effect of PEGylation.[62]

Prior to 2001, patients with chronic hepatitis C infections were treated with interferon-α (IFN-α) in combination with the antiviral drug ribavirin[64]. However, IFN-α has a serum half-life of 4-6 hours, requiring injections three times per week. In addition to the pain, inconvenience, and cost of this method of dosing, the fluctuations of serum concentrations of the antiviral medication allowed the virus to develop mutations to counteract the therapy, reducing its long-term effectiveness[11]. Two PEGylated drugs, peginterferon-α2a (Peg-IFN- α2a) and peginterferon-α2b (Peg-IFN- α2b) (marketed as Pegasys® and PegIntron®, respectively) were developed, consisting of IFN- α conjugated to a 40kDa branched PEG polymer[11,65] and a 12kDa linear PEG polymer,[66,67] respectively. IFN-α has a molecular weight of ~19kDa, and is rapidly cleared from the blood by renal filtration ($t_{1/2} \approx 4$h). Both PEGylated versions show significantly improved pharmacokinetics, with Peg-IFN- α2b showing a 10-fold increase in serum half-life, and Peg-IFN- α2a showing a hundred-fold increase.

For protein therapeutics prescribed to treat chronic conditions, maximizing serum half-life (all other things being equal) is usually considered desirable. However, for other applications of medicinal proteins, such as for radiolabeled antibodies for tumor imaging, it is desirable to match the serum half-life to the radioactive half-life of the conjugated radioisotope,[68,69] since overly-long circulation times result in poor contrast.[70] Diabodies (antibody fragments with MW=55 kDa) have a serum half-life of approximately 0.5 h in rats. In contrast, ^{64}Cu, a common PET imaging radioisotope, has a much longer half-life ($t_{1/2}$=12 h). In 2011, Li et al.[71] attached DOTA-chelated ^{64}Cu to diabodies using PEG oligomer linkers of various sizes (comprising 0, 12, 24, and 48 ethylene glycol units, respectively), affording DOTA-PEG-protein conjugates with apparent molecular weights of 50, 60, 70, and 80kDa, as determined by size-exclusion chromatography. Consistent with the observation that kidney filtration is reduced as size increases (up to approximately 70kDa), Li et al. observed that kidney uptake was uniformly reduced as PEG length increased, with 48-unit PEG affording a serum half-life of 6 hours. These results suggest

the possibility of fine-tuning the pharmacokinetic properties of pharmacologically relevant PEG-protein conjugates by modifying the size of the PEG polymer.

1.5.3 PEGylation Protects Against Proteases

In addition to their immunogenicity and rapid kidney clearance, protein therapeutics can be degraded by proteases. The major organs responsible for proteolysis are the liver and kidneys, but endothelial cells also play a major role; as a result, protein therapeutics are subject to proteolytic degradation regardless of their tissue distribution.[72] PEGylation is thought to protect therapeutic proteins from proteolytic degradation by much the same mechanism as it reduces immunogenicity: because of its bulk, the PEG conjugate reduces protease access to the PEGylated protein.[11]

Human growth hormone (hGH) has a very short serum half-life (~1/2 hour in rats), and is cleared primarily through the kidney. When hGH is modified with increasing numbers PEG polymers of 500kDa each, the serum half-life increased to up to ~10 hours (for hGH with 5 conjugated polymers).[73] However, previous research has shown that renal filtration is only responsible for ~67% to the clearance of hGH in rats.[74] This means that, while PEGylation undoubtedly decreases kidney clearance, it could, at most, increase the serum half-life by a factor of two if it were operating solely on renal filtration. PEGylation must also be acting by a second mechanism to prolong serum circulation; the most likely candidate is via proteolytic protection.

Further evidence of this mechanism can be seen in the results of PEGylating incretins, a family of small peptide hormones involved in insulin regulation. Glucagon-like peptide -1-(7-36) (GLP-1) is a small hormone, comprising 29 amino acids, with a very short circulating half-life (<2 min)[75]. In contrast with hGH, the primary method of clearance for GLP-1 is proteolytic degradation by the enzyme dipeptidyl peptidase[76]. When GLP-1 is PEGylated with a single 2kDa polymer, its half-life increases to 33 minutes, a greater than 10-fold increase.

Glucose-dependent insulinotropic polypeptide (GIP) is another incretin closely related to GLP-1, with potential as a treatment for type 2 diabetes. However, like GLP-1, it also suffers from a short half-

life *in vivo* (5-7 minutes) due to the action of dipeptidylpeptidase-IV (DPP-IV)[77]. In their efforts to improve the pharmacokinetics of GIP, Gault et al.[78] departed from the traditional approach of employing large (>50 ethylene oxide units) PEG polymers, instead opting for an ethylene oxide trimer (mini-PEG, or $mPEG_3$) attached at the C-terminus. Although the benefits of PEGylation are generally thought to derive from the size of PEG, which is thought to shield the protein from proteases and antibodies, this mini-PEGylated showed significant resistance to proteolytic degradation *in vitro* ($t_{1/2}$=24 hours, compared to 2.2 hours for the non-PEGylated peptide), even though the site of PEGylation is removed from the cleavage site for DPP-IV. Although no pharmacokinetic studies are available yet to indicate the result of mini-PEGylation on serum half-life *in vivo*, it is significant that mice treated with mPEG-GIP prior to an intraperitoneal glucose injection have significantly reduced serum glucose concentrations relative to mice treated with unmodified GIP.

Proteases are not the only circulating proteins which present challenges for protein-based therapeutics. Bovine pancreatic ribonuclease (RNaseA) is unusual among potential protein therapeutics in that it is capable of traversing the cell membrane, and is thus able to act on targets (in this case RNA) within a cell; unfortunately, ribonucleases are highly regulated in humans due to the action of the circulating ribonuclease inhibitor (RI) protein, which binds to mammalian ribonucleases with extremely high affinity (K_I=44 fM),[79] rendering any treatment with RNaseA inneffectual. Amphibian ribonucleases are not inhibited by RI, but they are also more immunogenic[80]. Fortuitously, the active site of RNaseA is on the opposite face of the protein as the site of RI binding; a 20 kDa PEG conjugate at the binding site successfully inhibited RI binding (K_I=37 nM) and reduced kidney clearance while minimally impacting RNase activity.[81]

1.6 Methods of PEGylation

A variety of strategies exist for appending PEG polymers to protein sufaces, including reacting the folded protein with chemically functionalized PEG polymers, incorporating non-natural amino acids into the protein co-translationally, and exploiting the cell's natural machinery for post-translational

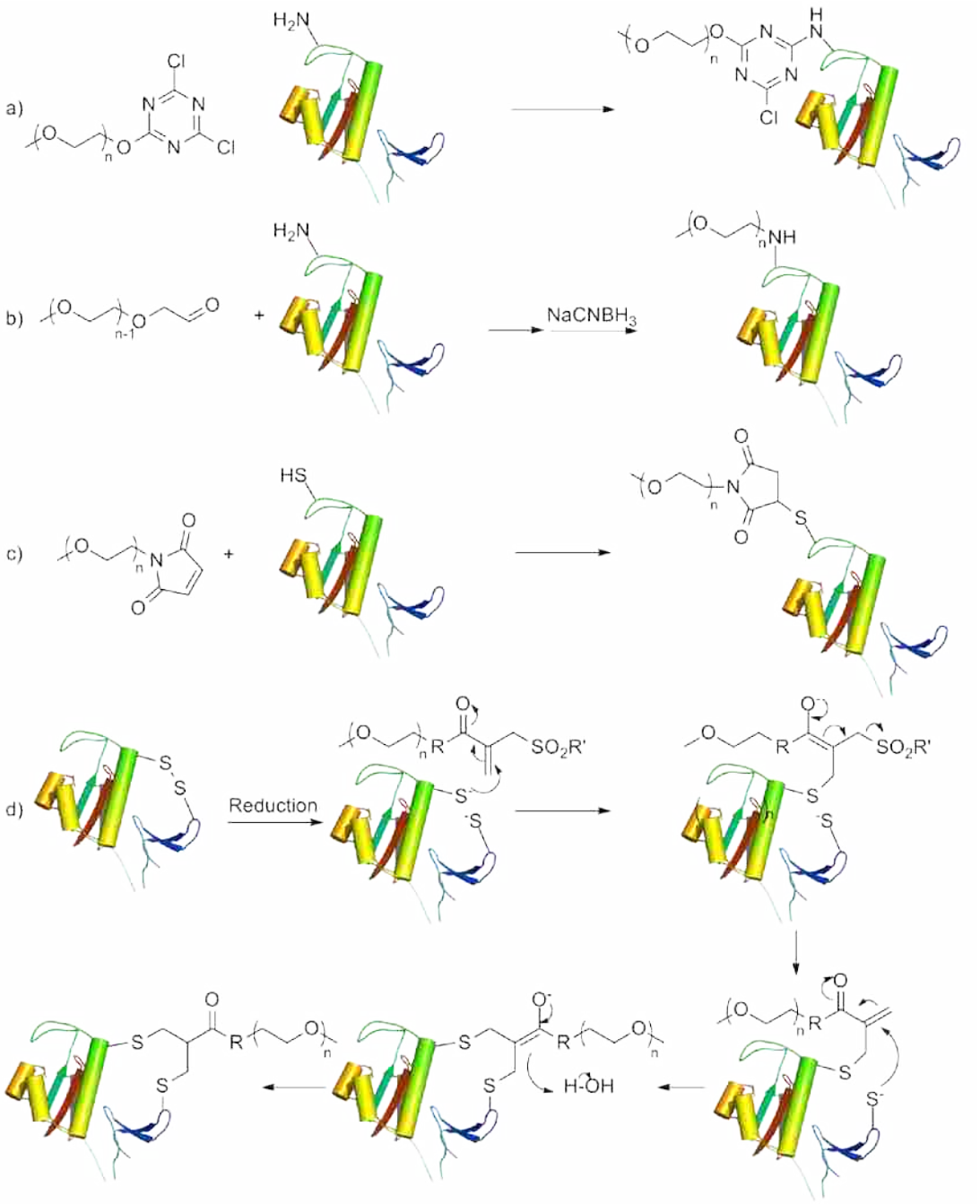
modification to install the PEG polymer. This section provides an overview of these methods, with examples of the more common reagents used. For further reading, consult refs 9,10 and 82.

Early PEGylation efforts utilized PEG polymers which were functionalized with electrophilic groups. They were added to aqueous protein solutions and reacted with surface exposed nucleophiles. Early examples included trichlorotriazine-[46,47,83] (Scheme 1a) and aldehyde-[84] (Scheme 1b) functionalized polymers. Since these methods afforded little control over the site or degree of modification, early PEGylated therapeutics were sold as mixtures of isomers, where the location and even the stoichiometry of PEGylation varied.

Recently, a number of chemical methods emerged to site-selctively PEGylate a protein surface. This approach requires that a unique chemical 'handle' be located at a single location on the protein surface. One of the earliest attempts at site-specific PEGylation involved the N-terminal amine of the protein[85-87]. Since the ε-amine of lysine differs in pKa from that of the α-amine of the N-terminus, the side-chain amines become protonated and non-nucleophilic under mildly acidic conditions, allowing PEGylation at a single site.

Because cysteine residues are chemically distinct from other amino acid side chains, as well as comparatively rare, a number of techniques have been developed to site-specifically PEGylate them. These include thiol PEGylation (Scheme 1c), in which a lone surface exposed cysteine is reacted with an electrophilic PE, usually functionalized with a maleiamide moiety[88,89], and bridging PEGylation (Scheme 1d),[90] in which a bi-functionalized PEG reactions with 2 cysteines in close proximity, forming a pseudo-disulfide bridge. If no convenient cysteines exist in the native protein, a mutant containing cysteine(s) at the desired position(s) can often be engineered with minimal loss of activity.[91,92]

Although naturally-occurring DNA sequences only code for 20 amino acids, it is often possible to incorporate non-natural amino acids into proteins using the amber stop codon[93,94] The amber stop codon normally signals an end to translation. However, it is possible to introduce tRNA which contain the corresponding anticodons linked to various non-natural amino acids. Using E. coli,[95-97] yeast,[98] or cell-free methods,[99,100] it is possible to site-specifically PEGylate virtually any position on the protein surface. This



Scheme 1 Chemical Methods for PEGylating Proteins. a) Trichlorotriazane functionalized PEG derivatives react with nucleophilic amines on the protein surface. b) Aldehyde functionalized PEG react with nucleophilic amines on the protein surface via reductive amination. c) Maleimide functionalized PEG derivatives react with nucleophilic thiols on the protein surface. d) (Adapted from ref. 90) Sulfone functionalized PEG derivatives react with reduced disulfide bonds on the protein surface

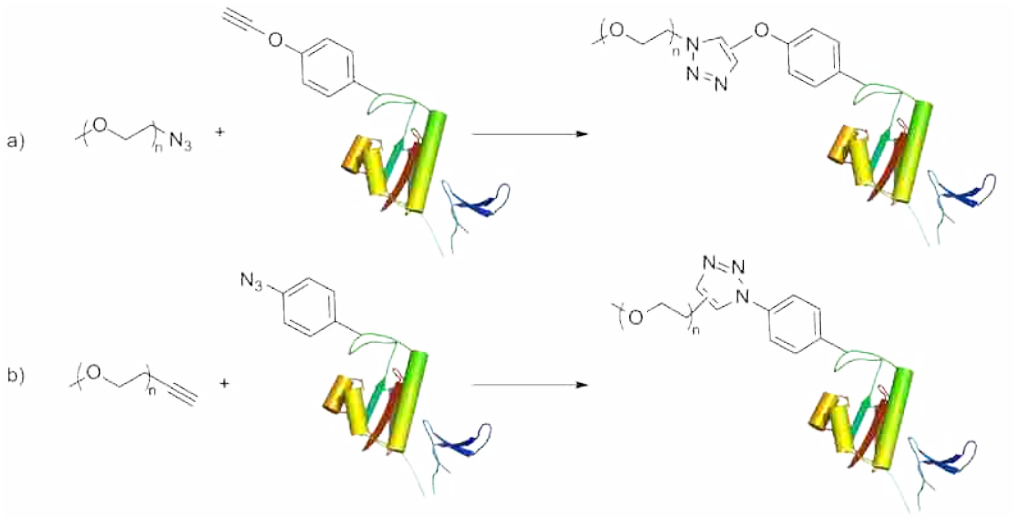
is done either by incorporating PEG directly onto the side-chain of a non-natural amino acid,[98] or by introducing an amino acid with unique chemical reactivity, such as a ketone,[95,99]azide,[99,101] or terminal alkyne.[100] In the latter case, an azide-functionalized PEG polymer can be added via a bioorthogonal [3+2] cycloaddition ("click") reaction (Scheme 2a).[102] This is also possible using an azide-functionalized peptide and a PEG polymer containing an alkyne (Scheme 2b).[103]

Recently, enzyme-driven PEGylation techniques have emerged, promising great specificity in future PEGylation. Glycopegylation (Scheme 3) harnesses existing cellular machinery to install a sugar to which a PEG polymer is attached.[104] This involves two steps: attachment of a GalNAc group using an O-GalNAc-transferase, and then appending the PEG enzymatically to the sugar. If there are multiple valid substrates for the transferase, polyPEGylation is possible,[105] but it is possible to screen multiple candidate transferases using fragments from the target protein sequence, to find one that will install the sugar in the desired spot.[104]

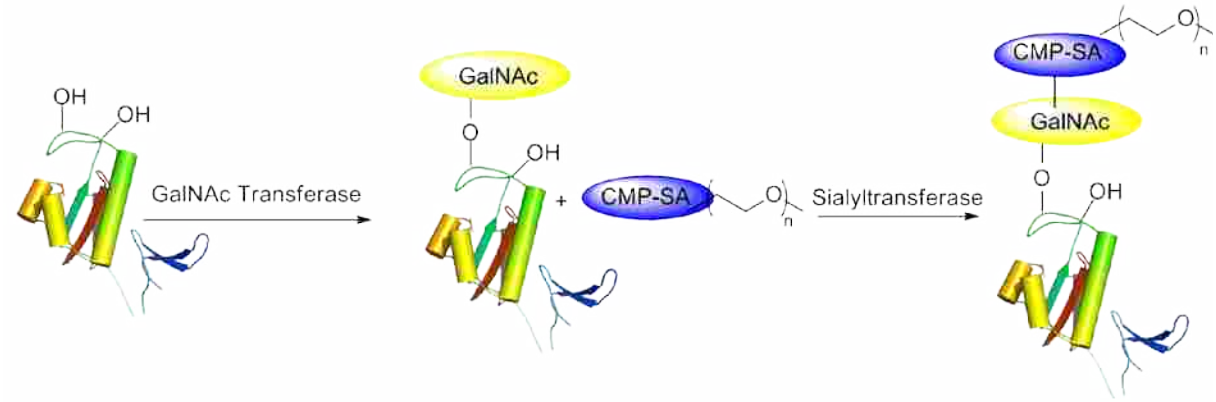
It is also possible to use enzymes to PEGylate the surface of a protein directly. In 2002, Sato[106] reported a procedure for installing PEG amines on surface exposed glutamines using transglutaminase. Although many peptides generally have several glutamine residues which are valid substrates for the transglutaminase,[107] the enzyme requires the glutamine to be located in a flexible loop.[108] Adding organic co-solvents (such as EtOH) can induce helicity in the target protein, reducing the number of valid substrates to a single glutamine, resulting in a single protein-PEG isoform (Scheme 4).[109]

1.7 Drawbacks to PEGylation

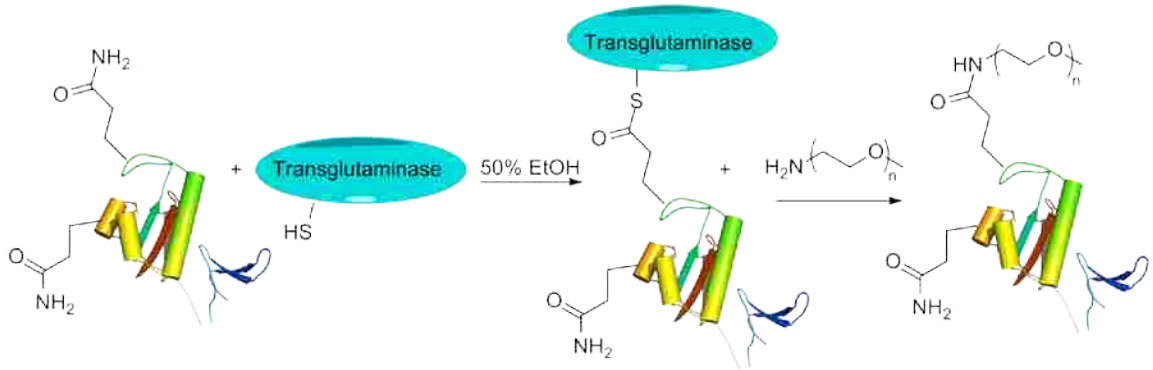
Despite the many advantages conferred upon protein therapeutics by PEGylation, there are still several drawbacks which, if addressed, might further increase the benefits of PEGylation. Usually, the biological activity of a protein is reduced upon PEGylation; while the prolonged exposure due to increased serum half-life often compensates for this,[11,73,91,110] it is not always the case.[111] There is some evidence to suggest that location of PEGylation strongly influences its effects on the biological activity of a protein drug.[112] Understanding the factors which lead to a decrease in activity might enable scientists to



Scheme 2: [3+2] cycloaddition ("click") reactions for installing PEG on protein surfaces.



Scheme 3: Method for O-Glycopegylation. N-Acetylgalactosamine(GalNAc) is installed on a surface Ser or Thr site-selectively by a GalNAc Transferase. A PEGylated sialic acid (CMP-SA) is then appended to the GalNAc using a sialyltransferase.



Scheme 4:(Adapted from ref. 109) Method for site-selective PEGylation using transglutaminase. Although a protein may contain multiple Gln residues which are valid targets for a given transglutaminase, addition of co-solvents (in this case EtOH) can alter the structure of the protein so that only one substrate can be modified.

circumvent them, leading to improved therapeutics with increased circulation time without sacrificing activity.

There are a variety of factors which may play a role in this loss of activity upon PEGylation. The most readily identified effect of PEGylation is that the large PEG polymer, which has been demonstrated to inhibit the action of antibodies and proteases by sterically shielding the protein, also limits access of the protein to its intended binding partner.[11] PEGylation may also introduce a conformational change either to the tertiary or quaternary structure of the protein. It should be noted, though, that while there are cases where this has been confirmed to have happened,[113,114] there are many indications that PEGylation often does not substantially perturb the structure of the protein.[115-117]

1.7.1 The Effect of PEGylation on Protein Thermodynamic Stability is Difficult to Predict.

Protein thermodynamic stability plays a key role in the biological activity of proteins; as mentioned above, thermodynamically unstable proteins populate the unfolded states(s) to a greater degree; unfolded proteins are generally biologically inactive, are more likely to generate an immune response, and are more prone to misfolding, aggregation and proteolysis. Although PEGylation generally decreases immunogenicity, there is some evidence that a thermodynamically destabilizing PEG conjugation may generate new epitopes (antigenic determinants), as a greater population of the protein is in the unfolded or partially unfolded state.[118] Furthermore, as mentioned earlier, protein aggregation (which is more common for thermodynamically unstable proteins) can exacerbate the immune response of proteins.[23,119,120]

Unfortunately, the thermodynamic consequences of PEGylation are unpredictable. PEGylation has been reported to stabilize some proteins,[115,121-124] and destabilize others,[117,125] while in some proteins PEGylation has no effect on protein thermodynamic stability.[83,126] It is interesting to note that in the case of lysozyme, PEGylation has been reported to stabilize the enzyme in some instances,[123,127] while in other instances it was reported to be destabilizing.[128,129] These conflicting reports may reflect differences in the

methods used for evaluating thermal stability; however, since each report employed a different method to PEGylate the protein, these discrepancies may also indicate that the location and/or linker chemistry are critical in determining the thermodynamic consequences of PEGylation.

A theoretical understanding of the thermodynamic consequences of PEGylation is important for the improvement of rationally-designed PEGylated proteins. Most studies on the site-specific impact of PEGylation have focused on different *regions* of the protein surface, such as avoiding PEGylation at the active site of a therapeutic enzyme. To our knowledge, no studies have currently been published examining the impact that PEGylation site has on the thermodynamic consequences of PEGylation. The discovery of engineering guidelines relating the site of PEGylation to its thermodynamic consequences would provide pharmaceutical chemists with an addition toolkit to optimize the pharmacokinetic and pharmacodynamics properties of therapeutic proteins.

1.8 Current Research into the Thermodynamic Consequences of PEGylation

To date, little research has been undertaken to probe the mechanism by which PEGylation alters the thermodynamic stability of proteins. This is partly because site-specific methods of PEGylation are still a developing field, and non-specific PEGylation strategies give mixtures of proteins which differ in number and location of attached PEG conjugates. In addition, most commercially available PEG polymers are polydisperse, which introduces further complications into the analytical measurements needed to probe these effects. No high-resolution crystal structures exist as yet of a PEGylated protein; beyond limited insights gained from low-resolution small angle scattering experiments,[130] the behavior of PEG must be derived inferentially.

1.8.1 PEG May Stabilize Proteins by Decreasing the Solvent Accessible Surface Area of the Folded State

Meng et al.[131] recently investigated the mechanism by which PEGylation stabilizes the SH3 domain, a small (~60 residues), highly-conserved, independently-folding domain which is found in over

300 proteins.[132,133] Subjecting both an N-Terminally MonoPEGylated protein and Tri-PEGylated protein to chemical denaturation, they observed that both species were more stable than the native protein. Fitting their curves to a two-state folding model, they extracted the m-value of the peptide, which is related to the slope of the stability/[denaturant] curve. Empirically, this value is correlated with the change in solvent-accessible surface area, or dASA.[134]

From this evidence, Meng et al. concluded that PEG stabilizes proteins by decreasing the solvent accessible surface area more in the folded state than in the unfolded state. Their model does not invoke any interactions between the PEG polymer and the protein; one drawback of their system is that it fails to explain why some proteins are stabilized upon PEGylation, and some are destabilized.

1.8.2 PEG May Act to Decrease the Conformational Dynamics of the Protein

An alternative, or perhaps complementary, hypothesis was proposed by Rodriguez-Martinez et al.[121] who probed the effects of PEGylation on α-chymotrypsin (α-CT). They noticed that increasing numbers of conjugated PEG polymers (of various molecular weights) stabilized the peptide to an increasing degree, but that the effect plateaued at four conjugated polymers. They also observed a similar effect on the structural dynamics of the protein: Using FTIR spectroscopy to calculate the rate of hydrogen/deuterium exchange, they observed an increasing degree of PEGylation corresponded with a more rigid protein, but the effect also plateaued at four polymers. Furthermore, both the stability and rigidity were largely unaffected by the size of the PEG polymer.

Since other research indicates that conformational rigidity may confer stability,[135] these results suggest that PEG may be stabilizing α-CT by decreasing the structural dynamics of the folded state. This model of Peg-induced stabilization assumes that the PEG polymer interacts with the protein surface in order to drive away water. There is some evidence of such an interaction. Molecular dynamics simulations of Peg-insulin,[136] PEG-hemoglobin,[137] and PEG-staphylokinase[52] conjugates show the PEG polymer wrapping itself around the surface of the protein; the ethylene moieties interact with the

hydrophobic surface, and the oxygens interact with the hydrophilic side chains. In addition, although no high-resolution crystal structure of a PEGylated protein exists as yet, a low resolution crystal structure obtained by Svergun et al[130] using small angle X-ray scattering shows that, in a PEG-hemoglobin conjugate, the PEG Moiety partially covers the protein surface. One possibility is that, when PEG interacts with the protein surface, it drives away water, which rigidifies the protein through dielectric shielding. This suggestion has been made elsewhere for the case of glycosylation[138], and may be the case for PEGylation as well.

While some evidence exists to support this model, it is far from conclusive. Like the dASA model, it fails to explain the unpredictable effect of PEGylation on thermodynamic stability. Further, recent work by Pai et al.[139] shows that, in hen lysozyme and human growth hormone, a 2kDa PEG adopts a compact, "dumbbell"-like shape, rather than wrapping itself around the protein surface.

1.8.3 Previous Work in the Price Lab

A complete theory explaining the thermodynamic consequences of PEGylation will explain not only the instances where PEG conjugation increases the thermodynamic stability of the protein, but also why it sometimes fails to do so. Our lab seeks engineering guidelines to inform the rational decisions guiding PEGylation, including conjugation site selection, in order to maximize the benefits of PEGylation while minimizing its drawbacks. Previously,[140] we demonstrated that when the WW domain (A small model peptide comprising a β-sheet with 3 β-strands) is PEGylated in a reverse turn, it is stabilized by both small PEG oligomers (1-6 ethylene oxide units) as well as moderately large PEG polymers (~45 ethylene oxide units), with oligomers of length 6 providing the greatest stabilization). This suggests that the origin of PEG-induced stability lies not in the action of the whole chain, but with the ethylene oxide units near the site of conjugation.

This thesis details the progress we have achieved in identifying the mechanism by which PEG oligomers stabilize the WW domain. Chapter 2 details a PEG scan, in which we determined that the

location of PEG conjugation strongly influences the thermodynamic consequences of PEGylation. Chapter 3 details our investigation into the mechanism by which PEGylation stabilizes the WW domain.

1.9 References

(1) Banting, F. G.; Best, C. H.; Collip, J. B.; Campbell, W. R.; Fletcher, A. A.: Pancreatic Extracts in the Treatment of Diabetes Mellitus. *Can. Med. Assoc. J.* **1922**, *12*, 141-6.
(2) Leader, B.; Baca, Q. J.; Golan, D. E.: Protein therapeutics: a summary and pharmacological classification. *Nat Rev Drug Discov* **2008**, *7*, 21-39.
(3) Goeddel, D. V.; Kleid, D. G.; Bolivar, F.; Heyneker, H. L.; Yansura, D. G.; Crea, R.; Hirose, T.; Kraszewski, A.; Itakura, K.; Riggs, A. D.: Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences* **1979**, *76*, 106-110.
(4) Walsh, G.: Biopharmaceutical benchmarks 2010. *Nat Biotech* **2010**, *28*, 917-924.
(5) Hopkins, A. L.; Groom, C. R.: The druggable genome. *Nat Rev Drug Discov* **2002**, *1*, 727-730.
(6) Clarke, J. B.: Mechanisms of adverse drug reactions to biologics. *Handbook of experimental pharmacology* **2010**, 453-74.
(7) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L.: Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239-1249.
(8) Soltero, R.: Oral Protein and Peptide Drug Delivery. In *Drug Delivery*; John Wiley & Sons, Inc., 2005; pp 189-200.
(9) Roberts, M. J.; Bentley, M. D.; Harris, J. M.: Chemistry for peptide and protein PEGylation. *Advanced Drug Delivery Reviews* **2002**, *54*, 459-476.
(10) Pasut, G.; Veronese, F. M.: State of the art in PEGylation: The great versatility achieved after forty years of research. *J. Controlled Release* **2012**, *161*, 461-472.
(11) Fishburn, C. S.: The pharmacology of PEGylation: balancing PD with PK to generate novel therapeutics. *J. Pharm. Sci.* **2008**, *97*, 4167-83.
(12) Krishnamurthy, R.; Manning, M. C.: The stability factor: importance in formulation development. *Curr. Pharm. Biotech.* **2002**, *3*, 361-371.
(13) Hermeling, S.; Crommelin, D.; Schellekens, H.; Jiskoot, W.: Structure-Immunogenicity Relationships of Therapeutic Proteins. *Pharm. Res.* **2004**, *21*, 897-903.
(14) Frokjaer, S.; Otzen, D. E.: Protein drug stability: A formulation challenge. *Nat. Rev. Drug. Discov.* **2005**, *4*, 298-306.
(15) Pace, C. N.; Shirley, B. A.; McNutt, M.; Gajiwala, K.: Forces contributing to the conformational stability of proteins. *The FASEB Journal* **1996**, *10*, 75-83.
(16) Daniel, R. M.; Cowan, D. A.; Morgan, H. W.; Curran, M. P.: A correlation between protein thermostability and resistance to proteolysis. *Biochem. J.* **1982**, *207*, 641-644.
(17) Parsell, D. A.; Sauer, R. T.: The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *J. Biol. Chem.* **1989**, *264*, 7590-7595.
(18) Klink, T. A.; Raines, R. T.: Conformational Stability Is a Determinant of Ribonuclease A Cytotoxicity. *J. Biol. Chem.* **2000**, *275*, 17463-17467.
(19) Ahmad, S.; Kumar, V.; Ramanand, K. B.; Rao, N. M.: Probing protein stability and proteolytic resistance by loop scanning: A comprehensive mutational analysis. *Protein Sci.* **2012**, *21*, 433-446.
(20) Imoto, T.; Yamada, H.; Ueda, T.: Unfolding rates of globular proteins determined by kinetics of proteolysis. *J. Mol. Biol.* **1986**, *190*, 647-649.
(21) Dobson, C. M.: Protein folding and misfolding. *Nature* **2003**, *426*, 884-890.
(22) Clark, P. L.: Protein folding in the cell: reshaping the folding funnel. *Trends Biochem. Sci.* **2004**, *29*, 527-534.
(23) *Aggregation of Therapeutic Proteins*; Wang, W.; Roberts, C. J., Eds.; Wiley: Hoboken, NJ, USA, 2010.
(24) Moore, W. V.; Leppert, P.: Role of Aggregated Human Growth Hormone (hGH) in Development of Antibodies to hGH. *J. Clin. Endocrinol. Metab.* **1980**, *51*, 691-697.
(25) Robbins, D. C.; Cooper, S. M.; Fineberg, S. E.; Mead, P. M.: Antibodies to Covalent Aggregates of Insulin in Blood of Insulin-Using Diabetic Patients. *Diabetes* **1987**, *36*, 838-841.

(26) Schernthaner, G.: Immunogenicity and Allergenic Potential of Animal and Human Insulins. *Diabetes Care* **1993**, *16*, 155-165.
(27) Braun, A.; Kwee, L.; Labow, M. A.; Alsenz, J.: Protein Aggregates Seem to Play a Key Role Among the Parameters Influencing the Antigenicity of Interferon Alpha (IFN-α) in Normal and Transgenic Mice. *Pharm. Res.* **1997**, *14*, 1472-1478.
(28) Ohkuri, T.; Nagatomo, S.; Oda, K.; So, T.; Imoto, T.; Ueda, T.: A Protein's Conformational Stability Is an Immunologically Dominant Factor: Evidence That Free-Energy Barriers for Protein Unfolding Limit the Immunogenicity of Foreign Proteins. *J. Immunol.* **2010**, *185*, 4199-4205.
(29) Thai, R.; Moine, G.; Desmadril, M.; Servent, D.; Tarride, J.-L.; Menez, A.; Leonetti, M.: Antigen Stability Controls Antigen Presentation. *J. Biol. Chem.* **2004**, *279*, 50257-50266.
(30) Carter, P. J.: Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp. Cell Res.* **2011**, *317*, 1261-9.
(31) Adessi, C.; Soto, C.: Converting a peptide into a drug: strategies to improve stability and bioavailability. *Curr. Med. Chem.* **2002**, *9*, 963-78.
(32) Taylor, E. M.; Otero, D. A.; Banks, W. A.; O'Brien, J. S.: Designing Stable Blood-Brain Barrier-Permeable Prosaptide Peptides for Treatment of Central Nervous System Neurodegeneration. *J. Pharmacol. Exp. Ther.* **2000**, *293*, 403-409.
(33) Vigneri, R.; Squatrito, S.; Sciacca, L.: Insulin and its analogs: actions via insulin and IGF receptors. *Acta Diabetol.* **2010**, *47*, 271-8.
(34) Howey, D. C.; Bowsher, R. R.; Brunelle, R. L.; Woodworth, J. R.: [Lys(B28), Pro(B29)]-human insulin. A rapidly absorbed analogue of human insulin. *Diabetes* **1994**, *43*, 396-402.
(35) Owens, D. R.; Coates, P. A.; Luzio, S. D.; Tinbergen, J. P.; Kurzhals, R.: Pharmacokinetics of 125I-labeled insulin glargine (HOE 901) in healthy men: comparison with NPH insulin and the influence of different subcutaneous injection sites. *Diabetes Care* **2000**, *23*, 813-819.
(36) Goebl, N. A.; Babbey, C. M.; Datta-Mannan, A.; Witcher, D. R.; Wroblewski, V. J.; Dunn, K. W.: Neonatal Fc Receptor Mediates Internalization of Fc in Transfected Human Endothelial Cells. *Mol. Biol. Cell* **2008**, *19*, 5490-5505.
(37) Lee, T.-Y.; Tjin Tham Sjin, R. M.; Movahedi, S.; Ahmed, B.; Pravda, E. A.; Lo, K.-M.; Gillies, S. D.; Folkman, J.; Javaherian, K.: Linking Antibody Fc Domain to Endostatin Significantly Improves Endostatin Half-life and Efficacy. *Clin. Cancer Res.* **2008**, *14*, 1487-1493.
(38) Bause, E.: Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **1983**, *209*, 331-6.
(39) Roitsch, T.; Lehle, L.: Structural requirements for protein N-glycosylation. Influence of acceptor peptides on cotranslational glycosylation of yeast invertase and site-directed mutagenesis around a sequon sequence. *Eur. J. Biochem.* **1989**, *181*, 525-9.
(40) Imperiali, B.; Shannon, K. L.: Differences between Asn-Xaa-Thr-containing peptides: a comparison of solution conformation and substrate behavior with oligosaccharyltransferase. *Biochemistry (Mosc).* **1991**, *30*, 4374-4380.
(41) 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.: Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat. Biotechnol.* **2003**, *21*, 414-21.
(42) Kiss, Z.; Elliott, S.; Jedynasty, K.; Tesar, V.; Szegedi, J.: Discovery and basic pharmacology of erythropoiesis-stimulating agents (ESAs), including the hyperglycosylated ESA, darbepoetin alfa: an update of the rationale and clinical impact. *Eur. J. Clin. Pharmacol.* **2010**, *66*, 331-40.
(43) Maeda, H.: SMANCS and polymer-conjugated macromolecular drugs: advantages in cancer chemotherapy. *Advanced Drug Delivery Reviews* **2001**, *46*, 169-185.
(44) Azzopardi, E. A.; Ferguson, E. L.; Thomas, D. W.: The enhanced permeability retention effect: a new paradigm for drug targeting in infection. *J. Antimicrob. Chemother.* **2013**, *68*, 257-274.
(45) Satchi, R.; Connors, T. A.; Duncan, R.: PDEPT: polymer-directed enzyme prodrug therapy. *Br. J. Cancer* **2001**, *85*, 1070-1076.

(46) Abuchowski, A.; McCoy, J. R.; Palczuk, N. C.; van Es, T.; Davis, F. F.: Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase. *J. Biol. Chem.* **1977**, *252*, 3582-6.
(47) Abuchowski, A.; van Es, T.; Palczuk, N. C.; Davis, F. F.: Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol. *J. Biol. Chem.* **1977**, *252*, 3578-81.
(48) Levy, Y.; Hershfield, M. S.; Fernandez-Mejia, C.; Polmar, S. H.; Scudiery, D.; Berger, M.; Sorensen, R. U.: Adenosine deaminase deficiency with late onset of recurrent infections: Response to treatment with polyethylene glycol-modified adenosine deaminase. *The Journal of Pediatrics* **1988**, *113*, 312-317.
(49) Jevsevar, S.; Kunstelj, M.; Porekar, V. G.: PEGylation of therapeutic proteins. *Biotechnology journal* **2010**, *5*, 113-28.
(50) Webster, R.; Elliott, V.; Park, B. K.; Walker, D.; Hankin, M.; Taupin, P.: PEG and PEG conjugates toxicity: towards an understanding of the toxicity of PEG and its relevance to PEGylated biologicals. In *PEGylated Protein Drugs: Basic Science and Clinical Applications*; Veronese, F., Ed.; Birkhäuser Basel, 2009; pp 127-146.
(51) Cheng, T. L.; Wu, P. Y.; Wu, M. F.; Chern, J. W.; Roffler, S. R.: Accelerated clearance of polyethylene glycol-modified proteins by anti-polyethylene glycol IgM. *Bioconjug. Chem.* **1999**, *10*, 520-8.
(52) Mu, Q.; Hu, T.; Yu, J.: Molecular Insight into the Steric Shielding Effect of PEG on the Conjugated Staphylokinase: Biochemical Characterization and Molecular Dynamics Simulation. *PLoS ONE* **2013**, *8*, e68559.
(53) Fineberg, S. E.; Galloway, J. A.; Fineberg, N. S.; Rathbun, M. J.; Hufferd, S.: Immunogenicity of recombinant DNA human insulin. *Diab tologia* **1983**, *25*, 465-9.
(54) Schellekens, H.: Immunogenicity of therapeutic proteins. *Nephrology Dialysis Transplantation* **2003**, *18*, 1257-1259.
(55) Palleroni, A. V.; Aglione, A.; Labow, M.; Brunda, M. J.; Pestka, S.; Sinigaglia, F.; Garotta, G.; Alsenz, J.; Braun, A.: Interferon immunogenicity: Preclinical evaluation of interferon-alpha 2a. *J. Interferon Cytokine Res.* **1997**, *17*, S23-S27.
(56) Schellekens, H.: Bioequivalence and the immunogenicity of biopharmaceuticals. *Nat Rev Drug Discov* **2002**, *1*, 457-462.
(57) Schellekens, H.: Immunogenicity of therapeutic proteins: Clinical implications and future prospects. *Clin. Ther.* **2002**, *24*, 1720-1740.
(58) Pieters, R.; Hunger, S. P.; Boos, J.; Rizzari, C.; Silverman, L.; Baruchel, A.; Goekbuget, N.; Schrappe, M.; Pui, C. H.: L-asparaginase treatment in acute lymphoblastic leukemia: a focus on Erwinia asparaginase. *Cancer* **2011**, *117*, 238-49.
(59) McCredie, K. B.; Ho, D. H. W.; Freireich, E. J.: L-asparaginase for the treatment of cancer. *CA. Cancer J. Clin.* **1973**, *23*, 220-227.
(60) Ho, D. H.; Brown, N. S.; Yen, A.; Holmes, R.; Keating, M.; Abuchowski, A.; Newman, R. A.; Krakoff, I. H.: Clinical pharmacology of polyethylene glycol-L-asparaginase. *Drug metabolism and disposition: the biological fate of chemicals* **1986**, *14*, 349-52.
(61) Peters, B. G.; Goeckner, B. J.; Ponzillo, J. J.; Velasquez, W. S.; Wilson, A. L.: Pegaspargase versus asparaginase in adult ALL: a pharmacoeconomic assessment. *Formulary* **1995**, *30*, 388-93.
(62) Davis, F. F.: The origin of pegnology. *Advanced Drug Delivery Reviews* **2002**, *54*, 457-458.
(63) Lote, C. J.: Principles of Renal Physiology. Springer, 2012; pp 37-38.
(64) Feld, J. J.; Hoofnagle, J. H.: Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* **2005**, *436*, 967-972.
(65) Zeuzem, S.; Welsch, C.; Herrmann, E.: Pharmacokinetics of Peginterferons. *Semin. Liver Dis.* **2003**, *23*, 023-028.
(66) Bukowski, R. M.; Tendler, C.; Cutler, D.; Rose, E.; Laughlin, M. M.; Statkevich, P.: Treating cancer with PEG Intron. *Cancer* **2002**, *95*, 389-396.

(67) Glue, P.; Fang, J. W. S.; Rouzier-Panis, R.; Raffanel, C.; Sabo, R.; Gupta, S. K.; Salfi, M.; Jacobs, S.: Pegylated interferon-[alpha]2b: Pharmacokinetics, pharmacodynamics, safety, and preliminary efficacy data[ast]. *Clin. Pharmacol. Ther.* **2000**, *68*, 556-567.
(68) Williams, L. E.; Wu, A. M.; Yazaki, P. J.; Liu, A.; Raubitschek, A. A.; Shively, J. E.; Wong, J. Y.: Numerical selection of optimal tumor imaging agents with application to engineered antibodies. *Cancer biotherapy & radiopharmaceuticals* **2001**, *16*, 25-35.
(69) Wu, A. M.; Olafsen, T.: Antibodies for molecular imaging of cancer. *Cancer journal (Sudbury, Mass.)* **2008**, *14*, 191-7.
(70) Holliger, P.; Hudson, P. J.: Engineered antibody fragments and the rise of single domains. *Nat Biotech* **2005**, *23*, 1126-1136.
(71) Li, L.; Crow, D.; Turatti, F.; Bading, J. R.; Anderson, A.-L.; Poku, E.; Yazaki, P. J.; Carmichael, J.; Leong, D.; Wheatcroft, M. P.; Raubitschek, A. A.; Hudson, P. J.; Colcher, D.; Shively, J. E.: Site-Specific Conjugation of Monodispersed DOTA-PEGn to a Thiolated Diabody Reveals the Effect of Increasing PEG Size on Kidney Clearance and Tumor Uptake with Improved 64-Copper PET Imaging. *Bioconjug. Chem.* **2011**, *22*, 709-716.
(72) Kontermann, R.: Therapeutic Proteins: Strategies to Modulate Their Plasma Half-lives. Wiley, 2012; pp 31-32.
(73) Clark, R.; Olson, K.; Fuh, G.; Marian, M.; Mortensen, D.; Teshima, G.; Chang, S.; Chu, H.; Mukku, V.; Canova-Davis, E.; Somers, T.; Cronin, M.; Winkler, M.; Wells, J. A.: Long-acting growth hormones produced by conjugation with polyethylene glycol. *J. Biol. Chem.* **1996**, *271*, 21969-77.
(74) Johnson, V.; Maack, T.: Renal extraction, filtration, absorption, and catabolism of growth hormone. *American Journal of Physiology - Renal Physiology* **1977**, *233*, F185-F196.
(75) Lee, S.-H.; Lee, S.; Youn, Y. S.; Na, D. H.; Chae, S. Y.; Byun, Y.; Lee, K. C.: Synthesis, Characterization, and Pharmacokinetic Studies of PEGylated Glucagon-like Peptide-1. *Bioconjug. Chem.* **2005**, *16*, 377-382.
(76) Deacon, C. F.; Nauck, M. A.; Toft-Nielsen, M.; Pridal, L.; Willms, B.; Holst, J. J.: Both subcutaneously and intravenously administered glucagon-like peptide I are rapidly degraded from the NH2-terminus in type II diabetic patients and in healthy subjects. *Diabetes* **1995**, *44*, 1126-31.
(77) Gault, V.; Parker, J.; Harriott, P.; Flatt, P.; O'Harte, F.: Evidence that the major degradation product of glucose-dependent insulinotropic polypeptide, GIP(3-42), is a GIP receptor antagonist in vivo. *J. Endocrinol.* **2002**, *175*, 525-533.
(78) Gault, V. A.; Kerr, B. D.; Irwin, N.; Flatt, P. R.: C-terminal mini-PEGylation of glucose-dependent insulinotropic polypeptide exhibits metabolic stability and improved glucose homeostasis in dietary-induced diabetes. *Biochem. Pharmacol.* **2008**, *75*, 2325-2333.
(79) Lee, F. S.; Shapiro, R.; Vallee, B. L.: Tight-binding inhibition of angiogenin and ribonuclease A by placental ribonuclease inhibitor. *Biochemistry (Mosc).* **1989**, *28*, 225-230.
(80) Matousek, J.; Soucek, J.; Slavik, T.; Tomanek, M.; Lee, J. E.; Raines, R. T.: Comprehensive comparison of the cytotoxic activities of onconase and bovine seminal ribonuclease. *Comparative biochemistry and physiology. Toxicology & pharmacology : CBP* **2003**, *136*, 343-56.
(81) Rutkoski, T. J.; Kink, J. A.; Strong, L. E.; Raines, R. T.: Site-specific PEGylation endows a mammalian ribonuclease with antitumor activity. *Cancer biology & therapy* **2011**, *12*, 208-14.
(82) Veronese, F. M.; Mero, A.: The impact of PEGylation on biological therapies. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy* **2008**, *22*, 315-29.
(83) Yang, Z.; Domach, M.; Auger, R.; Yang, F. X.; Russell, A. J.: Polyethylene glycol-induced stabilization of subtilisin. *Enzyme Microb. Technol.* **1996**, *18*, 82-89.
(84) Chamow, S. M.; Kogan, T. P.; Venuti, M.; Gadek, T.; Harris, R. J.; Peers, D. H.; Mordenti, J.; Shak, S.; Ashkenazi, A.: Modification of CD4 Immunoadhesin with Monomethoxypoly(Ethylene Glycol) Aldehyde via Reductive Alkylation. *Bioconjug. Chem.* **1994**, *5*, 133-140.
(85) Kinstler, O.; Molineux, G.; Treuheit, M.; Ladd, D.; Gegg, C.: Mono-N-terminal poly(ethylene glycol)–protein conjugates. *Advanced Drug Delivery Reviews* **2002**, *54*, 477-485.

(86) Wang, J.; Hu, T.; Liu, Y.; Zhang, G.; Ma, G.; Su, Z.: Kinetic and stoichiometric analysis of the modification process for N-terminal PEGylation of staphylokinase. *Anal. Biochem.* **2011**, *412*, 114-116.
(87) Hu, J.; Sebald, W.: N-terminal specificity of PEGylation of human bone morphogenetic protein-2 at acidic pH. *Int. J. Pharm.* **2011**, *413*, 140-146.
(88) Veronese, F. M.; Mero, A.; Caboi, F.; Sergi, M.; Marongiu, C.; Pasut, G.: Site-Specific Pegylation of G-CSF by Reversible Denaturation. *Bioconjug. Chem.* **2007**, *18*, 1824-1830.
(89) Goodson, R. J.; Katre, N. V.: Site-Directed Pegylation of Recombinant Interleukin-2 at its Glycosylation Site. *Nat Biotech* **1990**, *8*, 343-346.
(90) Brocchini, S.; Godwin, A.; Balan, S.; Choi, J.-w.; Zloh, M.; Shaunak, S.: Disulfide bridge based PEGylation of proteins. *Advanced Drug Delivery Reviews* **2008**, *60*, 3-12.
(91) Vanwetswinkel, S.; Plaisance, S.; Zhi-yong, Z.; Vanlinthout, I.; Brepoels, K.; Lasters, I.; Collen, D.; Jespers, L.: Pharmacokinetic and thrombolytic properties of cysteine-linked polyethylene glycol derivatives of staphylokinase. *Blood* **2000**, *95*, 936-942.
(92) Nesbitt, A.; Stephens, S.; Chartash, E.: Certolizumab pegol: a PEGylated anti-tumour necrosis factor alpha biological agent. In *PEGylated Protein Drugs: Basic Science and Clinical Applications*; Veronese, F., Ed.; Birkhäuser Basel, 2009; pp 229-254.
(93) Xie, J.; Schultz, P. G.: A chemical toolkit for proteins [mdash] an expanded genetic code. *Nat Rev Mol Cell Biol* **2006**, *7*, 775-782.
(94) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G.: Expanding the Genetic Code of Escherichia coli. *Science* **2001**, *292*, 498-500.
(95) Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A.-M. H.; Kraynov, V. S.; Sim, B.-C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; GC, B.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E.: Optimized clinical performance of growth hormone with an expanded genetic code. *Proceedings of the National Academy of Sciences* **2011**, *108*, 9060-9065.
(96) Chen, H.; Lu, Y.; Fang, Z.; Liu, J.; Tian, H.; Gao, X.; Yao, W.: High-level production of uricase containing keto functional groups for site-specific PEGylation. *Biochemical Engineering Journal* **2011**, *58–59*, 25-32.
(97) Chen, M.; Cai, L.; Fang, Z.; Tian, H.; Gao, X.; Yao, W.: Site-specific incorporation of unnatural amino acids into urate oxidase in Escherichia coli. *Protein Sci.* **2008**, *17*, 1827-1833.
(98) Tada, S.; Andou, T.; Suzuki, T.; Dohmae, N.; Kobatake, E.; Ito, Y.: Genetic PEGylation. *PLoS ONE* **2012**, *7*, e49235.
(99) Goerke, A. R.; Swartz, J. R.: High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnol. Bioeng.* **2009**, *102*, 400-416.
(100) Bundy, B. C.; Swartz, J. R.: Site-Specific Incorporation of p-Propargyloxyphenylalanine in a Cell-Free Environment for Direct Protein−Protein Click Conjugation. *Bioconjug. Chem.* **2010**, *21*, 255-263.
(101) Deiters, A.; Cropp, T. A.; Summerer, D.; Mukherji, M.; Schultz, P. G.: Site-specific PEGylation of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5743-5745.
(102) Tornøe, C. W.; Christensen, C.; Meldal, M.: Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *The Journal of Organic Chemistry* **2002**, *67*, 3057-3064.
(103) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G.: Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193.
(104) DeFrees, S.; Wang, Z.-G.; Xing, R.; Scott, A. E.; Wang, J.; Zopf, D.; Gouty, D. L.; Sjoberg, E. R.; Panneerselvam, K.; Brinkman-Van der Linden, E. C. M.; Bayer, R. J.; Tarp, M. A.; Clausen, H.: GlycoPEGylation of recombinant therapeutic proteins produced in Escherichia coli. *Glycobiology* **2006**, *16*, 833-843.
(105) Negrier, C.; Knobe, K.; Tiede, A.; Giangrande, P.; Møss, J.: Enhanced pharmacokinetic properties of a glycoPEGylated recombinant factor IX: a first human dose trial in patients with hemophilia B. *Blood* **2011**, *118*, 2695-2701.

(106) Sato, H.: Enzymatic procedure for site-specific pegylation of proteins. *Advanced Drug Delivery Reviews* **2002**, *54*, 487-504.
(107) Mero, A.; Spolaore, B.; Veronese, F. M.; Fontana, A.: Transglutaminase-Mediated PEGylation of Proteins: Direct Identification of the Sites of Protein Modification by Mass Spectrometry using a Novel Monodisperse PEG. *Bioconjug. Chem.* **2009**, *20*, 384-389.
(108) Spolaore, B.; Raboni, S.; Ramos Molina, A.; Satwekar, A.; Damiano, N.; Fontana, A.: Local Unfolding Is Required for the Site-Specific Protein Modification by Transglutaminase. *Biochemistry (Mosc).* **2012**, *51*, 8679-8689.
(109) Mero, A.; Schiavon, M.; Veronese, F. M.; Pasut, G.: A new method to increase selectivity of transglutaminase mediated PEGylation of salmon calcitonin and human growth hormone. *J. Controlled Release* **2011**, *154*, 27-34.
(110) Cox, G. N.; Rosendahl, M. S.; Chlipala, E. A.; Smith, D. J.; Carlson, S. J.; Doherty, D. H.: A long-acting, mono-PEGylated human growth hormone analog is a potent stimulator of weight gain and bone growth in hypophysectomized rats. *Endocrinology* **2007**, *148*, 1590-7.
(111) Wu, L.; Ho, S. V.; Wang, W.; Gao, J.; Zhang, G.; Su, Z.; Hu, T.: N-terminal mono-PEGylation of growth hormone antagonist: Correlation of PEG size and pharmacodynamic behavior. *Int. J. Pharm.* **2013**, *453*, 533-540.
(112) Foser, S.; Weyer, K.; Huber, W.; Certa, U.: Improved biological and transcriptional activity of monopegylated interferon-[alpha]-2a isomers. *Pharmacogenomics J* **2003**, *3*, 312-319.
(113) Pradhananga, S.; Wilkinson, I.; Ross, R.: Pegvisomant: structure and function. *J. Mol. Endocrinol.* **2002**, *29*, 11-14.
(114) Hu, T.; Li, D.; Manjula, B. N.; Brenowitz, M.; Prabhakaran, M.; Acharya, S. A.: PEGylation of Val-1(α) Destabilizes the Tetrameric Structure of Hemoglobin†. *Biochemistry (Mosc).* **2009**, *48*, 608-616.
(115) Hinds, K. D.; Kim, S. W.: Effects of PEG conjugation on insulin properties. *Advanced Drug Delivery Reviews* **2002**, *54*, 505-530.
(116) Digilio, G.; Barbero, L.; Bracco, C.; Corpillo, D.; Esposito, P.; Piquet, G.; Traversa, S.; Aime, S.: NMR Structure of Two Novel Polyethylene Glycol Conjugates of the Human Growth Hormone-Releasing Factor, hGRF(1−29)−NH2. *J. Am. Chem. Soc.* **2003**, *125*, 3458-3470.
(117) Plesner, B.; Fee, C. J.; Westh, P.; Nielsen, A. D.: Effects of PEG size on structure, function and stability of PEGylated BSA. *Eur. J. Pharm. Biopharm.* **2011**, *79*, 399-405.
(118) Caliceti, P.; Veronese, F. M.: Pharmacokinetic and biodistribution properties of poly(ethylene glycol)–protein conjugates. *Advanced Drug Delivery Reviews* **2003**, *55*, 1261-1277.
(119) Wang, W.: Protein aggregation and its inhibition in biopharmaceutics. *Int. J. Pharm.* **2005**, *289*, 1-30.
(120) Wang, W.; Singh, S. K.; Li, N.; Toler, M. R.; King, K. R.; Nema, S.: Immunogenicity of protein aggregates—Concerns and realities. *Int. J. Pharm.* **2012**, *431*, 1-11.
(121) Rodriguez-Martinez, J. A.; Sola, R. J.; Castillo, B.; Cintron-Colon, H. R.; Rivera-Rivera, I.; Barletta, G.; Griebenow, K.: Stabilization of alpha-chymotrypsin upon PEGylation correlates with reduced structural dynamics. *Biotechnol. Bioeng.* **2008**, *101*, 1142-9.
(122) Topchieva, I. N.; Efremova, N. V.; Khvorov, N. V.; Magretova, N. N.: Synthesis and Physicochemical Properties of Protein Conjugates with Water-Soluble Poly(alkylene oxides). *Bioconjug. Chem.* **1995**, *6*, 380-388.
(123) Longo, M. A.; Combes, D.: Thermostability of modified enzymes: a detailed study. *J. Chem. Technol. Biotechnol.* **1999**, *74*, 25-32.
(124) Gaertner, H. F.; Puigserver, A. J.: Increased activity and stability of poly(ethylene glycol)-modified trypsin. *Enzyme Microb. Technol.* **1992**, *14*, 150-155.
(125) García-Arellano, H.; Valderrama, B.; Saab-Rincón, G.; Vazquez-Duhalt, R.: High Temperature Biocatalysis by Chemically Modified Cytochrome c. *Bioconjug. Chem.* **2002**, *13*, 1336-1344.
(126) Monfardini, C.; Schiavon, O.; Caliceti, P.; Morpurgo, M.; Harris, J. M.; Veronese, F. M.: A Branched Monomethoxypoly(ethylene glycol) for Protein Modification. *Bioconjug. Chem.* **1995**, *6*, 62-69.

(127) Nodake, Y.; Yamasaki, N.: Some Properties of a Macromolecular Conjugate of Lysozyme Prepared by Modification with a Monomethoxypolyethylene Glycol Derivative. *Biosci. Biotechnol. Biochem.* **2000**, *64*, 767-774.
(128) Gokarn, Y. R.: Hydrodynamic Behavior and Thermal Stability of a PEGylated Protein: Studies with Hen Egg Lysozyme. Ph.D. Dissertation, University of New Hampshire, 2003.
(129) So, T.; Ueda, T.; Abe, Y.; Nakamata, T.; Imoto, T.: Situation of Monomethoxypolyethylene Glycol Covalently Attached to Lysozyme. *The Journal of Biochemistry* **1996**, *119*, 1086-1093.
(130) Svergun, D. I.; Ekström, F.; Vandegriff, K. D.; Malavalli, A.; Baker, D. A.; Nilsson, C.; Winslow, R. M.: Solution Structure of Poly(ethylene) Glycol-Conjugated Hemoglobin Revealed by Small-Angle X-Ray Scattering: Implications for a New Oxygen Therapeutic. *Biophys. J.* **2008**, *94*, 173-181.
(131) Meng, W.; Guo, X.; Qin, M.; Pan, H.; Cao, Y.; Wang, W.: Mechanistic Insights into the Stabilization of srcSH3 by PEGylation. *Langmuir* **2012**, *28*, 16133-16140.
(132) Carducci, M.; Perfetto, L.; Briganti, L.; Paoluzi, S.; Costa, S.; Zerweck, J.; Schutkowski, M.; Castagnoli, L.; Cesareni, G.: The protein interaction network mediated by human SH3 domains. *Biotechnology Advances* **2012**, *30*, 4-15.
(133) Karkkainen, S.; Hiipakka, M.; Wang, J.-H.; Kleino, I.; Vaha-Jaakkola, M.; Renkema, G. H.; Liss, M.; Wagner, R.; Saksela, K.: Identification of preferred protein interactions by phage-display of the human Src homology-3 proteome. *EMBO Rep* **2006**, *7*, 186-191.
(134) Myers, J. K.; Nick Pace, C.; Martin Scholtz, J.: Denaturant m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4*, 2138-2148.
(135) Solá, R. J.; Griebenow, K.: Chemical glycosylation: New insights on the interrelation between protein structural mobility, thermodynamic stability, and catalysis. *FEBS Lett.* **2006**, *580*, 1685-1690.
(136) Yang, C.; Lu, D. N.; Liu, Z.: How PEGylation Enhances the Stability and Potency of Insulin: A Molecular Dynamics Simulation. *Biochemistry (Mosc).* **2011**, *50*, 2585–2593.
(137) Manjula, B. N.; Tsai, A.; Upadhya, R.; Perumalsamy, K.; Smith, P. K.; Malavalli, A.; Vandegriff, K.; Winslow, R. M.; Intaglietta, M.; Prabhakaran, M.; Friedman, J. M.; Acharya, A. S.: Site-Specific PEGylation of Hemoglobin at Cys-93(β): Correlation between the Colligative Properties of the PEGylated Protein and the Length of the Conjugated PEG Chain. *Bioconjug. Chem.* **2003**, *14*, 464-472.
(138) Sola, R. J.; Griebenow, K.: Influence of modulated structural dynamics on the kinetics of alpha-chymotrypsin catalysis. Insights through chemical glycosylation, molecular dynamics and domain motion analysis. *The FEBS journal* **2006**, *273*, 5303-19.
(139) Pai, S. S.; Hammouda, B.; Hong, K.; Pozzo, D. C.; Przybycien, T. M.; Tilton, R. D.: The Conformation of the Poly(ethylene glycol) Chain in Mono-PEGylated Lysozyme and Mono-PEGylated Human Growth Hormone. *Bioconjug. Chem.* **2011**, *22*, 2317-2323.
(140) Pandey, B. K.; Smith, M. S.; Torgerson, C.; Lawrence, P. B.; Matthews, S. S.; Watkins, E.; Groves, M. L.; Prigozhin, M. B.; Price, J. L.: Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length. *Bioconjug. Chem.* **2013**, *24*, 796-802.

Chapter 2: The Thermodynamic Consequences of PEGylation Depend on the Site of Poly(Ethylene Glycol) Attachment

2.1 Introduction

The impact of PEGylation on protein conformation stability (i.e., the difference in free energy between the folded and unfolded conformations of a protein, see Figure 1) is incompletely understood. PEGylation can increase,[1-26] decrease,[27,28] or have no effect on protein conformational stability,[9,29-33] and the molecular basis for these differences is unknown, in part because of the prevalence of non-specific PEGylation strategies that attach one or more PEG oligomers at multiple locations on the protein surface. Although increasing numbers of site-specific PEGylation strategies can enable chemists to conjugate PEG to a specific residue on the protein surface, most studies on the site-specific impact of PEGylation have focused on different *regions* of the protein, such as avoiding PEGylation at the active site of a therapeutic enzyme. It is still unclear whether the location of a PEG conjugate has a significant effect on protein conformational stability. It seems reasonable to expect that some PEGylation sites would be better than others and if so, it would be important to understand why, with the goal of using such insight to identify optimal PEGylation sites a priori. To our knowledge, no studies have been published examining the impact that PEGylation site has on the thermodynamic consequences of PEGylation.

In 2011, Kelly and coworkers showed that a short PEG oligomer can substantially increase the conformational stability of the human protein Pin 1 (hereafter called WW). [34] They attached an oligomer comprising four ethylene oxide units to the side-chain amide nitrogen of an Asn residue at position 19 of the WW domain of the WW (Figure 1a), and observed an increase of -0.86 ± 0.05 kcal mol^{-1} in thermodynamic stability over the non-PEGylated peptide. Our lab recently demonstrated that this stabilization is due to an increase in folding rate and decrease in unfolding rate upon PEGylation[35]. The increased conformational stability associated with PEGylation depends strongly on the length of the attached PEG oligomer, with maximum stability achieved at four ethylene oxide units. This observation suggests the possibility that stabilizing interactions between the PEG oligomer and specific protein side-chain or backbone groups lower the free energy of the native state and the transition state relative to the unfolded state (Figure 1b). It is interesting to note that the PEG trimer and tetramer are significantly more

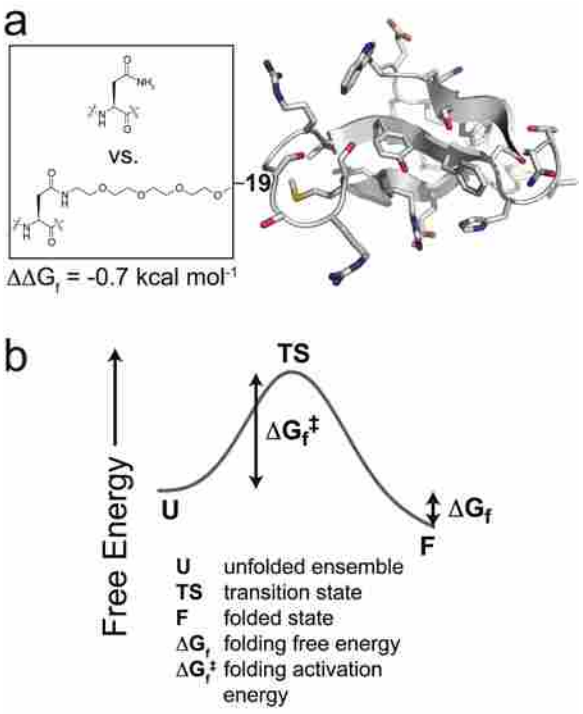


Figure 1(a) The energetic impact of PEGylating an Asn residue within a reverse turn in the WW domain, shown as a ribbon diagram, with side chains shown as sticks (PDB ID: 1PIN). (b) A two-state folding energy diagram approximates the folding free energy landscape of WW

stabilizing than the PEG dimer or monomer. This may indicate that the third ethylene glycol unit is able to reach a critical interacting partner that is out of reach for the dimer. If this is the case, we would expect the thermodynamic consequences of PEGylation to depend on the site to which PEG is conjugated.

PEGylation may stabilize some proteins and destabilize others because, in the latter case, PEG has conjugated to a sub-optimal site on the protein surface. As methods of site-specific PEGylation continue to improve, understanding the impact of PEGylation site on the thermodynamic consequences of PEGylation will facilitate the rational design of improved PEGylated proteins. We seek to understand the relationship between PEGylation site and protein conformational stability.

2.2 Results and Discussion

2.2.1 Initial PEG Scan

In 2010, Price et al.[36] determined that the thermodynamic consequences of glycosylation depend strongly upon the site of glycan conjugation. In a subsequent communication,[37] they demonstrated that this site-dependence was due to the presence of specific nearby side chains with which the glycan interacted in the folded state. We hypothesized that the thermodynamic impact of PEG might also depend on the site of PEGylation. We generated proteins **14**, **17**, **18**, **19**, and **23**, in which wild-type residues at positions 14, 17, 18, 19 and 23, respectively, have been changed to Asn. Asn already occupies positions 26 and 30 in the unmodified protein **WW**. In proteins **14p**, **17p**, **18p**, **19p**, **23p**, **26p**, and **30p**, the Asn residues at positions 14, 17, 18, 19, 23, 26 and 30, respectively, have been replaced by Asn-PEG_4 a PEGylated Asn derivative in which a PEG oligomer comprising four ethylene oxide units has been attached to the side-chain amide nitrogen of Asn. The sequences of these peptides, as well as the structure of Asn-PEG_4, are shown in Figure 2.

We used variable temperature circular dichroism and temperature jump kinetic experiments to assess the conformational stability, folding rate, and unfolding rate of PEGylated proteins **14p**, **17p**, **18p**, **19p**, **23p**, **26p**, and **30p** and their non-PEGylated counterparts **14**, **17**, **18**, **19**, **23**, and **WW**. All peptides were analyzed at 100 μM concentration in 20 mM aqueous sodium phosphate (pH 7.0). The results of

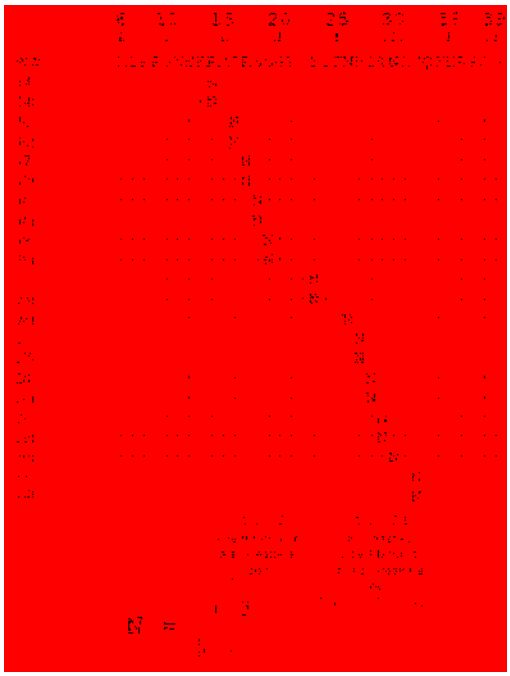


Figure 2: Amino acid sequences for PEGylated and non-PEGylated WW variants of the parent protein **WW**.

these analyses appear in Figure 3 and Table 1. PEGylation substantially increases WW conformational stability at positions 19 and 26 and moderately increases WW conformational stability at position 17.

Position 17 is located in the N-terminal reverse turn of WW; this reverse turn adopts an unusual conformation consisting of a four-residue type II β-turn embedded in a six-residue hydrogen bonded loop.[35,38] PEGylated protein **17p** (T_m = 55.6 ± 0.5 °C) is -0.18 ± 0.05 kcal mol^{-1} more stable than non-PEGylated protein **17** (T_m = 53.6 ± 0.4 °C). This moderate increase in conformational stability appears to come from a small increase in folding rate and a small decrease in unfolding rate (Table 1). This is consistent with simultaneous stabilization of the folded state and transition state, with the folded state being stabilized more.

Position 19 is also located in the N-terminal reverse turn of WW. PEGylated protein **19p** (T_m = 63.2 ± 0.3 °C), is -0.69 ± 0.05 kcal mol^{-1} more stable than non-PEGylated protein **19** (T_m = 55.6 ± 0.3 °C). The increase in conformational stability associated with PEGylation of Asn_{19} comes from a slightly accelerated folding rate and a reduced unfolding rate: PEGylated protein **19p** (k_f = 8.0 ± 0.5 × 10^3 s^{-1}, k_u = 2.8 ± 0.2 × 10^3 s^{-1}) folds 1.3 ± 0.1 faster than **19** (k_f = 6.2 ± 0.2 × 10^3 s^{-1}) and unfolds 2.2 ± 0.2 times more slowly than **19** (k_u = 6.2 ± 0.4 × 10^3 s^{-1}), corresponding to a -0.17 ± 0.05 kcal mol^{-1} decrease in folding activation energy, and a 0.52 ± 0.07 kcal mol^{-1} increase in unfolding activation energy (Table 1).

Position 26 is located in the C-terminal reverse turn of WW. This reverse turn contains two nested type I β-turns: Asn26, His27 and Ile 28 occupy the i, i+1 and i+2 positions, respectively of the first type I β-turn, but also occupy the i and i+1 positions, respectively, of the second β-turn [35,38]. PEGylated protein **26p** (T_m = 63.9 ± 0.1 °C) is -0.50 ± 0.08 kcal mol^{-1} more stable than non-PEGylated **WW** (T_m = 58.3 ± 0.8 °C). This increase in conformational stability comes mostly from an accelerated folding rate: PEGylated protein **26p** (k_f = 19.6 ± 0.4 × 10^3 s^{-1}) folds 2.2 ± 0.2 times faster than its non-PEGylated counterpart **WW** (k_f = 9.1 ± 0.8 ×10^3 s^{-1}), corresponding to a -0.51 ± 0.06 kcal mol^{-1} decrease in folding activation energy. In contrast, the unfolding rates of **26p** and **WW** (k_u = 9.1 ± 0.3 × 10^3 s^{-1} and 9.1 ± 1.3 × 10^3 s^{-1}, respectively) are indistinguishable (Table 1).

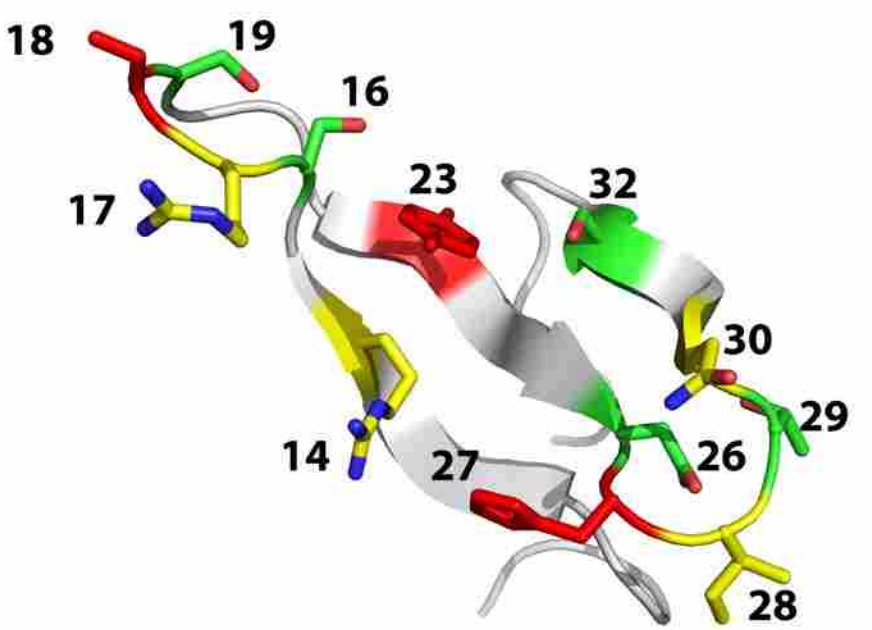


Figure 3: The energetic impact of PEGylating an Asn residue at various positions within the WW domain, shown as a ribbon diagram, with side chains shown as sticks (PDB ID: 1PIN). Side chains which were not tested are omitted for clarity. PEGylation sites are indicated by bold numbers corresponding to the position of each site within the WW primary sequence. The change in folding free energy upon PEGylation ($\Delta\Delta G_f$) at each position is indicated by the shading color at each position: at positions colored in red, $\Delta\Delta G_f$ is greater than +0.2 kcal mol^{-1}; at positions colored in yellow, $\Delta\Delta G_f$ is between +0.1 and -0.2 kcal mol^{-1}; at positions colored in green, $\Delta\Delta G_f$ is less than -0.3 kcal mol^{-1}.

Table 1: Folding free energies, folding and unfolding rates, and changes in folding and unfolding activation energies for PEGylated and non-PEGylated derivatives of WW.[a]

Protein	T_m (°C)	$\Delta\Delta G_f$ (kcal/mol)	k_f ($\times 10^3$ s^{-1})	k_f ratio	$\Delta\Delta G_f^{\ddagger}$ (kcal/mol)	k_u ($\times 10^3$ s^{-1})	k_u ratio	$\Delta\Delta G_u^{\ddagger}$ (kcal/mol)
14	33.7 ± 0.8		1.8 ± 0.10			1.8 ± 0.2		
14p	34.2 ± 3.2	-0.03 ± 0.23	1.2 ± 0.2	0.69 ± 0.11	0.23 ± 0.09	1.2 ± 0.5	0.7 ± 0.3	0.26 ± 0.25
16	54.8 ± 0.2		---	---	---	---	---	---
16p	62.3 ± 0.2	-0.71 ± 0.03	---	---	---	---	---	---
17	53.6 ± 0.4		5.7 ± 0.2			5.7 ± 0.3		
17p	55.6 ± 0.5	-0.18 ± 0.05	6.2 ± 0.2	1.08 ± 0.05	-0.05 ± 0.03	4.7 ± 0.3	0.82 ± 0.08	0.13 ± 0.06
18	59.7 ± 0.8		7.7 ± 0.4			7.7 ± 0.8		
18p	57.0 ± 0.7	0.26 ± 0.09	4.8 ± 0.3	0.63 ± 0.05	0.31 ± 0.05	7.2 ± 0.8	0.94 ± 0.14	0.04 ± 0.10
19	55.6 ± 0.3		6.2 ± 0.2			6.2 ± 0.4		
19p	63.2 ± 0.3	-0.69 ± 0.05	8.0 ± 0.5	1.3 ± 0.1	-0.17 ± 0.05	2.8 ± 0.2	0.45 ± 0.05	0.52 ± 0.07
20[b]	---	---	---	---	---	---	---	---
20p	57.1 ± 1.1	---	---	---	---	---	---	---
21[b]	---	---	---	---	---	---	---	---
21p[b]	---	---	---	---	---	---	---	---
23[b]	28.6 ± 0.4		---			---		
23p[b]	23.2 ± 1.0	0.42 ± 0.08	---	---	---	---	---	---
WW	58.3 ± 0.8		9.1 ± 0.8			9.1 ± 1.3		
26p	63.9 ± 0.1	-0.50 ± 0.07	19.6 ± 0.4	2.2 ± 0.2	-0.51 ± 0.06	9.1 ± 0.3	1.0 ± 0.2	0.0 ± 0.1
27	54.9 ± 0.2		---	---	---	---	---	---
27p	51.0 ± 0.4	0.37 ± 0.04	---	---	---	---	---	---
28	53.2 ± 0.5		---	---	---	---	---	---
28p	53.2 ± 0.5	0.00 ± 0.06	---	---	---	---	---	---
29	48.4 ± 0.3		---	---	---	---	---	---
29p	53.5 ± 0.2	-0.40 ± 0.03	---	---	---	---	---	---
WW	58.3 ± 0.8		9.1 ± 0.8			9.1 ± 1.3		
30p	58.3 ± 0.2	0.00 ± 0.08	7.7 ± 0.2	0.84 ± 0.08	0.11 ± 0.06	7.8 ± 0.4	0.9 ± 0.1	0.1 ± 0.1
32	45.0 ± 0.2		---	---	---	---	---	---
32p	50.3 ± 0.2	-0.46±0.03	---	---	---	---	---	---

[a]Tabulated Data are given as mean ± standard error 100 μM solutions of WW variants in 20 mM sodium phosphate buffer (pH 7) at the melting temperatures of the corresponding non-PEGylated proteins. ΔΔGf, ΔΔGf‡, and ΔΔGu‡, and folding and unfolding rate ratios for PEGylated proteins are relative to the corresponding sequence-matched non-PEGylated proteins. [b]Aggregation and/or thermal unfolding behavior inconsistent with a two-state folding/unfolding model precluded characterization of these proteins.

At the other positions investigated, PEGylation has a minimal impact on WW conformational stability (positions 14 and 30) or substantially decreases WW conformational stability (positions 18 and 23). The impact of PEGylation depends strongly upon the site of PEGylation.

2.2.2 PEGylation of Reverse Turns

Having determined that the site of PEG conjugation strongly influences the thermodynamic impact of PEGylation, we next sought to identify structural features in the folded state which make PEGylation stabilizing. Positions 19 and 26, which are both strongly stabilizing sites, are both located within reverse turns. We wondered whether PEGylation of Asn residues at other positions within the N- and C-terminal reverse turns of WW might also increase WW conformational stability. To test this hypothesis, we prepared proteins **16**, **20**, **21**, **27**, **28**, **29** in which wild-type residues at positions 16, 20, 21, 27, 28, and 29, respectively, have been replaced by Asn and PEGylated proteins **16p**, **20p**, **21p**, **27p**, and **28p**, **29p** in which wild-type residues at positions 16, 20, 21, 27, 28, and 29, respectively, have been replaced by Asn-PEG_4.

Position 16 is located in the N-terminal reverse turn of WW (Figure 3). PEGylated protein **16p** (T_m = 62.3 ± 0.2 °C) is -0.71 ± 0.03 kcal mol^{-1} more stable than non-PEGylated protein **16** (T_m = 54.8 ± 0.2 °C). Attempts to assess the impact of Asn PEGylation at positions 20 and 21 in the six-residue N-terminal reverse turn of WW were unsuccessful due to aggregation and/or thermal unfolding behavior in proteins **20**, **21**, and **21p** that was inconsistent with the two-state model that we use to extract thermodynamic information from our variable temperature CD Data.

Positions 27, 28, and 29 are all located C-terminal revese turn of WW (Figure 3). PEGylation at position 27 decreases WW conformational stability (compare proteins **27p** and **27**, $\Delta\Delta G_f$ = 0.39 ± 0.05 kcal mol^{-1}), whereas PEGylation at position 28 has no effect (compare proteins **28p** and **28**, $\Delta\Delta G_f$ = 0.00 ± 0.06 kcal mol^{-1}). However, PEGylation at position 29 substantially increases WW conformational stability (compare proteins **29p** and **29**, $\Delta\Delta G_f$ =-0.40 ± 0.03 kcal mol^{-1}).

2.2.3 PEGylation of a β-Strand

The five stabilizing sites identified thus far, positions 16, 19, 26, and 19, have two features in common: They are located in reverse turns, and they are positioned such that the side chain is oriented back towards the surface of the protein. We wondered whether either of these factors, or both, were important for PEG-induced stabilization. In the native WW sequence, position 32 is oriented back toward the protein surface, but is not located in a reverse turn. To test whether PEG needed to be located in a reverse turn to stabilize WW we generated proteins **32** and **32p**, in which the native serine at position 32 was replaced by Asn and Asn-PEG_4, respectively. PEGylated protein **32p** (T_m = 50.3 ± 0.2 °C) is -0.46 ± 0.02 kcal mol^{-1} more stable than non-PEGylated protein **32** (T_m = 45.0 ± 0.2 °C). This indicates that PEGylation can be stabilizing if the PEG moiety is oriented correctly, even if the site of PEGylation is not located within a reverse turn.

2.3 Conclusions

As we initially hypothesized, the thermodynamic consequences of PEGylation appear to depend strongly upon the location of PEG conjugation. We observe that PEGylation at positions 16, 19, 26, 29, or 32 increases WW conformational stability substantially, but that PEGylation at positions 14, 17, 18, 23, 27, 28, and 30 do not. Given the role that thermodynamic stability plays in protein aggregation, immunogenicity, and susceptibility to proteolysis (all of which reduce the effectiveness of protein therapeutics and may introduce new side effects), developers of PEGylated protein drugs may want to consider how the selection of PEGylation site affects the thermodynamic properties of a therapeutic protein.

Having identified several locations in WW variants where PEGylation increases WW conformational stability, it is interesting to consider what these locations have in common. In the crystal structure of the wild-type WW domain, positions 16 and 19 appear to point back toward the surface of the protein. Though we have changed the identities of these side-chains to Asn or PEGylated Asn in proteins **16**, **16p**, **19**, and **19p**, the similarity of the CD spectra of these WW variants to that of the parent protein

WW, suggests that this change has not resulting in dramatic structural rearrangement. Kelly and coworkers[39] observed that the residues in loop 1 of WW were particularly tolerant of mutations; they speculated that this may be due to the fact that loop one was evolutionarily optimized for ligand binding, not structural stability. Furthermore, recent solution structures of glycosylated WW mutants[40] demonstrate that when serine is mutated to asparagine, or even a glycosylated asparagine, at position 19, the residue is still oriented in much the same way was the original serine, as shown in Figure 4. It seems reasonable that an attached PEG oligomer at positions 16 and 19 would point back toward the surface of the protein (Figure 3), perhaps in the vicinity of side-chains with hydrogen-bond-donating groups. The previously observed[35] strong dependence of the observed PEG-associated increase in WW conformational stability and folding rate on the length of the PEG oligomer at position 19 suggests the possibility that the PEG oligomer at position 19 is interacting with nearby surface residues. In the second reverse turn, positions 26 is oriented into a hydrophobic pocket on the protein surface, while 29 points across the β-sheet toward the C-terminus. Likewise, position 32 is oriented toward the N-terminal reverse turn. All of these observations, along with the observation that the neutral side chains (where PEGylation did not significantly affect the thermodynamic stability of WW) project out into solution, suggest that PEGylation is most effective when the PEG can be directed back toward the protein surface.

There are several possible mechanisms by which PEG could stabilize the WW domain that account for these observations. Every stabilizing residue is pointed back toward at least one surface residue that is capable of hydrogen bonding. Given the high affinity of PEG polymers for water, [41] PEG may engage in hydrogen bonding interactions which lower the enthalpy of the folded state. It is possible that the electronegative oxygen atoms of the PEG oligomer are engaging in hydrogen bonding interactions with side-chain or backbone atoms on the protein surface, thereby partially desolvating the protein surface and releasing water molecules to the bulk solvent, increasing the entropy of the system.[1,21,22]

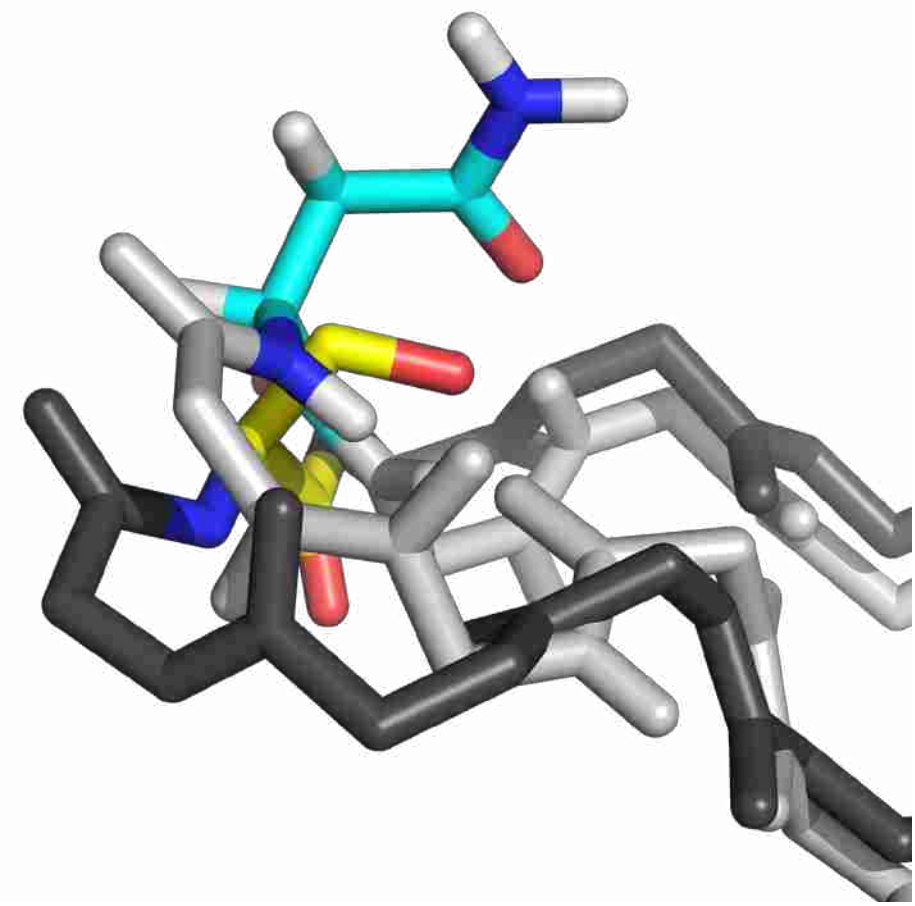


Figure 4: Structures of WW domain (Dark Gray, PDB code PIN1), an asparagine mutant (Light Gray, PDB code 2M9I). The sequences were aligned using PyMol's CEAlign function (RMSD =1.99 Å). The residue at position 19 is shown (Yellow – WW; Cyan – asparagine mutant). While some changes in the structure of loop 1 are observed, the orientation of the residue remains relatively unchanged. WW mutant structures are taken from ref. 34.

2.4 Supporting Information

2.4.1 Protein Synthesis

Proteins **WW, 14, 14p, 16, 16p, 17, 17p, 18, 18p, 19, 19p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32,** and **32p,** were synthesized as C-terminal acids, by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc Nα protection strategy. The amino acid sequences of all peptides appear in Figure 2. Amino acids were activated by 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU, purchased from Advanced ChemTech) and N-hydroxybenzotriazole hydrate (HOBt, purchased from Advanced ChemTech). Fmoc-Gly-loaded Novasyn WANG resin and all Fmoc-protected α-amino acids with acid-labile side-chain protecting groups were purchased from EMD Biosciences, except for Fmoc-Asn(PEG4)-OH (N^2-fluorenylmethyoxycarbonyl-N^4-[11-methoxy-3,6,9-trioxaundecyl]-L-asparagine), which was synthesized as described previously.[42] 1,2 Piperidine and N,N-diisopropylethylamine (DIEA) were purchased from Aldrich, and N-methyl pyrrolidinone (NMP) was purchased from Applied Biosystems.

All peptides were synthesized on a 25 μmol scale. A general protocol for manual solid phase peptide synthesis follows: Fmoc-Gly-loaded NovaSyn WANG resin (69.4 mg, 25 μmol at 0.38 mmol/g resin loading) was aliquotted into a fritted polypropylene syringe and allowed to swell first in CH_2Cl_2 , and then in dimethylformamide (DMF). [note: avoid using trityl resins in the microwave or at high temperatures. The peptide can spontaneously cleave from trityl resin at the high reaction temperatures used here, leading to considerably diminished yields]. Solvent was drained from the resin using a vacuum manifold.

To remove the Fmoc protecting group on the resin-linked amino acid, 1.25 ml of 20% piperidine in DMF was added to the resin, and the resulting mixture was allowed to sit at room temperature for 1 minute. The deprotection solution was then drained from the resin with a vacuum manifold. Then, an additional 2.5 mL of 20% piperidine in DMF was added to the resin, and the reaction vessel was placed in the microwave. The temperature was ramped from rt to 80°C over the course of 2 minutes, and held at

80°C for 2 minutes. The deprotection solution was drained from the resin using a vacuum manifold, and the resin was rinsed five times with DMF.

For coupling of an activated amino acid, we prepared a coupling solution of 100 mL NMP, 3.17 g HBTU (0.01mol, 0.1M) and 1.53 g HOBt (0.01 mol, 0.1M). The resulting solution was therefore 0.1M HBTU and 0.1M HOBt. The desired Fmoc-protected amino acid (125 μmol, 5 eq) was dissolved by vortexing in 1.25 mL coupling solution (125 μmol, 5 eq HBTU; 125 μmol, 5 eq HOBt). To the dissolved amino acid solution was added 44 μL DIEA (250μmol, 10eq). [Only 3 eq were used during the coupling of Fmoc-Asn(PEG)-OH monomer, and the required amounts of HBTU, HOBt, and DIEA were adjusted accordingly.] The resulting mixture was vortexed briefly and allowed to react for at least 1 min. The activated amino acid solution was then added to the resin, and the reaction vessel was placed in the microwave. The temperature was ramped from rt to 70°C over 2 minutes, and held at 70°C for 4 minutes. Following the coupling reaction, the activated amino acid solution was drained from the resin with a vaccum manifold, and the resin was subsequently rinsed five times with DMF. The cycles of deprotection and coupling were alternately repeated to give the desired full-length protein.

Acid-labile side-chain protecting groups were globally removed and proteins were cleaved from the resin by stirring the resin for ~4h in a solution of phenol (0.125 g), water (125 μL), thioanisole (125 μL), ethanedithiol 62.5 μL) and triisopropylsilane (25 μL) in trifluoroacetic acid (TFA, 2 mL). Following the cleavage reaction, the TFA solution was drained from the resin, the resin was rinsed with additional TFA. Proteins were precipitated from the concentrated TFA solution by addition of diethyl ether (~40 mL). Following centrifugation, the ether was decanted, and the pellet was dissolved in ~40mL 1:1 H_2O/MeCN, frozen and lyophilized to remove volatile impurities. The resulting powder was stored at -20°C until purification.

2.4.2 Purification and Characterization

Immediately prior to purification, the crude protein was dissolved in 1:1 H_2O/MeCN. Proteins were purified by preparative reverse-phase HPLC on a C18 column using a linear gradient of water in

acetonitrile with 0.1% v/v TFA. HPLC fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by electrospray ionization time of flight mass spectrometrey (ESI-TOF, spectra appear below in Figures 5-31), and purity was analyzed by Analytical HPLC (Figures 32-57)

2.4.3 ESI-TOF

ESI-TOF spectra for proteins **WW, 14, 14p, 16, 16p, 17, 17p, 18, 18p, 19, 19p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32,** and **32p** are shown in Figures 5-31

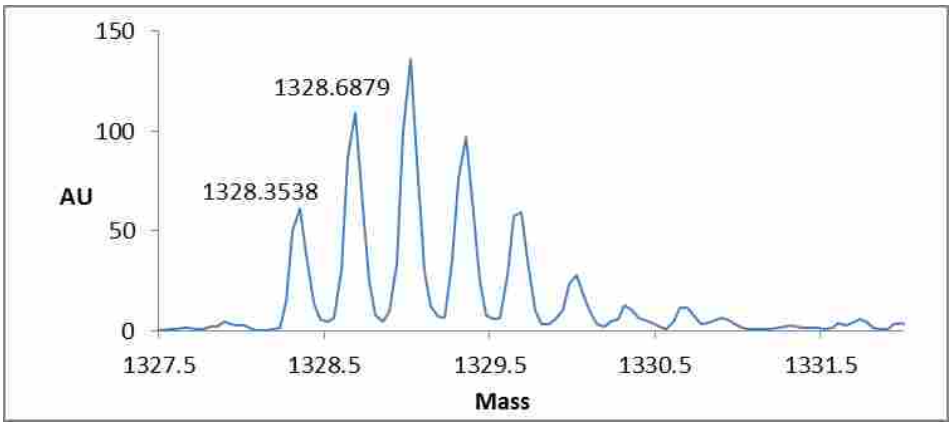


Figure 5: ESI-TOF spectrum for the **Pin WW** domain protein. Expected $[M+3H]^{3+}/3$ = 1328.3351 Da. Observed $[M+3H]^{3+}/3$ = 1328.3538 Da.

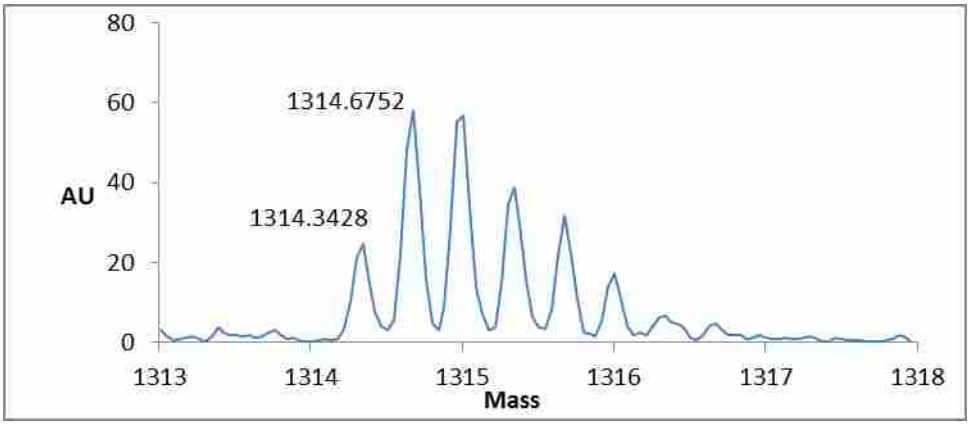


Figure 6: ESI-TOF spectrum for the Pin WW domain protein **14**. Expected $[M+3H]^{3+}/3$ = 1314.3157 Da. Observed $[M+3H]^{3+}/3$ = 1314.3428 Da.

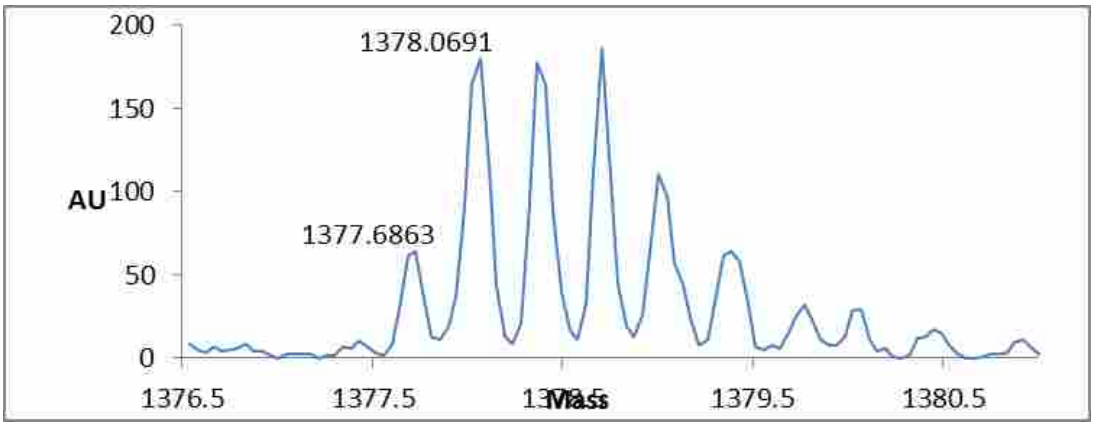


Figure 7: ESI-TOF spectrum for the Pin WW domain protein **14p**. Expected $[M+3H]^{3+}/3$ = 1377.6892 Da. Observed $[M+3H]^{3+}$= 1377.6863 Da.

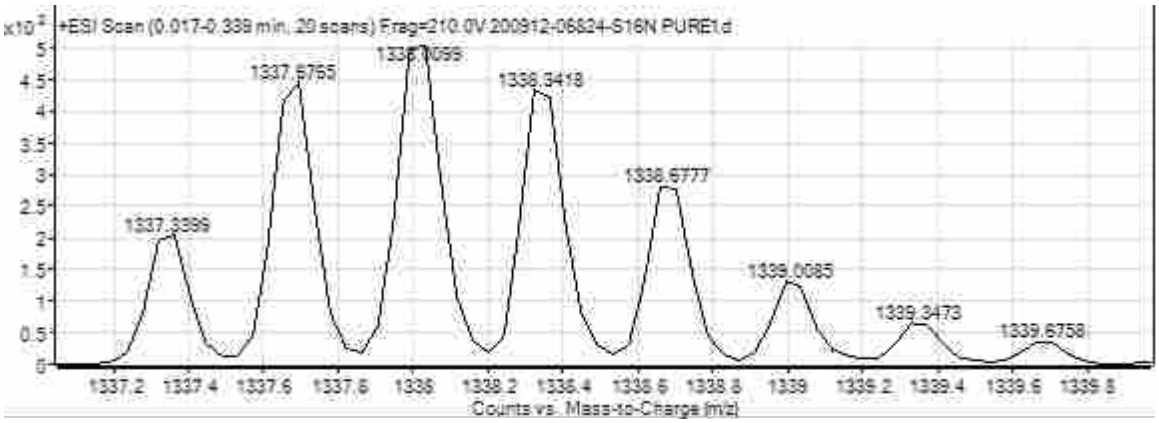


Figure 8: ESI-TOF spectrum for the Pin WW domain protein **16**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3389 Da.

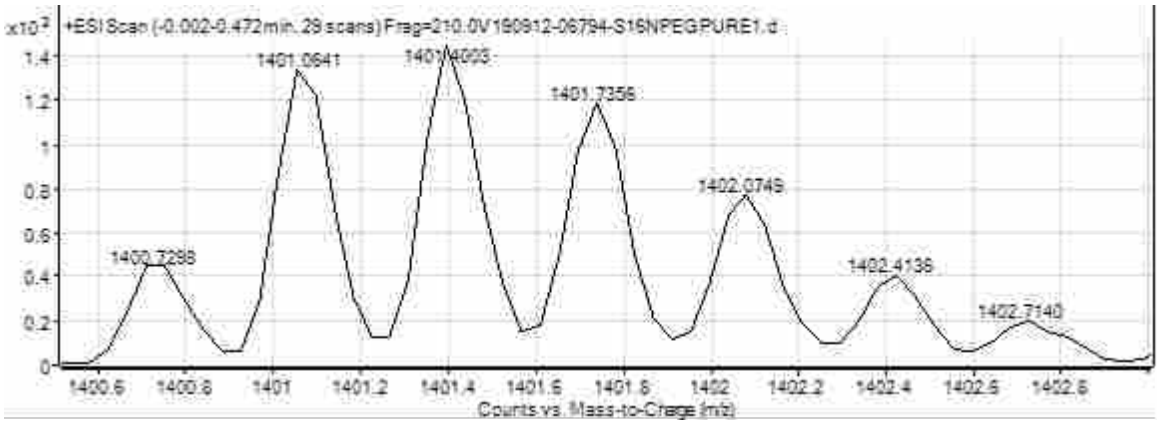


Figure 9: ESI-TOF spectrum for the Pin WW domain protein **16p**. Expected $[M+3H]^{3+}/3$ =1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.7298 Da.

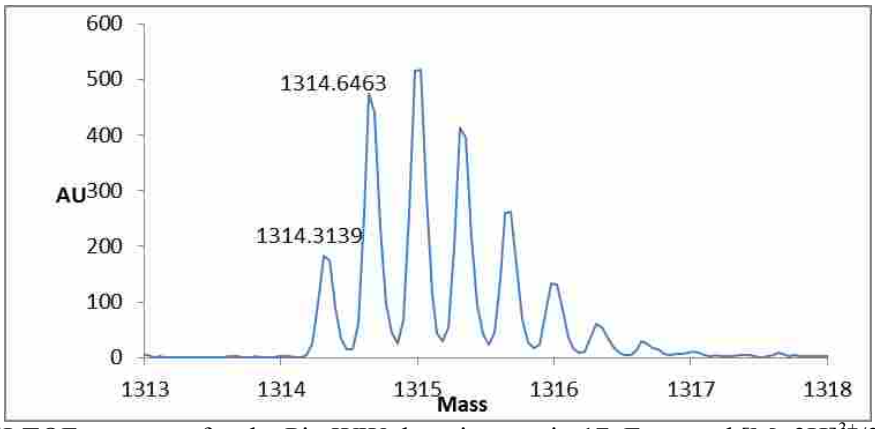


Figure 10: ESI-TOF spectrum for the Pin WW domain protein **17**. Expected $[M+3H]^{3+}/3$ = 1314.3157 Da. Observed $[M+3H]^{3+}/3$ = 1314.3139 Da

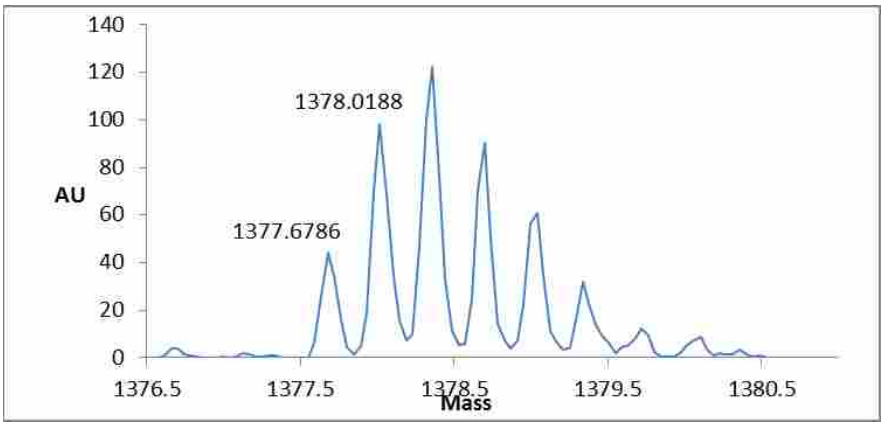


Figure 11: ESI-TOF spectrum for the Pin WW domain protein **17p**. Expected $[M+3H]^{3+}/3$ = 1377.6892 Da. Observed $[M+3H]^{3+}/3$ = 1377.6786 Da.

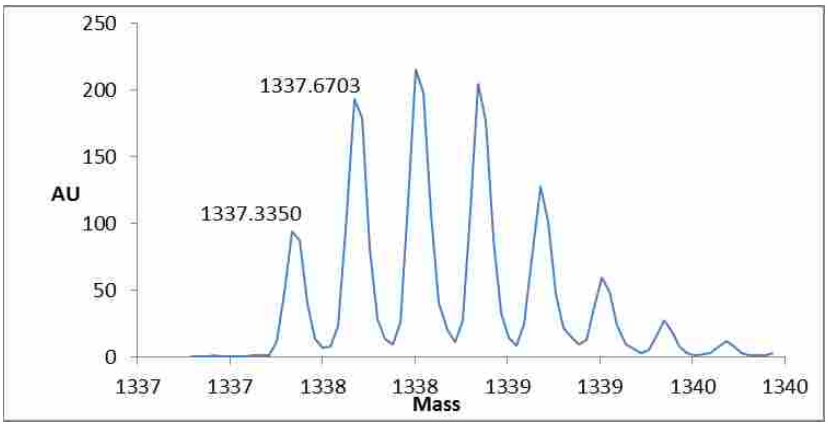


Figure 12: ESI-TOF spectrum for the Pin WW domain protein **18**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3350 Da.

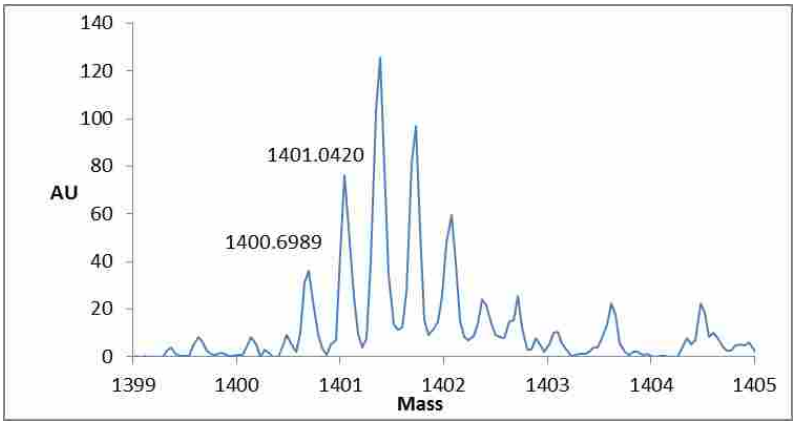


Figure 13: ESI-TOF spectrum for the Pin WW domain protein **18p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.6989 Da.

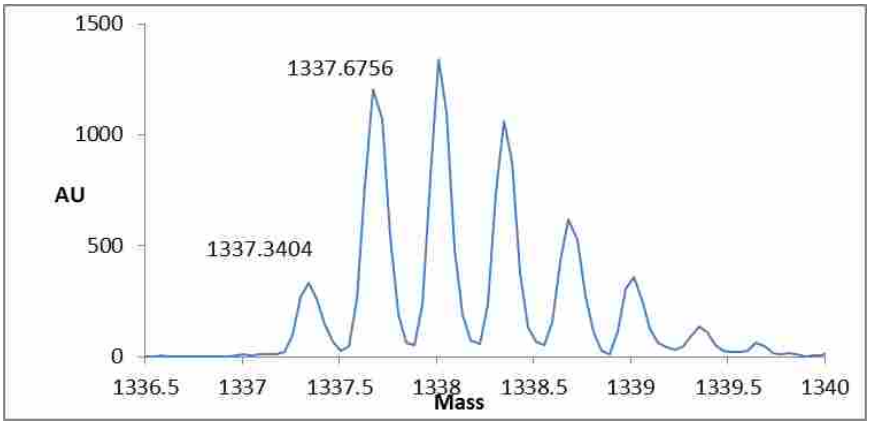


Figure 14: ESI-TOF spectrum for the Pin WW domain protein **19**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1377.3404Da.

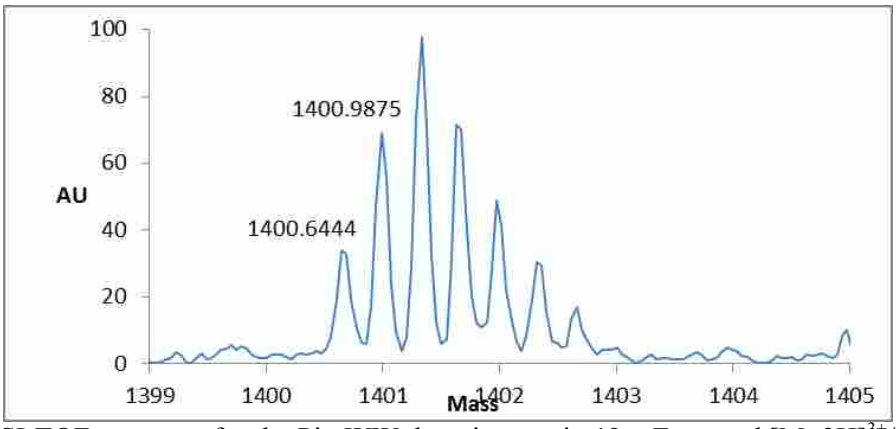


Figure 15: ESI-TOF spectrum for the Pin WW domain protein **19p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.6444 Da.

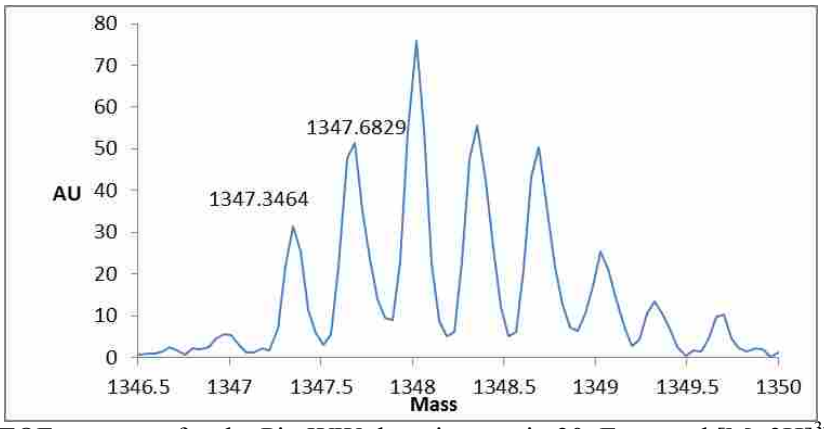


Figure 16: ESI-TOF spectrum for the Pin WW domain protein **20**. Expected $[M+3H]^{3+}/3$ = 1347.3423 Da. Observed $[M+3H]^{3+}/3$ = 1347.3464 Da.

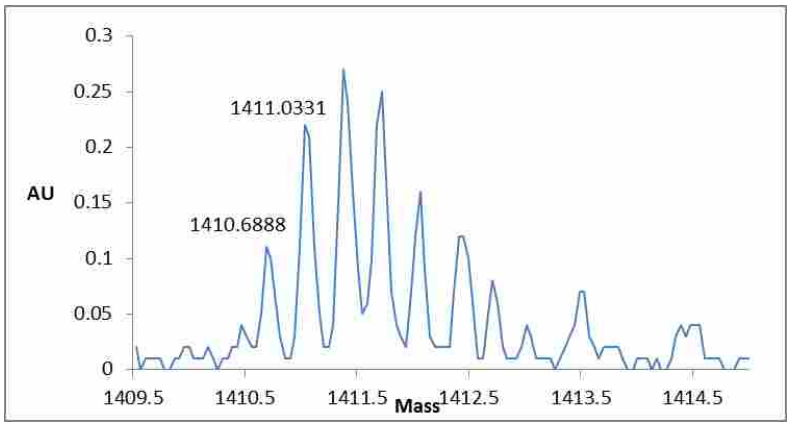


Figure 17: ESI-TOF spectrum for the Pin WW domain protein **20p**. Expected $[M+3H]^{3+}/3$ = 1410.7158 Da. Observed $[M+3H]^{3+}/3$ = 1410.6888 Da.

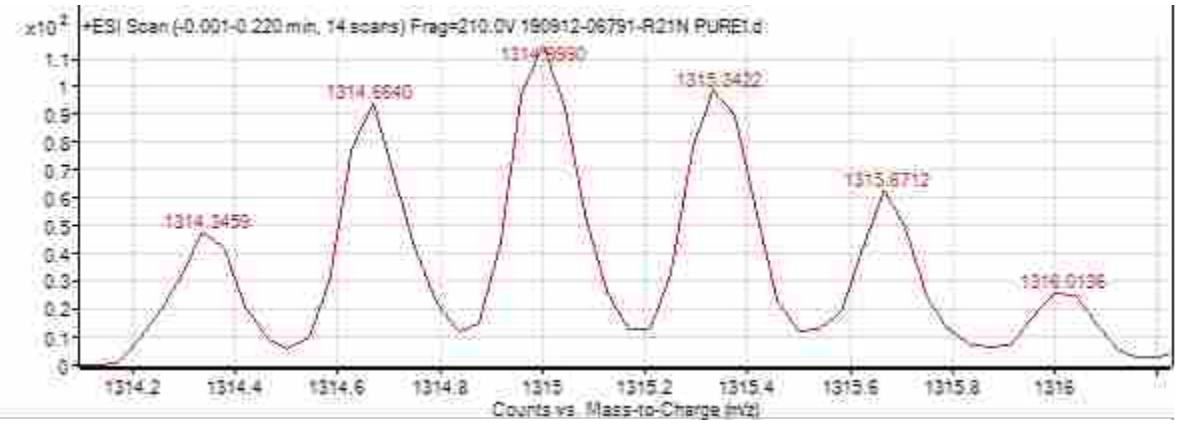


Figure 18: ESI-TOF spectrum for the Pin WW domain protein **21**. Expected $[M+3H]^{3+}/3$ = 1314.3157 Da. Observed $[M+3H]^{3+}/3$ = 1314.3459 Da.

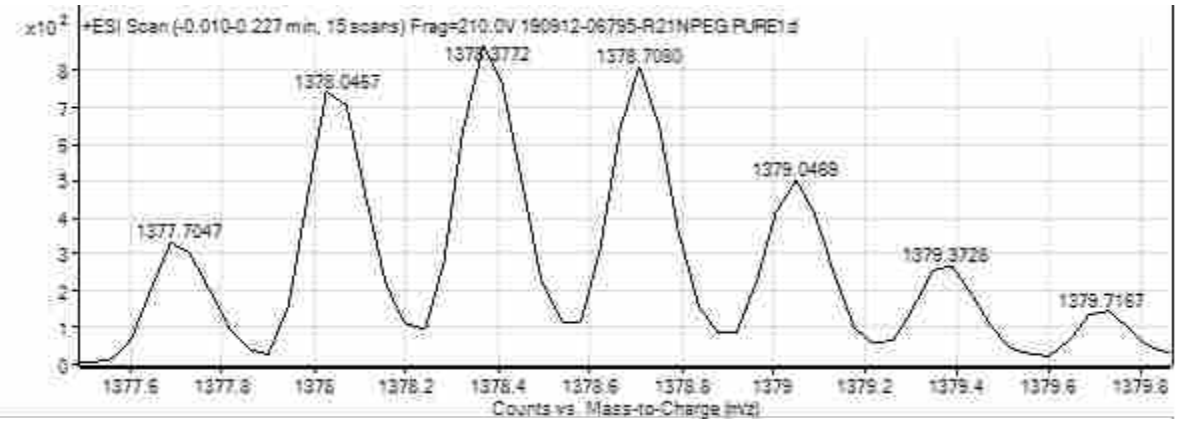


Figure 19: ESI-TOF spectrum for the Pin WW domain protein **21p**. Expected $[M+3H]^{3+}/3$ = 1377.6892 Da. Observed $[M+3H]^{3+}/3$ = 1377.7047 Da

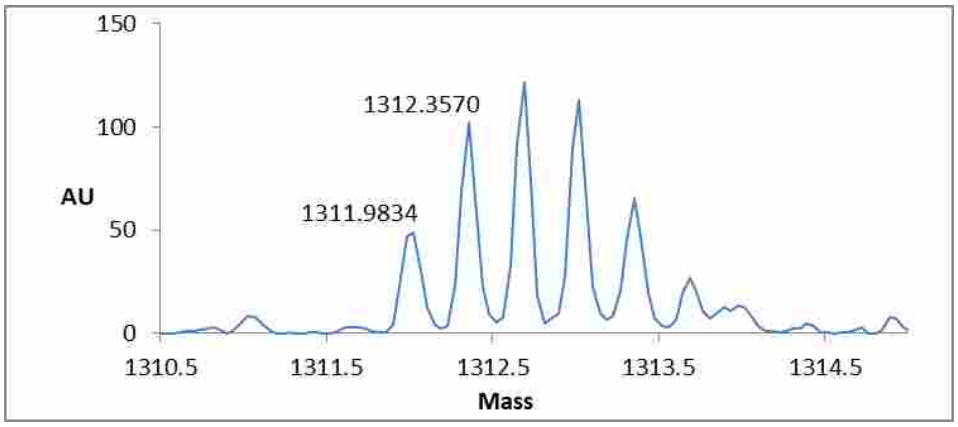


Figure 20: ESI-TOF spectrum for the Pin WW domain protein **23**. Expected $[M+3H]^{3+}/3$ = 1311.9950 Da. Observed $[M+3H]^{3+}/3$ = 1311.9834 Da.

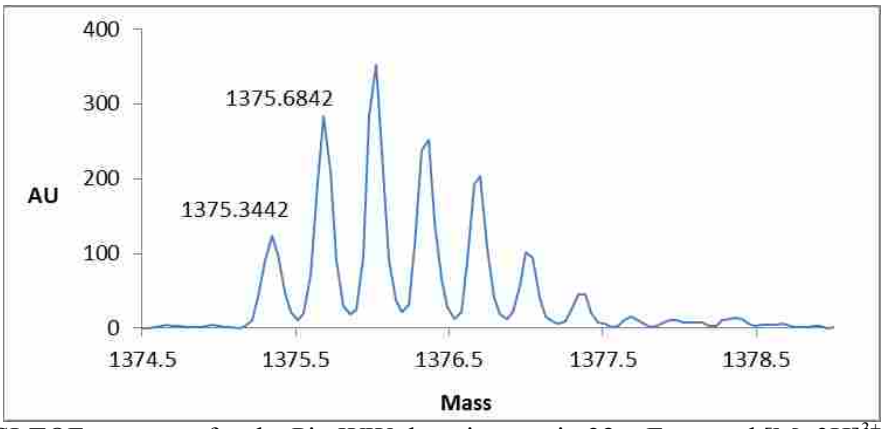


Figure 21: ESI-TOF spectrum for the Pin WW domain protein **23p**. Expected $[M+3H]^{3+}/3$ = 1375.3685 Da. Observed $[M+3H]^{3+}/3$ = 1375.3442 Da.

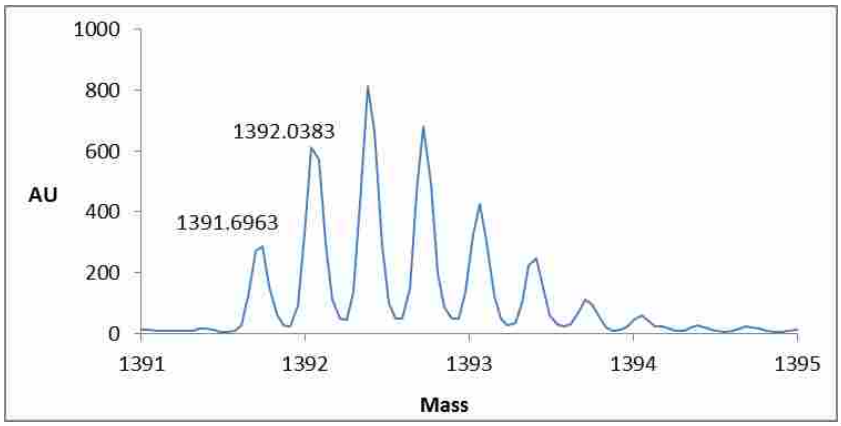


Figure 22: ESI-TOF spectrum for the Pin WW domain protein **26p**. Expected $[M+3H]^{3+}/3$ = 1391.7086 Da. Observed $[M+3H]^{3+}/3$ = 1391.6963 Da.

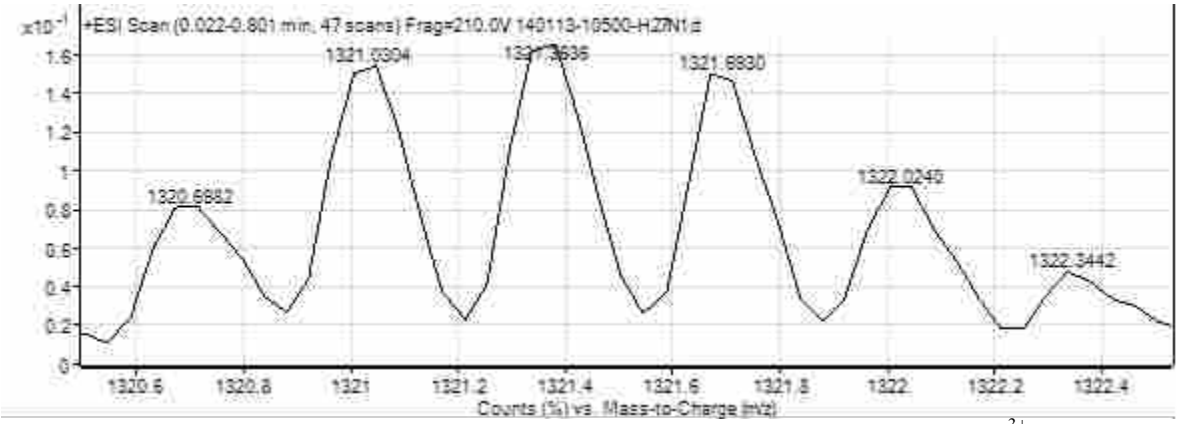


Figure 23: ESI-TOF spectrum for the Pin WW domain protein **27**. Expected $[M+3H]^{3+}/3$ = 1320.6631 Da. Observed $[M+3H]^{3+}/3$ = 1320.6982 Da.

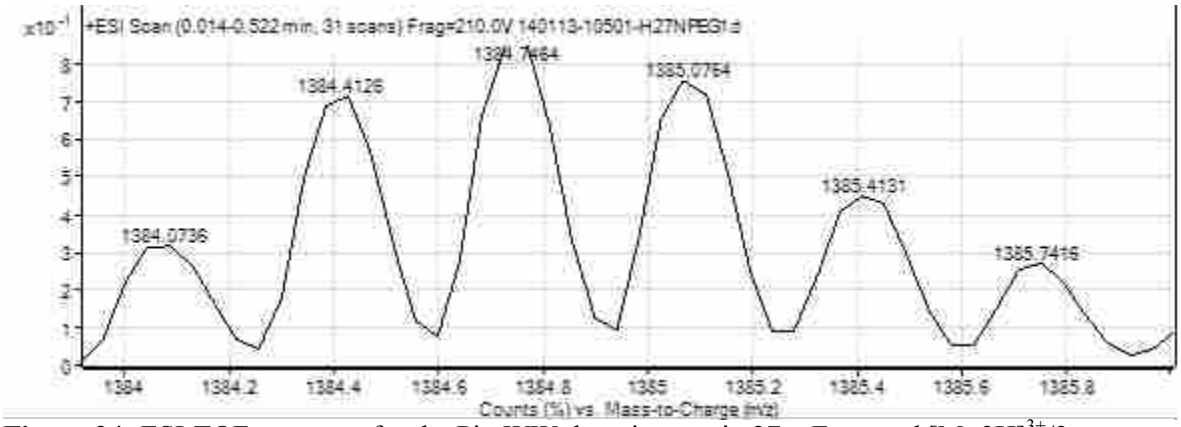


Figure 24: ESI-TOF spectrum for the Pin WW domain protein **27p**. Expected $[M+3H]^{3+}/3$ = 1384.0366 Da. Observed $[M+3H]^{3+}/3$ = 1384.0736 Da.

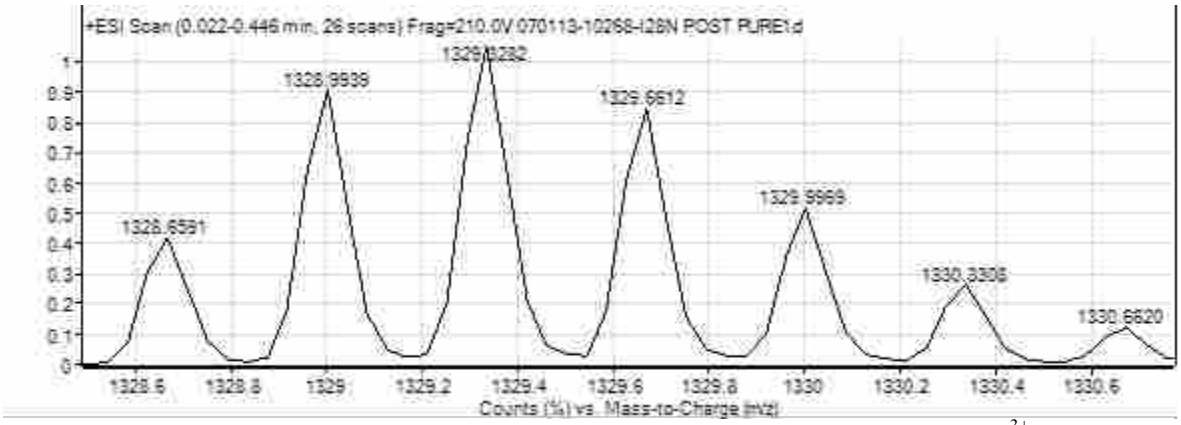


Figure 25: ESI-TOF spectrum for the Pin WW domain protein **28**. Expected $[M+3H]^{3+}/3$ = 1328.6547 Da. Observed $[M+3H]^{3+}/3$ = 1328.6591 Da.

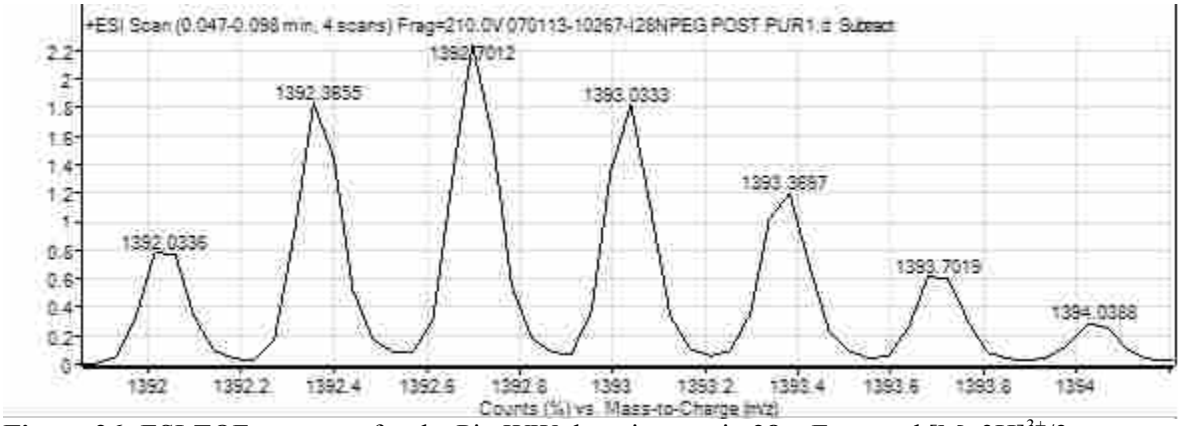


Figure 26: ESI-TOF spectrum for the Pin WW domain protein **28p**. Expected $[M+3H]^{3+}/3$ = 1392.0282 Da. Observed $[M+3H]^{3+}/3$ = 1392.0336 Da.

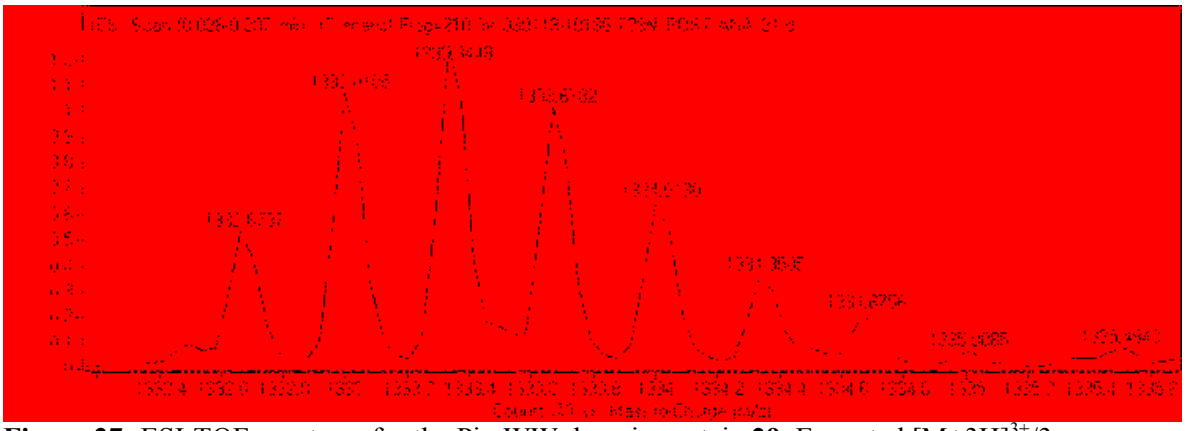


Figure 27: ESI-TOF spectrum for the Pin WW domain protein **29**. Expected $[M+3H]^{3+}/3$ = 1332.6668 Da. Observed $[M+3H]^{3+}/3$ = 1332.6737 Da.

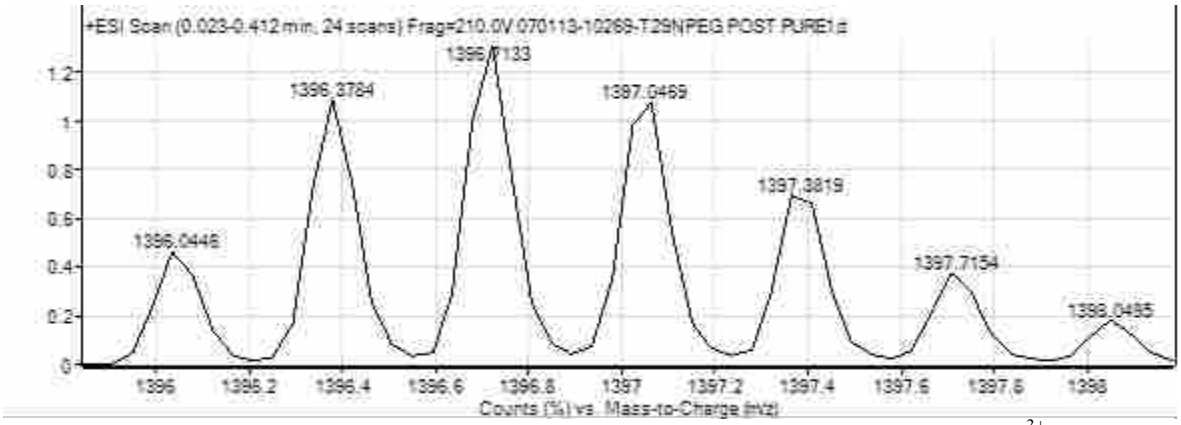


Figure 28: ESI-TOF spectrum for the Pin WW domain protein **29p**. Expected $[M+3H]^{3+}/3$ = 1396.0404 Da. Observed $[M+3H]^{3+}/3$ = 1396.0446 Da.

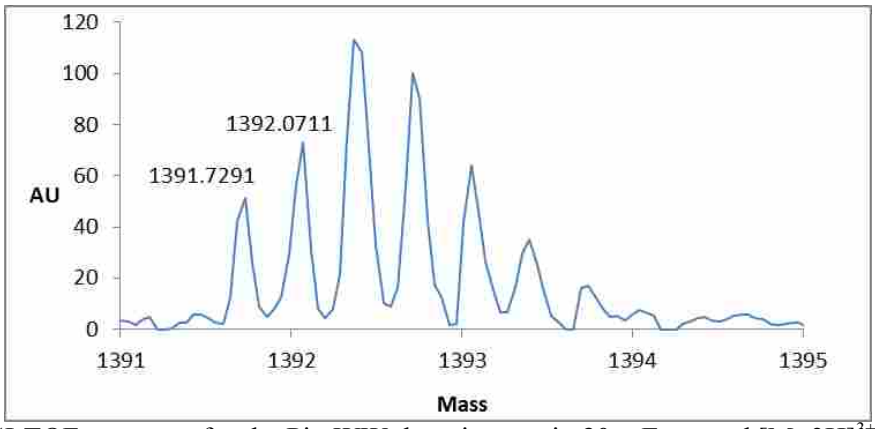


Figure 29: ESI-TOF spectrum for the Pin WW domain protein **30p**. Expected $[M+3H]^{3+}/3$ = 1391.7086 Da. Observed $[M+3H]^{3+}/3$ = 1391.7291 Da.

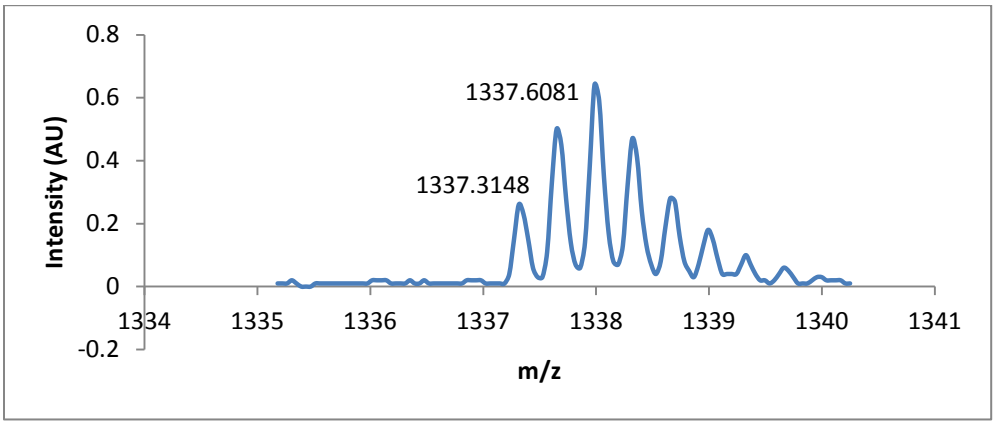


Figure 30: ESI-TOF spectrum for the Pin WW domain protein **32**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3148 Da.

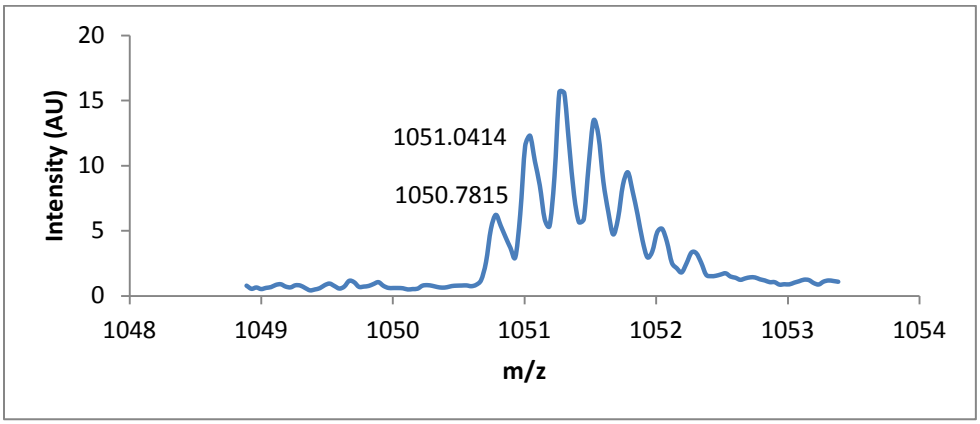


Figure 31: ESI-TOF spectrum for the Pin WW domain protein **32p**. Expected $[M+4H]4^{+}/4$ = 1050.7861 Da. Observed $[M+4H]4^{+}/4$ = 1387.0508 Da.

2.4.4 HPLC

HPLC traces for proteins **WW, 14, 14p, 16, 16p, 17, 17p, 18, 18p, 19, 19p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32, and 32p** are shown in figures 32-57

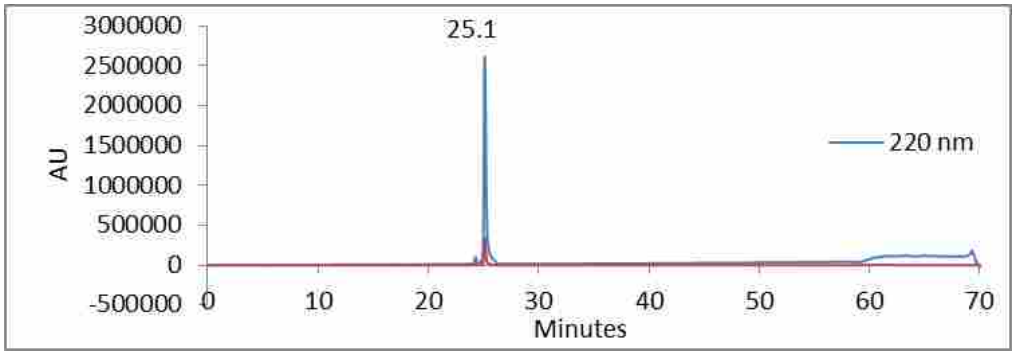


Figure 32: Analytical HPLC Data for **Pin WW**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

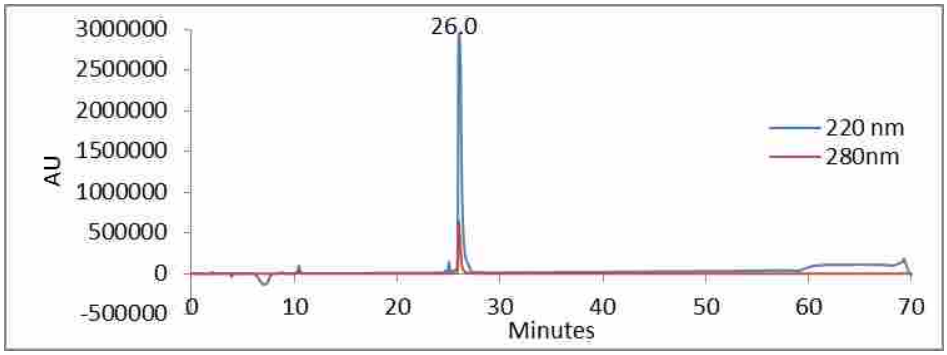


Figure 33: Analytical HPLC Data for Pin WW domain protein **14**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

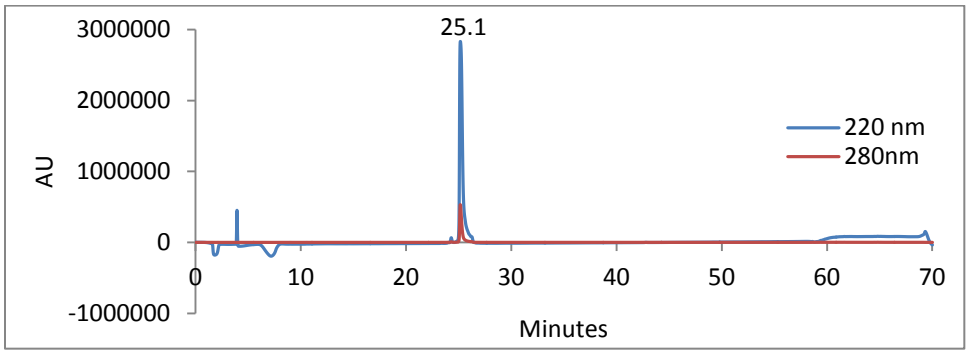


Figure 34: Analytical HPLC Data for Pin WW domain protein **14p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

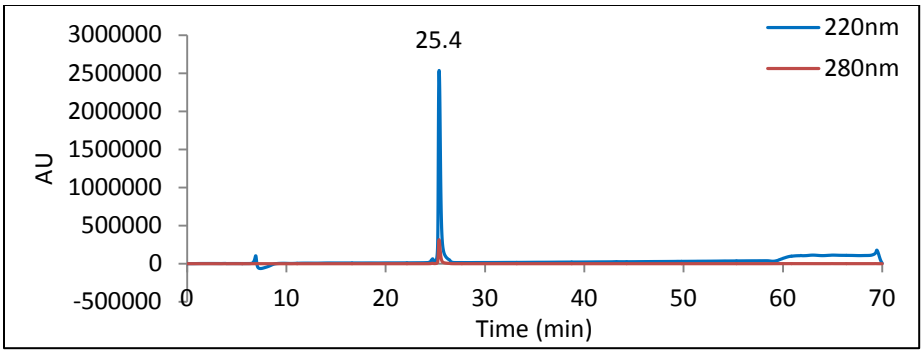


Figure 35: Analytical HPLC Data for Pin WW domain protein **16**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

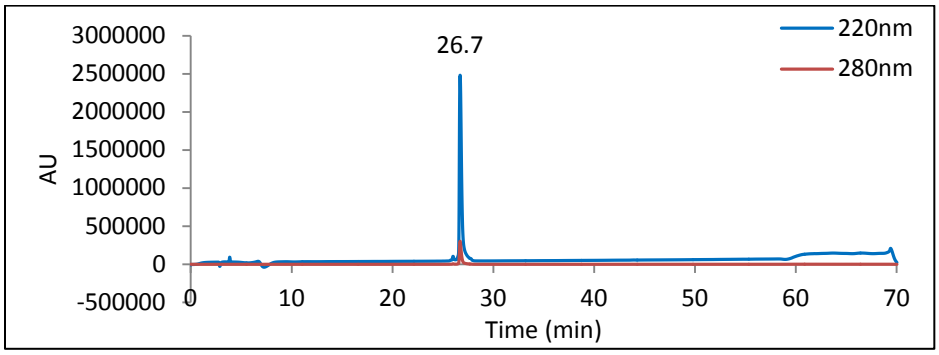


Figure 36: Analytical HPLC Data for Pin WW domain protein **16p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

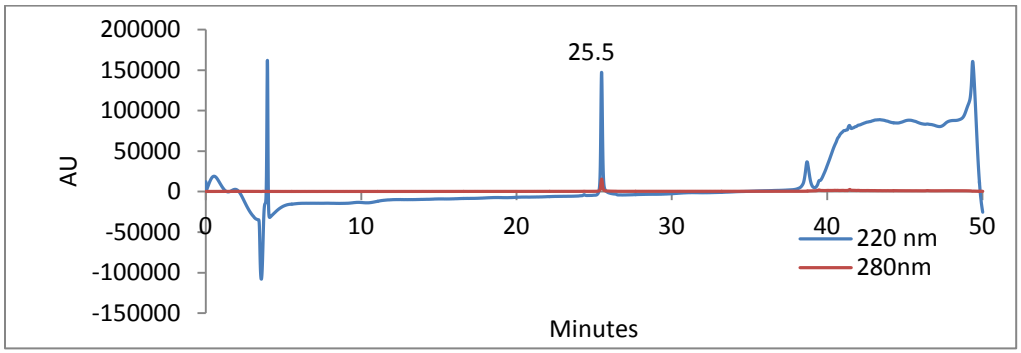


Figure 37: Analytical HPLC Data for Pin WW domain protein **17**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min

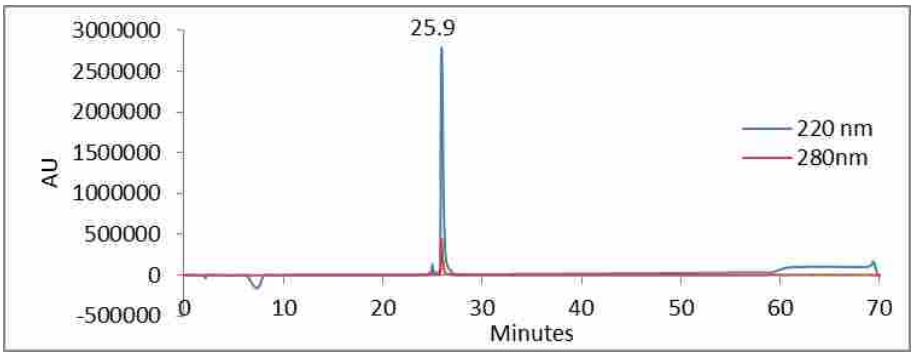


Figure 38: Analytical HPLC Data for Pin WW domain protein **17p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

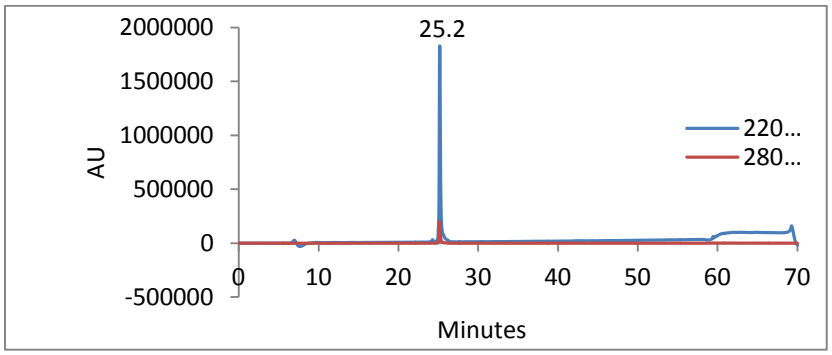


Figure 39: Analytical HPLC Data for Pin WW domain protein **18**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

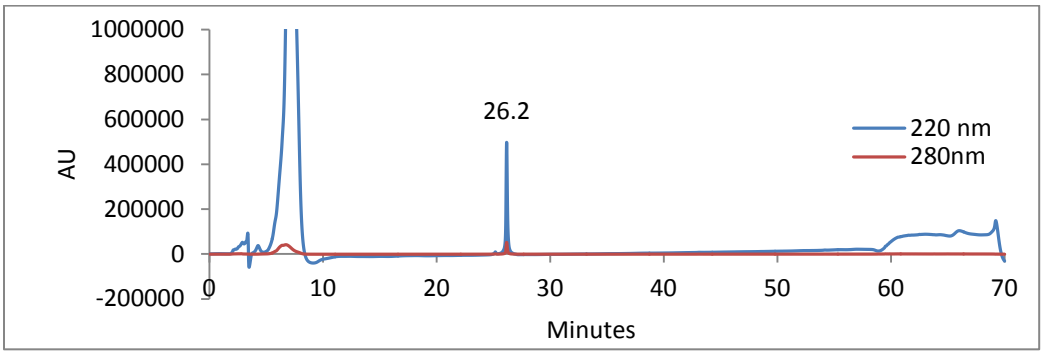


Figure 40: Analytical HPLC Data for Pin WW domain protein **18p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

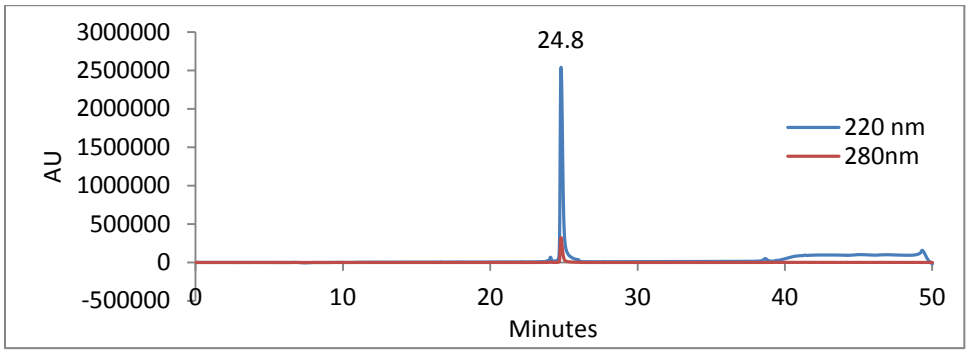


Figure 41: Analytical HPLC Data for Pin WW domain protein **19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

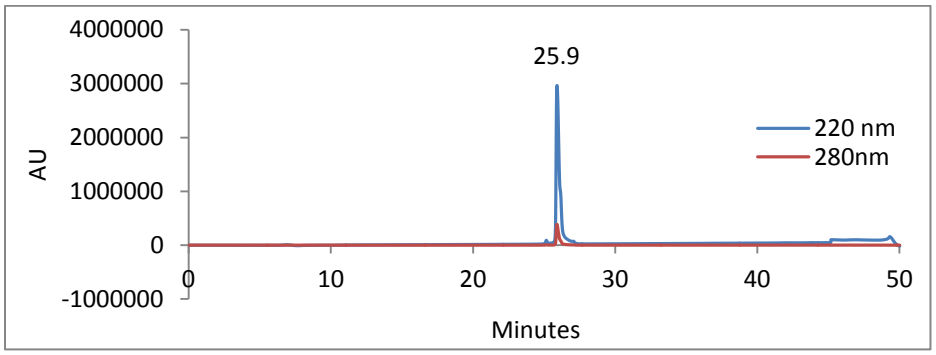


Figure 42: Analytical HPLC Data for Pin WW domain protein **19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

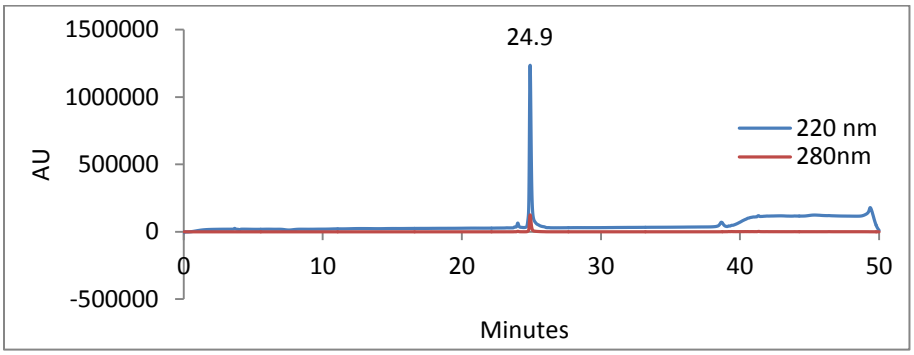


Figure 43: Analytical HPLC Data for Pin WW domain protein **20**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

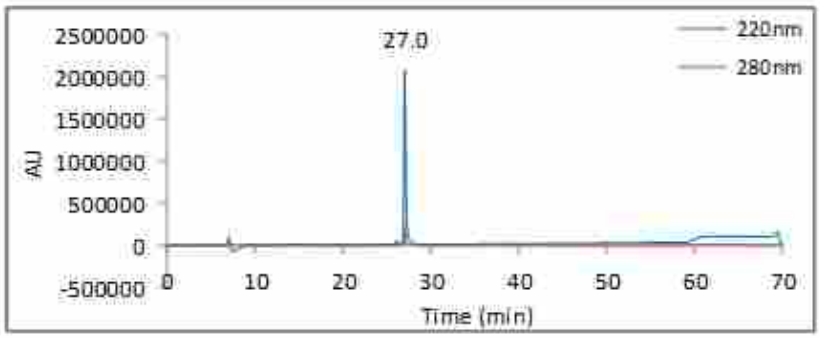


Figure 44: Analytical HPLC Data for Pin WW domain protein **21**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

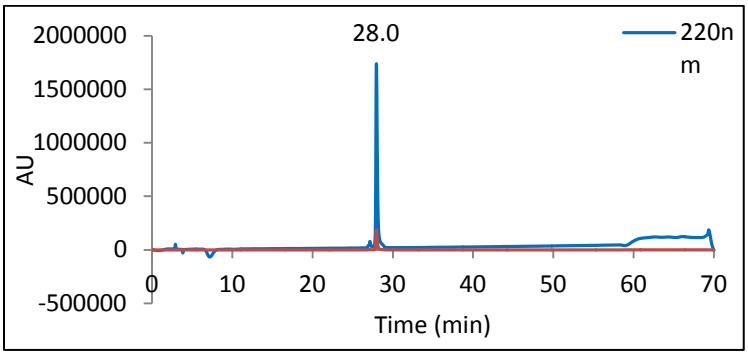


Figure 45: Analytical HPLC Data for Pin WW domain protein **21p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

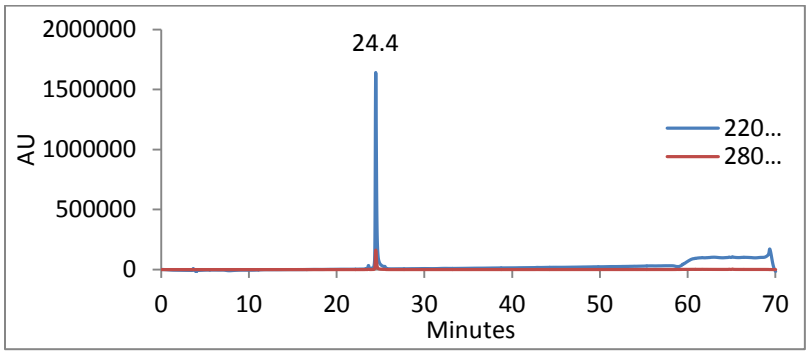


Figure 46: Analytical HPLC Data for Pin WW domain protein **23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

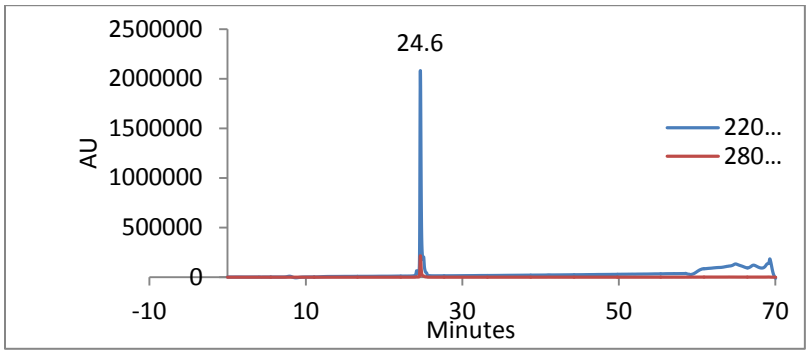


Figure 47: Analytical HPLC Data for Pin WW domain protein **23p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

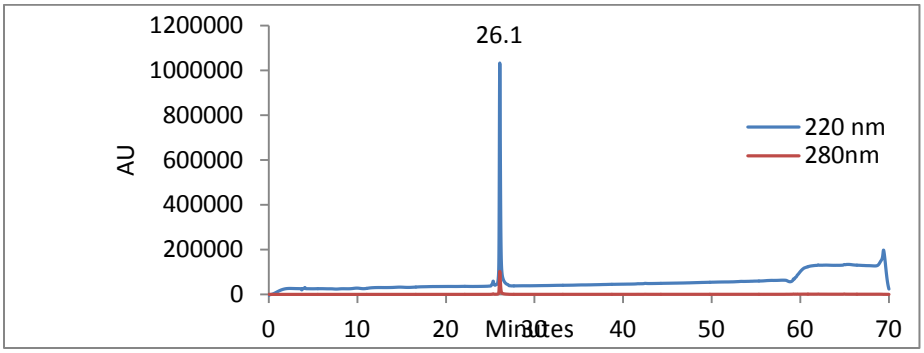


Figure 48: Analytical HPLC Data for Pin WW domain protein **26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

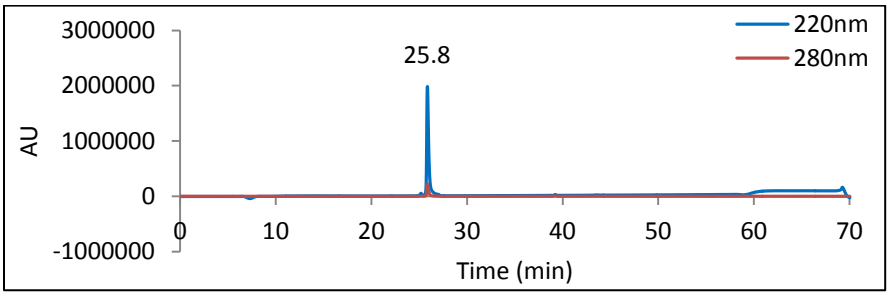


Figure 49: Analytical HPLC Data for Pin WW domain protein **27**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

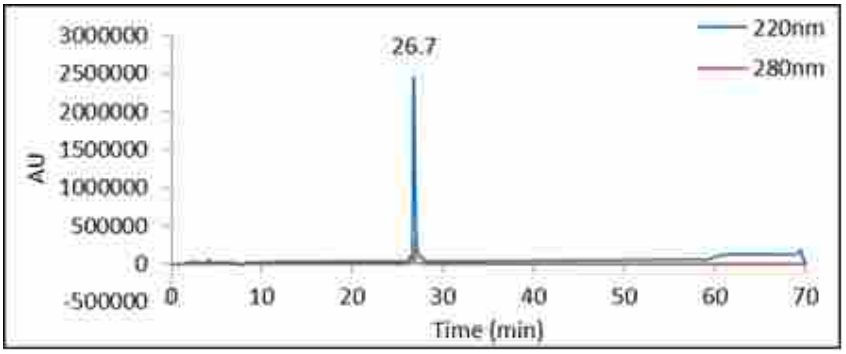


Figure 50: Analytical HPLC Data for Pin WW domain protein **27p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

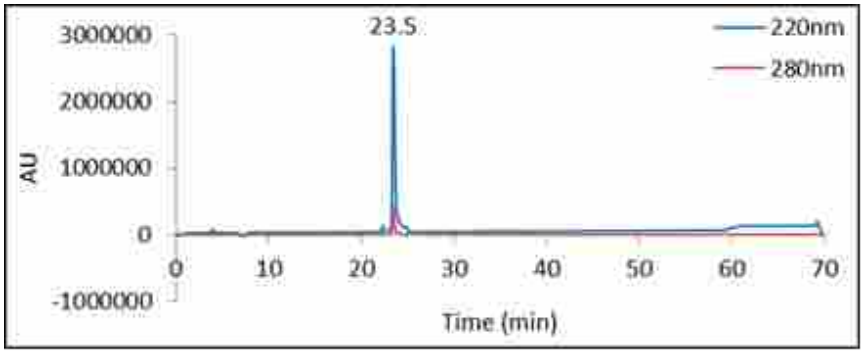


Figure 51: Analytical HPLC Data for Pin WW domain protein **28**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

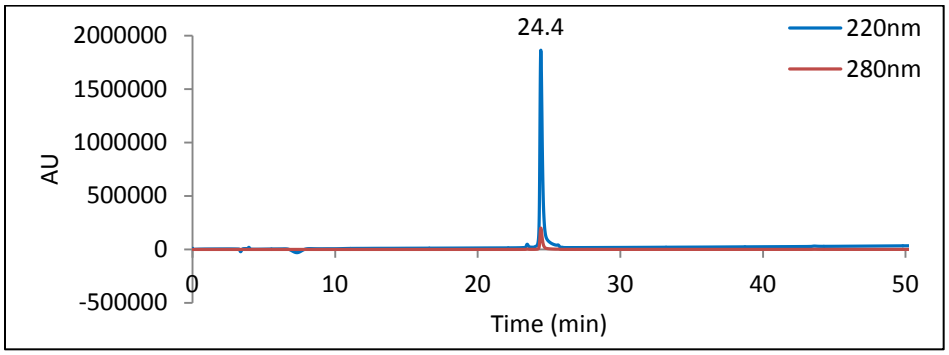


Figure 52: Analytical HPLC Data for Pin WW domain protein **28p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. Data collection was truncated after 50 minutes (when rinse began).

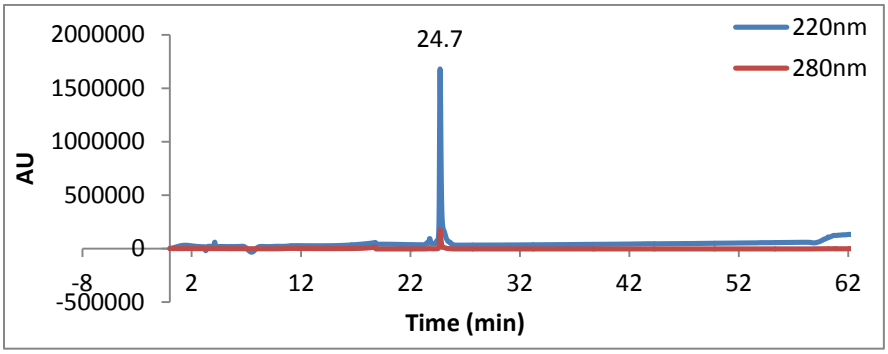


Figure 53: Analytical HPLC Data for Pin WW domain protein **29**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. Data collection was truncated after 62 minutes (during column re-equilibration).

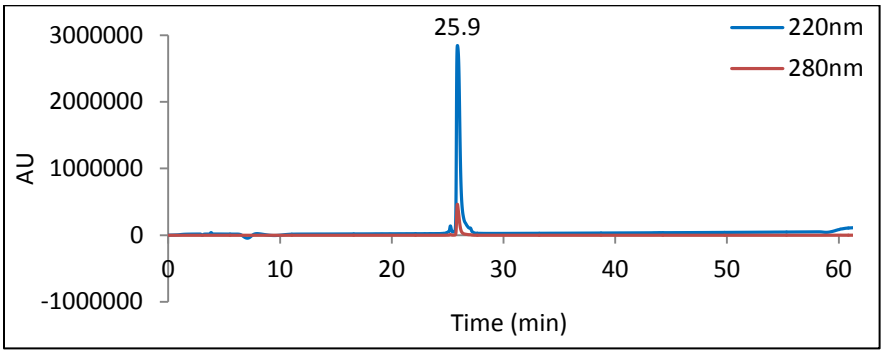


Figure 54: Analytical HPLC Data for Pin WW domain protein **29p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. Data collection was truncated after 61 minutes (during column re-equilibration).

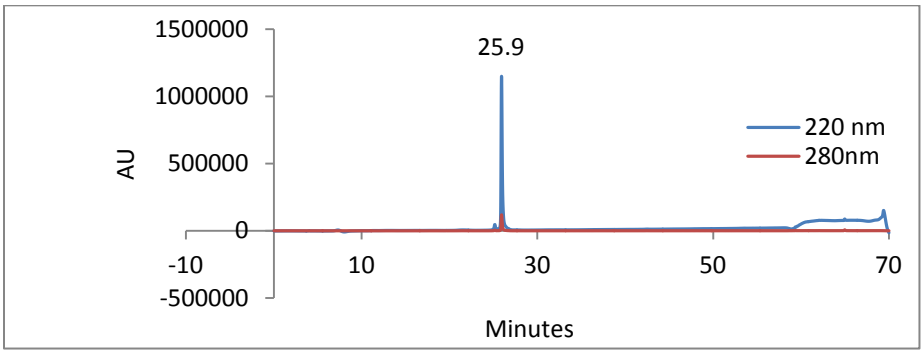


Figure 55: Analytical HPLC Data for Pin WW domain protein **30p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

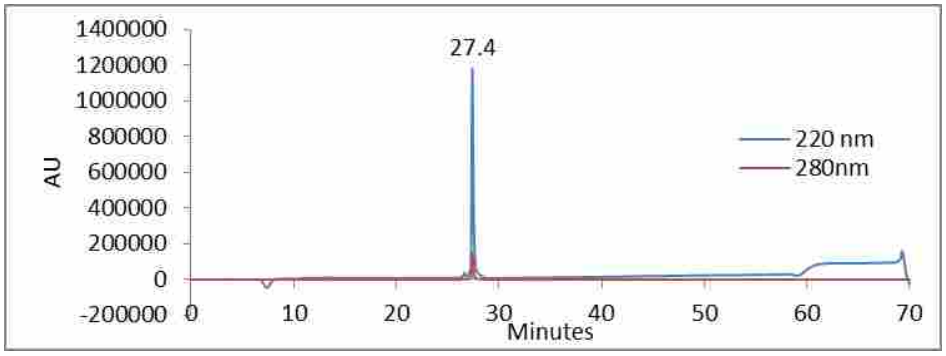


Figure 56: Analytical HPLC Data for Pin WW domain protein **32.** Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

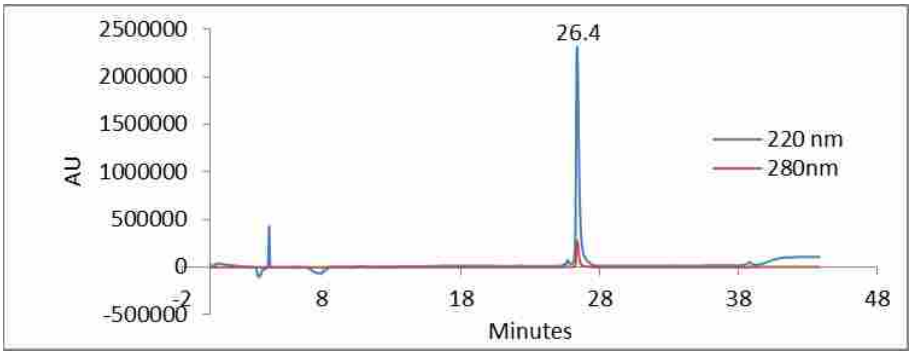


Figure 57: Analytical HPLC Data for Pin WW domain protein **32p.** Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-50% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 40 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. Analysis was truncated after 42 minutes (during column rinse)

2.4.5 Analysis of Thermal and Kinetic Parameters

2.4.6 Circular Dichroism Spectropolarimetry

Measurements were made with an Aviv 420 Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of 0.1 cm. Protein solutions were prepared in 20 mM sodium phosphate buffer, pH 7, and protein concentrations were determined spectroscopically based on tyrosine and tryptophan absorbance at 280 nm in 6 M guanidine hydrochloride + 20 mM sodium phosphate (εTrp = 5690 M-1cm-1, εTyr = 1280 M-1cm-1) [43]. CD spectra of 100 μM solutions were obtained from 340 to 200 nm at 25ºC. Variable temperature CD data were obtained at least in triplicate for 50 or 100 μM solutions of **WW, 14, 14p, 16, 16p, 17, 17p, 18, 18p, 19, 19p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32,** and **32p** in 20 mM sodium phosphate (pH 7) by monitoring molar ellipticity at 227 nm from 1 to 95°C at 2 °C intervals, with 120 s equilibration time between data points and 30 s averaging times. Variable temperature data for proteins **WW, 14, 14p, 16, 16p, 17, 17p, 18, 18p, 19, 19p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32,** and **32p** appear in Figures **58**-**69**.

2.4.7 Laser Temperature Jump Experiments

PEGylated peptides **14p, 17p, 18p,** and **19p**, as well as corresponding non-PEGylated peptides **14, 17, 18,** and **19** (50 μM in 20 mM sodium phosphate, pH 7) were also subjected to a rapid laser-induced temperature jump of ~10-11 °C using a nanosecond laser temperature jump apparatus as described previously,[44-47] at each of several temperatures. Following each temperature jump, the approach of the protein to equilibrium at the new temperature (i.e. relaxation) was monitored using the fluorescence decay of a Trp residue in the protein as a probe.

Each relaxation trace shown in Figures **64**-**69** represents the average of as many as 60 replicate temperature-jump experiments, and was obtained by fitting the shape f of each fluorescence decay at time t to a linear combination of the fluorescence decay shapes before f_1 and after f_2 the temperature jump:

$$f(t) = a_1(t) \cdot f_1 + a_2(t) \cdot f_2 \quad (1)$$

where $a_1(t)$ and $a_2(t)$ are the coefficients of the linear combination describing the relative contributions of f_1 and f_2 to the shape of the fluorescence decay at time t. The relaxation of the protein to equilibrium following the laser-induced temperature jump can then be represented as $\chi_1(t)$:

$$\chi_1(t) = \frac{a_1(t)}{a_1(t) + a_2(t)} \tag{2}$$

which is plotted as a function of time for each protein at several temperatures. Kinetic Data for proteins **14, 14p, 17, 17p, 18, 18p, 19,** and **19p** are presented alongside thermal Data in figures

2.4.8 Global Fitting of Variable Temperature CD Data

For **WW, 16, 16p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32,** and **32p** data from variable temperature CD were fit globally to the equations indicated below, to generate internally consistent temperature-dependent estimates of the folding free energy $\Delta G_f(T)$ and the folding activation energy $\Delta G^{\ddagger}_f(T)$

Data from the three (or more) replicate variable temperature CD experiments on each protein were fit to the following model for two-state thermally induced unfolding transitions:

$$\left[\theta\right] = \frac{(D_0 + D_1 \cdot T) + K_f(N_0 + N_1 \cdot T)}{1 + K_f} \tag{3}$$

where T is temperature in Kelvin, D_0 is the *y*-intercept and D_1 is the slope of the post-transition baseline; N_0 is the *y*-intercept and N_1 is the slope of the pre-transition baseline; and K_f is the temperature-dependent folding equilibrium constant. For a given protein, each replicate variable temperature CD experiment had distinct pre- and post-transition baselines (i.e., N_0, N_1, D_0, D_1).

K_f is related to the temperature-dependent free energy of folding $\Delta G_f(T)$ according to the following equation:

$$K_f = \exp\left[\frac{-\Delta G_f(T)}{RT}\right] \tag{4}$$

where R is the universal gas constant (0.0019872 kcal/mol/K). The midpoint of the thermal unfolding transition (or melting temperature T_m) was calculated by fitting $\Delta G_f(T)$ to the following equation:

$$\Delta G_f(T) = \Delta G_1(T-T_m) + \Delta G_2(T-T_m)^2 \quad (5)$$

where ΔG_1, ΔG_2, and T_m are parameters of the fit.

The parameters for equations 3-5 were used to calculate the values of the folding free energy ΔG_f for proteins at 333.15 K, that were used to generate Table 1. Figures **58**-**63** show the results of variable temperature CD and laser temperature jump experiments for proteins **WW, 16, 16p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32, and 32p** along with the parameters of equations 3-5 that were used to generate global fits for each compound. The standard error for each fitted parameter is also shown. These standard parameter errors were used to estimate the uncertainty in the average thermodynamic values given in the main text by propagation of error.

2.4.9 Global Fitting of Variable Temperature CD and Laser Temperature Jump Data

For peptides **14, 14p, 17, 17p, 18, 18p, 19,** and **19p**, data from variable temperature CD and laser temperature jump experiments were fit globally to the equations indicated below, to generate internally consistent temperature-dependent estimates of the folding free energy $DG_f(T)$ and the folding activation energy $DG^{\ddagger}_f(T)$

Data from the three (or more) replicate variable temperature CD experiments on each protein were fit to the following model for two-state thermally induced unfolding transitions according to equation 3, above. Each relaxation trace from the laser temperature jump experiments on a given protein was fit to the following equation:

$$x_1 = A_0 \cdot exp\left[\frac{k_f(1+k_f)}{K_f} \cdot t\right] + y_0, \quad (S6)$$

where t is time, A_0 is the intial value of c_1 at $t = 0$, y_0 is the value of c_1 at $t = \infty$, K_f is the temperature dependent folding equilibrium constant, and k_f is the folding rate. For a given protein, each

replicate variable temperature CD experiment had distinct pre- and post-transition baselines (i.e., N_0, N_1, D_0, D_1), and each relaxation trace has distinct A_0 and y_0 values, but K_f and k_f were constrained to be the same across all experiments for that protein.

K_f is related to the temperature-dependent free energy of folding $\Delta G_f(T)$ according to the following equation:

$$K_f = \exp\left[\frac{-\Delta G_f(T)}{RT}\right] \tag{4}$$

where R is the universal gas constant (0.0019872 kcal/mol/K). The midpoint of the thermal unfolding transition (or melting temperature T_m) was calculated by fitting $\Delta G_f(T)$ to the following equation:

$$\Delta G_f(T) = \Delta G_1(T-T_m) + \Delta G_2(T-T_m)^2 \tag{5}$$

where ΔG_1, ΔG_2, and T_m are parameters of the fit.

The folding rate k_f is related to the temperature-dependent folding activation energy $\Delta G^{\ddagger}_f(T)$ according to the following Kramers[48-50] model equation:

$$k_f(T) = \nu(59\ ^\circ C)\cdot\frac{\eta(59\ ^\circ C)}{\eta(T)}\exp\left[-\frac{\Delta G^{\dagger}_0 + \Delta G^{\dagger}_1\cdot(T-T_m) + \Delta G^{\dagger}_2\cdot(T-T_m)^2}{RT}\right], \tag{7}$$

in which $\Delta G^{\ddagger}_f(T)$ is represented as a second order Taylor series expansion about T_m, and $\Delta G^{\dagger}_0$, $\Delta G^{\dagger}_1$, $\Delta G^{\dagger}_2$, and T_m are parameters of the fit (T_m is constrained to be the same in equations 6 and 7). The pre-exponential term in equation S7 represents the viscosity-corrected frequency ν of the characteristic diffusional folding motion at the barrier[51,52] (at 59 °C, $\nu = 5 \times 10^5\ s^{-1}$).[53] η(59 °C) is the solvent viscosity at 59 °C and η(T) is the solvent viscosity at temperature T, both calculated with equation 8:

$$\eta(T) = A\cdot 10^{\frac{B}{T-C}}, \tag{8}$$

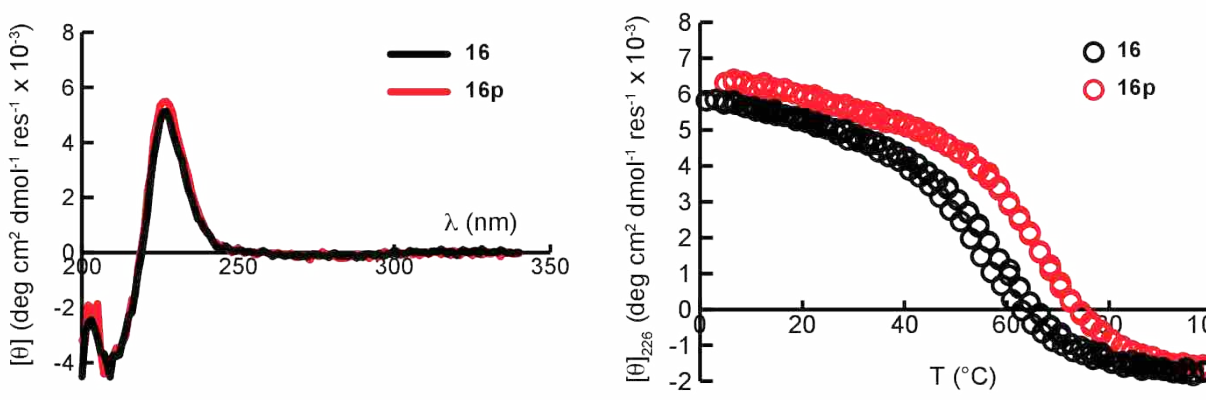
where $A = 2.41 \times 10^5$ Pa·s, B = 247.8 K, and C = 140 K.[54]

The parameters for equations 3-8 were used to calculate the values of the folding free energy ΔG_f, folding rate k_f, unfolding rate k_u ($k_u = k_f / K_f$), folding activation energy $\Delta G^{\ddagger}_f$, and unfolding activation

$\Delta G^{\ddagger}_{u}$ energy rates for proteins **14, 14p, 17, 17p, 18, 18p, 19,** and **19p** at 333.15 K, that were used to generate Table 1. The results of the laser-induced temperature jump experiments are shown alongside the thermal denaturation data (for proteins **14, 14p, 17, 17p, 18, 18p, 19,** and **19p)** in figures 64-69, along with the parameters of equations 3-8 that were used to generate the global fits for each compound. The standard error for each fitted parameter is also shown. These standard parameter errors were used to estimate the uncertainty in the average thermodynamic and kinetic values given in the main text by propagation of error.

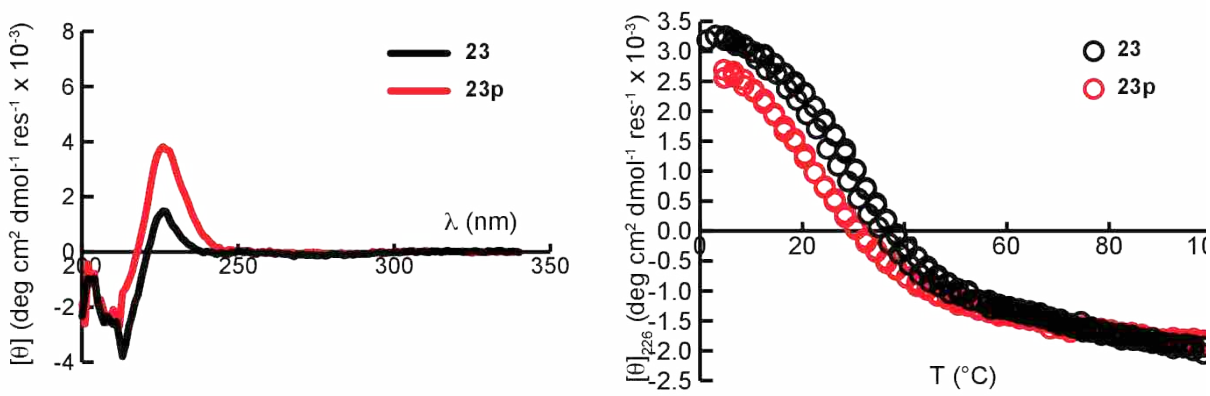
2.4.10 CD Spectra and Thermal Denaturation Plots

CD spectra and thermal denaturation plots for **WW, 16, 16p, 20, 20p, 21, 21p, 23, 23p, 26p, 27, 27p, 28, 28p, 29, 29p, 30p, 32,** and **32p,**are shown in figures **58-63**, along with fitting parameters for determining T_m and G_f.



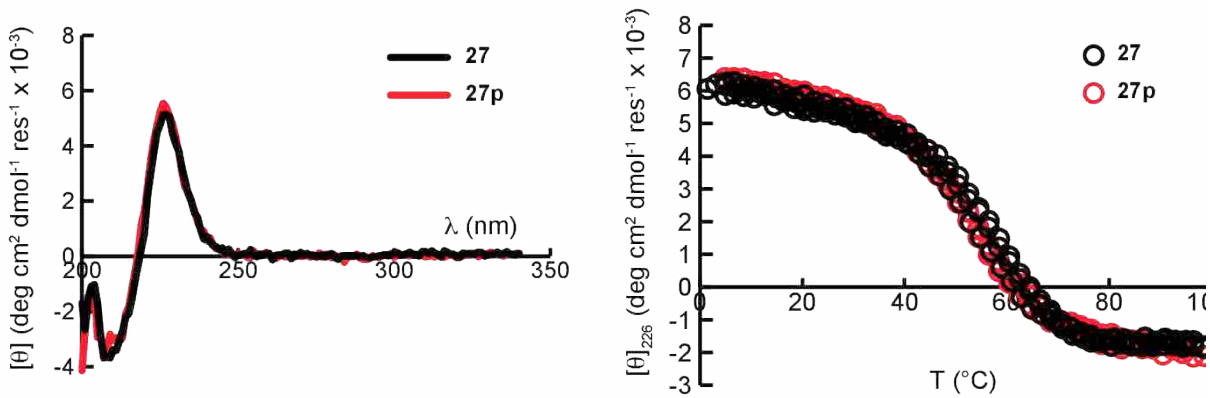
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
16	327.96 ± 0.23	0.0945 ± 0.0020	0.001016 ± 0.000155	0.99968	0.05479
16p	335.40 ± 0.17	0.1049 ± 0.0022	0.001266 ± 0.000059	0.99973	0.05156

Figure 58: CD spectra and variable temperature CD data for Pin WW domain protein **16** (which has Asn at position 16) and Pin WW domain protein **16p** (which has an Asn-linked poly(ethylene glycol) residue at position 16) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations 3-5) appear in the table, along with parameter standard errors.



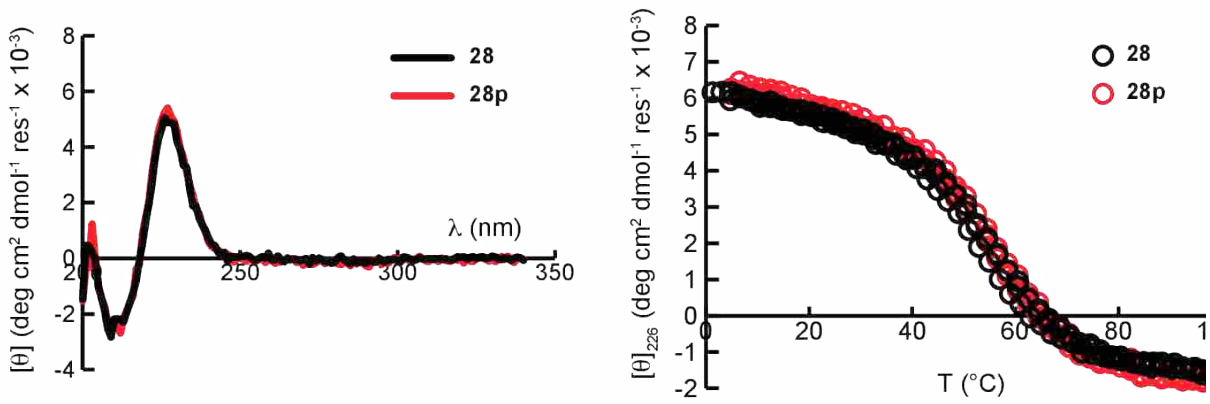
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
23	301.79 ± 0.43	0.0816 ± 0.0021	-0.000701 ± 0.000120	0.99955	0.04006
23p	296.33 ± 0.97	0.0824 ± 0.0033	-0.000899 ± 0.000138	0.99903	0.04725

Figure 59: CD spectra and variable temperature CD data for Pin WW domain protein **23** (which has Asn at position 23) and Pin WW domain protein **23p** (which has an Asn-linked poly(ethylene glycol) residue at position 23) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations 3-5) appear in the table, along with parameter standard errors.



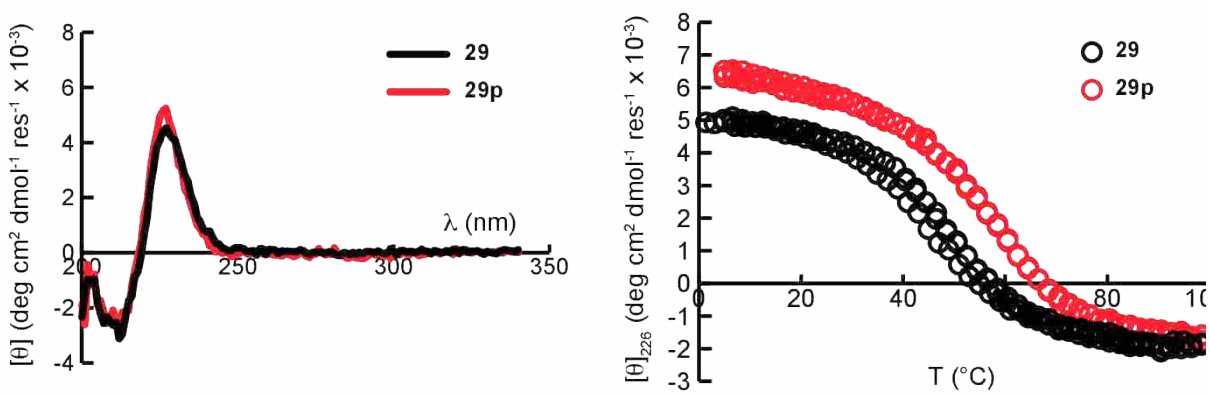
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
27	328.08 ± 0.18	0.0966 ± 0.0016	0.001026 ± 0.000133	0.99973	0.05384
27p	324.20 ± 0.37	0.0953 ± 0.0016	0.000162 ± 0.000212	0.99977	0.05220

Figure 60: CD spectra and variable temperature CD data for Pin WW domain protein **27** (which has Asn at position 27) and Pin WW domain protein **27p** (which has an Asn-linked poly(ethylene glycol) residue at position 27) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations 3-5) appear in the table, along with parameter standard errors.



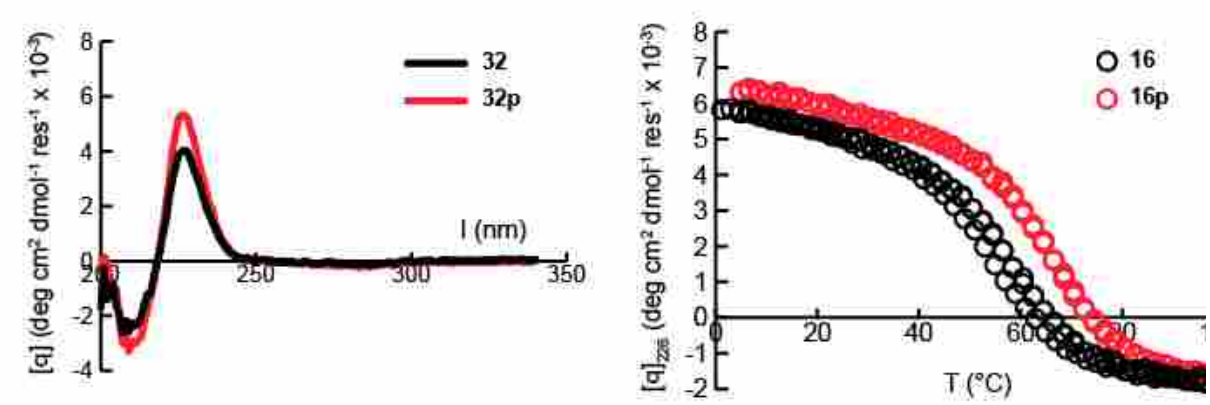
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd erro
28	326.37 ± 0.47	0.0933 ± 0.0016	0.000458 ± 0.000233	0.99977	0.04614
28p	326.39 ± 0.52	0.0915 ± 0.0017	0.000381 ± 0.000241	0.99973	0.05310

Figure 61: CD spectra and variable temperature CD data for Pin WW domain protein **28** (which has Asn at position 28) and Pin WW domain protein **28p** (which has an Asn-linked poly(ethylene glycol) residue at position 28) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations 3-5) appear in the table, along with parameter standard errors.



Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
29	321.53 ± 0.35	0.0810 ± 0.0025	-0.001163 ± 0.000085	0.99960	0.05743
29p	326.63 ± 0.17	0.0840 ± 0.0014	0.000941 ± 0.000088	0.99982	0.04300

Figure 62: CD spectra and variable temperature CD data for Pin WW domain protein **29** (which has Asn at position 29) and Pin WW domain protein **29p** (which has an Asn-linked poly(ethylene glycol) residue at position 29) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations 3-5) appear in the table, along with parameter standard errors.

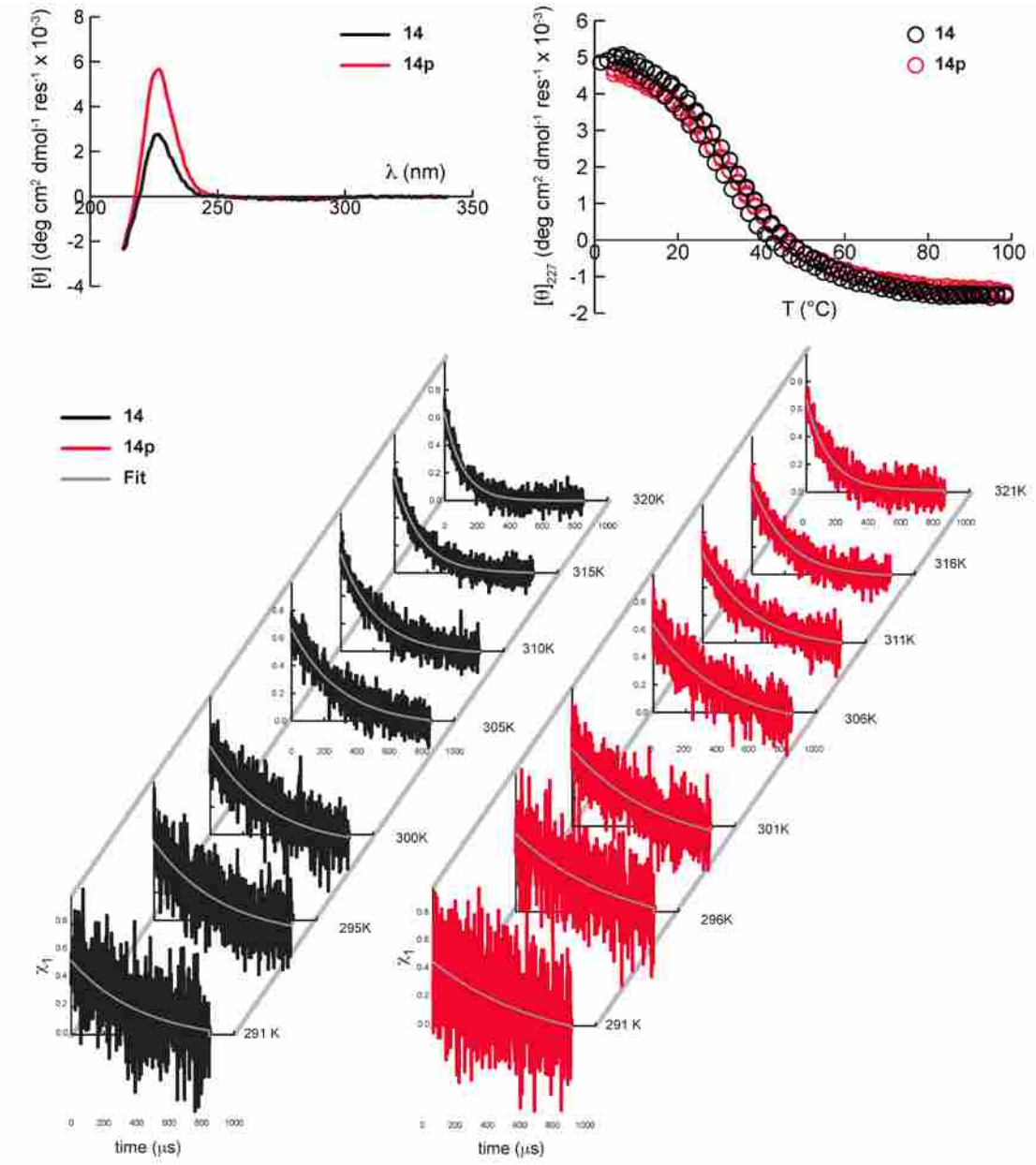


Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd err
32	318.14 ± 0.22	0.084 ± 0.0015	0.0010 ± 0.000102	0.99991	0.02397
32p	323.48 ± 0.23	0.0897 ± 0.0009	0.0006 ± 0.000099	0.9999606	0.02135

Figure 63: CD spectra and variable temperature CD data for Pin WW domain protein **32** (which has Asn at position 32) and Pin WW domain protein **32p** (which has an Asn-linked poly(ethylene glycol) residue at position 33) in 20 mM sodium phosphate, pH 7. Fit parameters (obtained by fitting the variable temperature CD data to equations 3-5) appear in the table, along with parameter standard errors.

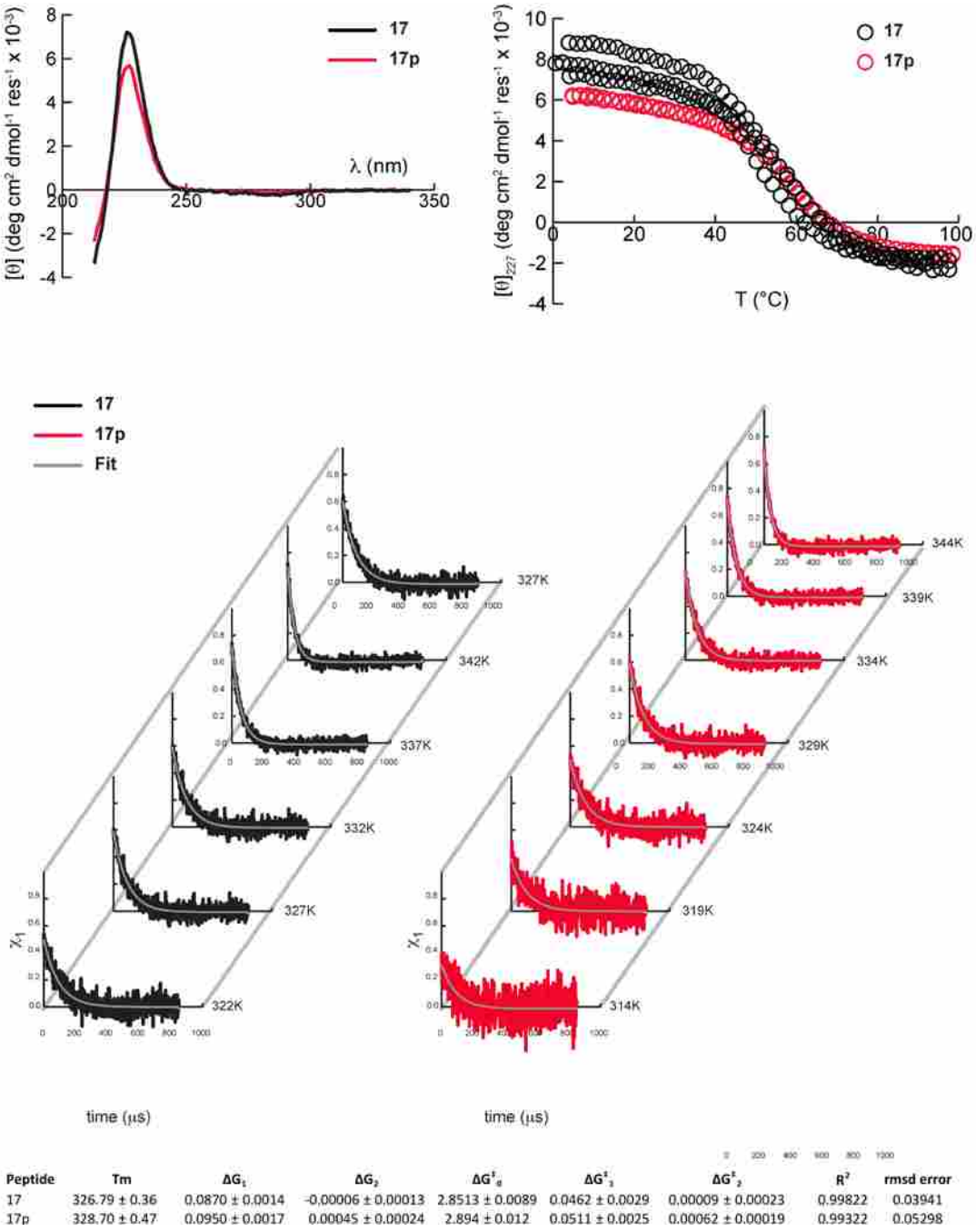
2.4.11 CD Spectra, Thermal Denaturation, and Laser-Induced Temperature Jump Plots

CD spectra, thermal denaturation, and laser-induced temperature plots for peptides **14, 14p, 17, 17p, 18, 18p, 19,** and **19p** are shown in Figures 64-69, along with fitting parameters for determining T_m, ΔG_f, k_f, and k_u.



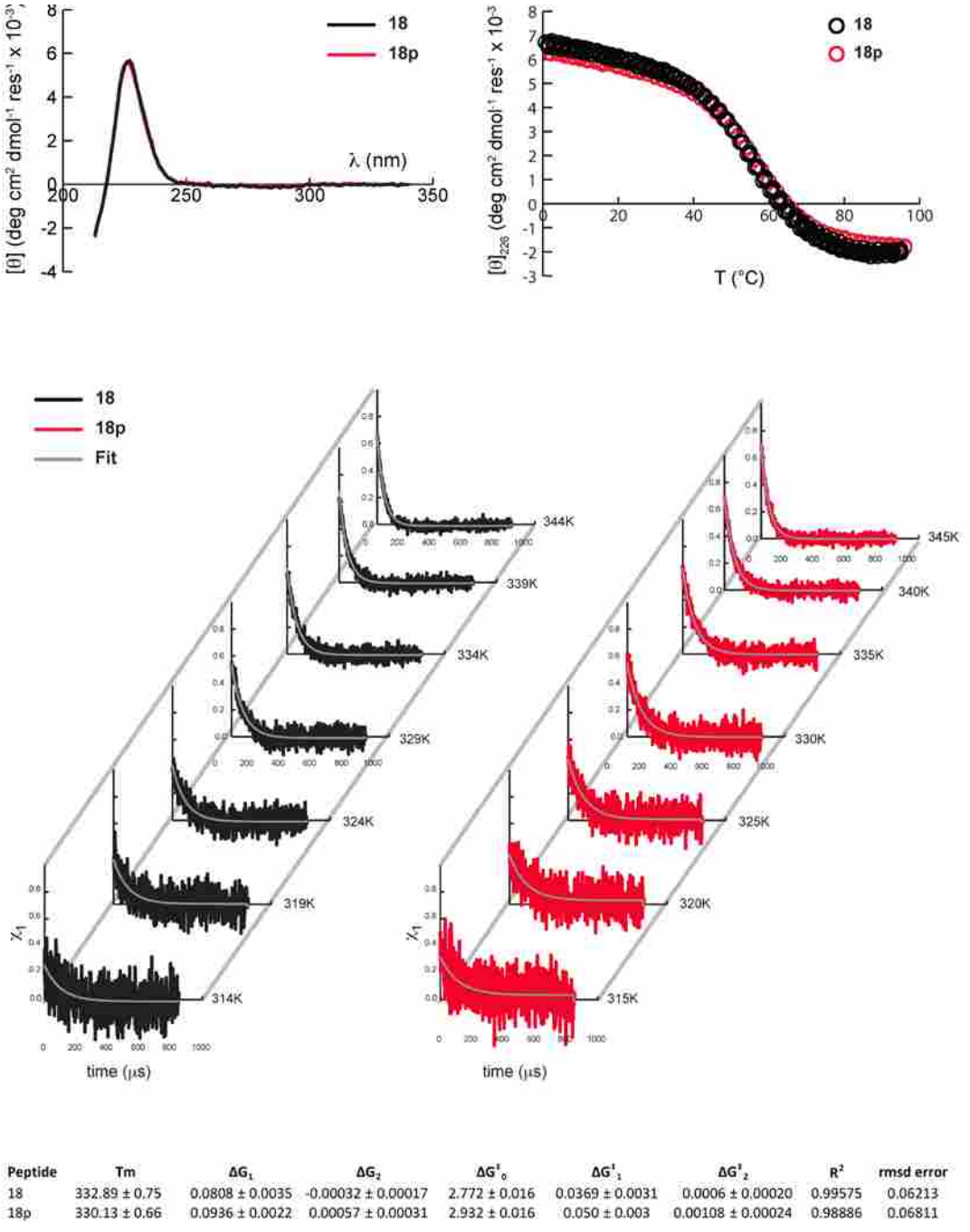
Peptide	Tm	ΔG_1	ΔG_2	ΔG^1_0	ΔG^1_1	ΔG^1_2	R^2	rmsd error
14	306.87 ± 0.85	0.0703 ± 0.0049	-0.00082 ± 0.00018	3.151 ± 0.026	0.023 ± 0.0032	-0.00037 ± 0.00025	0.92971	0.11793
14p	307.3 ± 3.2	0.0691 ± 0.0077	-0.00057 ± 0.00073	3.389 ± 0.057	0.021 ± 0.012	-0.00036 ± 0.00052	0.88253	0.14272

Figure 64: CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for solutions of proteins **14** and **14p (**in which position 14 was replaced by Asn and Asn-PEG, respectively) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data to equations 3-8, using the indicated parameters.



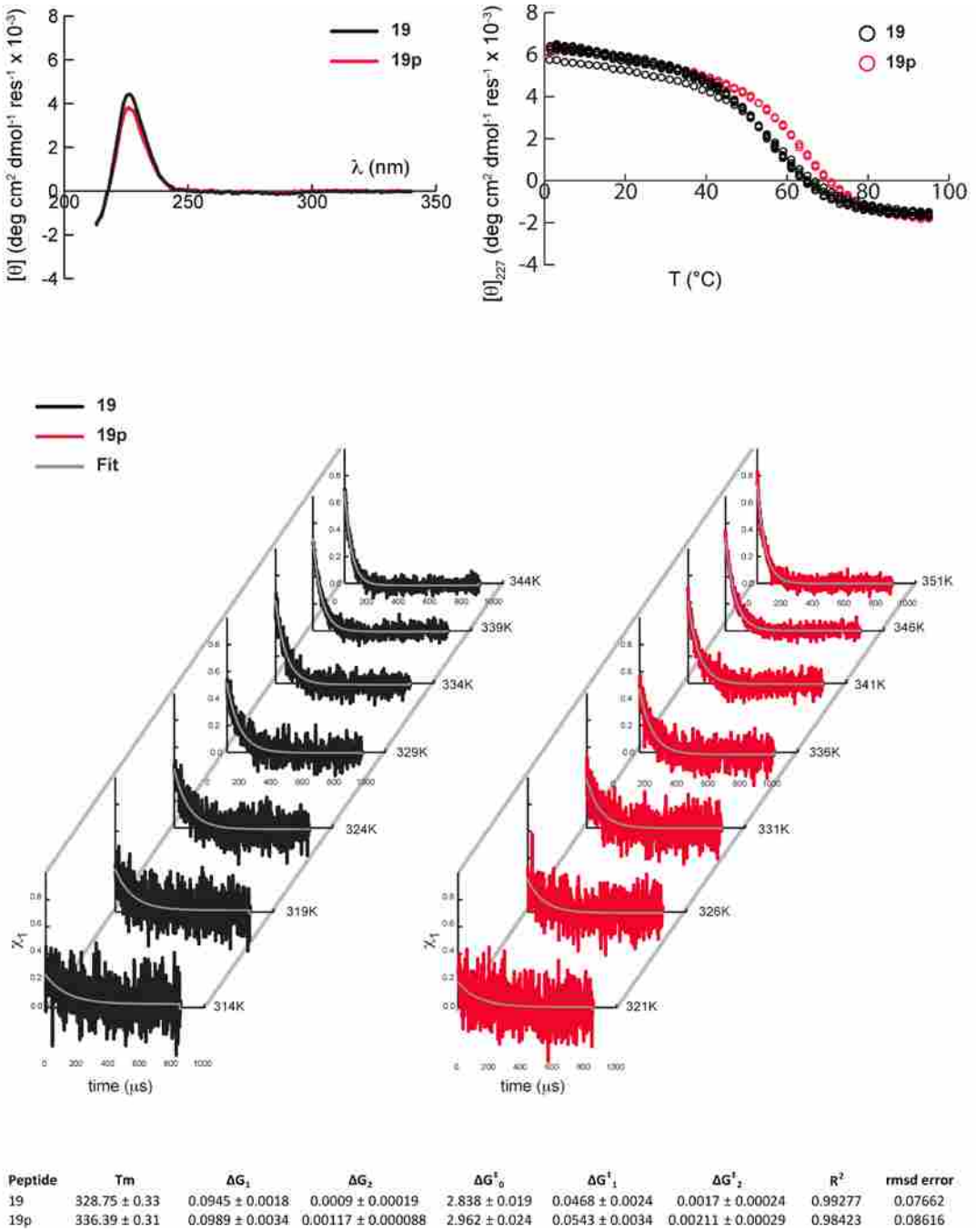
Peptide	Tm	ΔG_1	ΔG_2	$\Delta G^{\ddagger}_{d}$	$\Delta G^{\ddagger}_{3}$	$\Delta G^{\ddagger}_{2}$	R^2	rmsd error
17	326.79 ± 0.36	0.0870 ± 0.0014	-0.00006 ± 0.00013	2.8513 ± 0.0089	0.0462 ± 0.0029	0.00009 ± 0.00023	0.99822	0.03941
17p	328.70 ± 0.47	0.0950 ± 0.0017	0.00045 ± 0.00024	2.894 ± 0.012	0.0511 ± 0.0025	0.00062 ± 0.00019	0.99322	0.05298

Figure 65: CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for solutions of proteins **17** and **17p (**in which position 17 was replaced by Asn and Asn-PEG, respectively) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data to equations 3-8, using the indicated parameters.



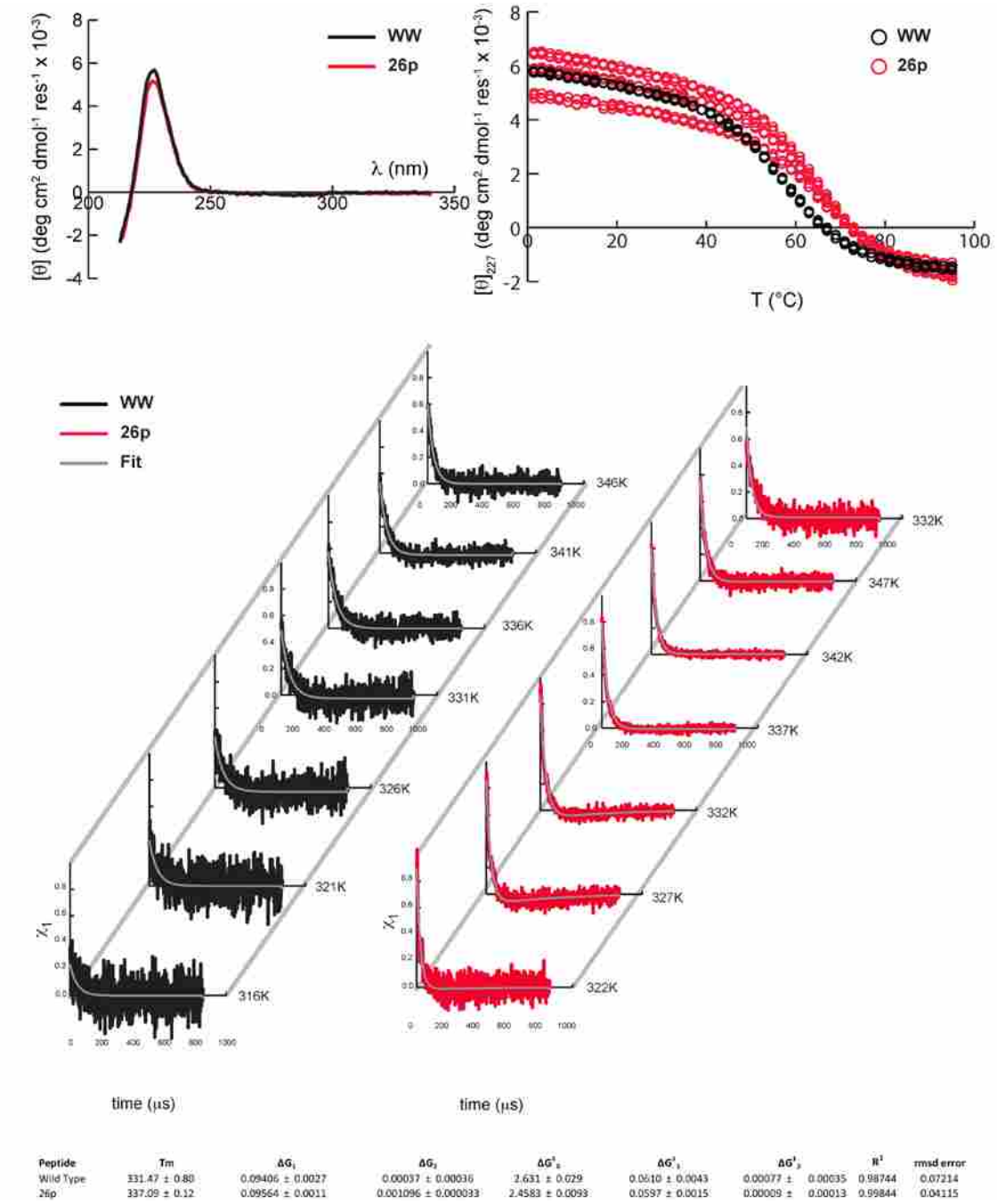
Peptide	Tm	ΔG_1	ΔG_2	$\Delta G^{\ddagger}_0$	$\Delta G^{\ddagger}_1$	$\Delta G^{\ddagger}_2$	R^2	rmsd error
18	332.89 ± 0.75	0.0808 ± 0.0035	-0.00032 ± 0.00017	2.772 ± 0.016	0.0369 ± 0.0031	0.0006 ± 0.00020	0.99575	0.06213
18p	330.13 ± 0.66	0.0936 ± 0.0022	0.00057 ± 0.00031	2.932 ± 0.016	0.050 ± 0.003	0.00108 ± 0.00024	0.98886	0.06811

Figure 66: CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for solutions of proteins **18** and **18p (**in which position 18 was replaced by Asn and Asn-PEG, respectively) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data to equations 3-8, using the indicated parameters.



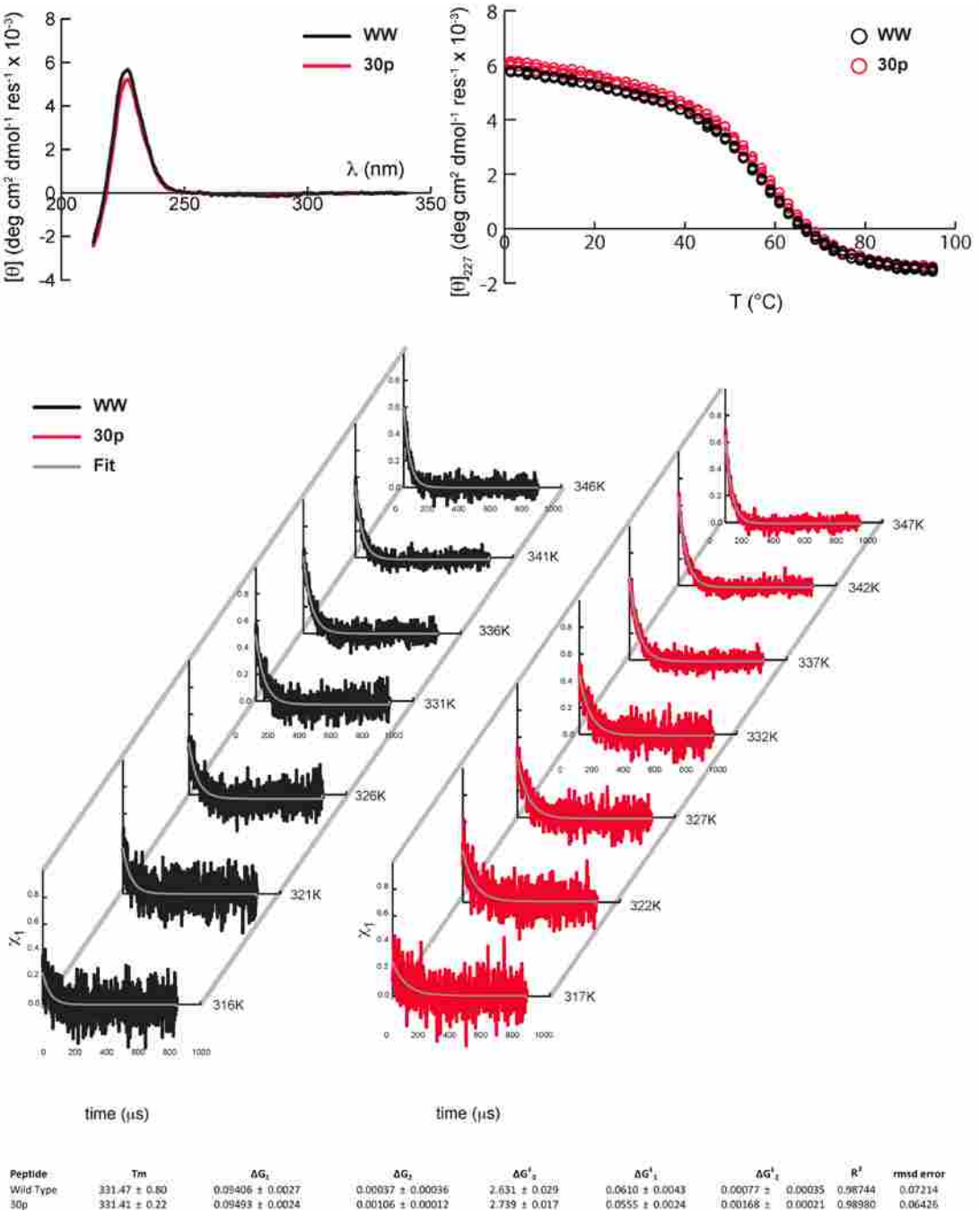
Peptide	Tm	ΔG_1	ΔG_2	$\Delta G^{\ddagger}_0$	$\Delta G^{\ddagger}_1$	$\Delta G^{\ddagger}_2$	R^2	rmsd error
19	328.75 ± 0.33	0.0945 ± 0.0018	0.0009 ± 0.00019	2.838 ± 0.019	0.0468 ± 0.0024	0.0017 ± 0.00024	0.99277	0.07662
19p	336.39 ± 0.31	0.0989 ± 0.0034	0.00117 ± 0.000088	2.962 ± 0.024	0.0543 ± 0.0034	0.00211 ± 0.00029	0.98423	0.08616

Figure 67: CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for solutions of proteins **19** and **19p (**in which position 19 was replaced by Asn and Asn-PEG, respectively) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data to equations 3-8, using the indicated parameters.



Peptide	Tm	ΔG_1	ΔG_2	ΔG^1_0	ΔG^1_1	ΔG^1_2	R^2	rmsd error
Wild Type	331.47 ± 0.80	0.09406 ± 0.0027	0.00037 ± 0.00036	2.631 ± 0.029	0.0510 ± 0.0043	0.00077 ±	0.00035	0.98744	0.07214
26p	337.09 ± 0.12	0.09564 ± 0.0011	0.001096 ± 0.000033	2.4583 ± 0.0093	0.0597 ± 0.0015	0.00009 ±	0.00013	0.99844	0.04115

Figure 68: CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for solutions of proteins **WW** and **26p (**in which position 26 was either left as an Asn or replaced with Asn-PEG, respectively) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data to equations 3-8, using the indicated parameters.



Peptide	Tm	ΔG_1	ΔG_2	$\Delta G^{\ddagger}_2$	$\Delta G^{\ddagger}_1$	$\Delta G^{\ddagger}_2$	R^2	rmsd error
Wild Type	331.47 ± 0.80	0.09406 ± 0.0027	0.00037 ± 0.00036	2.631 ± 0.029	0.0610 ± 0.0043	0.00077 ± 0.00035	0.98744	0.07214
30p	331.41 ± 0.22	0.09493 ± 0.0024	0.00106 ± 0.00012	2.739 ± 0.017	0.0555 ± 0.0024	0.00168 ± 0.00021	0.98980	0.06426

Figure 69: CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for solutions of PEGylated protein **30p (**in which position 26 was replaced with Asn-PEG) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data to equations 3-8, using the indicated parameters. For comparison, the data for protein **WW** is repeated here.

2.5 References

(1) Shu, J. Y.; Lund, R.; Xu, T.: Solution Structural Characterization of Coiled-Coil Peptide–Polymer Side-Conjugates. *Biomacromolecules* **2012**, *13*, 1945-1955.
(2) Shu, J. Y.; Tan, C.; DeGrado, W. F.; Xu, T.: New Design of Helix Bundle Peptide–Polymer Conjugates. *Biomacromolecules* **2008**, *9*, 2111–2117.
(3) Jain, A.; Ashbaugh, H. S.: Helix Stabilization of Poly(ethylene glycol)–Peptide Conjugates. *Biomacromolecules* **2011**, *12*, 2729-2734.
(4) Baillargeon, M.; Sonnet, P.: Polyethylene glycol modification *Candida rugosa* lipase. *J. Am. Oil Chem. Soc.* **1988**, *65*, 1812-1815.
(5) Basri, M.; Ampon, K.; Yunus, W. M. Z. W.; Razak, C. N. A.; Salleh, A. B.: Synthesis of fatty esters by polyethylene glycol-modified lipase. *J. Chem. Tech. Biotechnol.* **1995**, *64*, 10-16.
(6) Longo, M. A.; Combes, D.: Thermostability of modified enzymes: a detailed study. *J Chem Technol Biot* **1999**, *74*, 25–32.
(7) Hernaiz, M. J.; Sanchez-Montero, J. M.; Sinisterra, J. V.: Modification of purified lipases from Candida rugosa with polyethylene glycol: A systematic study. *Enzyme Microb. Technol.* **1999**, *24*, 181–190.
(8) Gaertner, H. F.; Puigserver, A. J.: Increased Activity and Stability of Poly(Ethylene Glycol)-Modified Trypsin. *Enzyme Microb. Technol.* **1992**, *14*, 150–155.
(9) Monfardini, C.; Schiavon, O.; Caliceti, P.; Morpurgo, M.; Harris, J. M.; Veronese, F. M.: A Branched Monomethoxypoly(Ethylene Glycol) for Protein Modification. *Bioconjugate Chem.* **1995**, *6*, 62–69.
(10) Zhang, Z.; He, Z.; Guan, G.: Thermal stability and thermodynamic analysis of native and methoxypolyethylene glycol modified trypsin. *Biotechnol. Tech.* **1999**, *13*, 781-786.
(11) Treetharnmathurot, B.; Ovartlarnporn, C.; Wungsintaweekul, J.; Duncan, R.; Wiwattanapatapee, R.: Effect of PEG molecular weight and linking chemistry on the biological activity and thermal stability of PEGylated trypsin. *Int. J. Pharm.* **2008**, *357*, 252-259.
(12) Chiu, K.; Agoubi, L. L.; Lee, I.; Limpar, M. T.; Lowe, J. W.; Goh, S. L.: Effects of Polymer Molecular Weight on the Size, Activity, and Stability of PEG-Functionalized Trypsin. *Biomacromolecules* **2010**, *11*, 3688-3692.
(13) Topchieva, I. N.; Efremova, N. V.; Khvorov, N. V.; Magretova, N. N.: Synthesis and Physicochemical Properties of Protein Conjugates with Water-Soluble Poly(Alkylene Oxides). *Bioconjugate Chem.* **1995**, *6*, 380–388.
(14) Castellanos, I. J.; Al-Azzam, W.; Griebenow, K.: Effect of the covalent modification with poly(ethylene glycol) on alpha-chymotrypsin stability upon encapsulation in poly(lactic-co-glycolic) microspheres. *J. Pharm. Sci.* **2005**, *94*, 327–340.
(15) Rodriguez-Martinez, J. A.; Solá, R. J.; Castillo, B.; Cintron-Colon, H. R.; Rivera-Rivera, I.; Barletta, G.; Griebenow, K.: Stabilization of alpha-Chymotrypsin Upon PEGylation Correlates With Reduced Structural Dynamics. *Biotechnol. Bioeng.* **2008**, *101*, 1142–1149.
(16) Rodríguez-Martínez, J.; Rivera-Rivera, I.; Solá, R.; Griebenow, K.: Enzymatic activity and thermal stability of PEG-α-chymotrypsin conjugates. *Biotechnol. Lett.* **2009**, *31*, 883-887.
(17) Garcia, D.; Ortéga, F.; Marty, J.-L.: Kinetics of thermal inactivation of horseradish peroxidase: stabilizing effect of methoxypoly(ethylene glycol). *Biotechnol. Appl. Biochem.* **1998**, *27*, 49-54.
(18) Lopez-Cruz, J. I.; Viniegra-Gonzalez, G.; Hernandez-Arana, A.: Thermostability of native and pegylated Myceliophthora thermophila laccase in aqueous and mixed solvents. *Bioconjugate Chem* **2006**, *17*, 1093–1098.
(19) Nie, Y.; Zhang, X.; Wang, X.; Chen, J.: Preparation and Stability of N-Terminal Mono-PEGylated Recombinant Human Endostatin. *Bioconjugate Chem* **2006**, *17*, 995-999.
(20) Hinds, K. D.; Kim, S. W.: Effects of PEG conjugation on insulin properties. *Adv. Drug Deliv. Rev.* **2002**, *54*, 505-530.

(21) Yang, C.; Lu, D. N.; Liu, Z.: How PEGylation Enhances the Stability and Potency of Insulin: A Molecular Dynamics Simulation. *Biochemistry (Mosc)*. **2011**, *50*, 2585–2593.
(22) Meng, W.; Guo, X.; Qin, M.; Pan, H.; Cao, Y.; Wang, W.: Mechanistic Insights into the Stabilization of srcSH3 by PEGylation. *Langmuir* **2012**, *Article ASAP*, doi: 10.1021/la303466w.
(23) Shu, J. Y.; Tan, C.; DeGrado, W. F.; Xu, T.: New Design of Helix Bundle Peptide−Polymer Conjugates. *Biomacromolecules* **2008**, *9*, 2111-2117.
(24) Dhalluin, C.; Ross, A.; Leuthold, L. A.; Foser, S.; Gsell, B.; Muller, F.; Senn, H.: Structural and biophysical characterization of the 40 kDa PEG-interferon-alpha(2a) and its individual positional isomers. *Bioconjugate Chem* **2005**, *16*, 504–517.
(25) 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.: Structure-Function Engineering of Interferon-β1b for Improving Stability, Solubility, Potency, Immunogenicity, and Pharmacokinetic Properties by Site-Selective Mono-PEGylation. *Bioconjugate Chem* **2006**, *17*, 618-630.
(26) Ramon, J.; Saez, V.; Baez, R.; Aldana, R.; Hardy, E.: PEGylated interferon-alpha 2b: A branched 40K polyethylene glycol derivative. *Pharm. Res.* **2005**, *22*, 1374–1386.
(27) Garcia-Arellano, H.; Valderrama, B.; Saab-Rincon, G.; Vazquez-Duhalt, R.: High temperature biocatalysis by chemically modified cytochrome c. *Bioconjugate Chem* **2002**, *13*, 1336–1344.
(28) Plesner, B.; Fee, C. J.; Westh, P.; Nielsen, A. D.: Effects of PEG size on structure, function and stability of PEGylated BSA. *Eur. J. Pharm. Biopharm.* **2011**, *79*, 399-405.
(29) Yang, Z.; Williams, D.; Russell, A. J.: Synthesis of Protein-Containing Polymers in Organic-Solvents. *Biotechnol. Bioeng.* **1995**, *45*, 10–17.
(30) Yang, Z.; Domach, M.; Auger, R.; Yang, F. X.; Russell, A. J.: Polyethylene glycol-induced stabilization of subtilisin. *Enzyme Microb. Technol.* **1996**, *18*, 82–89.
(31) Callahan, W. J.; Narhi, L. O.; Kosky, A. A.; Treuheit, M. J.: Sodium chloride enhances the storage and conformational stability of BDNF and PEG-BDNF. *Pharm. Res.* **2001**, *18*, 261–266.
(32) Plesner, B.; Westh, P.; Nielsen, A. D.: Biophysical characterisation of GlycoPEGylated recombinant human factor VIIa. *Int. J. Pharm.* **2011**, *406*, 62–68.
(33) Rodríguez-Martínez, J. A.; Rivera-Rivera, I.; Griebenow, K.: Prevention of benzyl alcohol-induced aggregation of chymotrypsinogen by PEGylation. *J. Pharm. Pharmacol.* **2011**, *63*, 800-805.
(34) Price, J. L.; Powers, E. T.; Kelly, J. W.: N-PEGylation of a Reverse Turn is Stabilizing in Multiple Sequence Contexts unlike N-GlcNAcylation. *ACS Chem Biol* **2011**, *6*, 1188–1192.
(35) Pandey, B. K.; Smith, M. S.; Torgerson, C.; Lawrence, P. B.; Matthews, S. S.; Watkins, E.; Groves, M. L.; Prigozhin, M. B.; Price, J. L.: Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length. *Bioconjug. Chem.* **2013**, *24*, 796-802.
(36) Price, J. L.; Shental-Bechor, D.; Dhar, A.; Turner, M. J.; Powers, E. T.; Gruebele, M.; Levy, Y.; Kelly, J. W.: Context-Dependent Effects of Asparagine Glycosylation on Pin WW Folding Kinetics and Thermodynamics. *J. Am. Chem. Soc.* **2010**, *132*, 15359-15367.
(37) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W.: Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts. *Proceedings of the National Academy of Sciences* **2011**, *108*, 14127-14132.
(38) Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P.: Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **1997**, *89*, 875–886.
(39) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M.: The folding mechanism of a β-sheet: the WW domain. *J. Mol. Biol.* **2001**, *311*, 373-393.
(40) Chen, W.; Enck, S.; Price, J. L.; Powers, D. L.; Powers, E. T.; Wong, C.-H.; Dyson, H. J.; Kelly, J. W.: Structural and Energetic Basis of Carbohydrate–Aromatic Packing Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 9877-9884.

(41) Mu, Q.; Hu, T.; Yu, J.: Molecular Insight into the Steric Shielding Effect of PEG on the Conjugated Staphylokinase: Biochemical Characterization and Molecular Dynamics Simulation. *PLoS ONE* **2013**, *8*, e68559.
(42) Price, J. L.; Powers, E. T.; Kelly, J. W.: N-PEGylation of a Reverse Turn Is Stabilizing in Multiple Sequence Contexts, unlike N-GlcNAcylation. *ACS Chemical Biology* **2011**, *6*, 1188-1192.
(43) Edelhoch, H.: Spectroscopic Determination of Tryptophan and Tyrosine in Proteins*. *Biochemistry (Mosc)*. **1967**, *6*, 1948-1954.
(44) Ballew, R. M.; Sabelko, J.; Gruebele, M.: Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl. Acad. Sci USA* **1996**, *93*, 5759–5764.
(45) Ballew, R. M.; Sabelko, J.; Reiner, C.; Gruebele, M.: A single-sweep, nanosecond time resolution laser temperature-jump apparatus. *Rev. Sci. Instrum.* **1996**, *67*, 3694–3699.
(46) Ervin, J.; Sabelko, J.; Gruebele, M.: Submicrosecond real-time fluorescence detection: application to protein folding. *J. Photochem. Photobiol. sect. B* **2000**, *54*, 1–15.
(47) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M.: The Folding Mechanism of a β-sheet: The WW Domain. *J. Mol. Biol.* **2001**, *311*, 373–393.
(48) Kramers, H. A.: Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **1940**, *7*, 284.
(49) Hänggi, P.; Talkner, P.; Borovec, M.: Reaction rate theory - 50 years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251–341.
(50) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J.: Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci USA* **2000**, *97*, 7220–7225.
(51) Bieri, O.; Wirz, J.; Hellrung, B.; Schutokowski, M.; Drewello, M.; Kiefhaber, T.: The speed limit of protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci USA* **1999**, *96*, 9597–9601.
(52) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A.: The role of solvent viscosity in the dynamics of protein conformational changes. *Science* **1992**, *256*, 1796–1798.
(53) Fuller, A. A.; Du, D.; Liu, F.; Davoren, J. E.; Bhabha, G.; Kroon, G.; Case, D. A.; Dyson, H. J.; Powers, E. T.; Wipf, P.; Gruebele, M.; Kelly, J. W.: Evaluating β-turn mimics as β-sheet folding nucleators. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11067–11072.
(54) Weast, R. C.: *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, 1982.

Chapter 3: Insights into the Mechanism by Which PEGylation Stabilizes the WW Domain

3.1 Introduction

Protein folding occurs because the native (folded) state is lower in free energy than the denatured (unfolded) state(s) and/or any metastable states. The folding landscape of the WW domain can be approximated by the two state energy diagram shown in Figure 1, in which an unfolded protein proceeds directly via a high energy transition state to the folded conformation without passing through any long-lived intermediates.[1] As illustrated in Figure 1(a), the stability of a folded protein is related to the change in free energy upon folding, or ΔG_f; the more negative the ΔG_f , the more stable the protein is. There are two possible ways by which PEGylation could stabilize a folded protein; that is, two possible ways that the ΔG_f could increase upon PEGylation. The first is by destabilizing the unfolded state, as shown in Figure 1(b). A variety of mechanisms could account for this; PEG might restrict the conformational entropy of the unfolded state, or it might disrupt transient interactions in the unfolded state, forcing it to adopt more extended conformations.[2] Alternatively, PEGylation might stabilize a protein by stabilizing the folded state, as shown in Figure 1(c). The free energy of folding (ΔG_f) can be parsed into enthalpic (ΔH_f) and entropic ($-T\Delta S_f$) components. If PEG engaged in favorable noncovalent interactions with the protein surface (such as hydrogen bonds with polar residues or van der Waals interactions with hydrophobic residues), it would increase the magnitude of ΔH_f; conversely if PEG decreased the solvent-accessible surface area of the protein, thereby releasing ordered water molecules into the bulk solvent, it would increase the magnitude of $-T\Delta S_f$.

It should be noted at this point that these alternatives (native vs. denatured state effects, enthalpic vs. entropic effects) are not necessarily mutually exclusive and may both contribute to the stabilizing impact of PEGylation. PEGylation likely impacts the thermodynamics of protein folding through a variety of subtle mechanisms, acting on both the folded and unfolded state through both enthalpic and entropic mechanisms. Our goal is to identify structural features which govern the thermodynamic impact of PEGylation, to enable the rational design of PEGylated therapeutics with optimal thermodynamic properties.

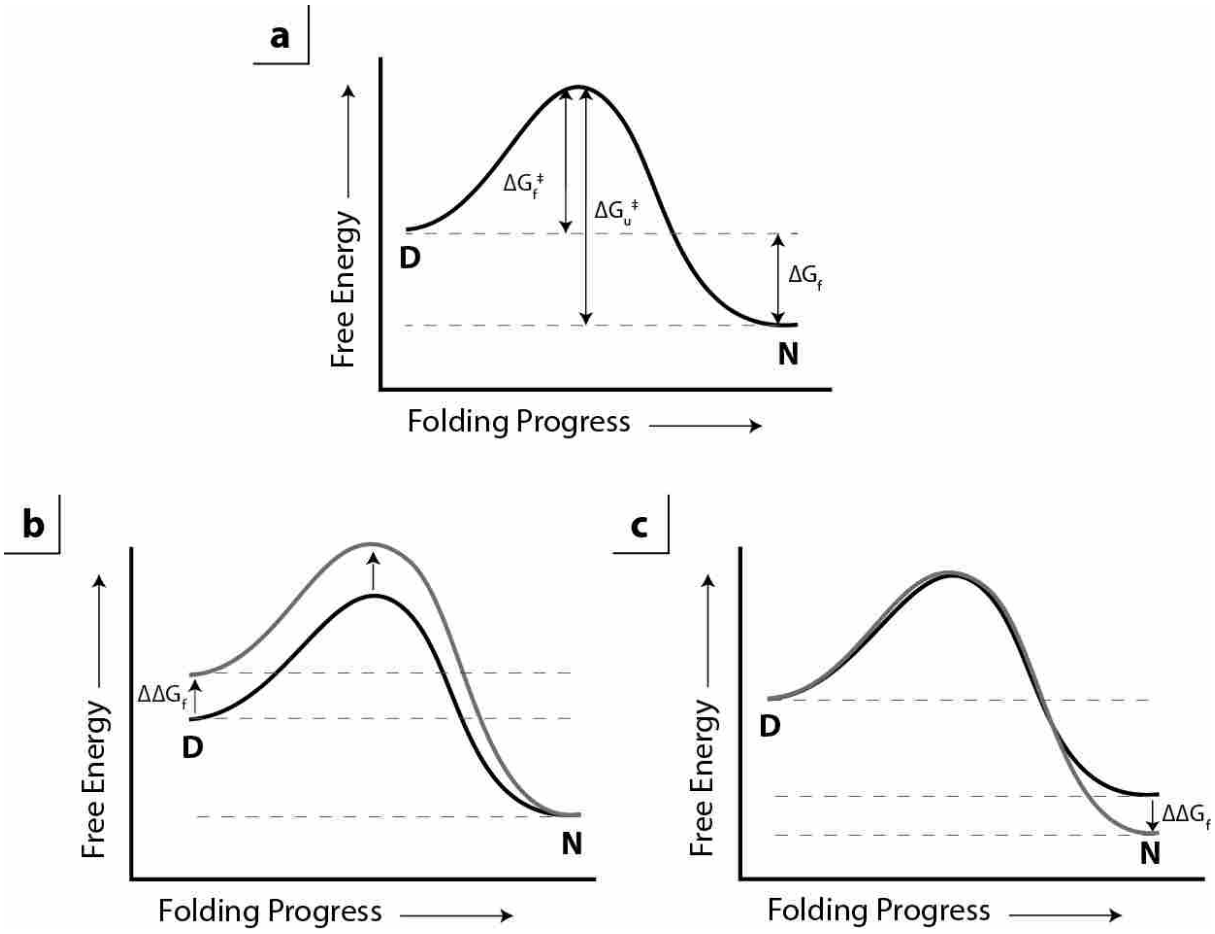


Figure 70: a) A simple model of WW domain folding thermodynamics. D = Denatured state. N= Native State. ΔG_f= free energy of folding. $\Delta G_f^{\ddagger}$=free energy of folding activation. $\Delta G_u^{\ddagger}$= free energy of unfolding activation b) PEGylation may stabilize (increase the magnitude of $\Delta\Delta G_f$) the WW domain by destabilizing the denatured state. c) Alternatively, PEGylation may stabilize the WW domain by stabilizing the folded state.

Two previous experiments suggest that the primary mechanism by which PEG stabilizes proteins is native-state stabilization. The thermodynamic impact of PEGylation on the WW domain depends strongly on the site of PEG conjugation, as illustrated in Chapter 2. Specifically, PEGylation is only stabilizing at sites where the side chain is oriented back toward the protein surface. This apparent dependence on folded-state orientation would be difficult to reconcile with a model in which PEG destabilizes the unfolded ensemble.

Further evidence for native-state stabilization may be found in Pandey et al.,[3] in which we demonstrated that, at position 19, the degree of PEG-induced stabilization depends strongly on the length of PEG polymer used, as shown in Figure 2. Very small oligomers (1 to 2 ethylene oxide units) stabilize WW significantly less than oligomers of slightly longer length (3 to 4 ethylene oxide units); longer PEG polymers (8 to ~45 ethylene oxide units) were less effective than oligomers of 4 ethylene oxide units, but still more effective than the very small oligomers. The sudden increase in stability upon adding a third ethylene oxide unit is consistent with the presence of a specific folded-state interaction accessible to the trimer but not to shorter oligomers.

3.2 Results and Discussion

If PEG were to increase ΔG_f by destabilizing the unfolded ensemble, it seems reasonable that the effects of PEGylation might depend on the identity of residues close in sequence to the site of PEGylation, but not to the identity of residues which are further away. On the other hand, if PEGylation acts primarily on the folded state, then the identity of residues which are close in space in the folded state, but not necessarily close in sequence, might affect the thermodynamic impact of PEGylation. By modifying these characteristics (through amino acid substitution, or by changing the direction that the PEG oligomer is oriented, for example), it should be possible to determine what structural features in the folded state, if any, are required for PEGylation to stabilize a protein.

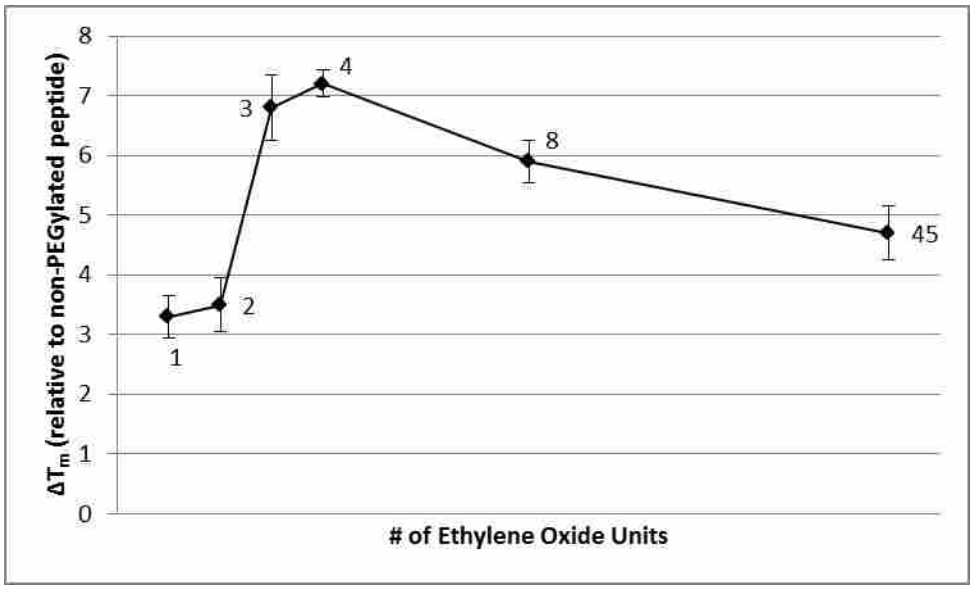


Figure 71: Effect of oligomer length on the stabilizing effects of PEGylation at position 19. PEG oligomers of various lengths were conjugated to the WW domain at position 19. Error bars represent standard errors for ΔT_m.

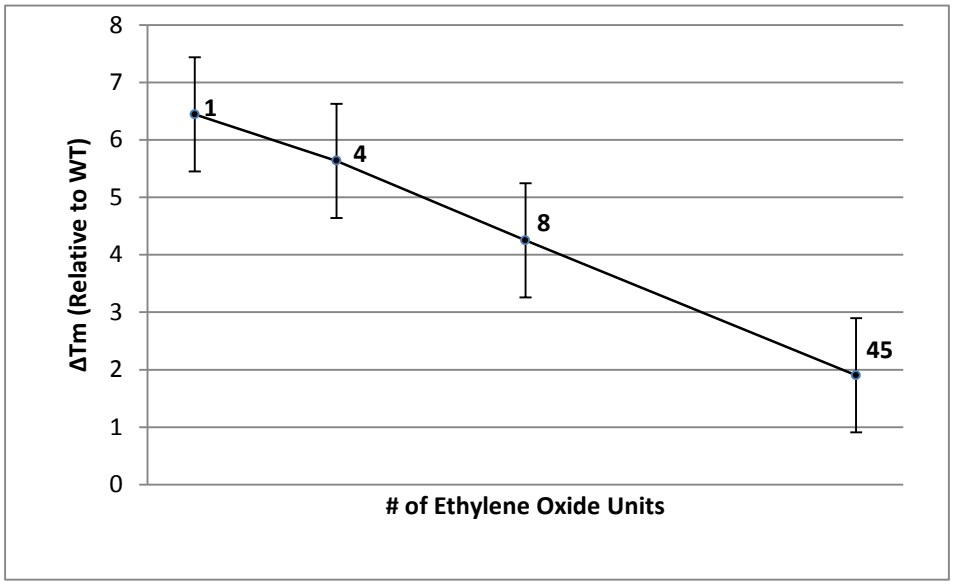


Figure 72: Effect of oligomer length on the stabilizing effects of PEGylation at position 26. PEG oligomers of various lengths were conjugated to the WW domain at position 26. Error bars represent standard errors for ΔT_m.

3.2.1 PEG Length Studies at Position 26

The results described by Pandey et al.[3] suggest that PEG polymers must be a certain length in order to interact with specific regions on the nearby protein surface. To further confirm this hypothesis, we wanted to test the effect of PEG length on stabilization at another stabilizing site. At position 26, addition of a PEG tetramer stabilizes the WW domain by -0.50±0.07 kcal mol^{-1} (compare peptides **WW** and **26p**, Chapter 2). To test the dependence of this effect on PEG length, we prepared peptides **26p(1),** and **26p(8)** in which PEG oligomers comprising 1 and 8 ethylene oxide units, respectively, were attached to the Asn residue at position 26. For protein synthesis and characterization procedures, see Section 107. We also prepared **26p(45)**, in which a polydisperse PEG polymer averaging 45 ethylene oxide units (average MW=2,000 Da) was attached to the Asn residue at position 26**.**

Table 1 shows the results of these experiments. PEGylation at position 26 with a PEG monomer (affording **26p(1)**, $\Delta\Delta G_f$= -0.58±0.07 kcal mol^{-1})was just as effective as PEGylation with a tetramer (affording **26p**, $\Delta\Delta G_f$= -0.50±0.07 kcal mol^{-1}). PEGylation with PEG of 8 units (affording **26p(8)**) was slightly less stabilizing ($\Delta\Delta G_f$= -0.36±0.07 kcal mol^{-1}), as was PEGylation with the 45-unit PEG ($\Delta\Delta G_f$= -0.18±0.11 kcal mol^{-1}). In contrast, PEGylation at position 19 with very short oligomers (1-2 units long) was significantly less effective than PEGylation with PEG trimers or tetramers.

The fact that shorter PEG chains are just as stabilizing as longer ones at position 26 (as shown in Figure 3) suggests that, if PEG stabilizes the folded state of the WW domain, then PEG polymers installed at position 26 may be interacting with nearby regions that are close enough that the PEG monomer is able to reach them.

3.2.2 D-Asn Mutagenesis

If the site-dependence of PEG-induced stability is indeed due to interactions with the protein surface, then changing the orientation of the side chain would likely change the effect of PEGylation at that site. One of the attractive features of the WW domain as a model system is its tolerance toward point

Table 2: Melting temperatures (T_m) and changes in folding free energy ($\Delta\Delta G_f$ / kcal mol^{-1}) for peptides in PEG length study at position 26[a]

Protein	Tm (°C)	$\Delta\Delta G_f$ (kcal/mol)
WW[b]	58.3 ± 0.8	
26p(1)	64.8 ± 0.2	-0.58 ± 0.07
26p[b]	63.9 ± 0.1	-0.50 ± 0.07
26p(8)	62.6 ± 0.2	-0.36 ± 0.07
26p(45)	60.2 ± 0.3	-0.18 ± 0.11

[a]Tabulated data are given as mean ± standard error 100 μM solutions of WW variants in 20 mM sodium phosphate buffer (pH 7) at the melting temperatures of the wild-type peptide WW. [b] Peptides from Chapter 2 are included for comparison

mutations, particularly in the first reverse turn.[4] Kaul et al[5] reported that residues 18 and 19 can be replaced with D-amino acids or β-turn mimetics without disrupting the protein structure.

Although positions 18 and 19 are close in space, they differ greatly in their native-state orientation. As shown in Figure 4, the position 19 is oriented such that the side chain projects back toward the protein surface, while the side chain of position 18 projects out into solution. Comparing the effects of PEGylation at positions 19 ($\Delta\Delta G_f$= -0.69±0.05 kcal mol^{-1}) and 18 ($\Delta\Delta G_f$= +0.26±0.09 kcal mol^{-1}) shows that changing the site of PEG conjugation by a single residue vastly alters its thermodynamic effects. One explanation for this observation could be that PEG stabilizes WW by interacting with the native protein surface, in which PEG would only stabilize the protein when it was oriented such that it could interact with the surface.

At position 19, the α-hydrogen is oriented away from the protein into solution, as indicated in Figure 5a. Incorporating D-Asn or D-Asn(PEG) (Figure 5b) at position 19 would invert the stereocenter; it is reasonable to suppose that doing so would re-orient the side chain such that it would project out into solution, rather than back toward the protein surface (provided that such an inversion does not significantly disrupt the folded state of the protein). Peptides **19(D)** and **19(D)p** are mutants of the WW domain in which residue 19 was replaced by a D-Asn or D-Asn-PEG, respectively (See Figure 5). Although no crystal structures exist for these mutants, the CD spectra of these peptides (see Figure 93 in the supporting information) indicate no disruption of the β-sheet structure. Furthermore, the melting temperatures of **19(D)** (Tm =55.4 ± 0.3) and **19** (55.6 ± 0.3) are identical, suggesting that no major structural perturbations occurred upon inversion of the stereocenter at position 19.

PEGylation of **19(D)** (affording **19(D)p**, Tm = 55.3 ± 0.3) did not stabilize the peptide ($\Delta\Delta G_f$= 0.01±0.04 kcal mol^{-1}). These data, summarized in Table 2, are strong evidence that the thermodynamic consequences of PEGylation are dependent on the orientation of the PEGylated side-chain, suggesting that PEG stabilizes the native state of WW via interactions with the protein surface.

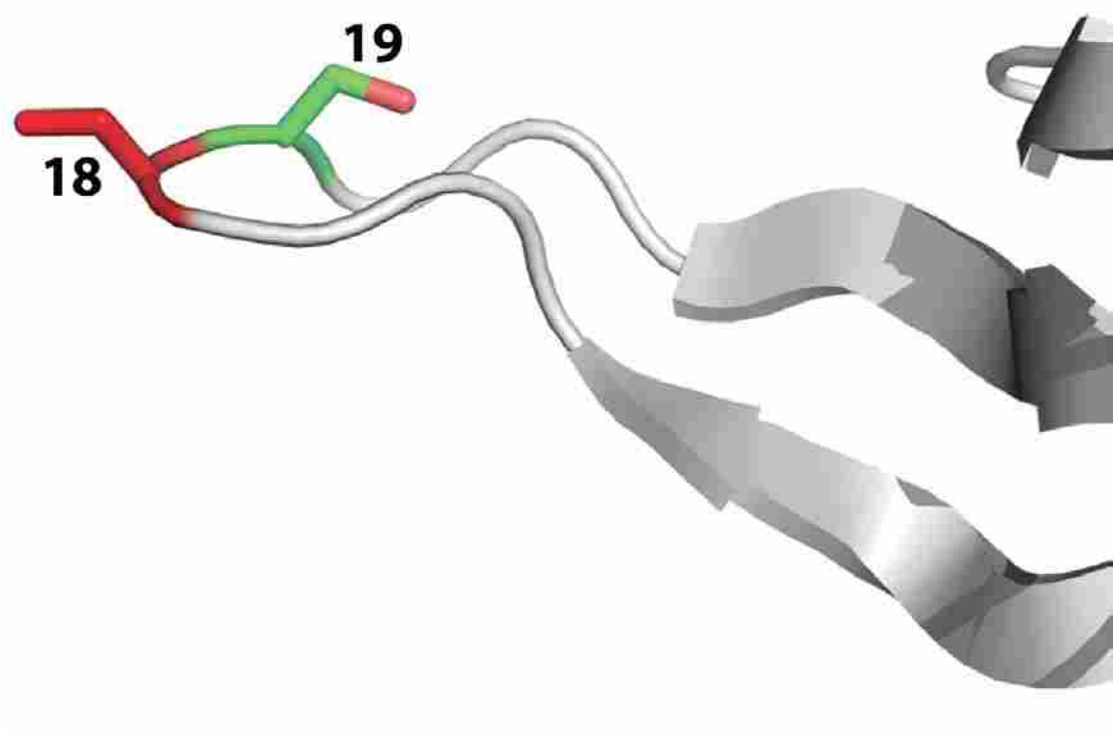


Figure 73: Difference in orientation between position 18 (red) and position 19(green). All other side chains have been removed for clarity.

Table 3: Melting temperatures (T_m) and changes in folding free energy ($\Delta\Delta G_f$ / kcal mol^{-1}) for D-Asn derivatives of WW[a]

Protein	Tm (°C)	$\Delta\Delta G_f$ (kcal/mol)
19[b]	55.6 ± 0.3	
19p[b]	63.2 ± 0.3	-0.69 ± 0.05
19(D)	55.4 ± 0.3	
19(D)p	55.3 ± 0.3	-0.01 ± 0.04

[a]Tabulated data are given as mean ± standard error 100 μM solutions of WW variants in 20 mM sodium phosphate buffer (pH 7) at the melting temperatures of the corresponding non-PEGylated proteins. [b] Peptides from Chapter 2 are included for comparison

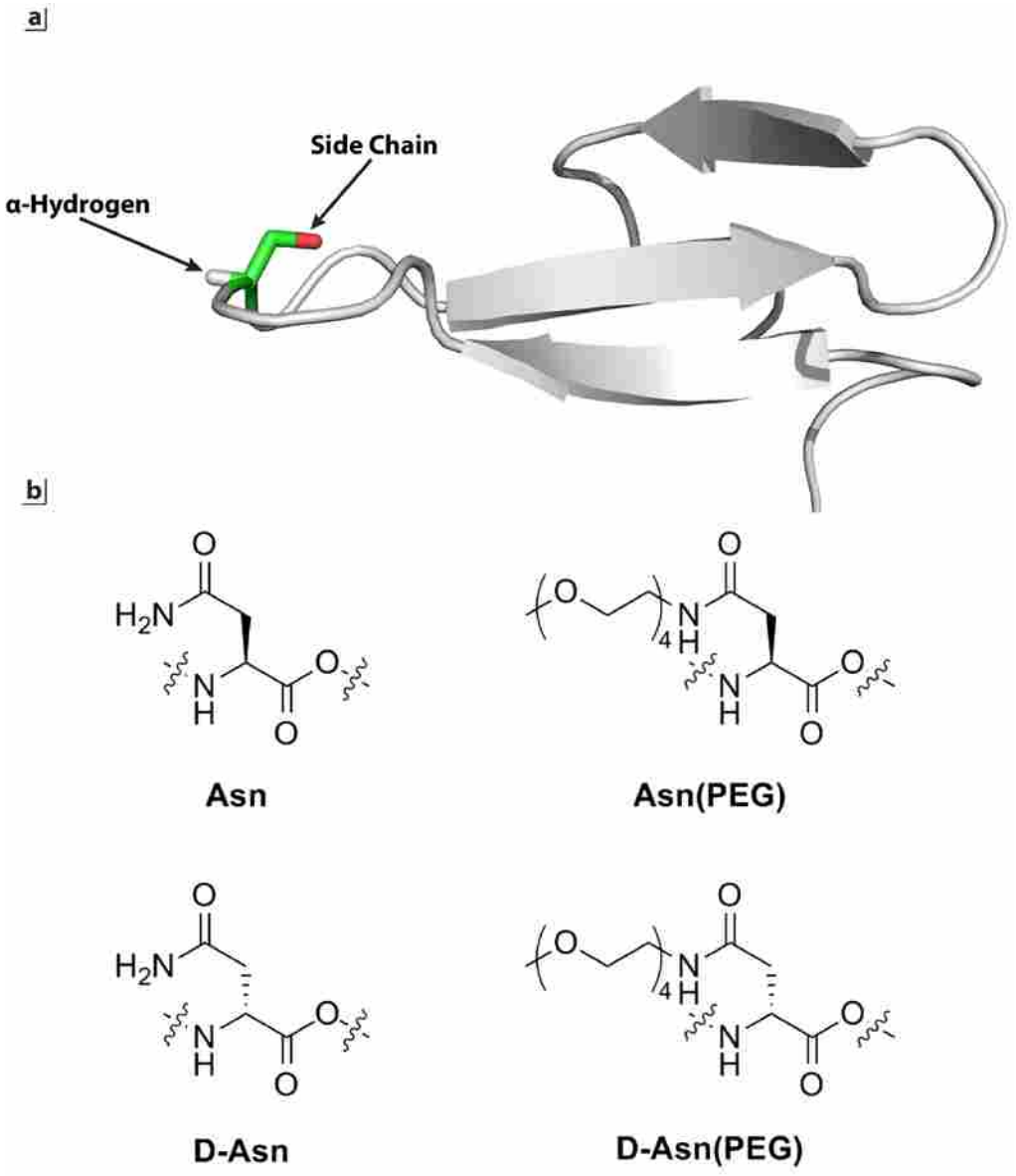


Figure 74: a) Orientations of the side chain and α-hydrogen of position 19 relative to the rest of the protein. b) Structures of Asn, Asn(PEG), D-Asn, and D-Asn(PEG)

3.2.3 Double PEGylation

If PEG stabilizes the native state of WW, it may be interacting with specific regions on the protein surface, or its interactions may be non-specific. As shown in Figure 6a, Positions 16 and 19 are located in the first reverse turn, and the side chains of both residues are oriented back toward the same face of the protein surface. PEGylation at both positions is stabilizing. If PEG stabilizes the WW domain by interacting with specific regions in the folded state, it is plausible that two PEG polymers, one attached at each site, might compete for the same stabilizing interactions, and the effects of PEGylating both sites would not be cumulative. To test this hypothesis we prepared WW mutants **16/19, 16p/19, 16/19p** and **16p/19p**. In these four mutants, positions 16 and 19 were both mutated to asparagine. **16/19** was not PEGylated. **16p/19**, **16/19p**, and **16p/19p** were PEGylated at 16, 19, and both 16 and 19, respectively. The sequences of these peptides, as well as all others detailed in this section, are shown in Figure 7.

These peptides enable us to probe the dependence of the impact of PEGylation at position 19 on the presence or absence of PEG at position 16. The results are summarized in Table 3. When **16/19** is PEGylated at position 19 (resulting in peptide **16/19p**), it is stabilized by -0.49±0.02 kcal mol^{-1} (for all peptides in this section, the free energy of folding is calculated at the melting temperature of the non-PEGylated peptide – in this case, 56.8°C, the T_m of **16/19**)

The difference in folding free energy between **16p/19** and **16p/19p** may be thought of as the stabilizing impact of PEGylation at 19 when 16 is already PEGylated. If the two PEG moieties were interacting with different sites, we would expect that PEGylation at one site would not alter the thermodynamic impact of the other. Instead, we see that, when 16 is already PEGylated, PEGylation at 19 only stabilizes the peptide by -0.16±0.02 kcal mol^{-1}. This anti-synergy (meaning the detrimental effect that PEGylation at one site has on effect of PEGylation at another) is consistent with the hypothesis that the two PEG moieties are competing for the same binding interactions, or alternatively, interfering with each other, preventing each other from interacting optimally with nearby regions of the protein surface.

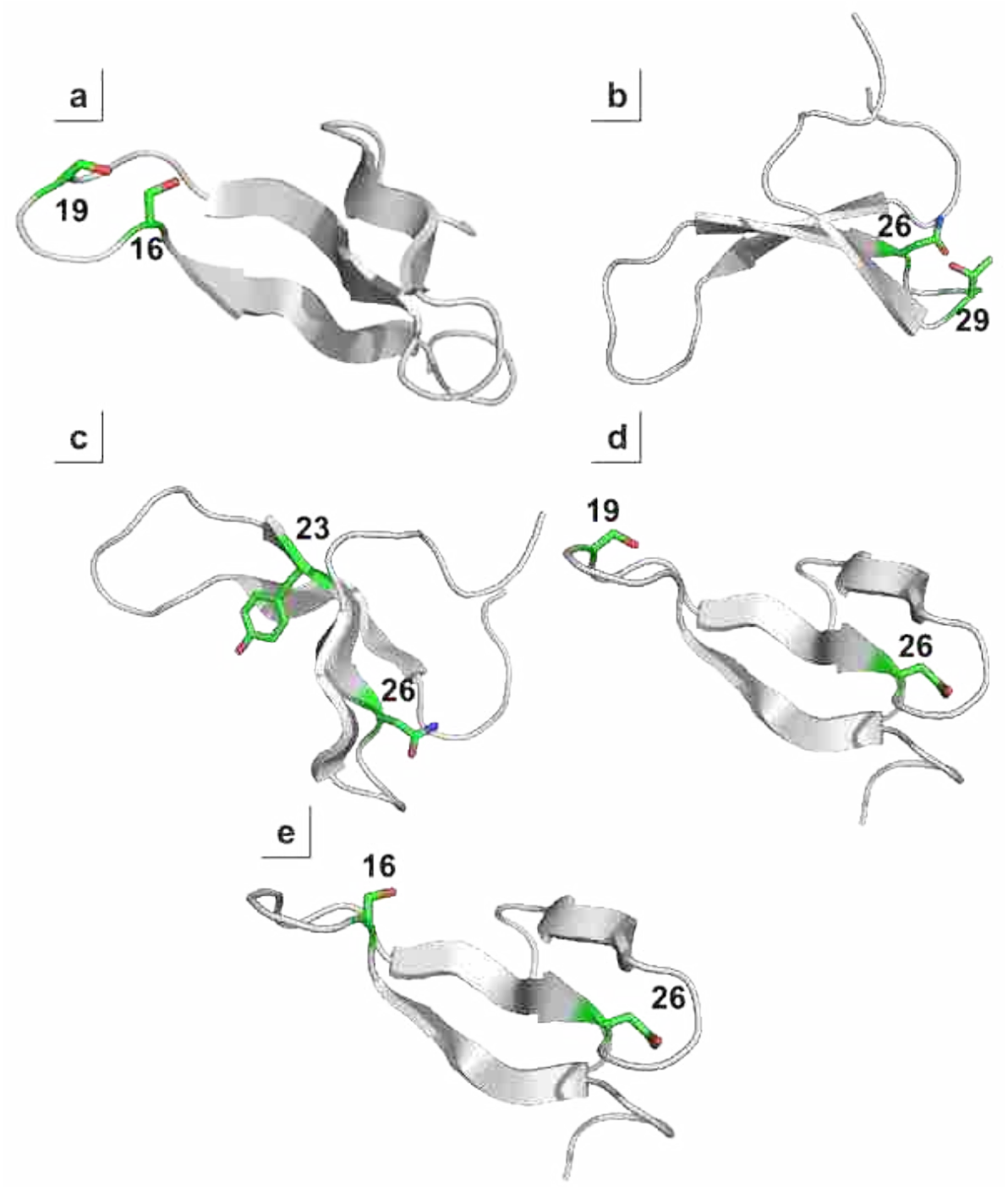


Figure 75: Orientation of side chains at sites probed by Double PEGylation studies. **a)** Positions 16 and 19 are close in sequence and in space, and their side chains are oriented back toward the same face of the protein surface. **b)** Positions 16 and 19 are likewise close in sequence and in space, and their side chains are oriented back toward the same face of the protein surface. **c)** 23 and 26 are close in sequence and space, but their side chains are oriented towards opposite faces of the protein surface. **d)** 19 and 26 are not close in sequence or space, and their side chains are oriented toward opposite faces of the protein surface. **e)** 16 and 26 are not close in sequence, but comparatively close in space. Their side chains are oriented toward opposite faces of the protein surface.

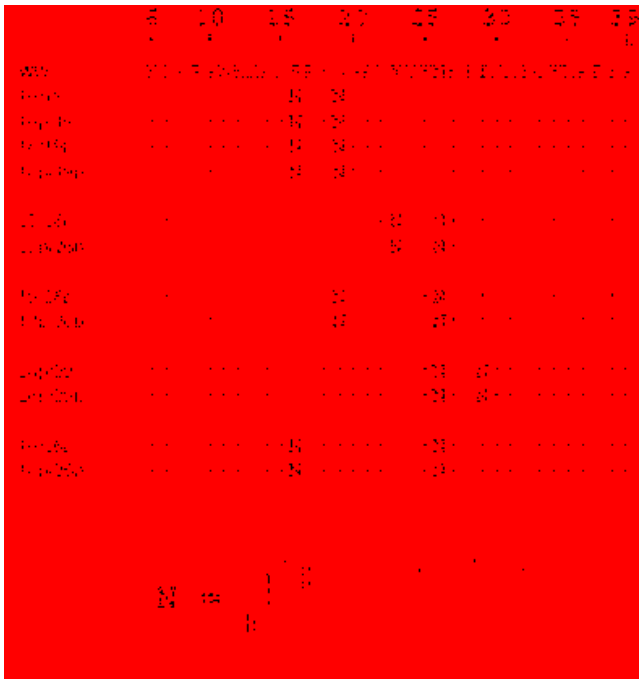


Figure 76: Sequences of **16/19, 16p/19, 16/19p, 16p/19p, 23/26p, 23p/26p, 19/26p, 19p/26p, 26p/29, 26p/29p, 16/26p, and 16p/26p**

Table 4: Melting temperatures (T_m) and changes in folding free energy ($\Delta\Delta G_f$ / kcal mol^{-1}) of double-PEGylated WW variants.[a]

Protein	T_m (°C)	$\Delta\Delta G_f$ (kcal/mol)[b]	Protein	T_m (°C)	$\Delta\Delta G_f$ (kcal/mol)
16/19	56.8 ± 0.2		**19**[c]	55.6 ± 0.2	
16/19p	62.2 ± 0.1	-0.49 ± 0.02	**19/26p**	62.2 ± 0.1	-0.57 ± 0.03
16p/19	63.1 ± 0.1		**19p**[c]	63.2 ± 0.3	
16p/19p	65.3 ± 0.1	-0.16 ± 0.02	**19p/26p**	69.6 ± 0.2	-0.55 ± 0.04
29[c]	48.4 ± 0.3		**16**[c]	54.8 ± 0.2	
26/29p	57.1 ± 0.3	-0.74 ± 0.04	**16/26p**	62.6 ± 0.3	-0.57 ± 0.03
29p[c]	53.5 ± 0.2		**16p**[c]	62.3 ± 0.2	
26p/29p	56.8 ± 0.3	-0.29 ± 0.03	**16p/26p**	67.1 ± 0.1	-0.55 ± 0.04
23[c]	28.6 ± 0.4				
23/26p	35.4 ± 0.9	-2.36 ± 0.15			
23p[c]	23.2 ± 1.0				
23p/26p	31.9 ± 1.2	-2.05 ± 0.24			

[a]Tabulated data are given as mean ± standard error 100 μM solutions of WW variants in 20 mM sodium phosphate buffer (pH 7). [b] $\Delta\Delta G_f$ is reported at the melting temperature of the corresponding non-PEGylated peptide (the first of each set of 4 peptides). [c] Peptides from Chapter 2 are included for comparison

Like positions 16 and 19, positions 26 and 29 also project onto the same face of the β-sheet (Figure 6b). PEGylation is stabilizing at both these positions ($\Delta\Delta G_f$=-0.50±0.07 kcal mol^{-1} for 26, and -0.40±0.03 kcal mol^{-1} for 29). WW mutants **26p/29** (in which residue 26 is replaced with Asn-PEG and 29 is replaced with Asn) and **26p/29p** (in which residues 26 and 29 are replaced with Asn-PEG), together with **29** and **29p** (from Chapter 2 – NOTE: **29** and **29p** already have Asn at position 26), allow us to probe the dependence of PEG-based stabilization at position 26 on the presence or absence of PEG at position 29. When **29** (which has Asn at positions 26 and 29) is PEGylated at position 26 (affording **26p/29**), it is stabilized by -0.74 ± 0.02 kcal mol^{-1}. When **29p** is PEGylated at 26 (affording **26p/29p**), it is stabilized by -0.29 ± 0.03 kcal mol^{-1}; this also indicates that two PEG oligomers are competing for the same binding interactions.

It is interesting to note that replacing Thr29 with Asn itself has an effect on the PEGylation of position 28 (As presented in Chapter 2, PEGylation of WW to afford **26p** stabilizes the protein by -0.50 ± 0.07 kcal mol^{-1}). This increase in stability upon Thr29Asn mutagenesis may indicate that PEG is interacting with the side chains of position 29. This possibility will be explored further in section 3.2.4.

In both of these experiments, the paired PEGylation sites (16/19 and 26/29) are not only orientated in the same direction in the folded protein, but they are also close in sequence. To verify that the observed anti-synergy was due to a native-state interaction, and not an interaction in the unfolded state, we conducted three additional experiments, with pairs of PEGylation sites more distant in sequence and/or in space.

Positions 23 and 26 project onto opposite faces of the protein surface, but they are close in sequence (Figure 6c). If PEG stabilized WW primarily through an unfolded state effect, we would expect that conjugation at position 23 to interfere with the stabilizing impact of 26. To test this hypothesis, we prepared WW mutants **23/26p** (in which residue 23 is replaced with Asn and 26 is replaced with Asn-PEG) and **23p/26p** (in which residues 23 and 26 are replaced with Asn-PEG), which we compared with with **23** and **23p** (from Chapter 2; NOTE: **23** and **23p** already have Asn at position 26) When **23** (which has Asn at positions 23 and 26) is PEGylated at position 26 (affording **23/26p**), it is stabilized

by -0.49±0.08 kcal mol^{-1}. When **23p** (which has Asn-PEG at position 23 and Asn at position 26) is PEGylated at 26 (affording **23p/26p**), it is stabilized by -0.66±0.12 kcal mol^{-1}. The difference in these two values is 0.16±0.14 kcal mol^{-1}, indicating that PEGylation at position 23 slightly *increases* the effect of PEGylation at position 26. The large errors of these measurements are due to the low melting temperature of the proteins (~30ºC). The stabilizing effects of PEGylation at 26 are not substantially affected by the presence or absence of PEG at position 23. It seems likely that the anti-synergy exhibited by PEGylation at positions 26 and 29 was due to special proximity of the two PEG polymers in the folded state; this is consistent with the hypothesis that PEG interacts with the surface of the protein near the site of conjugation.

The side chains of positions 19 and 26 project onto opposite faces of the β-sheet (Figure 6d). Like position 26, PEGylation at 19 is stabilizing (-0.69±0.05 kcal mol^{-1}). WW mutants **19/26p** (in which residue 19 is replaced with Asn and 26 is replaced with Asn-PEG) and **19p/26p** (in which residues 19 and 26 are replaced with Asn-PEG), which we compared with peptides **19** and **19p** (from Chapter 2; NOTE: peptides **19** and **19p** already have Asn at position 26). When **19** (which has Asn at positions 19 and 26) is PEGylated at position 26 (affording **19/26p**), it is stabilized by -0.57±0.03 kcal mol^{-1}. When **19p** (which has Asn-PEG at position 19 and Asn at position 26) is PEGylated at 26 (affording **23p/26p**), it is stabilized by -0.55±0.04 kcal mol^{-1}. This further indicates that two PEG moieties exhibit anti-synergy only when they are oriented toward the same region on the protein surface, which is consistent with a folded-state effect.

The examples considered thus far suggest that PEGylation stabilizes the folded state of WW by interacting with nearby regions of the protein surface; when two PEG polymers are installed nearby on the same face, they interfere with each other, whereas when they project onto opposite faces, their effects are independent. In one case, however, out data are inconsistent with this trend. 16 and 26 are far apart in both sequence and space (although they are closer in space than 19 and 26) and project onto opposite faces of the β-sheet (Figure 6e). **16/26p** (which residue 16 is replaced with Asn and 26 is replaced with

Asn-PEG) and **16p/26p** (in which 16 and 26 are replaced Asn-PEG), together with **16** and **16p** (from Chapter 2) allow us to probe interactions between PEG moieties conjugated at 16 and 26.

When **16** (which has Asn at positions 16 and 26) is PEGylated at 26 (affording **16/26p**), it is stabilized by -0.68±0.03 kcal mol^{-1}. However, when **16p** (which has Asn-PEG at position 16 and Asn at position 26) is PEGylated at 26 (affording **16p/26p**), it is only stabilized by -0.35±0.02 kcal mol^{-1}. This is anti-synergistic effect is similar to that observed between PEG moieties conjugated at positions 16 and 19; however in this case, the orientations of the two side-chains make it unlikely that the two PEG moieties are trying to interact with the same region on the protein.

This is a puzzling result, but paradoxically it also lends credence to the hypothesis that PEG acts primarily on the folded state. Comparing the results of the 19/26 double mutant cycle with 16/26 results, we see that PEG at 19 does not affect the action of PEG at 26, but PEG at 16 does. Figure 6 shows the relative locations in the folded peptide of these three residues. Although, in sequence 16 is further from 26 than is 19, it is actually closer in space in the folded state. If PEG is acting on the folded peptide, then two PEG residues would only interfere with one another if they were close in the folded state. If, as Rodriguez-Martinez et al.[6] have proposed, PEG stabilizes folded proteins by reducing conformational dynamics, then two PEG moieties, located near in space but on opposite faces of a protein, may be stabilizing the same region, in this case from opposite sides. If one site is already PEGylated, a second PEG conjugated nearby on the same face of the peptide would be sterically excluded from interacting with the same site, and would not stabilize the protein to the same degree as it otherwise would. This may serve to explain the observation that adding additional PEG polymers (using a non-site specific method) to α-Chymotrypsin increased the stability of the protein, but that each successive PEG polymer increased the stability less than the previous. This is consistent with a model in which PEG polymers are competing to stabilize the same regions of the protein, perhaps through a reduction in protein conformational dynamics.

3.2.4 Probing for Specific Contacts

Although it seems likely that PEG stabilizes the folding of WW by lowering the free energy of the folding state, the mechanism by which this is accomplished is still elusive. One possibility is that PEG is hydrogen bonding with nearby side-chain OH groups. Surface hydrogen bonding is not thought to be a major driving force of protein folding,[7] since the formation of a protein-protein hydrogen bond displaces a (roughly) equivalent water-protein hydrogen bond. However, there is some evidence that these interactions might play a minor role in protein folding thermodynamics. Pokkuluri et al.[8] demonstrated that, for human immunoglobulin light chain variable domain (V_L), addition of hydrogen bonding pairs to loops in the protein surface increased the thermodynamic stability by as much as 2.7 kcal mol^{-1}, while Yamagata et al.[9] have reported contributions of up to 1.8 kcal mol^{-1} for surface hydrogen bonds.

PEG hydrogen bonds readily to water, so it is plausible that it might also bond to nearby polar side chains on the protein surface. If this is the mechanism by which PEG stabilizes the WW domain, it should be possible to remove the relevant hydrogen-bond contacts through mutagenesis, and observe a corresponding decrease in the effectiveness of PEGylation.

Position 19 is located in the N-terminal reverse turn, and is oriented back toward the protein surface. Four plausible hydrogen bonding partners for a PEG polymer installed at position 19 are Ser16, Tyr23, Arg21, and Ser32, as shown in Figure 8a. Price et al.[10] tested the hypothesis that PEG was interacting with the Arg at position 21 by replacing it with Thr. The resulting protein was stabilized by -0.68 ± 0.10 kcal mol^{-1} upon PEGylation. Recently, our lab determined that replacing this residue with Ala or Leu caused PEGylation to stabilize the protein by -0.96 ± 0.12 kcal mol^{-1} and -0.68 ± 0.11 kcal mol^{-1}, respectively.[3] These experiments seem to indicate that PEG does not engage in hydrogen-bonding interactions with the side chain of residue 21. However, it should be noted that changing Arg to Thr, Ala, or Leu involves more than simply removing a hydrogen bond contact, and it would be difficult to isolate the effects of hydrogen bonding with any other effects which may be present.

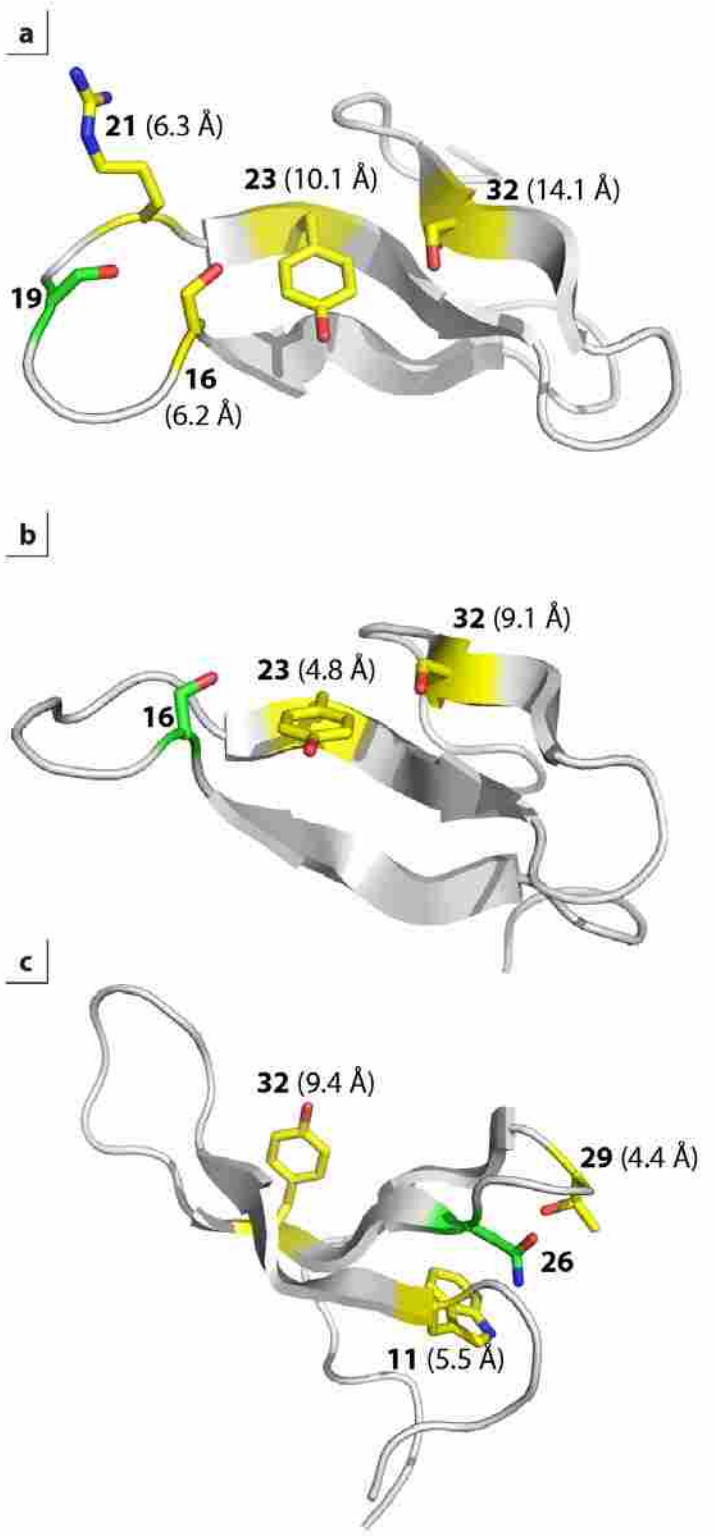


Figure 77: Sites at which PEGylation is stabilizing (green) and nearby polar residues (yellow) which may be hydrogen-bonding with PEG. Numbers in parentheses indicate the distance to the stabilizing residue, measured from the α-carbon to the –OH or $–NH_3$

To probe the possible interactions between the PEG polymer at 19 and the –OH groups of the remaining three residues, we made mutants of the WW domain in which these residues were replaced with nonpolar analogues. Figure 9 shows the sequences of WW mutants generated to probe for hydrogen bond contacts between PEG and nearby polar residues. **19/16(S→A)** and **19/32(S→A)** are analogous to **19**, but with Ala replacing Ser at positions 16 and 32, respectively. **19/23(Y→F)** is likewise analogous to **19**, but with Tyr replacing Phe at position 23. **19p/16(S→A), 19p/32(S→A)**, and **19p/23(Y→F)** are the corresponding PEGylated peptides. The sequences for these peptides, as well as all others detailed in this section, are shown in Figure 9.

Table 4 shows the results of these mutations. PEGylation at 19 (cf **19** vs **19p**) stabilizes the WW domain by -0.74±0.05 kcal mol^{-1} (at 60°C; because these experiments deal with multiple non-PEGylated peptides, a fixed temperature was chosen to evaluate all free energies of folding in this section). When Ser16 is mutated to Ala, this stabilization is reduced to ΔG_f=-0.55±0.03 kcal mol^{-1}. Mutating Tyr32 to Phe likewise attenuates the effects of PEGylation (-0.41±0.04 kcal mol^{-1}).

In contrast, replacing Ser32 with Ala has no effect on the stabilization of PEG (ΔG_f=-0.74±0.03 kcal mol^{-1}). This is consistent with the earlier observation that a PEG trimer is significantly more stabilizing than a dimer. Figure 10 shows the lengths of PEG monomers, dimers, trimers, and tetramers conjugated to Asn. The dimer is only likely to interact directly with residues within 9 Å of the side chain to which it's attached, while the trimer can reach up to 13 Å. Since the tetramer and all larger PEG polymers are equally or less stabilizing than the trimer, we can conclude that any regions which interact with PEG (attached at position 19) are probably within this 13 Å radius of position 19. This is what we observe. Polar side chains within this radius (Ser16, 6.2 Å; Tyr23, 10.1 Å, see Figure 8a) affect the magnitude of PEG-induced stabilization, while Ser32, which is 14.1 Å away, does not.

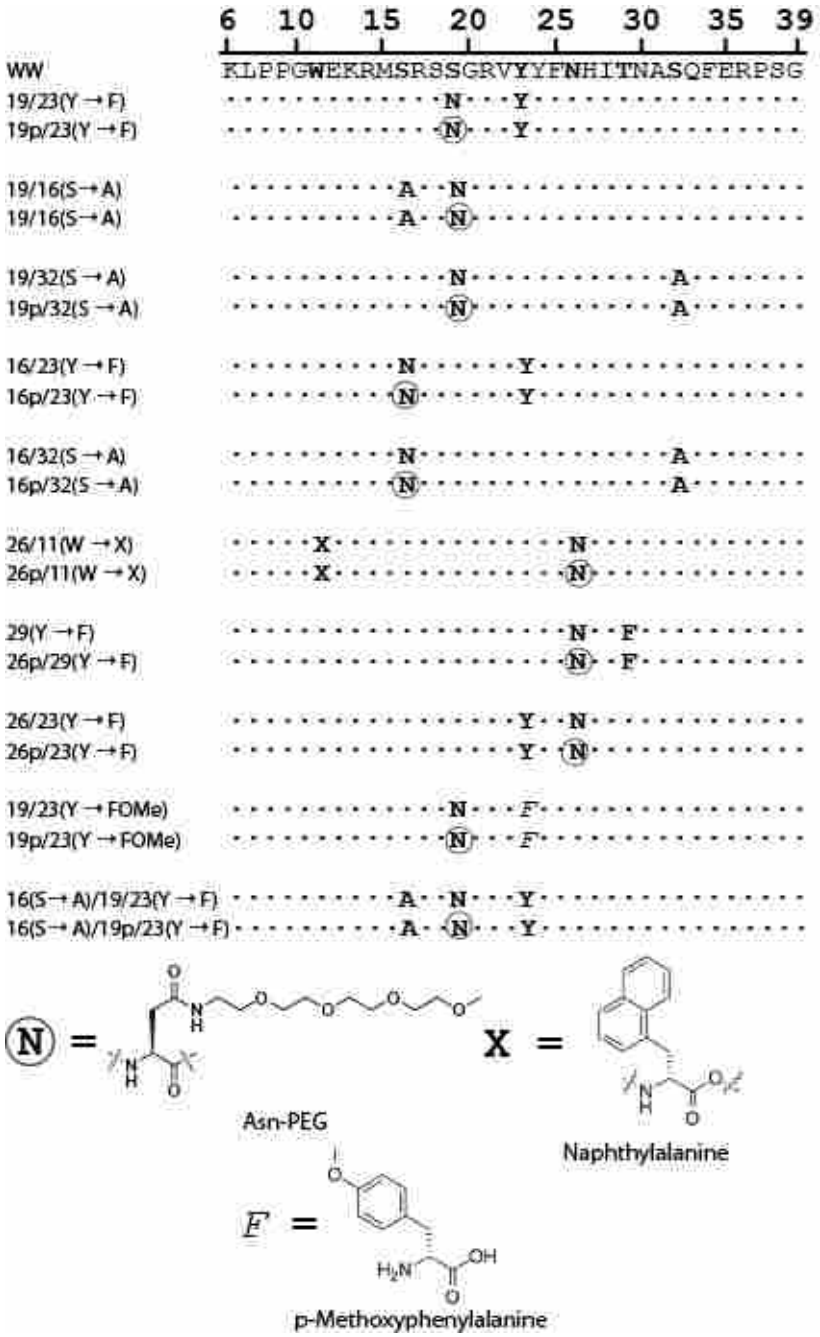


Figure 78: Sequences of **19/23(Y→F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 19/32(S→A), 19p/32(S→A), 16/23(Y→F), 16p/23(Y→F), 16/32(S→A), 16p/32(S→A), 11(W→X), 26p/11(W→X), 29(T→A), 26p/29(T→A), 23(Y→F), 26p/23(Y→F), 16(S→A)/19/23(Y→F),**

Table 5: Melting temperatures (T_m) and changes in folding free energy ($\Delta\Delta G_f$ / kcal mol^{-1}) PEGylated- and non-PEGylated- WW variants containing amino acid substitutions near the site of PEGylation.[a]

Protein	T_m (°C)	ΔG_f (kcal/mol)[b]	Protein	T_m (°C)	ΔG_f (kcal/mol)
19[c]	55.6 ± 0.3		**16**[c]	54.8 ± 0.2	
19p[c]	63.2 ± 0.3	-0.74 ± 0.05	**16p**[c]	62.3 ± 0.2	-0.75 ± 0.03
19/23(Y→F)	51.4 ± 0.4		**16/23(Y→F)**	50.7 ± 0.7	
19p/23(Y→F)	55.0 ± 0.2	-0.41 ± 0.04	**16p/23(Y→F)**	56.2 ± 0.3	-0.48 ± 0.08
19/16(S→A)	51.0 ± 0.2		**16/32(S→A)**	55.1 ± 0.3	
19p/16(S→A)	56.8 ± 0.2	-0.55 ± 0.03	**16p/32(S→A)**	61.4 ± 0.1	-0.58 ± 0.03
19/32(S→A)	54.4 ± 0.3				
19p/32(S→A)	62.7 ± 0.1	-0.74 ± 0.03	**WT**[c]	58.3 ± 0.8	
19/23(Y→FOMe)	55.0 ± 0.2		**26p**[c]	64.5 ± 0.2	-0.54 ± 0.08
19p/23(Y→FOMe)	60.7 ± 0.2	-0.53 ± 0.03	**29(T→A)**	40.4 ± 0.7	
16(S→A)/19/23(Y→F)	45.8 ± 1.1		**26p/29(T→A)**	45.1 ± 0.4	-0.33 ± 0.11
16(S→A)/19p/23(Y→F)	53.7 ± 0.3	-0.63 ± 0.13	**23(Y→F)**	53.2 ± 0.3	
			26p/23(Y→F)	59.8 ± 0.6	-0.63 ± 0.06
			11(W→X)	52.0 ± 0.4	
			26p/11(W→X)	58.4 ± 0.2	-0.61 ± 0.04

[a]Tabulated data are given as mean ± standard error 100 μM solutions of WW variants in 20 mM sodium phosphate buffer (pH 7). [b] ΔG_f is reported at 60°C. [c] Peptides from Chapter 2 are included for comparison

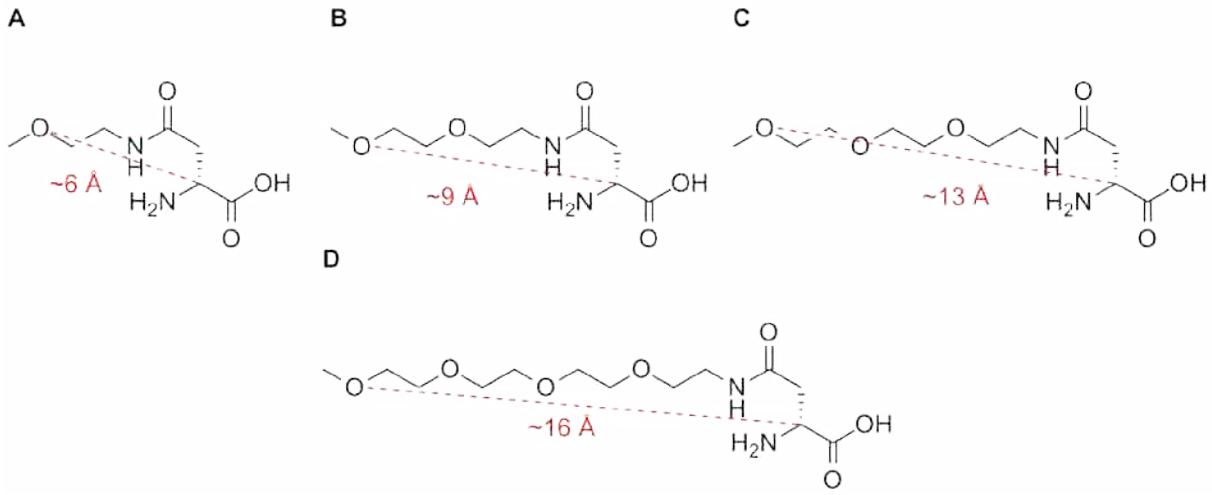


Figure 79: Length of a) Asn(PEG_1), b) Asn(PEG_2), c) Asn(PEG_3), and d) Asn(PEG_4), measured from the α-carbon of the amino acid to the most distant oxygen in the PEG chain, as indicated in red.

PEGylation at position 16 is also stabilizing ($\Delta\Delta G_f$=-0.75±0.03 kcal mol^{-1}), and Tyr23 and Ser32 may contribute to this (Figure 8b). **16/32(S→A)** and **16/23(Y→F)** are analogues of **16** in which Ser32 has been replaced with Ala, and Tyr23 has been replaced with Phe, respectively. **16p/32(S→A)**, and **16p/23(Y→F)** are the corresponding PEGylated compounds. The results of these mutations are shown in Table 4. As with PEGylation at position 19, PEGylation at position 16 is less effective when Tyr23 is replaced by Phe. When Tyr23 is replaced with Phe, PEGylation at 16 only stabilizes the peptide by -0.48±0.08 kcal mol^{-1}. This suggests that the –OH at position 23 plays a role in PEG-mediated stabilization.

While PEGylation at position 19 was not affected by the Ser32Ala mutant, 16 is closer to position 32 (9.1 Å) than is 19. When **16/32(S→A)** is PEGylated (affording **16p/32(S→A)**), the change in folding free energy is -0.58±0.03 kcal mol^{-1}, indicating that PEG, when conjugated at position 16, interacts with the –OH at position 32.

The third stabilizing site which we investigated was position 26. In the wild-type peptide, **WW**, residue 26 is Asn. PEGylating this position stabilizes the peptide by -0.50 ± 0.07 kcal mol^{-1}. Because a PEG monomer stabilizes this position as much as a tetramer, any interacting partners must be closer than 6 Å. There are there are two polar residues within 6 Å of position 26: Trp11 and Thr29. **11(W→X)** and **29(T→A)** are analogues of **WW** in which Trp11 has been replaced with naphthylalanine, (a non-natural tryptophan analog, see Figure 9) and Thr29 has been replaced with Ala, respectively. Table 4 summarizes the results of these mutants. The W11X mutation does not affect the stabilizing properties of PEGylation at position 26 ($\Delta\Delta G_f$=-0.61±0.04 kcal mol^{-1}), while the T29A mutation does ($\Delta\Delta G_f$=-0.33±0.11 kcal mol^{-1}). The hydroxyl group at position 29 appears to play an important role in PEG-induced stability at position 26

The stabilizing effect of PEGylation appears to depend on the identity of nearby hydroxyl groups; removing the group(s) reduced the effect of PEGylation. We wanted to know if the hydroxyl group needed to be on the face as the PEG oligomer; that is, whether the hydroxyl group needed to be physically

accessible to the PEG oligomer, or if removing the –OH somehow affected the local structure of the peptide such that PEG was less intrinsically stabilizing.

Tyr23 and Asn26 are close in both sequence and space in the folded peptide, but their respective side chains project onto opposite faces of the β-sheet, as shown in Figure 8c. **23(Y→F)** is analogous to the wild type peptide **WW**, but with Tyr23 replaced with Phe. **26p/23(Y→F)** is the corresponding PEGylated peptide. As indicated above, **WW** and **26p** differ in free energy by -0.54±0.08 kcal mol^{-1}. The change in folding free energy upon PEGylation of **23(Y→F)** is -0.63±0.06 kcal mol^{-1} (a difference that is not statistically significant), indicating that removing a nearby hydroxyl group reduces the effect of PEGylation only when the hydroxyl group is on the same face as the PEG. This suggests that PEG is interacting with the side chains of polar residues, although the nature of such an interaction is unclear.

From these experiments, it seems likely that PEG increases WW folding stability by engaging in favorable interactions with the surface of the protein in the folded state. The hydroxyl groups of nearby polar residues appear to be important - every stabilizing site is oriented back toward a residue with a hydroxyl group, and the removal of these hydroxyl groups attenuates the stabilizing effect of PEGylation. However, it's unclear whether PEG is directly hydrogen bonding with surface hydroxyl groups, or whether the interaction is more complicated - for example if it were water-mediated.

One alternative explanation for the Y23F mutants is that removing the hydroxyl moiety altered the electronic properties of the phenyl ring. To see if this were the case, we prepared two more WW mutants, **19/23(Y→FOMe)** and **19p/23(Y→FOMe)**. In both cases, Tyr23 is replaced with p-methoxyphenylalanine (a non-natural analogue of tyrosine, but with a methyl ether instead of a hydroxyl moiety, as shown in Figure 8. When **19/23(Y→FOMe)** is PEGylated, it is stabilized by -0.53±0.03 kcal mol^{-1}. The Y23FOMe mutation reduces the effect of PEGylation to a similar degree as does the Y23F mutation. This suggests that PEGylation is most effective when the residue at position 26 is capable of hydrogen bonding.

3.2.5 Triple Mutant Cycle Analysis

It seems evident from the previous experiments that PEG needs to be oriented toward nearby hydroxyl groups; removal of these groups reduces the thermodynamic impact of PEGylation. We wondered if PEG was hydrogen bonding directly with specific hydroxyl groups, or if the presence of specific hydroxyl groups altered the stereoelectronic properties of the protein surface; as has been suggested, PEG may sweep water away from the protein surface, and the energetic impact of this dehydration may depend on the presence of nearby hydroxyl groups.

If we assume that PEG hydrogen bonds directly with hydroxyl groups on the protein surface, we can use triple mutant cycle analysis[10,11] to isolate the contributions to thermal stability from each interaction. On the other hand, if the effects of PEG are more complicated than direct hydrogen bonding with specific side chains, then the individual contributions of specific hydrogen bonding partners would not be easily parsed into binary or ternary interactions.

16(S→A)/19/23(Y→F) is a WW mutant in which Ser16 is replaced by Ala, Ser19 is replaced by Asn, and Tyr23 is replaced by Phe. This peptide represents the "baseline" of the triple mutant cycle, as it is the peptide without PEG and with no hydroxyl groups of interest. **16(S→A)/19p/23(Y→F)** is analogous to **16(S→A)/19/23(Y→F)**, but with Asn-PEG at position 19. As indicated in Table 4, the difference in folding free energy between these two peptides is -0.63±0.13, which is statistically indistinguishable from the stabilization observed upon PEGylation of **19** (to afford **19p**, $\Delta\Delta G_f$=-0.74±0.05 kcal mol^{-1}).

The contributions from single mutations (i.e. the effect of adding a single –OH at positions or PEGylating position 19), binary interactions (i.e. hydrogen bonds between a single –OH and PEG), and a ternary interaction between two –OH groups and PEG can be expressed according to the equation:

$$\Delta G_f = \Delta G_f^o + W_S \cdot C_{A\rightarrow S} + W_{\mathbf{N}} \cdot C_{N\rightarrow \mathrm{N}} + W_Y \cdot C_{F\rightarrow Y} +$$

$$(W_S W_{\mathbf{N}}) \cdot C_{S,\mathbf{N}} + (W_{\mathbf{N}} W_Y) \cdot C_{\mathbf{N},Y} + (W_S W_Y) \cdot C_{S,Y} + (W_S W_{\mathbf{N}} W_Y) \cdot C_{S,\mathbf{N},Y} \quad (1)$$

In Equation 1, ΔG_f is the folding free energy for a given variant of **19** and ΔG_f^o is the average folding free energy of **16(S→A)/19/23(Y→F).** $C_{S\to A}$, $C_{N\to \mathbf{N}}$, and $C_{Y\to F}$ represent the contribution to thermodynamic stability made by mutating Ser16 to Ala, Asn19 to Asn(PEG), and Phe23 to Tyr. $C_{S,\mathbf{N}}$, $C_{\mathbf{N},Y}$, and $C_{S,Y}$ represent the contribution to thermodynamic stability due to binary interactions between Ser16 and PEG, PEG and Tyr23, and Ser16 and Tyr23, respectively. $C_{S,\mathbf{N},Y}$ represents the contribution to thermodynamic stability due to a ternary interactions between Ser16, PEG, and Tyr23. W_S = 0 when position 16 is Ala or 1 when it is Ser. $W_{\mathbf{N}}$ = 0 when position 19 is Asn or 1 when it is Asn(PEG). W_Y= 0 when position 23 is Phe or 1 when it is Tyr.

Figure 12a shows how **16(S→A)/19/23(Y→F)**, **16(S→A)/19p/23(Y→F),** along with **19/23(Y→F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 19**, and **19p** comprise a triple mutant cycle, the results of which are summarized in Figure 12b. Full fitting parameters are found in Table 5

One PEGylation is intrinsically stabilizing; however, there are *unfavorab*le binary interactions, both between PEG and Ser16 and between PEG at Tyr23, which are almost completely cancelled out by a favorable three-way interaction between PEG, Ser16, and Tyr23. This would imply that PEGylation is inherently stabilizing, but that at some sites nearby side chains interact with PEG to disrupt this stabilizing effect.

An alternative explanation is that one or more of the assumptions underlying the analysis are incorrect. Equation 1 assumes that Ala16, Asn19, and Phe23 have no significant interactions, either with each other or with PEG. It also assumes that any thermodynamic consequences of PEGylation are due to either to intrinsic effects of PEGylation or to binary or ternary interactions with nearby polar side chains. If PEG acts by a more complicated mechanism, such as by releasing water from the solvation shell into the bulk solvent, its effects could not be parsed into such simple interactions.

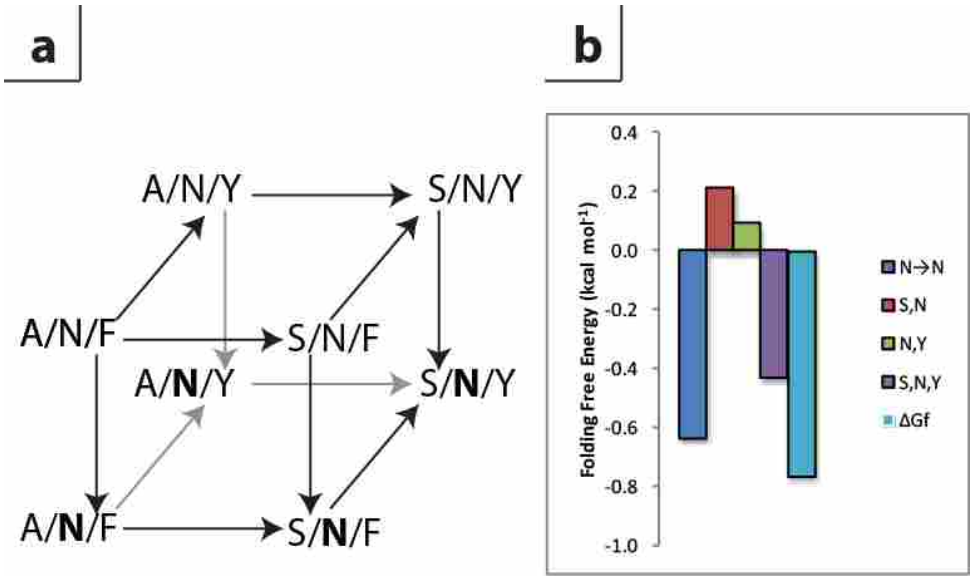


Figure 80: a) Graphical representation of a triple mutant cycle analysis. **16(S→A)/19/23(Y→F)** contains Ala at position 16 , Asn at position 19, and Phe at position 23. Replacing Ala16 with Ser and Phe23 with Tyr in every possible combination affords **19/16(S→A), 19/23(Y→F),** and **19.(S→A)/19p/23(Y→F), 19p/16(S→A), 19p/23(Y→F),** and **19p** are the corresponding proteins PEGylated at position 19. B) Summary of triple mutant cycle results. The dark blue bar represents the intrinsic energetic consequences of PEGylation. The red, and green bars represent the energetic consequences of two-way interactions between PEG and Ser16 and between PEG and Tyr23, respectively. The purple bar represents the energetic consequences of a three-way interaction between PEG, Ser16 and Tyr23. The light blue bar represents the overall change in free energy of folding upon PEGylation.

Table 6: Triple mutant cycle analysis parameters

Standard Parameters:[a]	**$\Delta G°_f$**	**$\Delta H°_f$**	**$-T\Delta S°_f$**	
	1.22	-30.0	31.2	
Parameter	**$\Delta\Delta G_f$**	**$\Delta\Delta H_f$**	**$-T\Delta\Delta S_f$**	**Physical Significance**
$C_{A\rightarrow S}$	-0.46	1.1	-1.6	intrinsic impact of S16A mutation
$C_{N\rightarrow N}$	-0.63	-0.9	0.2	intrinsic impact of PEGylation of Asn19
$C_{F\rightarrow Y}$	-0.37	-1.1	0.8	intrinsic impact of F23Ymutation
$C_{S,N}$	0.21	-2.4	2.7	two-way interaction between Ser16 and PEG
$C_{S,Y}$	0.05	-4.3	4.3	two-way interaction between Ser16 and Tyr23
$C_{N,Y}$	0.09	-0.3	0.4	two-way interaction between Tyr23 and PEG
$C_{S,N,Y}$	-0.43	6.8	-7.3	three-way interaction between Ser16, Tyr23, and PEG

[a]Standard parameters are for **16(S→A)/19/23(Y→F)**, which has no –OH groups at positions 16 or 23.

3.2.6 Entropic and Enthalpic Components of PEG-Induced Stabilization

Further insight into the mechanism by which PEG stabilizes the WW domain can be gained by understating how PEGylation affects entropic and enthalpic components of the free energy of folding. The entropic and enthalpic components of ΔG_f for **19, 19p,19/23(Y→F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 16(S→A)/19/23(Y→F),** and **16(S→A)/19p/23(Y→F)** were globally fit to obtain ΔH_f, $-T\Delta S_f$, and T_m, as described in section 3.7.7. The results are shown in Table 6 (see section 3.7.7 for a complete list of parameters obtained from the fits. Peptide **19** has a Ser at position 16 and a Tyr at position 13. When both –OH groups are present, PEGylation at position 19 (to give **19p**) is enthalpically disfavored ($\Delta\Delta H_f$= 3.2 ± 1.4 kcal mol^{-1}) and entropically favored ($-T\Delta\Delta S_f$= -4.0±1.4 kcal mol^{-1}). However, when either of the –OH groups, or both, are removed, PEG appears to stabilize the peptide through a different mechanism; in all cases, addition of PEG is entropically disfavorable and enthalpically favorable.

This experiment demonstrates that the thermodynamic consequences of PEGylation originate from mechanisms more complex than simple binary or ternary interactions with nearby side chains. It is clear that PEG alters both the entropic and enthalpic properties of the folded state, and that the identity of nearby side chains strongly influences the balance of these two effects. One possible mechanism which we are currently exploring is that Ser16 and Tyr23 bind water molecules to the protein surface. When the peptide is PEGylated, the polymer may sweep away these molecules, releasing them to the bulk solvent. This is consistent with the model proposed by Meng et al,[12] who suggested that PEG acts to decrease the solvent accessible surface area of the SH3 domain.

Table 7: Entropic and Enthalpic Components of Folding Free Energies

	ΔG_f [a] (kcal mol^{-1})	$\Delta\Delta G_f$ [a] (kcal mol^{-1})	ΔH_f [a] (kcal mol^{-1})	$\Delta\Delta H_f$ [a] (kcal mol^{-1})	$-T\Delta S_f$ [a] (kcal mol^{-1})	$-T\Delta\Delta S_f$ [a] (kcal mol^{-1})	ΔH_f or $-T\Delta S_f$ Predominates?
19	0.44±0.02		-34.3±0.7		34.7±0.7		
19p	-0.32±0.03	-0.76±0.04	-31.0±1.2	3.2±1.4	30.7±1.2	-4.0±1.4	Entropy
19/23(Y→F)	0.75±0.04		-28.9±0.9		29.6±0.9		
19p/23(Y→F)	0.33±0.01	-0.42±0.04	-32.2±0.4	-3.3±1.0	32.5±0.4	2.9±1.0	Enthalpy
19/16(S→A)	0.85±0.03		-31.1±0.7		32.0±0.7		
19p/16(S→A)	0.31±0.01	-0.54±0.03	-32.3±0.3	-1.2±0.8	32.6±0.3	0.6±0.8	Enthalpy
16(S→A) /19/23(Y→F)	1.22±0.13		-30.0±2.7		31.2±2.8		
16(S→A) /19p/23(Y→F)	0.58±0.03	-0.63±0.13	-30.9±0.6	-0.9±2.8	31.5±0.6	0.2±2.9	Enthalpy

[a]Tabulated data are given as mean ± standard error 100 μM solutions of WW variants in 20 mM sodium phosphate buffer (pH 7) at 60°C.

3.3 Conclusions

Improvements in site-specific methods of PEGylation have generated an increased interest in choosing the optimum site for PEG conjugation. Most efforts to site-specifically PEGylate a protein have focused on avoiding PEGylation near the active site; one questions which has not yet been asked is how the choice of PEGylation site affects the thermodynamic consequences of PEGylation. Many challenges faced by protein therapeutics are inherently related to protein thermodynamic stability; an understanding of the mechanism by which PEGylation stabilizes proteins and the structural determinants of PEG-induced stabilization will enable chemists to choose PEGylation sites which are maximally stabilizing.

In order to increase the free energy of folding of a protein (meaning the difference in free energy between the native and denatured state(s)), PEG must either increase the energy of the denatured state or decrease the energy of the native state. Here, we have demonstrated that PEGylation likely stabilizes the WW domain by lowering the free energy of the native state.

As detailed in Chapter 2, PEGylation stabilizes the WW domain in a site-dependent manner. The sites at which PEGylation is stabilizing are all oriented back toward the protein surface. When position 19 is replaced with a D-amino acid (which likely projects out into solution instead of pointing back toward the protein surface) PEGylation ceases to be stabilizing. It is likely that PEG must be oriented correctly in the folded peptide to stabilize; this is inconsistent with a model where PEG acts primarily on the denatured state. Furthermore, when two PEG oligomers are installed, they interfere with each other in a pattern that is consistent with their relative positions and orientations in the folded state, but not with their relative sequences in the denatured state.

Finally, the effect of PEGylation depends strongly upon the identity of nearby surface residues. When nearby –OH groups are removed, PEGylation is less effective. However, when –OH groups that are close in sequence but not in spatial orientation are removed, the effect of PEGylation is unperturbed. This indicates that PEG stabilizes the peptide by interacting with the specific regions of the surface. This model is consistent with the work of Svergun et al.[13] who observed that PEG tends to at least partially

cover the surface of hemoglobin. From the results we have gathered thus far, it seems clear that PEG must be conjugated to a residue that points back toward the protein surface, and that nearby surface must contain residues with –OH moieties.

There are several possible mechanisms by which PEG could be stabilizing the native state. A triple mutant cycle analysis revealed that it is unlikely that PEG is engaging in simple binary or ternary interactions with hydroxyl groups on the protein surface. One possibility, suggested by Meng et al,[12] is that PEG liberates water in the hydration shell into the bulk solvent. This would be consistent with our observation that PEGylation of **19** (to afford **19p**) is entropically favorable. Why nearby –OH groups would contribute to this stabilizing effect is unclear, but they may strongly bind water molecules which are released upon PEGylation.

3.4 Experimental Procedures

3.4.1 Protein Synthesis

All proteins were synthesized as C-terminal acids by microwave-assisted solid-phase peptide synthesis[14] using a standard Nα protection strategy. Amino acids were activated by 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU, purchased from Advanced ChemTech) and *N*-hydroxybenzotriazole hydrate (HOBt, purchased from Advanced ChemTech). Fmoc-Gly-loaded Wang LL resin was purchased from EMD Biosciences. Fmoc-protected α-amino acids (with acid-labile side-chain protecting groups) were purchased from Advanced ChemTech, except for Fmoc-Asn(PEG4)-OH, which was synthesized as reported in Chapter 2. and Fmoc-D-Asn(PEG)-OH, which was synthesized as described in Section 3.7.2. Piperidine, *N,N*-diisopropylethylamine (DIEA), and *N*-methyl pyrrolidinone (NMP) were purchased from Advanced ChemTech.

Acid-labile side-chain protecting groups were globally removed and proteins were cleaved from the resin by stirring 50 μmol resin for ~4 h in a solution of phenol (250 mg), water (250 μL), thioanisole (250 μL), ethanedithiol (125 μL), and triisopropylsilane (50 μL) in trifluoroacetic acid (TFA, 4 mL). Proteins were purified by preparative reverse-phase high-performance liquid chromatography (HPLC) on

a C18 column using a linear gradient of water in acetonitrile with 0.1% v/v TFA. Proteins were characterized by electrospray-ionization time-of-flight mass spectrometry (See Figures 24-56) and protein purity was confirmed by analytical HPLC (See Figures Figures 60-91).

3.4.2 Circular Dichroism

Variable temperature CD measurements were made with an Aviv 420 Circular Dichroism Spectropolarimeter at 227 nm, from 1 to 95 °C (at 2 °C intervals), with 120 s equilibration time between data points and 30 s averaging times, using quartz cuvettes with a path length of 0.1 cm (See Figures 93-108). Protein concentrations were determined spectroscopically based on tyrosine and tryptophan absorbance at 280 nm in 6 M guanidine hydrochloride in 20 mM sodium phosphate.[15] The melting temperature and free energy of folding data were obtained by globally fitting the variable temperature CD data to equations for two-state thermal unfolding transitions (see Chapter 2 Supporting Information for details).

3.5 Supporting Information

Protein synthesis, purification and characterization (including HPLC, ESI-TOF MS, and CD spectropolarimetry were performed using identical conditions to those reported in Chapter 2.

3.5.1 Synthesis of PEGylated Fmoc-Protected Asparagine

The synthesis of Fmoc-Asn(PEG_4)-OH was described in Chapter 2. The syntheses of Fmoc-Asn-(PEG_1)-OH, Fmoc-Asn-(PEG_8)-OH, and Fmoc-Asn-(PEG_{45})-OH are described in ref 3.

3.5.2 Synthesis of PEGylated Fmoc-Protected D-Asparagine

*(R)-3-((((*9H-*fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert-*butoxy)-4-oxobutanoic acid* **1** (Fmoc-Asn(PEG)-OtBu)

Procedure

(Fmoc-D-Asn(PEG)-OtBu) was synthesized following a procedure analogous to that of Herzner and Kunz[16]: to a solution of (R)-3-((((*9H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert*-butoxy)-4-oxobutanoic acid (Fmoc-D-Asp-OtBu, 1.0g, 2.430 mmol) in dry dichloromethane (50 mL) was added isobutyl (2-isobutoxy)-1,2-dihydroquinoline-1-carboxylate (IIDQ, 1.1g, 3.645 mmol), and the resulting mixture was stirred for 15 min at room temperature under an argon atmosphere. Then, 2-(2-(2-methoxyethoxy)ethoxy)ethanamine (0.5g, 2.430 mmol) was added, and stirring was continues for 24h. The reaction was then quenched with brine (50 mL) and washed with water (50 mL), and the organic extracts were dried with $MgSO_4$, filtered through celite, and concentrated by rotary evaporation to afford a yellow oil. The desired product was purified by flash chromatography over silica gel using ethyl acetate/hexanes (3:7 for ~1000 mL), followed by acetic acid (~2000 mL) then acetic acid/ethyl acetate (1:99 for ~1000 mL, 1:9 for ~2500 mL) as eluents. The product was concentrated via rotary evaporation (chloroform and benzene were employed to remove residual ethyl acetate and acetic acid) and dried in vacuo to give a thick oily solid (0.90 g, 1.6 mmol, 67% yield). Rf = 0.15 (1:100 acetic acid/ethyl acetate).

Analyticial Data

1H NMR (500 MHz, $CDCl_3$): δ 7.75(2H, d, J = 7.5 Hz, Fmoc aryl C-**H**); 7.62 (2H, t, J = 6.25 Hz, Fmoc aryl C-**H**); 7.39 (2H, t, J = 7.5 Hz, Fmoc aryl C-**H**); 7.30 (2H, t, J = 7.5Hz, Fmoc aryl C-**H**); 6.83 (1H, broad s, -CON**H**-CH_2-CH_2-O-); 6.24 (1H, d, J = 8.5 Hz, -CON**H**-CαH(COOH)-); 4.50 (1H, broad s, -CONHCα**H**(COOH)-CβH_2-); 4.40 (1H, dd, J = 10.5 Hz, 7.0 Hz, Fmoc Ar2CH-C$\mathbf{H_{(a)}}H_{(b)}$-O-); 4.30 (H, apparent t, Fmoc Ar2CH-$CH_{(a)}\mathbf{H_{(b)}}$-O-); 4.23 (1H, t, J = 7.0 Hz, Fmoc Ar2C**H**-CH_2-O-); 3.51-3.67 (14H, m, -CONH-CH_2-C$\mathbf{H_2}$-O-C$\mathbf{H_2}$-C$\mathbf{H_2}$-O-C$\mathbf{H_2}$-C$\mathbf{H_2}$-O--C$\mathbf{H_2}$-C$\mathbf{H_2}$-O-); 3.45 (2H, m, -CONH-C$\mathbf{H_2}$-CH_2-O-); 3.36 (3H, s, -O-C$\mathbf{H_3}$); 2.85-2.92 (1H, m, -CαH(COOH)-Cβ(**Ha**)Hb-CONH-); 2.72 (1H, dd, J = 15.5 Hz, 4.0 Hz, -CαH(COOH)-Cβ(Ha)**Hb**-CONH); 1.47 (9H, s, -O-C-(C$\mathbf{H_3}$)$_3$. 2.89, 2.97, 8.03 (dimethyl formamide contamination), 2.08 (ethyl acetate contamination), 7.36 (benzene). The full ^{1}H NMR spectrum for **1** is shown in **Figure 13.**

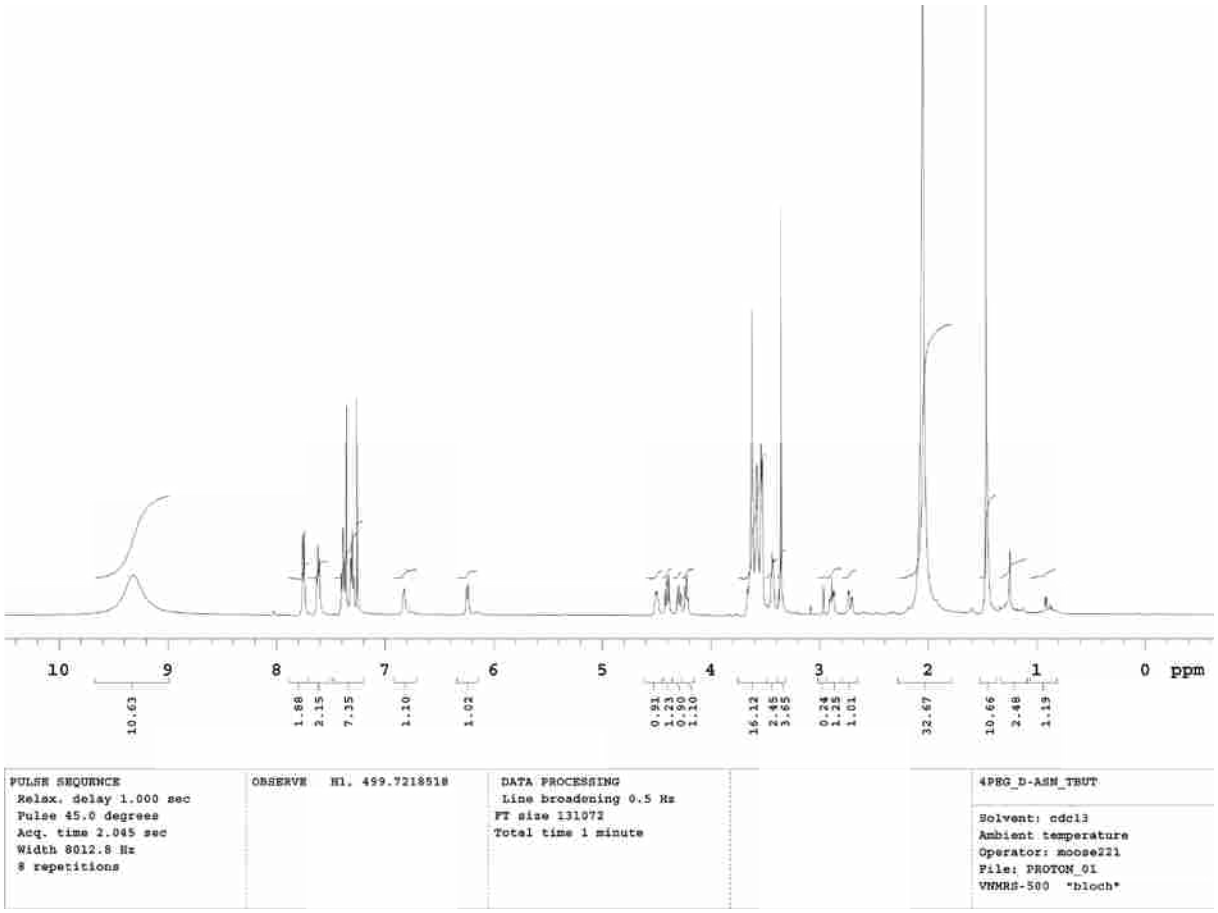


Figure 81: ^{1}H NMR for Fmoc-D-Asn(PEG)-OtBu

13C NMR (126 MHz, CDCl3): δ 170.21 (-NH-CαH(**C**OOH)-CβH_2- and/or -CβH_2-**C**ONH-CH_2-); 144.00, 141.25 (Fmoc aryl **ipso** C's); 128.30, 127.64, 127.05,125.12, 119.89 (Fmoc Ar **C**-H); 82.14 (-O-**C**$(CH_3)_3$); 71.74, 70.33, 70.28, 70.06, 69.75 (-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-); 67.10 (Fmoc Ar$_2$CH-**C**H_2-O-); 58.91 (-O-**C**H_3); 51.52 (-NH-**C**αH(COOH)-CβH_2-); 47.15 (Fmoc Ar2**C**H-CH_2-O-); 39.28 (CONH-**C**H_2-CH_2-O-); 37.79 (-CαH(COOH)-**Cβ**H_2-CONH-); 27.90 (-O-C(**C**$H_3)_3$). 21.18 (ethyl acetate). The full ^{13}C NMR spectrum for **1** is shown in Figure 14.

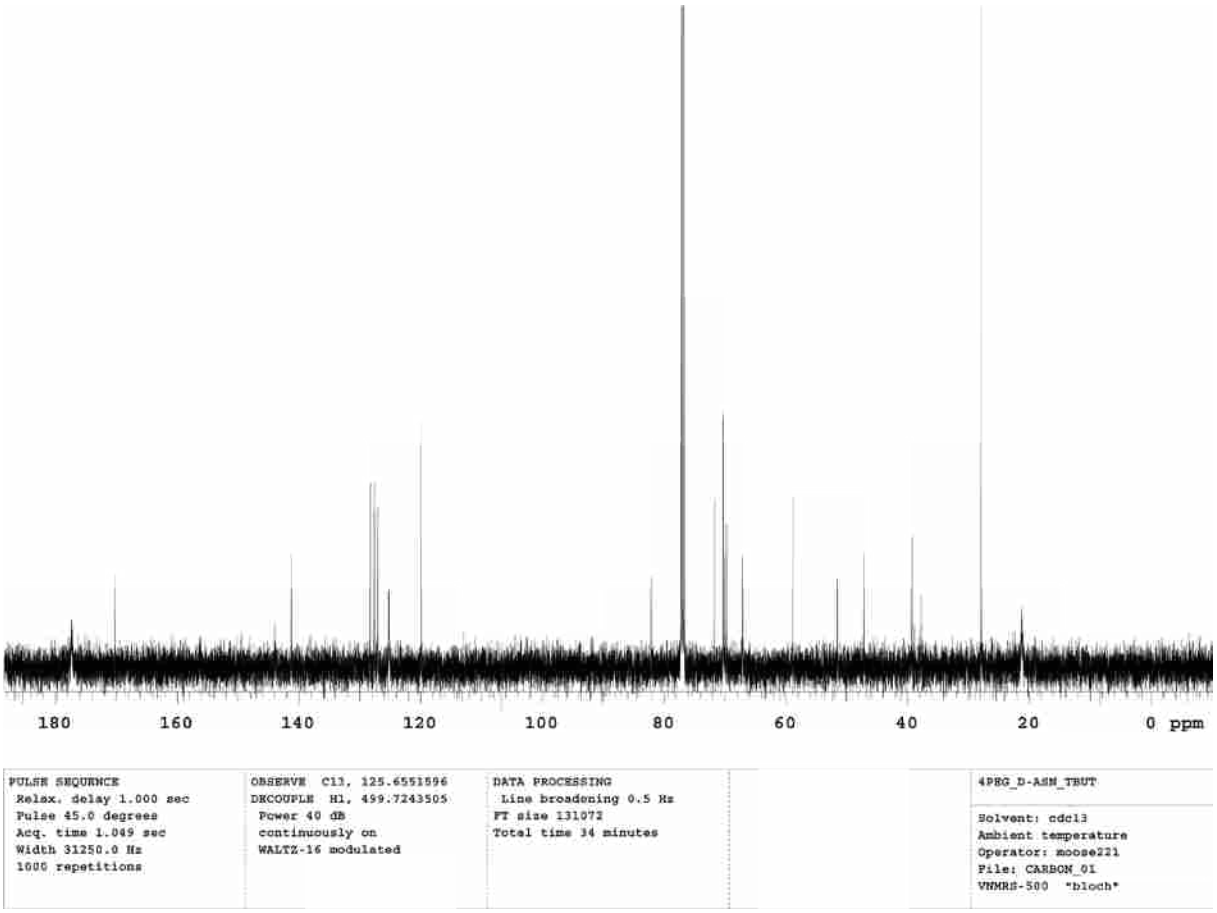


Figure 82: ^{13}C NMR for Fmoc-Asn(PEG)-OtBu

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-D-Asn(PEG)-OH were made using a 2D HSQC experiment (See Figure 15) to indentify the one-bond C-H correlations shown in Table 7 and by analogy with published spectral data for related compounds[3,10].

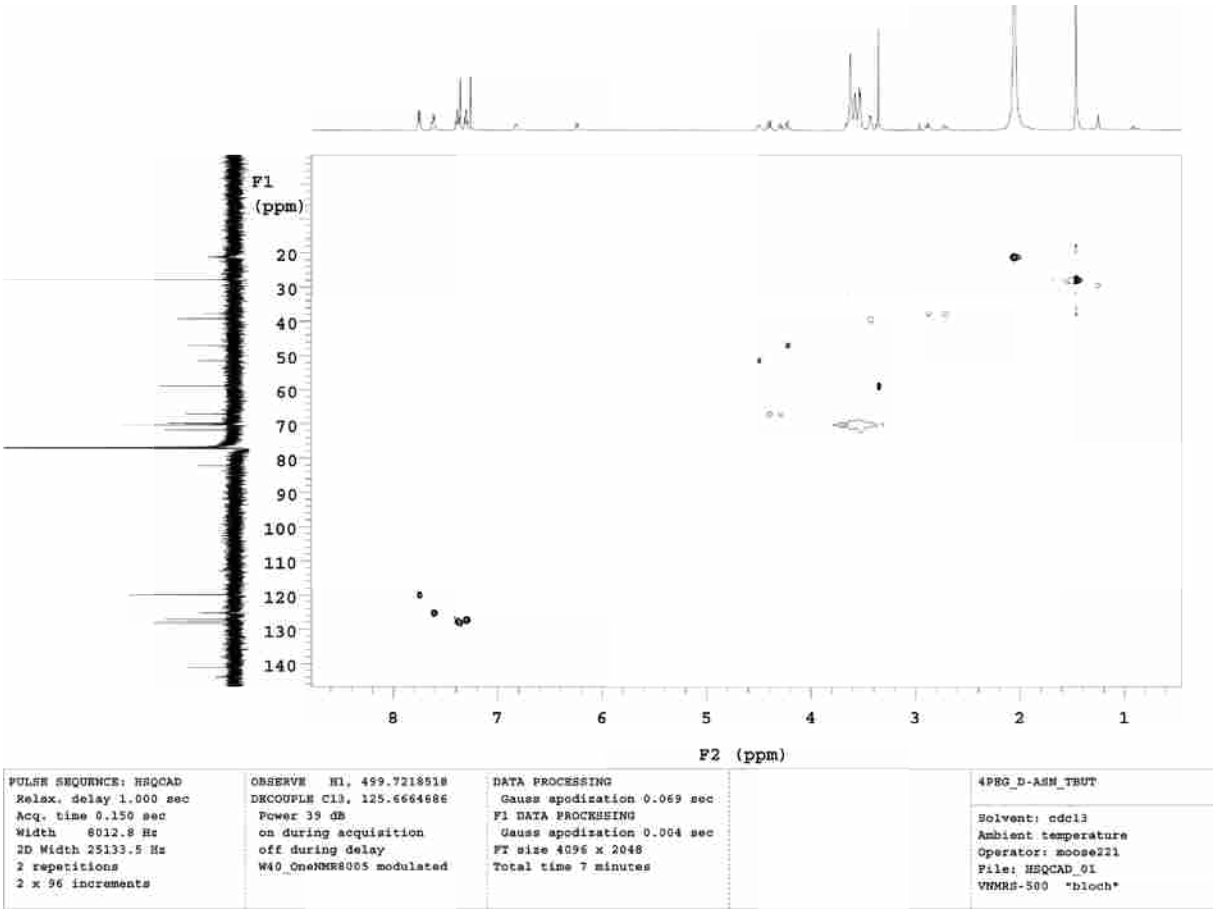


Figure 83: 2D HSQC spectrum of Fmoc-D-Asn(PEG)-OH

Table 8: One-Bond C-H correlations identified from HSQC experiment on Fmoc-D-Asn-PEG-OtBu **1**

^{1}H δ	^{1}C δ	Assignment
7.75	120.1	Fmoc aryl C-H
7.61	125.3	Fmoc aryl C-H
7.31	127.8	Fmoc aryl C-H
7.30	127.2	Fmoc aryl C-H
4.50	51.52	-CONH**CαH**(COOH)-CβH2-
4.40, 4.29	67.22	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
4.22	47.20	Fmoc Ar2**CH**-CH2-O-
3.51-3.67	69.10-72.70	C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O--C**H$_2$**-C**H$_2$**-O-
3.43	39.48	CONH-**CH$_2$**-CH$_2$-O
3.35	59.24	-O-**CH$_3$**
2.88, 2.72	37.81	-CαH(COOH)-**CβH$_2$**-
1.47	28.05	-C(**CH$_3$**)$_3$

High-resolution electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) is shown in Figure 16:

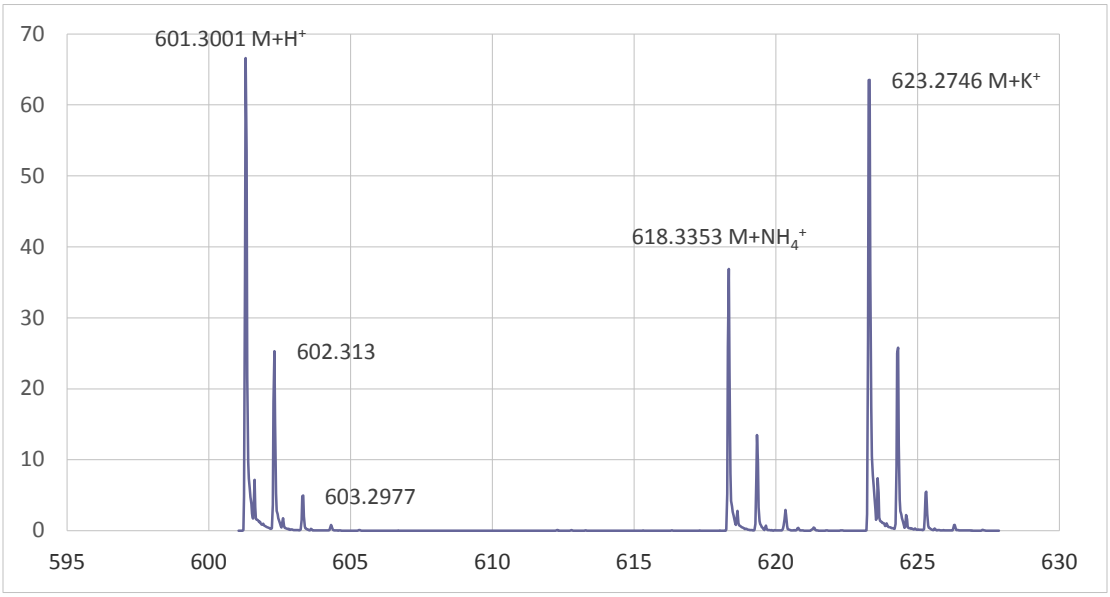


Figure 84: ESI-TOF MS data for (Fmoc-D-Asn(PEG)-OtBu) **1.** Calculated m/z for $C_{32}H_{44}N_2O_9$ ($M+H^+$) is 601.31, found 601.30

*(R)-14-((((*9H*-fluoren-9-yl)methoxy)carbonyl)amino)-12-oxo-2,5,8-trioxa-11-azapentadecan-15-oic acid*

2 (Fmoc-Asn(PEG)-OH)

Procedure

To a solution of TFA (95% in water, 50ml) was added 0.89g **1**, and the solution was stirred for 4 h under an argon atmosphere. The product was concentrated by rotary evaporation, and used without further purification.

1H NMR (500 MHz, CDCl3): δ 7.75(2H, d, J = 7.5 Hz, Fmoc aryl C-**H**);7.60 (2H, t, J = 8.75 Hz, Fmoc aryl C-**H**); 7.39 (2H, t, J = 7.5 Hz, Fmoc aryl C-**H**); 7.30 (2H, t, J = 7.5Hz, Fmoc aryl C-**H**); 6.23 (1H, apparent d, -CON**H**-CαH(COOH)-, or -CON**H**-CH2-CH2-O-); 4.56 (1H, broad s, -CONHCα**H**(COOH)-CβH_2-); 4.39 (1H, apparent t, Fmoc Ar2CH-C$\mathbf{H}_{(a)}H_{(b)}$-O-); 4.31 (H, apparent t, Fmoc Ar2CH-$CH_{(a)}\mathbf{H}_{(b)}$-O-); 4.21 (1H, t, J = 5.5 Hz, Fmoc Ar2C**H**-CH2-O-); 3.51-3.70 (14H, m, -CONH-CH2-C$\mathbf{H}_2$-O-C$\mathbf{H}_2$-C$\mathbf{H}_2$-O-C$\mathbf{H}_2$-C$\mathbf{H}_2$-O--C$\mathbf{H}_2$-C$\mathbf{H}_2$-O-); 3.45 (2H, m, -CONH-C$\mathbf{H}_2$-CH_2-O-); 3.32 (3H, s, -O-C$\mathbf{H}_3$); 2.94 (1H, apparent d, -CαH(COOH)-Cβ(**Ha**)Hb-CONH-); 2.73 (1H, dd, J = 15.5 Hz, 7.5

Hz, -CαH(COOH)-Cβ(Ha)**Hb**-CONH).5.59, 1.25 (*t*-butyl alcohol contamination). The full ^{1}H NMR spectrum for **2** is shown in Figure 17.

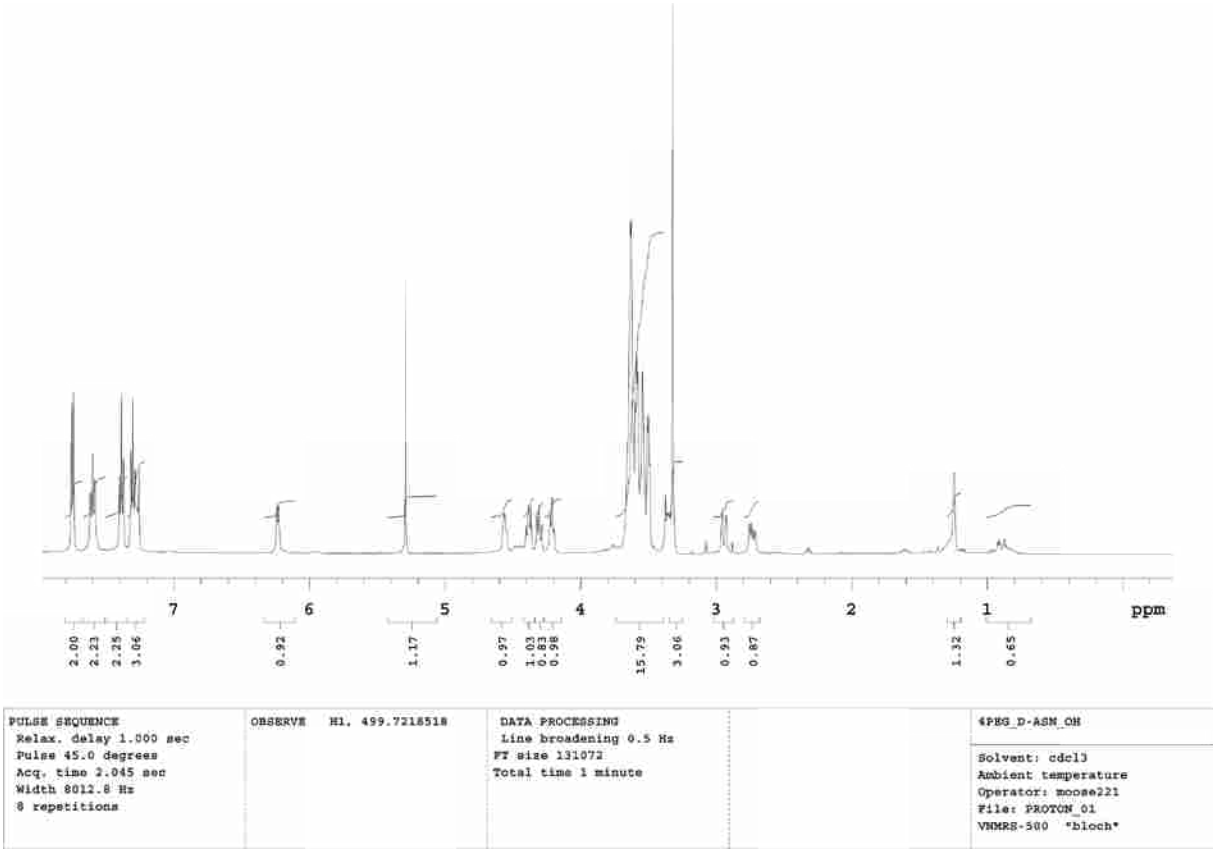


Figure 85: ^{1}H NMR for Fmoc-Asn(PEG)-OH

13C NMR (126 MHz, CDCl3): δ 172.22, 171.43 (-NH-CαH(**C**OOH)-CβH2-, -CβH2-**C**ONH-CH2-); 155.95 (Fmoc-O-**C**ONH-); 143.90, 143.72, 141.26, 141.23 (Fmoc aryl **ipso** C's); 127.71, 127.10, 125.24,125.16, 119.95 (Fmoc Ar **C**-H); 71.75, 70.60, 70.37, 70.24, 69.98 (-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-); 67.23 (Fmoc Ar_2CH-**C**H_2-O-); 58.86 (-O-**C**H_3); 50.73 (-NH-**C**αH(COOH)-CβH_2-); 47.04 (Fmoc Ar_2**C**H-CH_2-O-); 39.72 (CONH-**C**H_2-CH_2-O-); 37.79 (-CαH(COOH)-**Cβ**H_2-CONH-), 53.44 *t*-butyl alcohol contamination. The full ^{13}C NMR spectrum for **2** is shown in Figure 18.

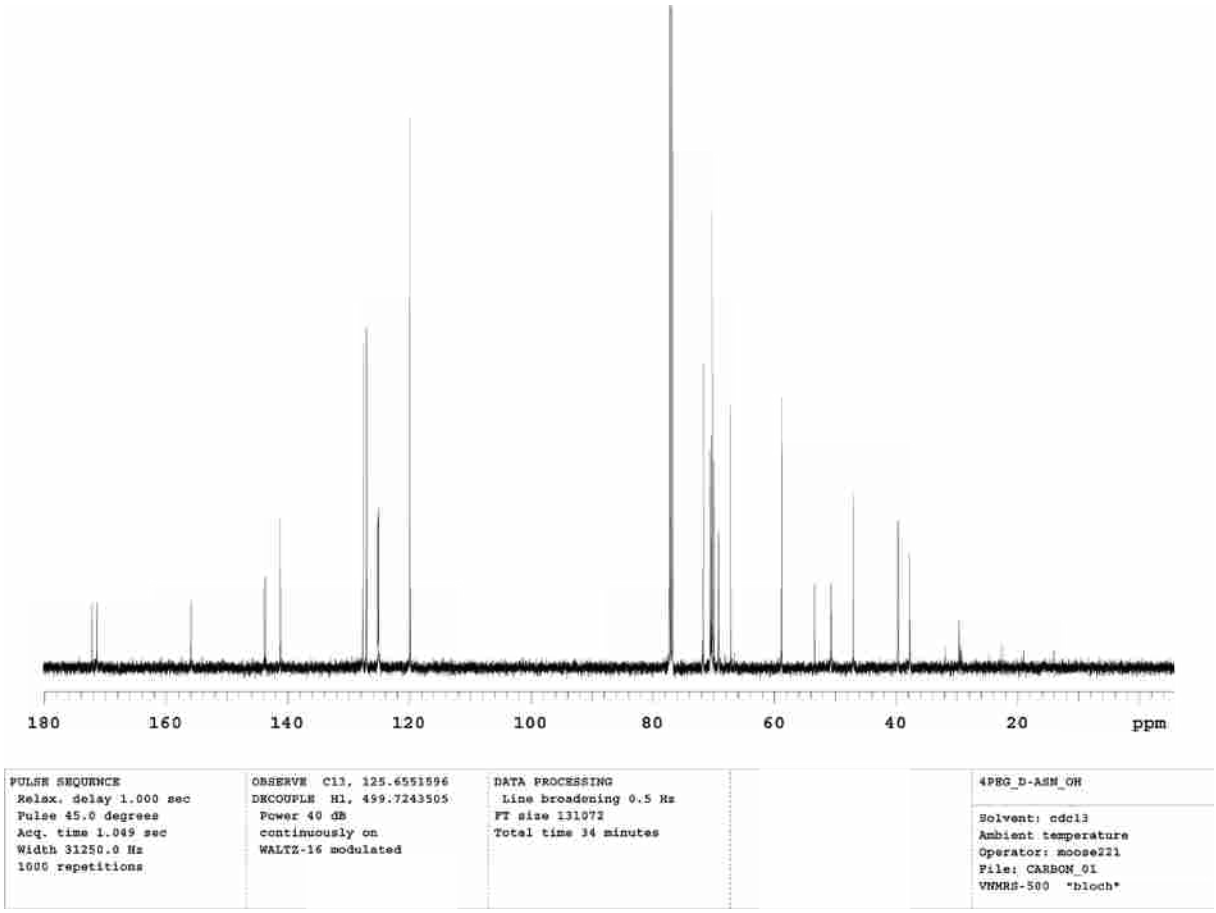


Figure 86: ^{13}C NMR for Fmoc-Asn(PEG)-OH

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-D-Asn(PEG)-OH were made using a 2D HSQC experiment (Figure 19) to indentify the one-bond C-H correlations shown Table 8 and by analogy with published spectral data for related compounds[3,10].

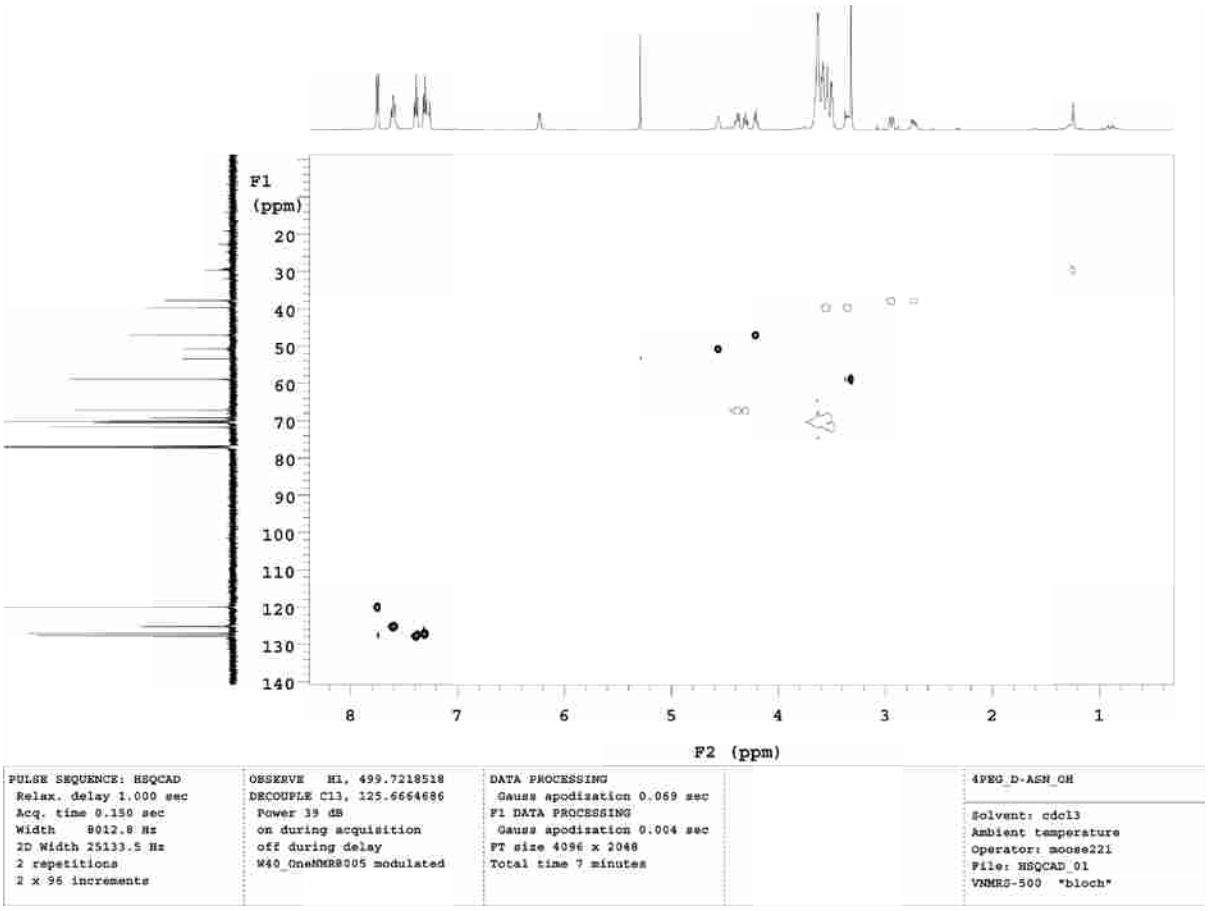


Figure 87: 2D HSQC spectrum of Fmoc-D-Asn(PEG)-OH

Table 9: One-Bond C-H correlations identified from HSQC experiment on Fmoc-D-Asn(PEG)-OH **2**

^{1}H δ	^{1}C δ	Assignment
7.75	119.95	Fmoc aryl C-H
7.60	125.16	Fmoc aryl C-H
7.39	127.71	Fmoc aryl C-H
7.30	127.10	Fmoc aryl C-H
4.56	50.73	-NH-**CαH**(COOH)-CβH_2-
4.39, 4.31	67.23	Fmoc Ar_2CH-**CH(a)H(b)**-O-
4.21	47.04	Fmoc Ar_2**CH**-CH_2-
3.50-3.70	71.75, 70.60, 70.37, 70.24, 69.98	-**CH_2**-O-**CH_2**-**CH_2**-O-**CH_2**-**CH_2**-O-**CH_2**-**CH_2**-O-
3.45	39.72	CONH-**CH_2**-CH_2-O-
3.32	58.86	O-**CH_3**
2.94, 2.73	37.79	CαH(COOH)**CβH_2**-CONH-

High-resolution electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) is shown in Figure 20.

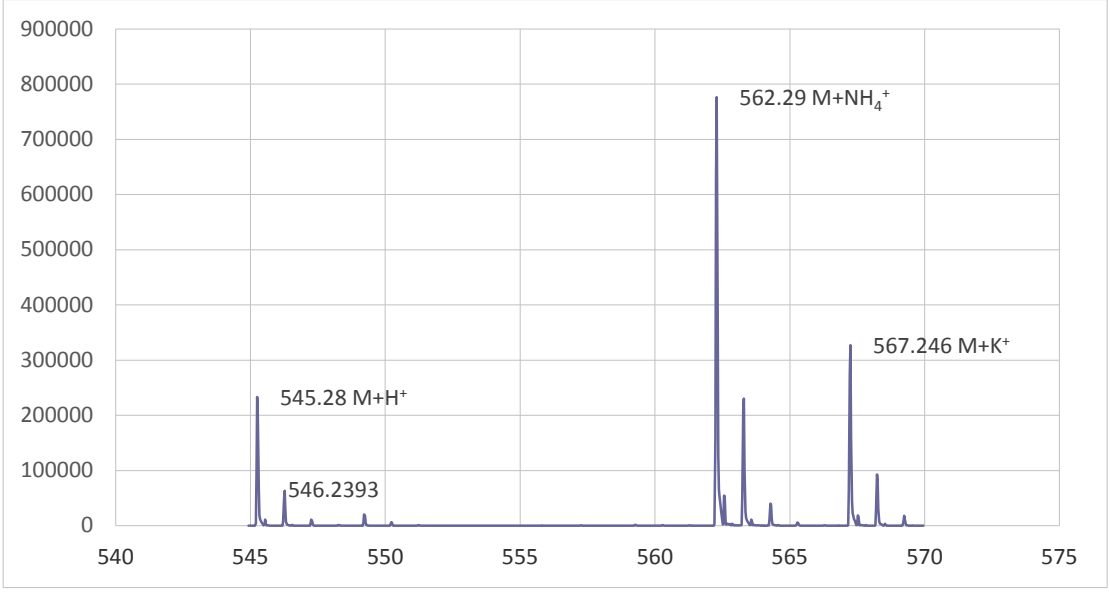


Figure 88: ESI-TOF MS data for (Fmoc-D-Asn(PEG)-OH) **2.** Calculated m/z for $C_{28}H_{36}N_2O_9$ ($M+H^+$) 545.25, found 545.28

3.5.3 Global Fitting of Variable Temperature CD Data

For peptides, **19(D), 19(D)p, 16/19, 16p/19, 16/19p, 16p/19p, 23/26p, 23p/26p, 19/26p, 19p/26p, 26p/29, 26p/29p, 16/26p, 16p/26p, 19/23(Y→F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 19/32(S→A), 19p/32(S→A), 16/23(Y→F), 16p/23(Y→F), 16/32(S→A), 16p/32(S→A), 11(W→X), 26p/11(W→X), 29(T→A), 26p/29(T→A), 23(Y→F), 26p/23(Y→F), 16(S→A)/19/23(Y→F), 16(S→A)/19p/23(Y→F), 19/23(Y→FOMe), and 19p/23(Y→FOMe)**, data from variable temperature CD were fit globally to the equations shown in Chapter 2.

3.5.4 ESI-TOF

ESI-TOF spectra for proteins **26p(1), 26p(8), 26p(45), 19(D), 19(D)p, 16/19, 16p/19, 16/19p, 16p/19p, 23/26p, 23p/26p, 19/26p, 19p/26p, 26p/29, 26p/29p, 16/26p, 16p/26p, 19/23(Y →F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 19/32(S→A), 19p/32(S→A), 16/23(Y→F), 16p/23(Y→F), 16/32(S→A), 16p/32(S→A), 11(W→X), 26p/11(W→X), 29(T→A), 26p/29(T→A), 23(Y→F), 26p/23(Y→F), 16(S→A)/19/23(Y→F), 16(S→A)/19p/23(Y→F), 19/23(Y→FOMe), and 19p/23(Y→FOMe)** are shown in Figures 21-56 .

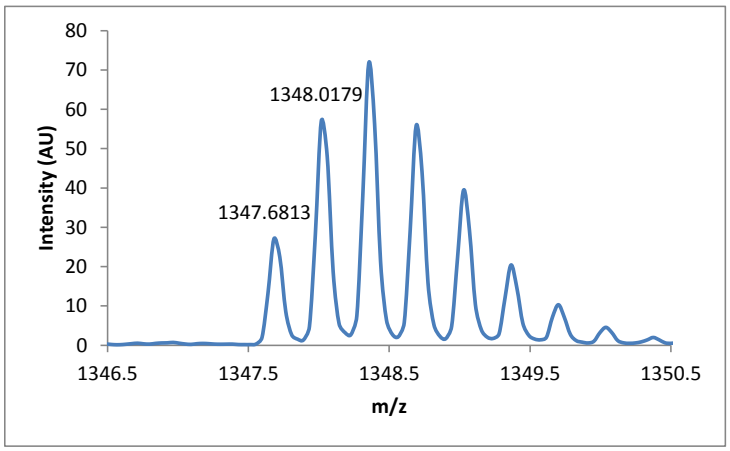


Figure 89: ESI-TOF spectrum for the Pin WW domain protein **26p(1)**. Expected $[M+3H]^{3+}/3$ = 1347.6824 Da. Observed $[M+3H]^{3+}/3$ = 1347.6813 Da.

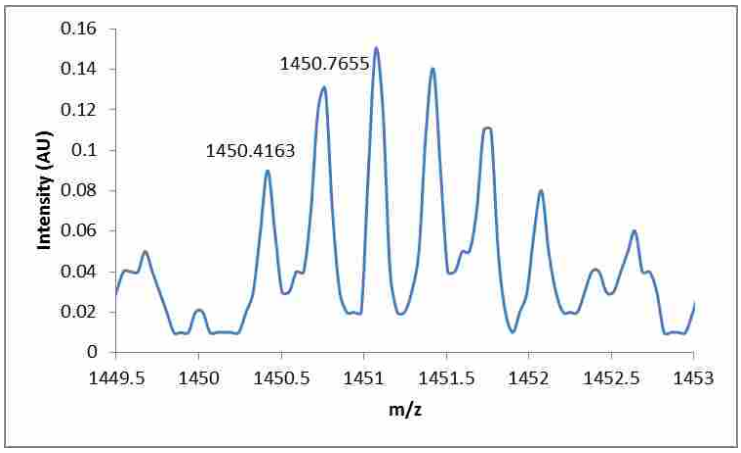


Figure 90: ESI-TOF spectrum for the Pin WW domain protein **26p(8).** Expected $[M+3H]^{3+}/3$ = 1450.4103 Da. Observed $[M+3H]^{3+}/3$ = 1450.4163 Da.

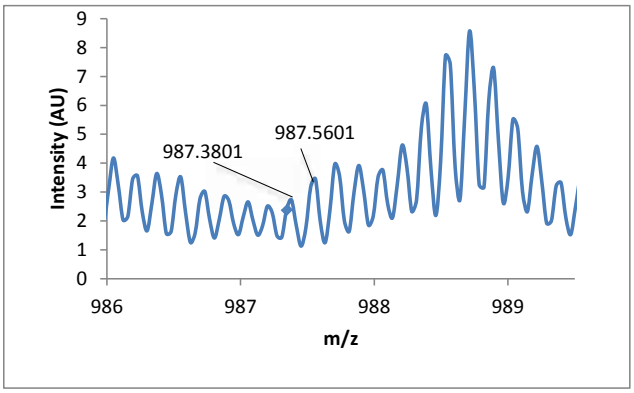


Figure 91: ESI-TOF spectrum for the Pin WW domain protein **26p(45)**. Expected $[M+3NH_3^{+}+1K^{+}]^{5+}/5$ = 987.3680 Da. Observed $[M+3NH_3^{+}+1K^{+}]^{5+}/5$ = 987.3801 Da.

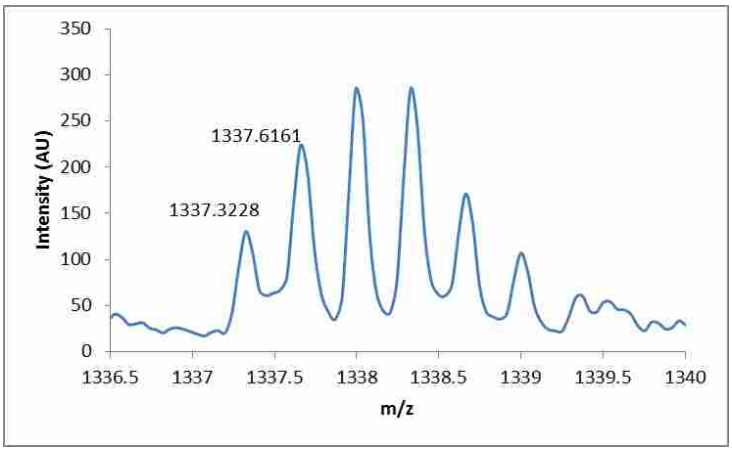


Figure 92: ESI-TOF spectrum for the Pin WW domain protein **19(D)**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3228 Da.

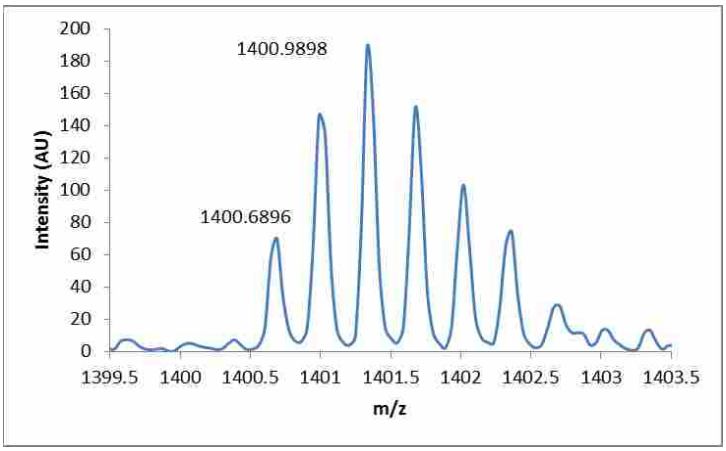


Figure 93: ESI-TOF spectrum for the Pin WW domain protein **19(D)p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.6896 Da.

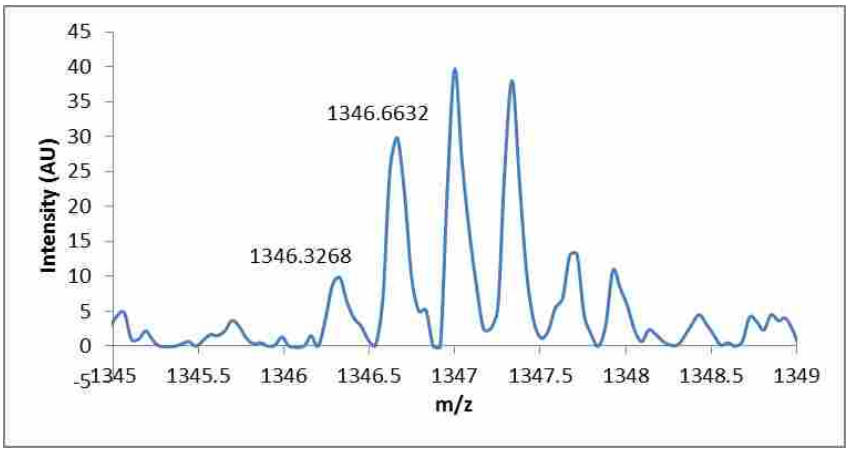


Figure 94: ESI-TOF spectrum for the Pin WW domain protein **16/19**. Expected $[M+3H]^{3+}/3$ = 1346.3424 Da. Observed $[M+3H]^{3+}/3$ = 1346.3268 Da.

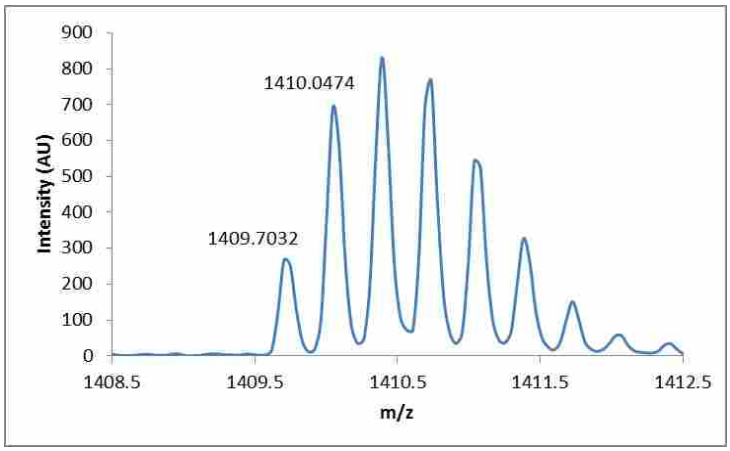


Figure 95: ESI-TOF spectrum for the Pin WW domain protein **16p/19**. Expected $[M+3H]^{3+}/3$ = 1409.7159 Da. Observed $[M+3H]^{3+}/3$ = 1409.7032 Da.

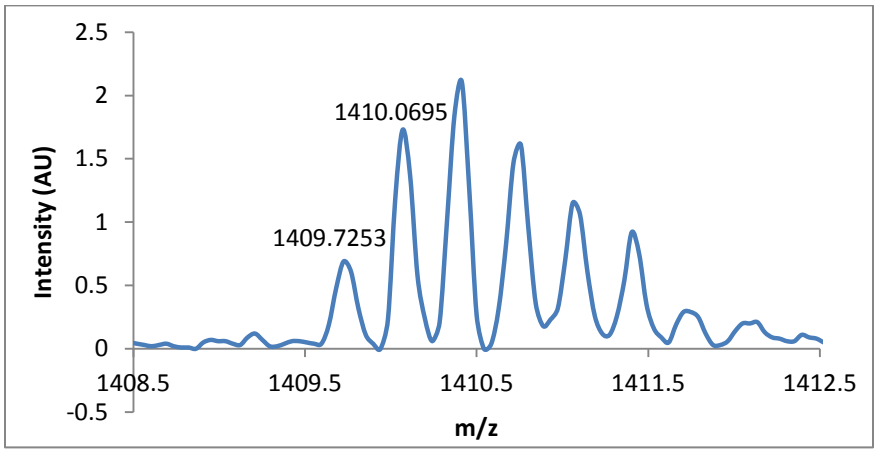


Figure 96: ESI-TOF spectrum for the Pin WW domain protein **16/19p**. Expected $[M+3H]^{3+}/3$ = 1409.7159 Da. Observed $[M+3H]^{3+}/3$ = 1409.7253 Da.

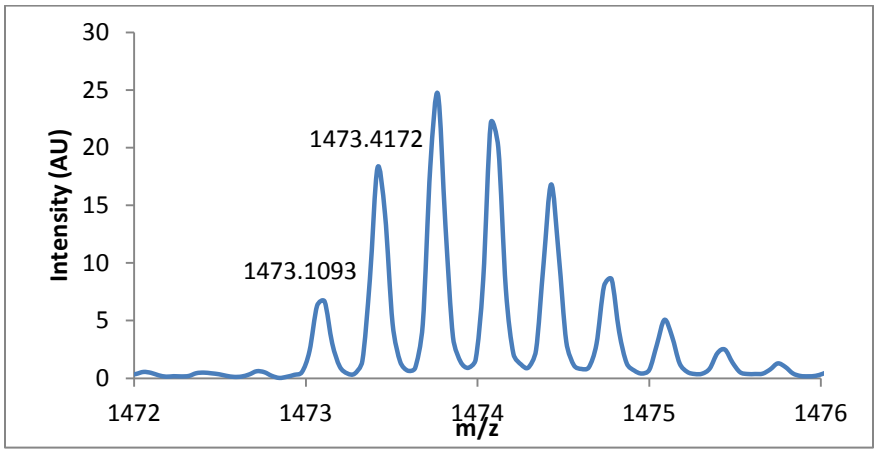


Figure 97: ESI-TOF spectrum for the Pin WW domain protein **16p/19p**. Expected $[M+3H]^{3+}/3$ = 1473.0894 Da. Observed $[M+3H]^{3+}/3$ = 1473.1093 Da.

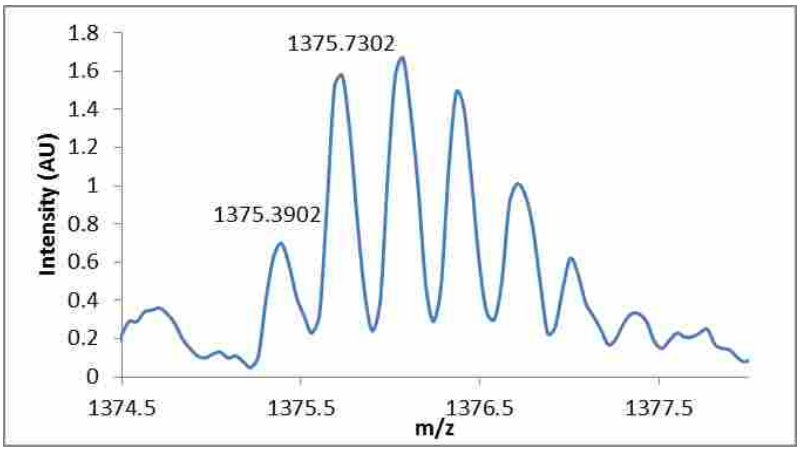


Figure 98: ESI-TOF spectrum for the Pin WW domain protein **23/26p**. Expected $[M+3H]^{3+}/3$ = 1375.3685 Da. Observed $[M+3H]^{3+}/3$ = 1473.1093 Da.

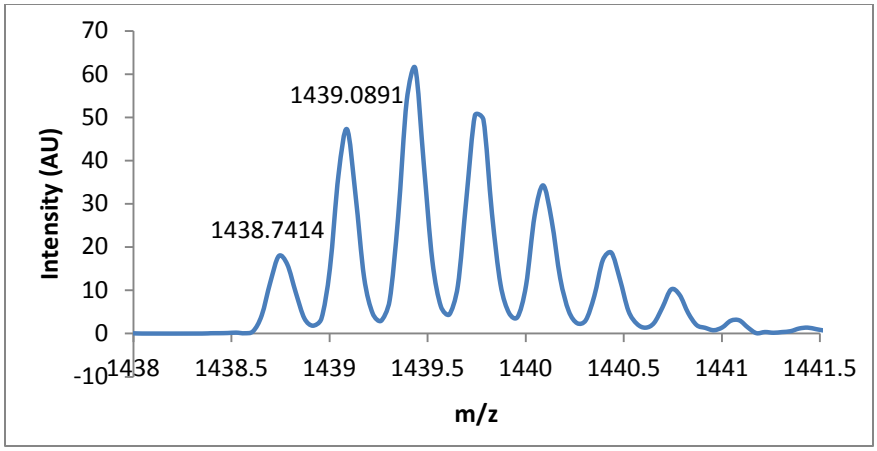


Figure 99: ESI-TOF spectrum for the Pin WW domain protein **23p/26p**. Expected $[M+3H]^{3+}/3$ = 1438.7420 Da. Observed $[M+3H]^{3+}/3$ = 1438.7414

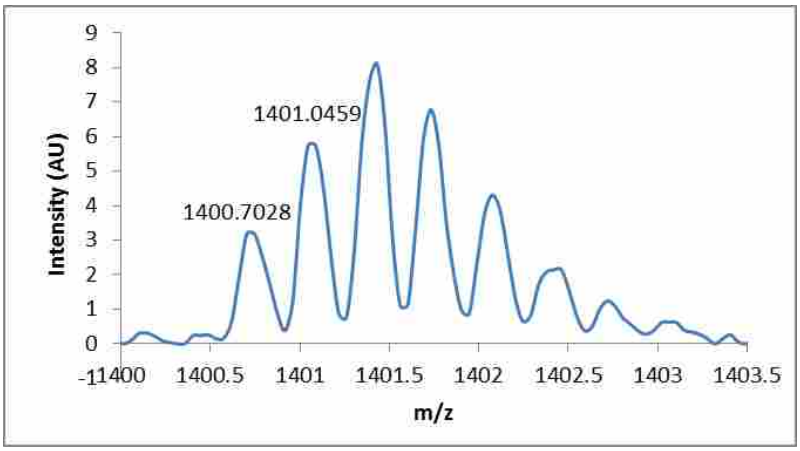


Figure 100: ESI-TOF spectrum for the Pin WW domain protein **19/26p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.7028

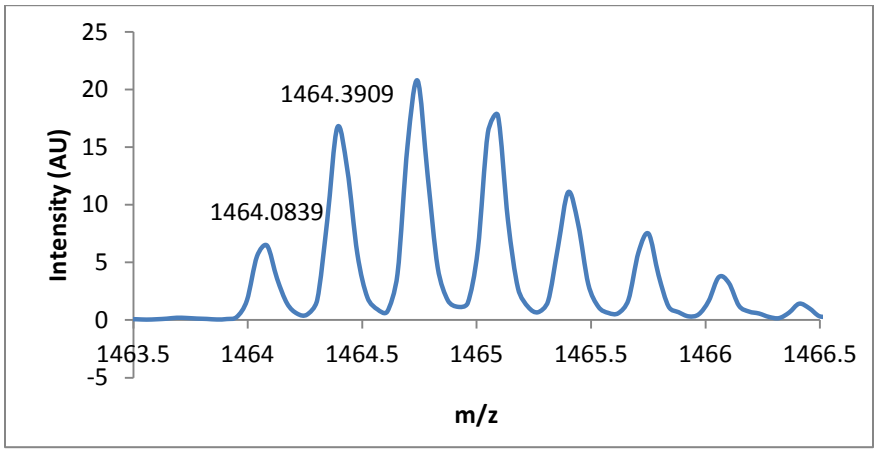


Figure 101: ESI-TOF spectrum for the Pin WW domain protein **19p/26p**. Expected $[M+3H]^{3+}/3$ = 1464.0858 Da. Observed $[M+3H]^{3+}/3$ = 1464.0839

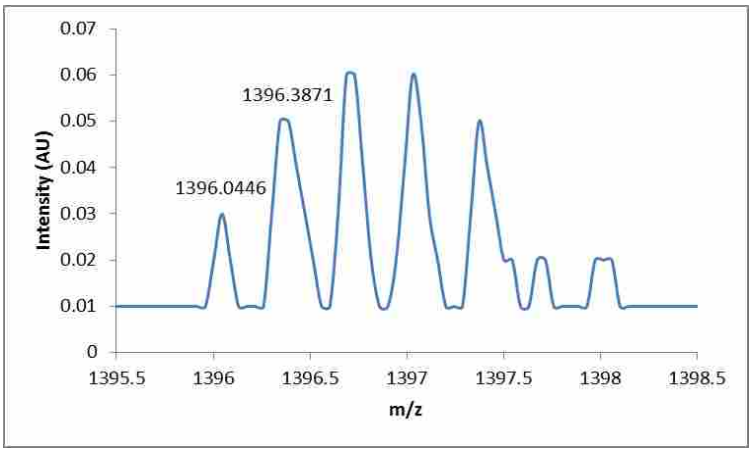


Figure 102: ESI-TOF spectrum for the Pin WW domain protein **26p/29**. Expected $[M+3H]^{3+}/3$ = 1396.0404 Da. Observed $[M+3H]^{3+}/3$ = 1396.0446

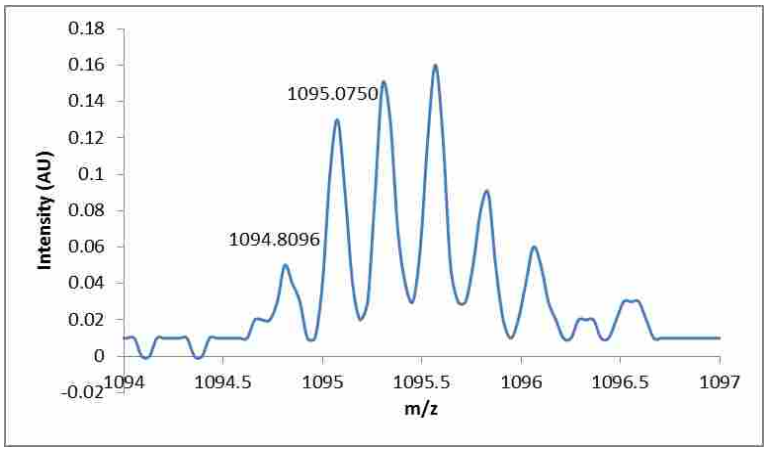


Figure 103: ESI-TOF spectrum for the Pin WW domain protein **26p/29p**. Expected $[M+4H]^{4+}/4$ = 1094.8124Da. Observed $[M+4H]^{4+}/4$ = 1094.8096

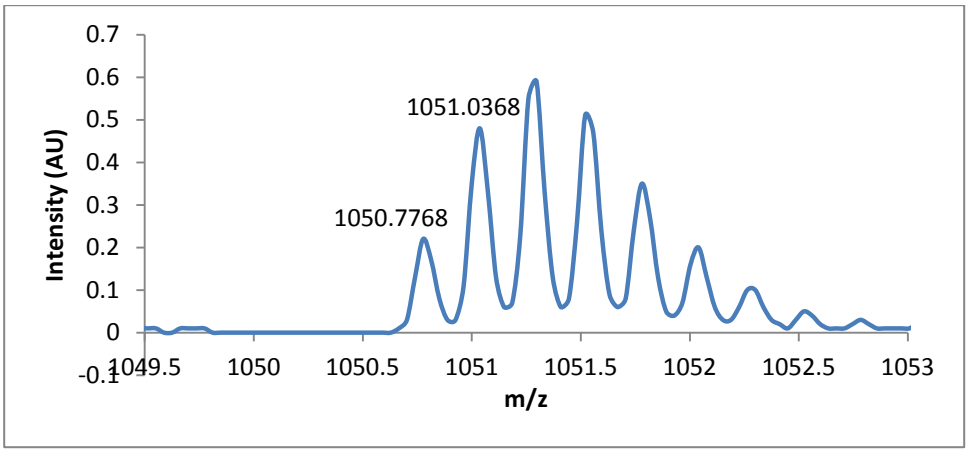


Figure 104: ESI-TOF spectrum for the Pin WW domain protein **16/26p**. Expected $[M+4H]^{4+}/4$ = 1050.7861Da. Observed $[M+4H]^{4+}/4$ = 1050.7768

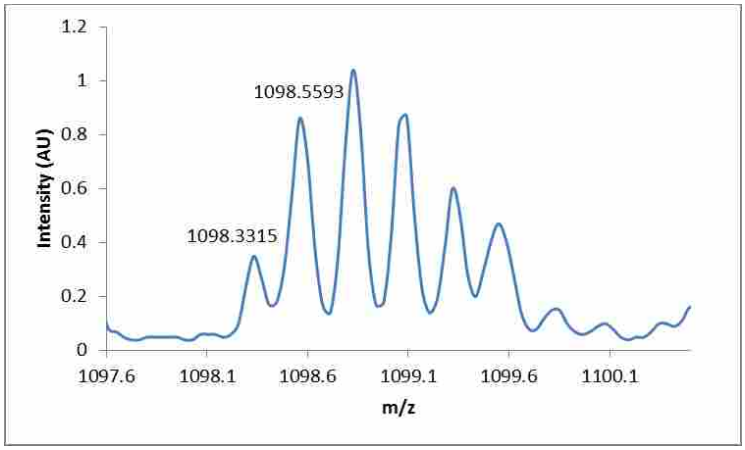


Figure 105: ESI-TOF spectrum for the Pin WW domain protein **16p/26p**. Expected $[M+4H]^{4+}/4$ = 1098.3163 Da. Observed $[M+4H]^{4+}/4$ = 1098.3315

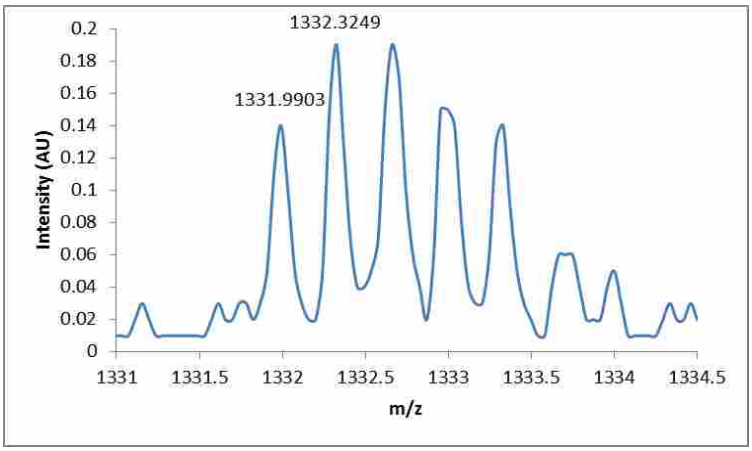


Figure 106: ESI-TOF spectrum for the Pin WW domain protein **19/23(Y→F)**. Expected $[M+3H]^{3+}/3$ = 1332.0071 Da. Observed $[M+3H]^{3+}/3$ = 1331.9903

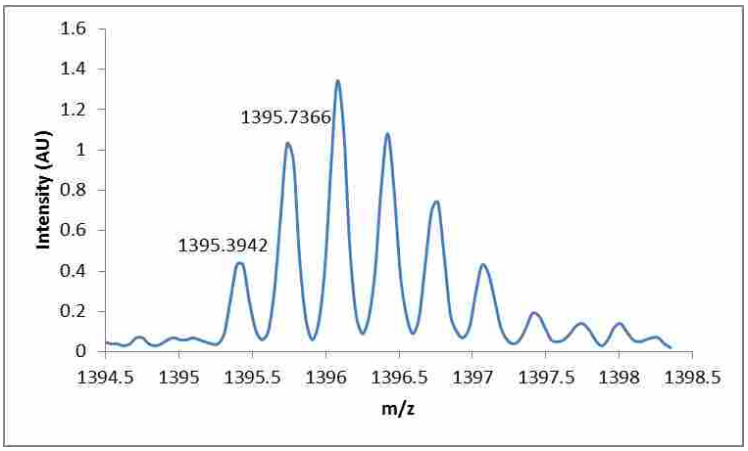


Figure 107: ESI-TOF spectrum for the Pin WW domain protein **19p/23(Y→F)**. Expected $[M+3H]^{3+}/3$ = 1395.3806 Da. Observed $[M+3H]^{3+}/3$ = 1395.3942

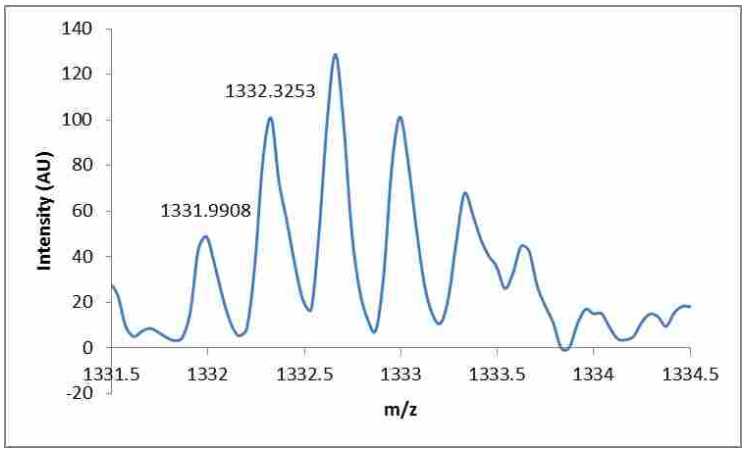


Figure 108: ESI-TOF spectrum for the Pin WW domain protein **19/16(S→A)**. Expected $[M+3H]^{3+}/3$ = 1332.0071 Da. Observed $[M+3H]^{3+}/3$ = 1331.9908

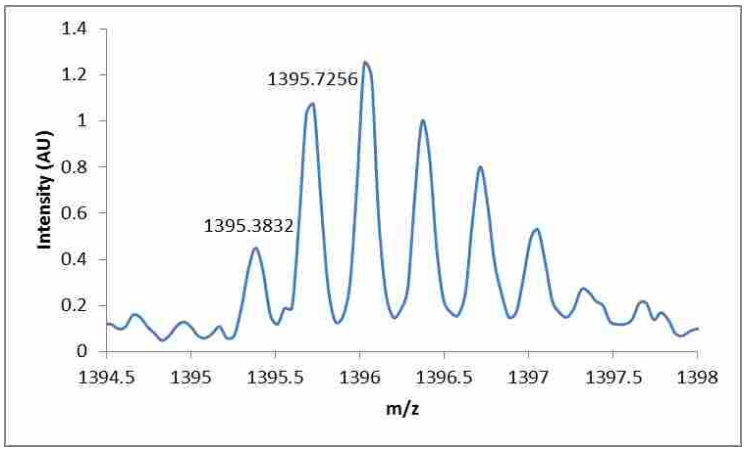


Figure 109: ESI-TOF spectrum for the Pin WW domain protein **19p/16(S→A)**. Expected $[M+3H]^{3+}/3$ = 1395.3806 Da. Observed $[M+3H]^{3+}/3$ = 1395.3832

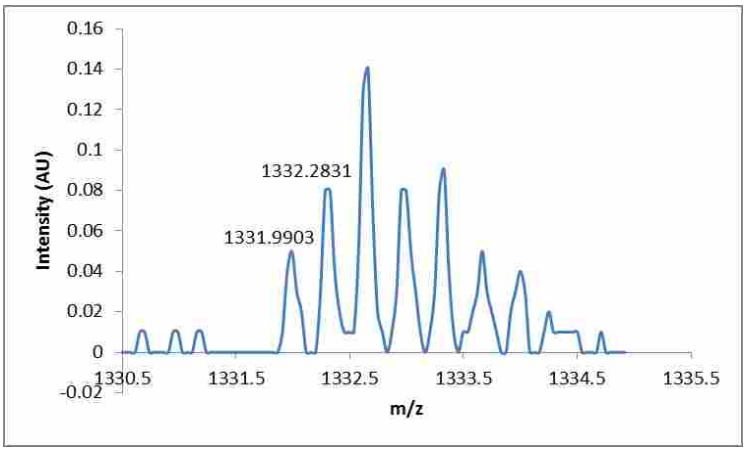


Figure 110: ESI-TOF spectrum for the Pin WW domain protein **19/32(S→A)**. Expected $[M+3H]^{3+}/3$ = 1332.0071 Da. Observed $[M+3H]^{3+}/3$ = 1331.9903

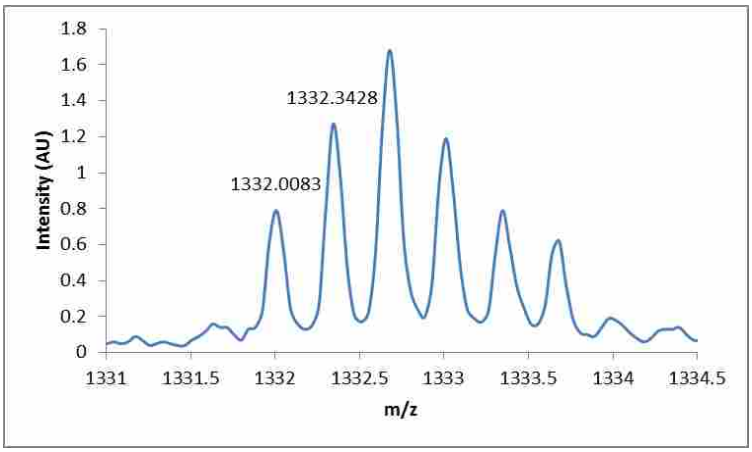


Figure 111: ESI-TOF spectrum for the Pin WW domain protein **16/23(Y→F)**. Expected $[M+3H]^{3+}/3$ = 1332.0071 Da. Observed $[M+3H]^{3+}/3$ = 1332.0083

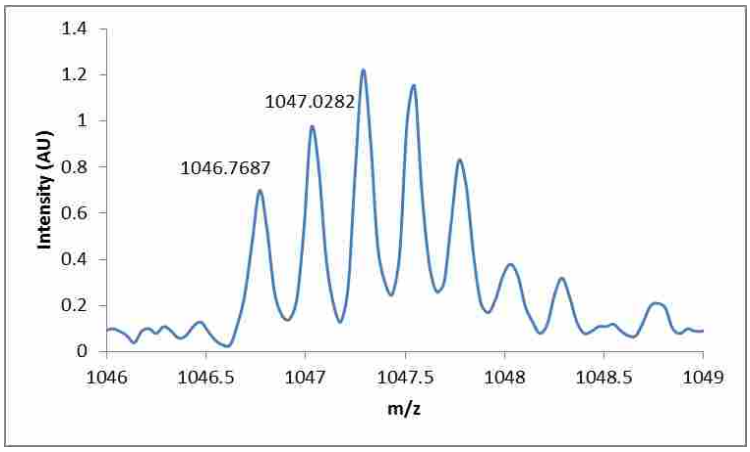


Figure 112: ESI-TOF spectrum for the Pin WW domain protein **16p/23(Y→F)**. Expected $[M+4H]^{4+}/4$ = 1046.7874 Da. Observed $[M+4H]^{4+}/4$ = 1046.7687

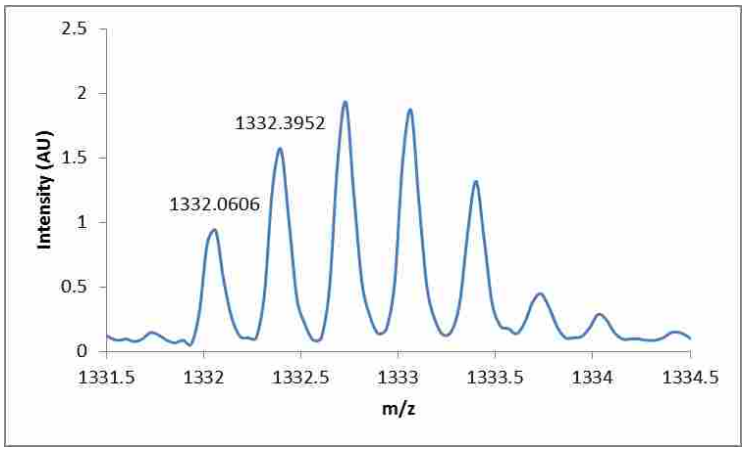


Figure 113: ESI-TOF spectrum for the Pin WW domain protein **16/32(S→A)**. Expected $[M+3H]^{3+}/3$ = 1332.0071 Da. Observed $[M+3H]^{3+}/3$ = 1332.0606

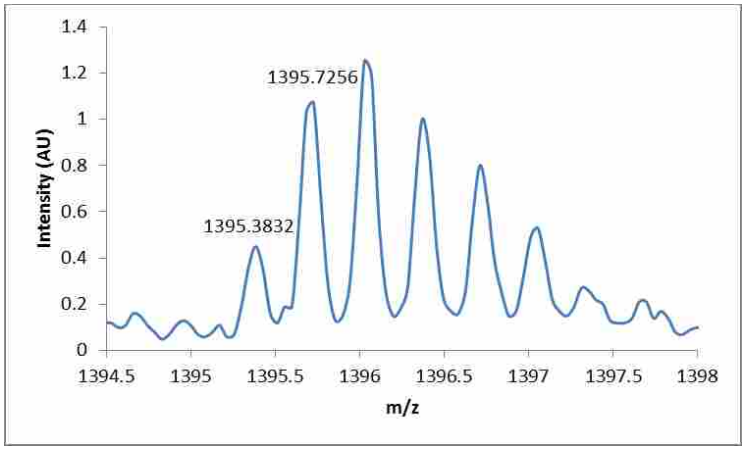


Figure 114: ESI-TOF spectrum for the Pin WW domain protein **16p/32(S→A)**. Expected $[M+3H]^{3+}/3$ = 1395.3806 Da. Observed $[M+3H]^{3+}/3$ = 1395.3832

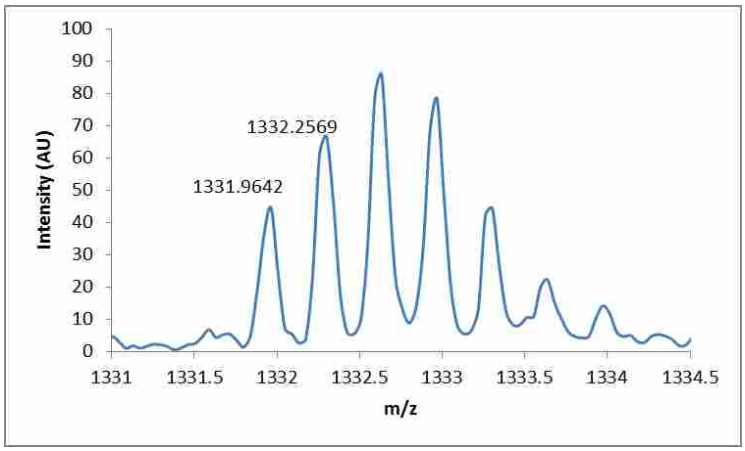


Figure 115: ESI-TOF spectrum for the Pin WW domain protein **26/11(W→AX)**. Expected $[M+3H]^{3+}/3$ = 1332.0033 Da. Observed $[M+3H]^{3+}/3$ = 1331.9642

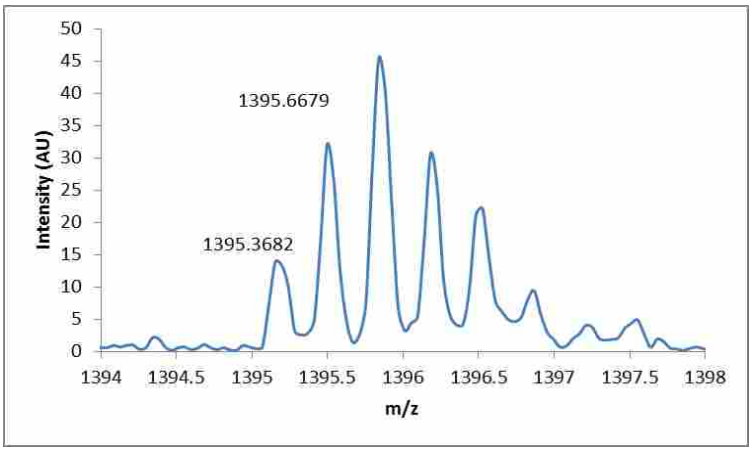


Figure 116: ESI-TOF spectrum for the Pin WW domain protein **26p/11(W→X)**. Expected $[M+3H]^{3+}/3$ = 1395.3769 Da. Observed $[M+3H]^{3+}/3$ = 1395.3682

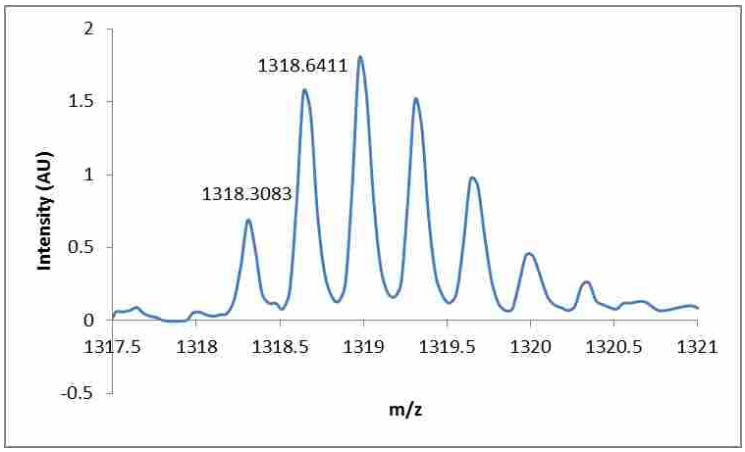


Figure 117: ESI-TOF spectrum for the Pin WW domain protein **29(T→A)**. Expected $[M+3H]^{3+}/3$ = 1318.3316 Da. Observed $[M+3H]^{3+}/3$ = 1318.3083

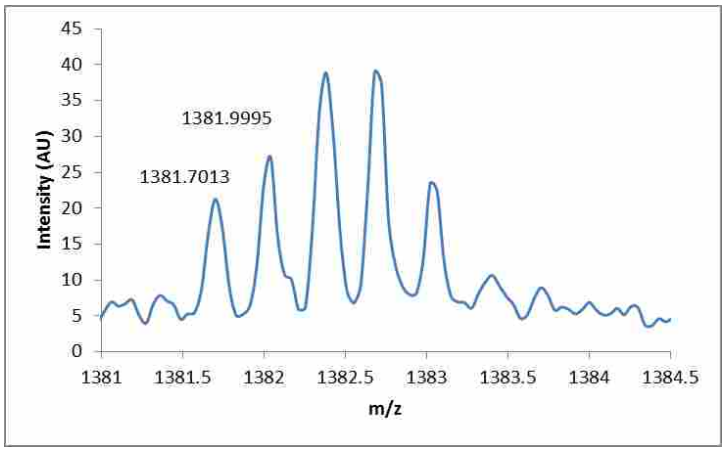


Figure 118: ESI-TOF spectrum for the Pin WW domain protein **26p/29(T→A)**. Expected $[M+3H]^{3+}/3$ = 1381.7051 Da. Observed $[M+3H]^{3+}/3$ = 1381.7013

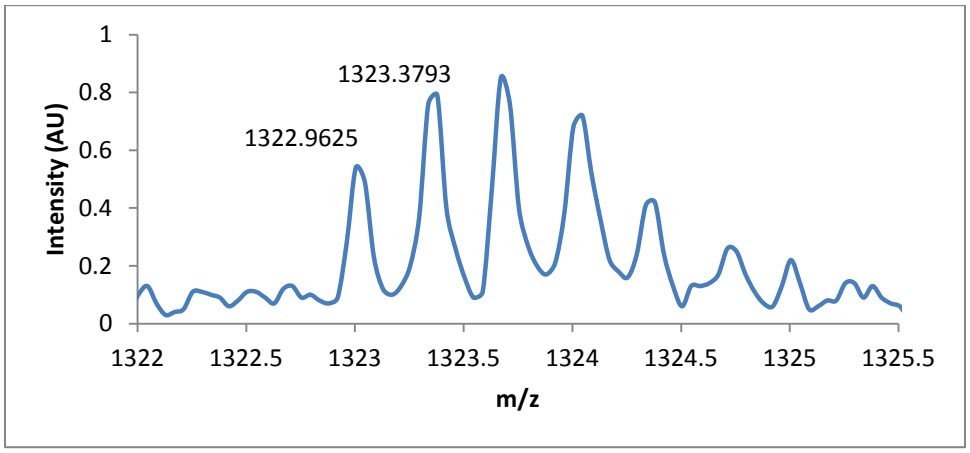


Figure 119: ESI-TOF spectrum for the Pin WW domain protein **23(Y→F)**. Expected $[M+3H]^{3+}/3$ = 1323.0035 Da. Observed $[M+3H]^{3+}/3$ = 1322.9625

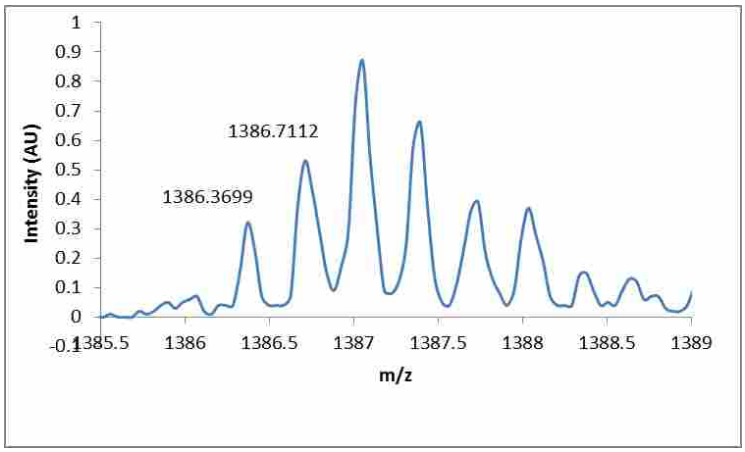


Figure 120: ESI-TOF spectrum for the Pin WW domain protein **26p/23(Y→F)**. Expected $[M+3H]^{3+}/3$ = 1386.3770 Da. Observed $[M+3H]^{3+}/3$ = 1386.3699

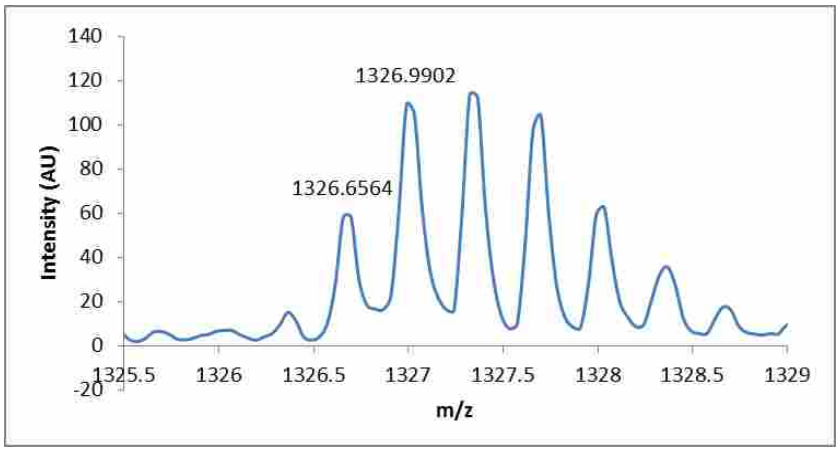


Figure 121: ESI-TOF spectrum for the Pin WW domain protein **16(S→A)/19/23(Y→F).** Expected $[M+3H]^{3+}/3$ = 1326.6755 Da. Observed $[M+3H]^{3+}/3$ = 1326.6564

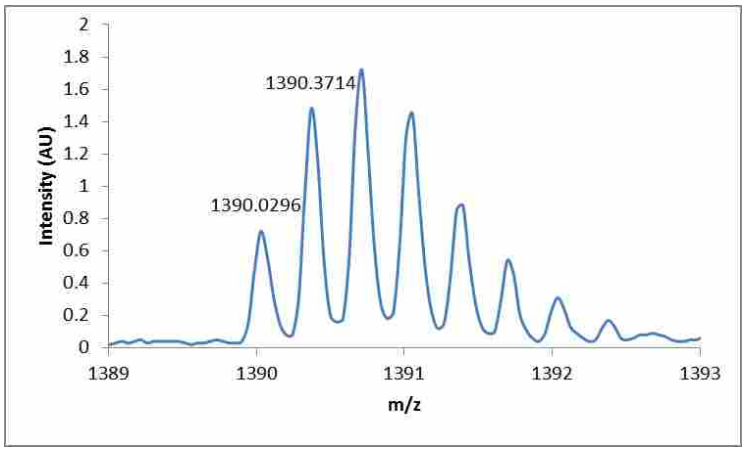


Figure 122: ESI-TOF spectrum for the Pin WW domain protein **16(S→A)/19p/23(Y→F).** Expected $[M+3H]^{3+}/3$ = 1390.0490 Da. Observed $[M+3H]^{3+}/3$ = 1390.0296

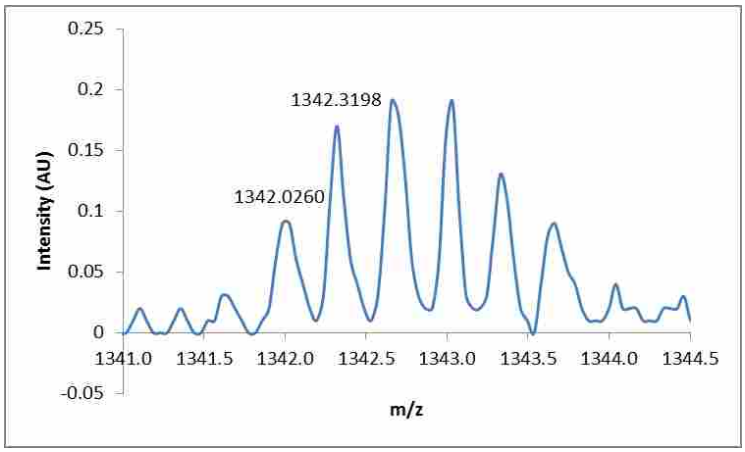


Figure 123: ESI-TOF spectrum for the Pin WW domain protein **19/23(Y→FOMe).** Expected $[M+3H]^{3+}/3$ = 1342.0143 Da. Observed $[M+3H]^{3+}/3$ = 1342.0260

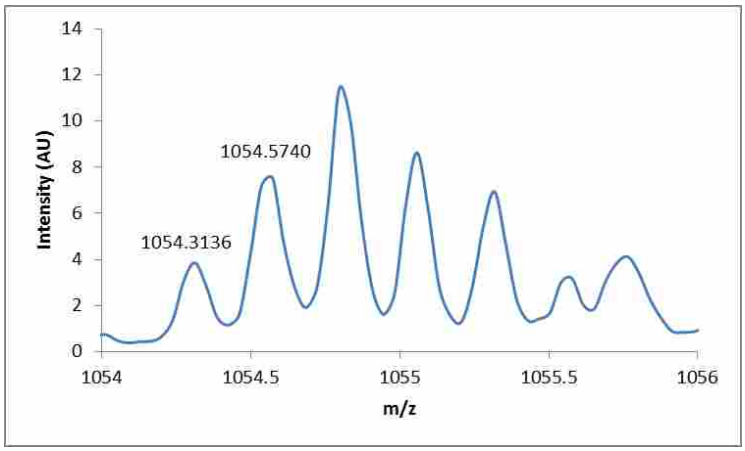


Figure 124: ESI-TOF spectrum for the Pin WW domain protein **19p/23(Y→FOMe).** Expected $[M+4H]^{4+}/4$ = 1054.2928 Da. Observed $[M+4H]^{4+}/4$ = 1054.3136

3.5.5 HPLC

HPLC traces for proteins **26p(1), 26p(8), 26p(45), 19(D), 19(D)p, 16/19, 16p/19, 16/19p, 16p/19p, 23/26p, 23p/26p, 19/26p, 19p/26p, 26p/29, 26p/29p, 16/26p, 16p/26p, 19/23(Y→F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 19/32(S→A), 19p/32(S→A), 16/23(Y→F), 16p/23(Y→F), 16/32(S→A), 16p/32(S→A), 11(W→X), 26p/11(W→X), 29(T→A), 26p/29(T→A), 23(Y→F), 26p/23(Y→F), 16(S→A)/19/23(Y→F), 16(S→A)/19p/23(Y→F), 19/23(Y→FOMe), and 19p/23(Y→FOMe)** are shown in Figures 57-91.

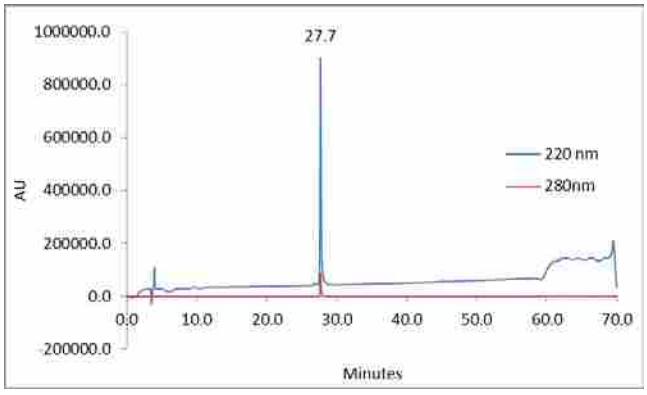


Figure 125: Analytical HPLC data for Pin WW domain protein **26p(1)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

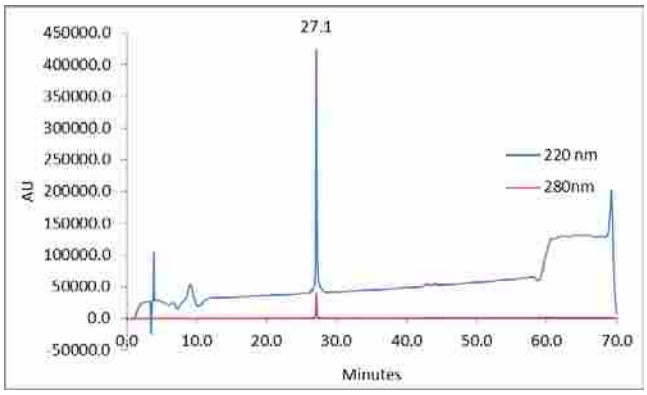


Figure 126: Analytical HPLC data for Pin WW domain protein **26p(8)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

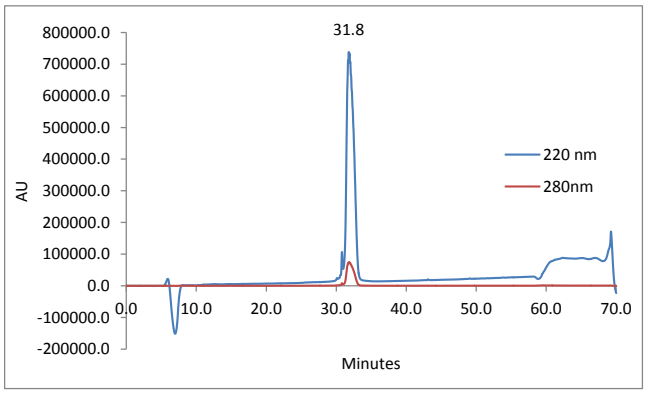


Figure 127: Analytical HPLC data for Pin WW domain protein **26p(45)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

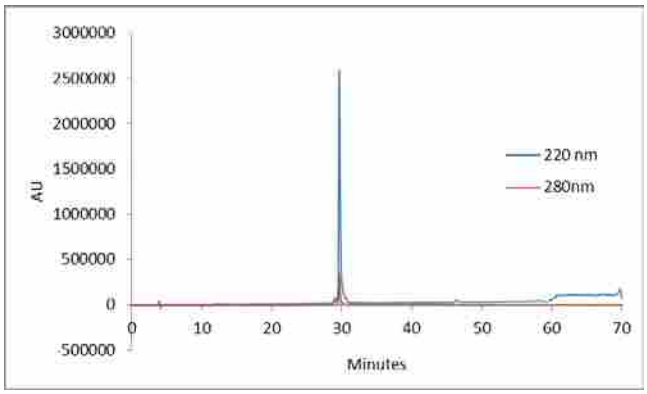


Figure 128: Analytical HPLC data for Pin WW domain protein **19(D)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

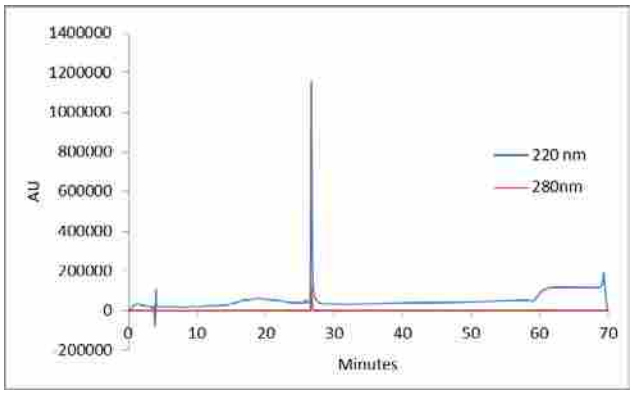


Figure 129: Analytical HPLC data for Pin WW domain protein **19(D)p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

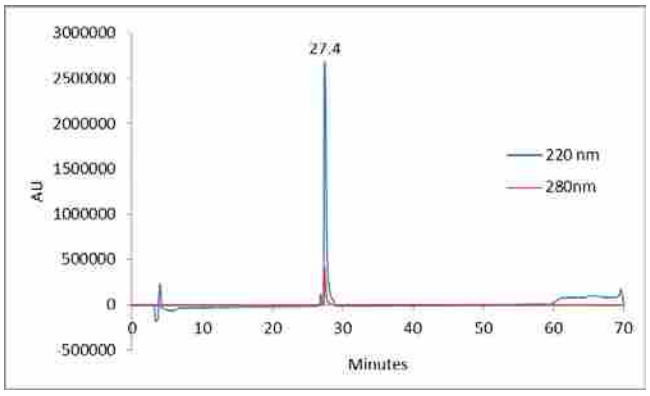


Figure 130: Analytical HPLC data for Pin WW domain protein **16/19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

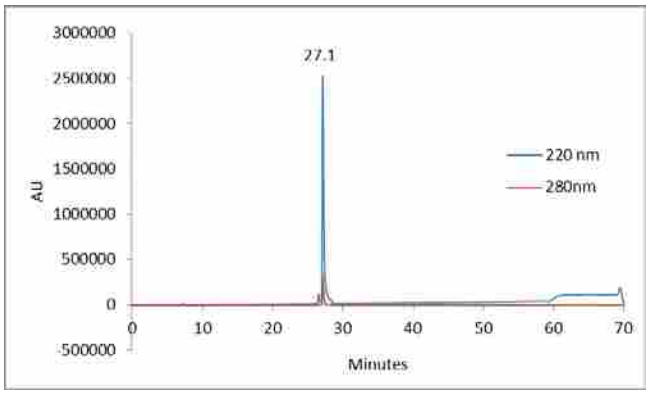


Figure 131: Analytical HPLC data for Pin WW domain protein **16p/19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

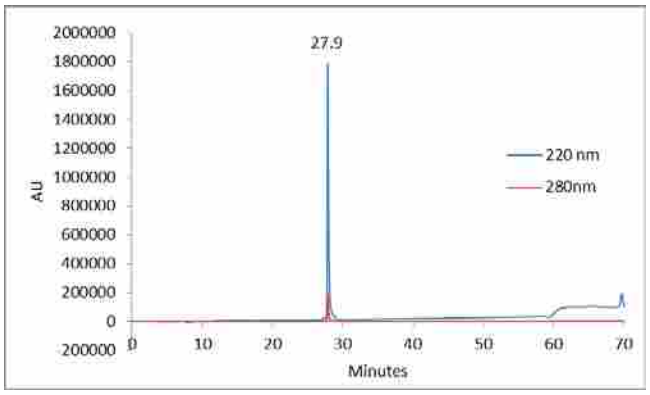


Figure 132: Analytical HPLC data for Pin WW domain protein **16/19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

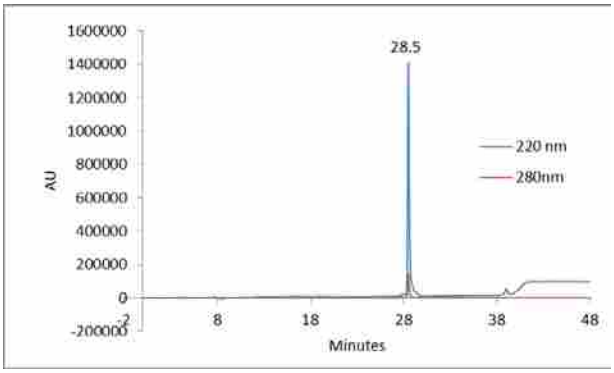


Figure 133: Analytical HPLC data for Pin WW domain protein **16p/19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 48 minutes (during column re-equilibration)

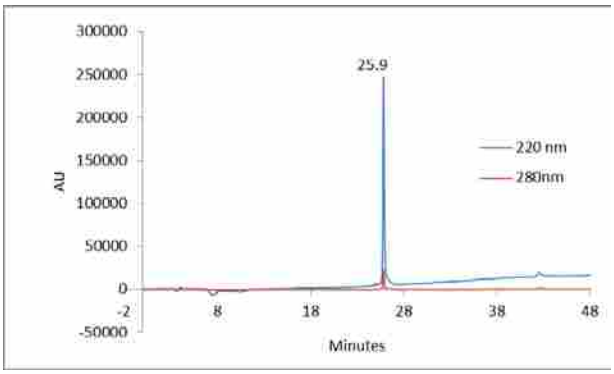


Figure 134: Analytical HPLC data for Pin WW domain protein **23/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 48 minutes (during column re-equilibration)

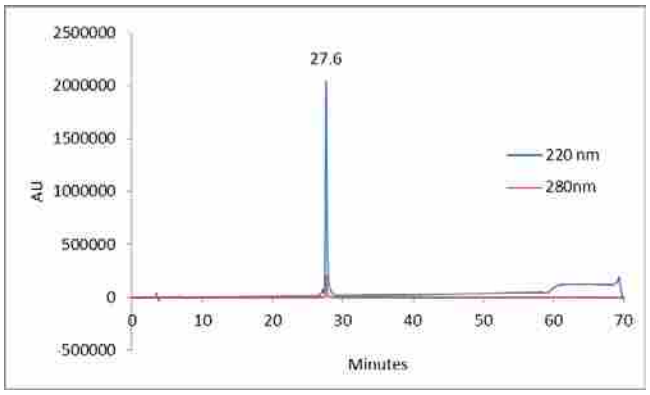


Figure 135: Analytical HPLC data for Pin WW domain protein **23p/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

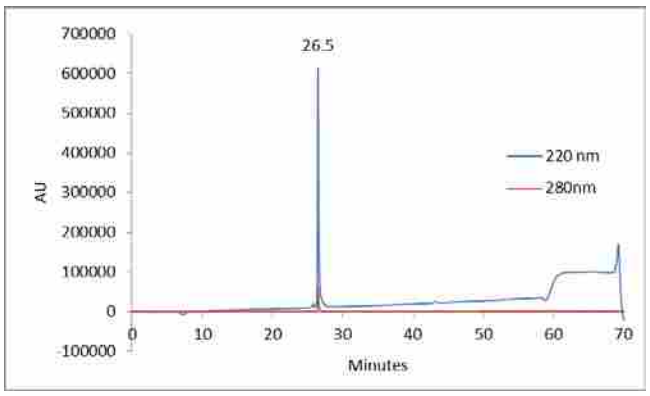


Figure 136: Analytical HPLC data for Pin WW domain protein **19/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

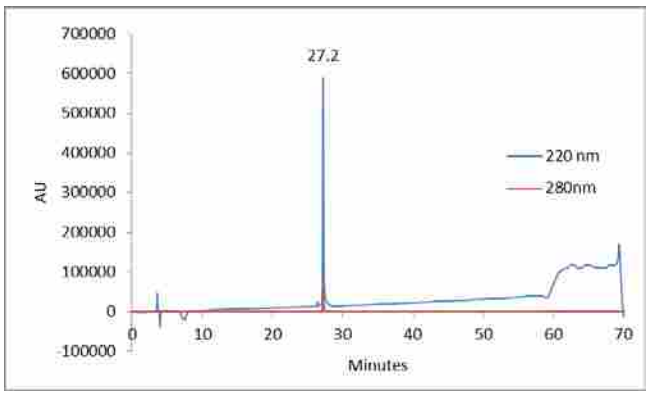


Figure 137: Analytical HPLC data for Pin WW domain protein **19p/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

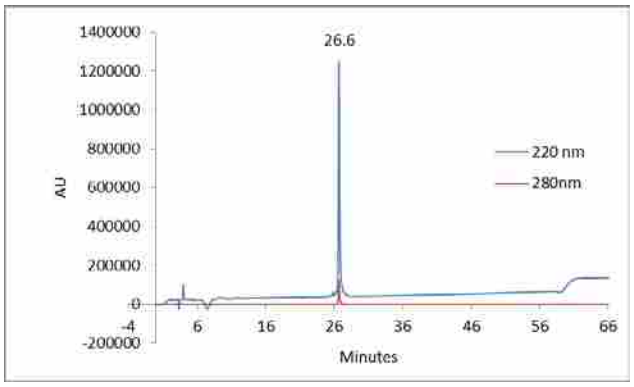


Figure 138: Analytical HPLC data for Pin WW domain protein **26p/29**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 66 minutes (during column re-equilibration).

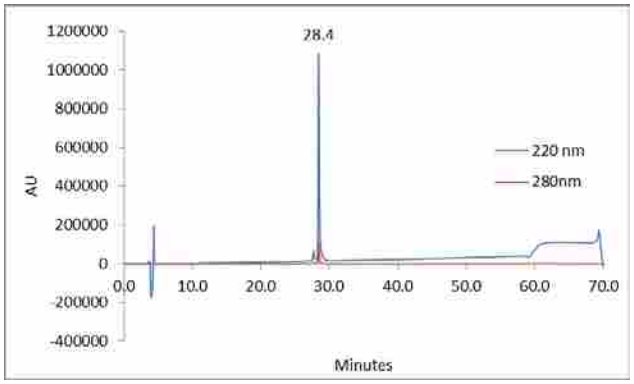


Figure 139: Analytical HPLC data for Pin WW domain protein **26p/29p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

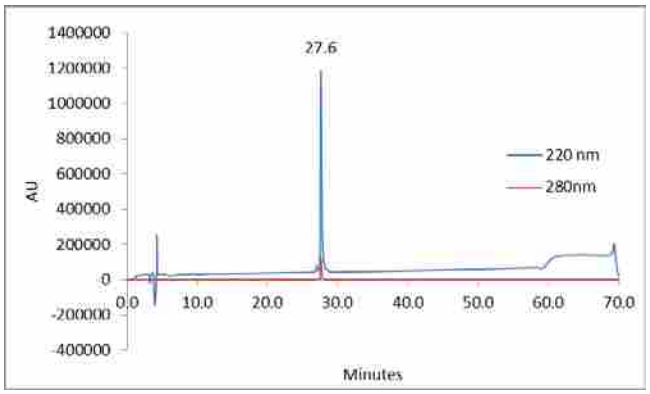


Figure 140: Analytical HPLC data for Pin WW domain protein **16/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

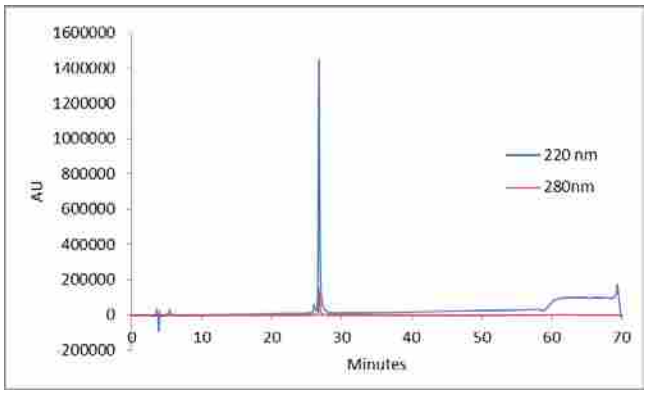


Figure 141: Analytical HPLC data for Pin WW domain protein **16p/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

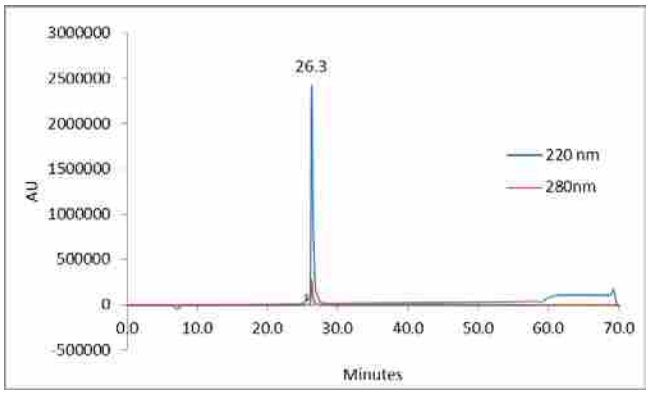


Figure 142: Analytical HPLC data for Pin WW domain protein **19/23(Y→F)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

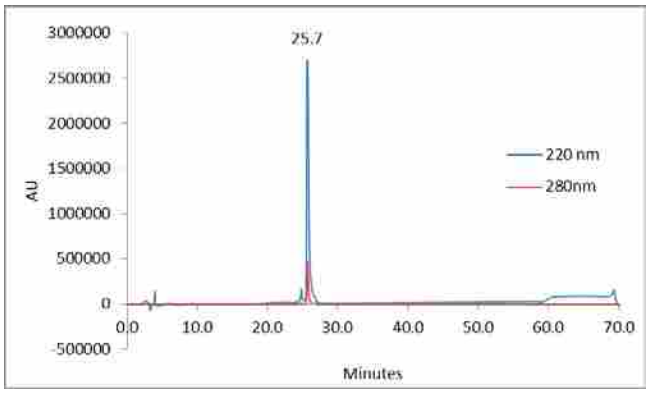


Figure 143: Analytical HPLC data for Pin WW domain protein **19/16(S→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

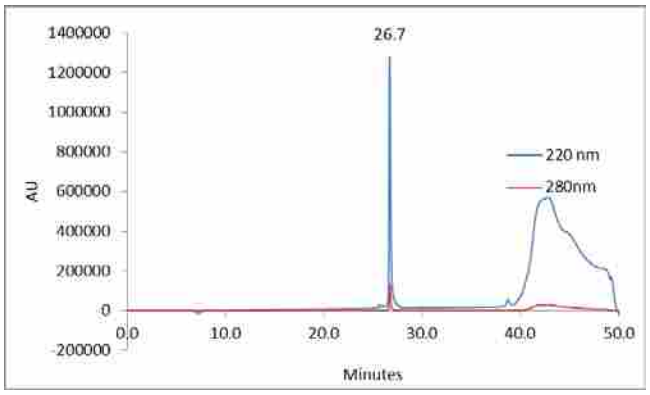


Figure 144: Analytical HPLC data for Pin WW domain protein **19p/16(S→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

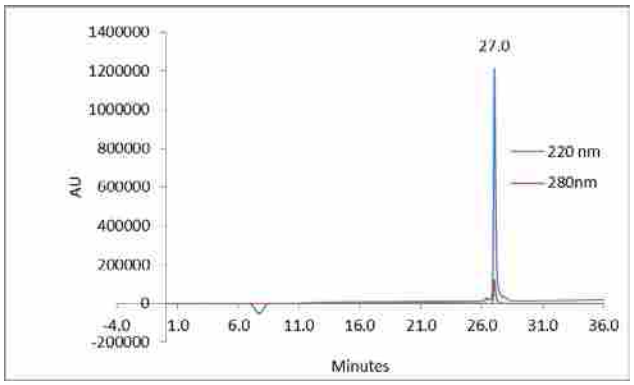


Figure 145: Analytical HPLC data for Pin WW domain protein **19/32(S→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 36 minutes.

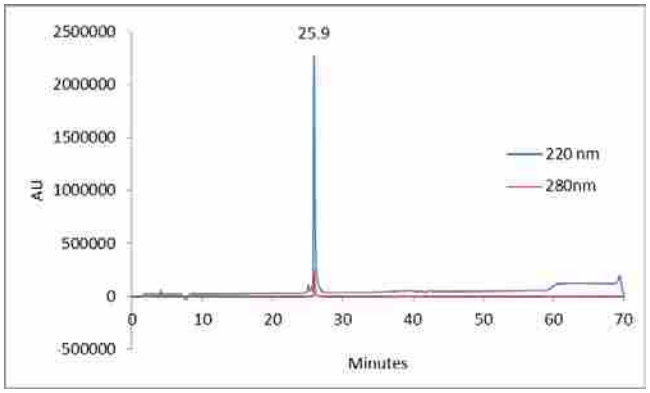


Figure 146: Analytical HPLC data for Pin WW domain protein **19p/32(S→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

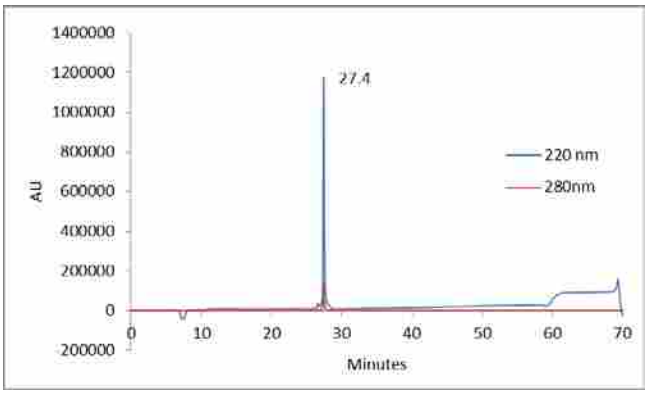


Figure 147: Analytical HPLC data for Pin WW domain protein **16/23(Y→F)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

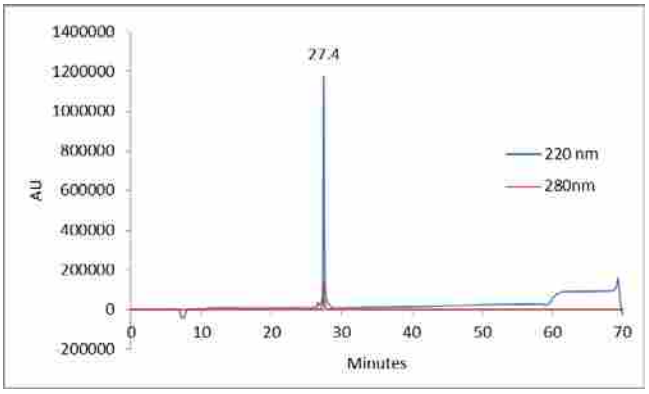


Figure 148: Analytical HPLC data for Pin WW domain protein **16p/23(Y→F)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

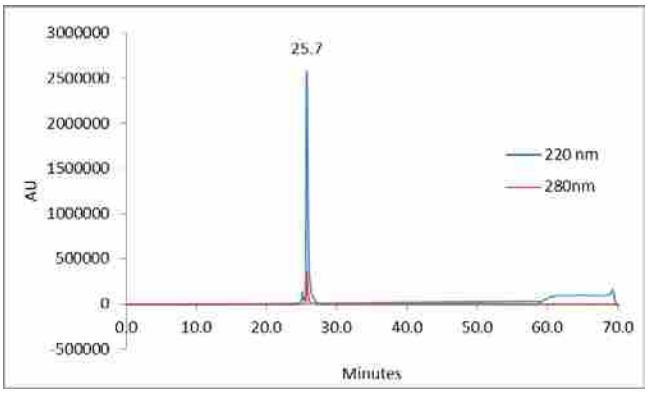


Figure 149: Analytical HPLC data for Pin WW domain protein **16/32(S→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

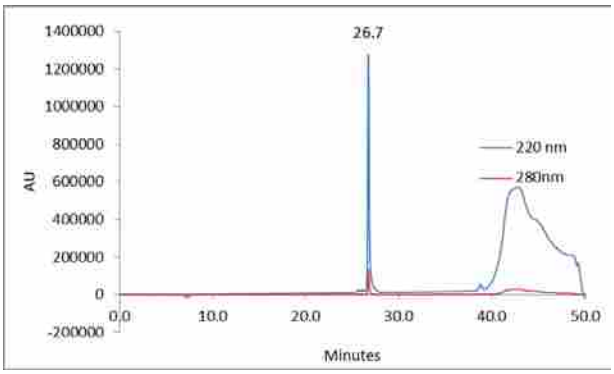


Figure 150: Analytical HPLC data for Pin WW domain protein **16p/32(S→A).** Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 36 minutes.

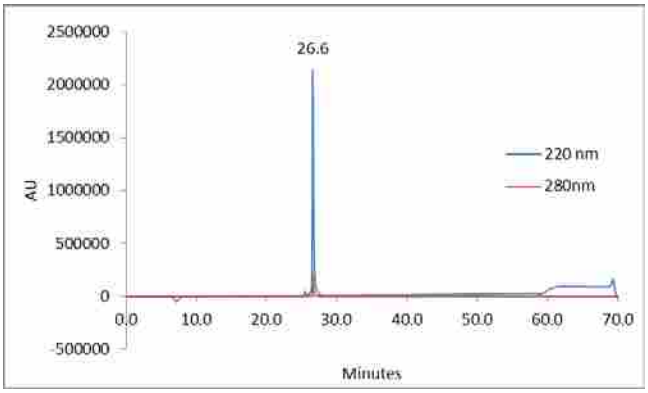


Figure 151: Analytical HPLC data for Pin WW domain protein **26/11(W→X)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

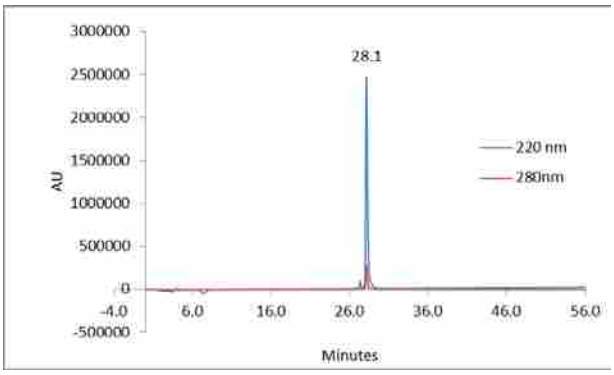


Figure 152: Analytical HPLC data for Pin WW domain protein **26p/11(W→X)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 56 minutes

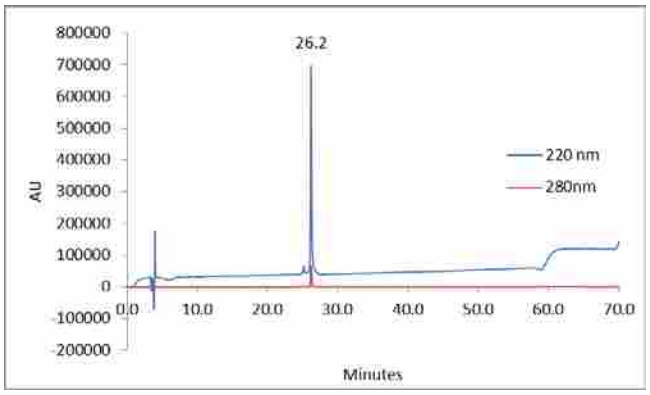


Figure 153: Analytical HPLC data for Pin WW domain protein **29(T→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

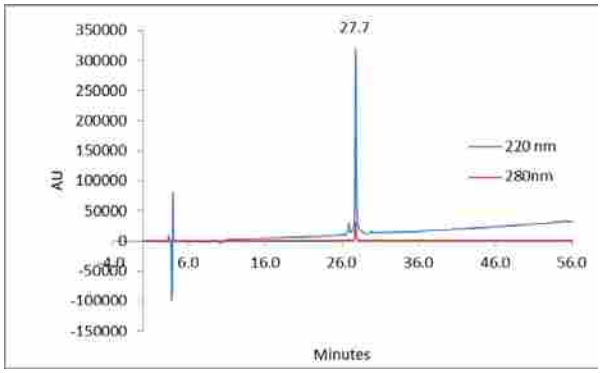


Figure 154: Analytical HPLC data for Pin WW domain protein **26p/29(T→A)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 56 minutes.

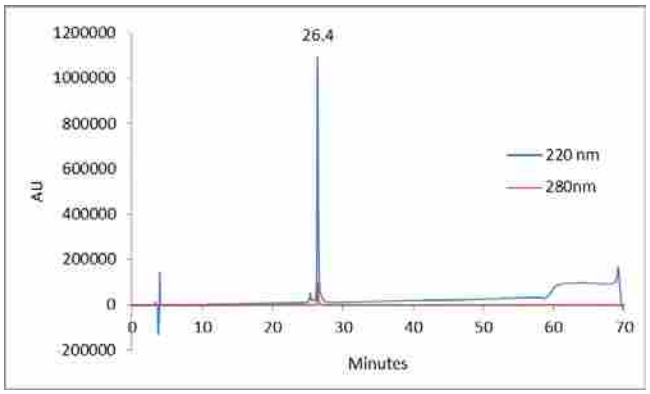


Figure 155: Analytical HPLC data for Pin WW domain protein **23(Y→F)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

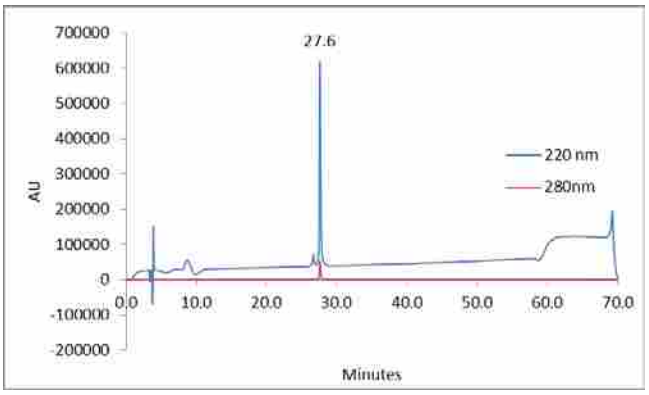


Figure 156: Analytical HPLC data for Pin WW domain protein **26p/23(Y→F)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

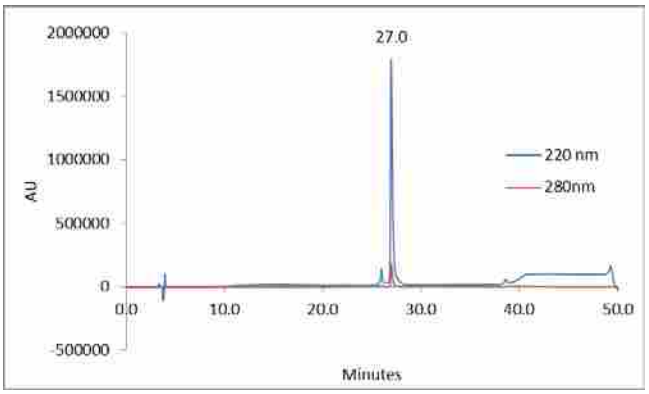


Figure 157: Analytical HPLC data for Pin WW domain protein **16(S→A)/19/23(Y→F).** Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

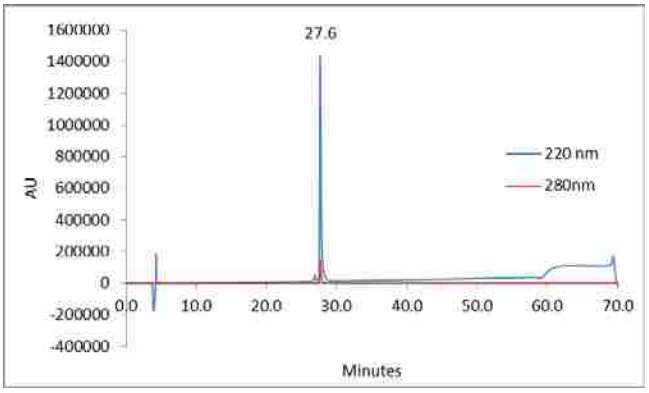


Figure 158: Analytical HPLC data for Pin WW domain protein **19/23(Y→FOMe)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

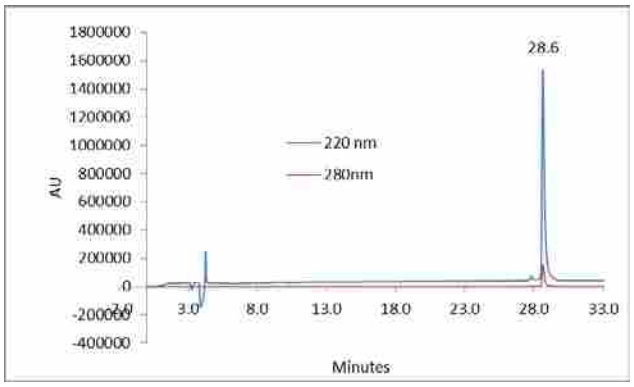
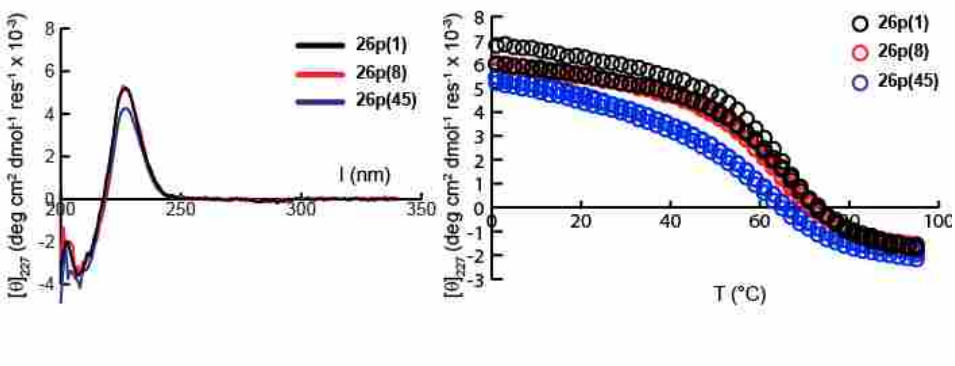


Figure 159: Analytical HPLC data for Pin WW domain protein **19p/23(Y→FOMe)**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated at 33 minutes.

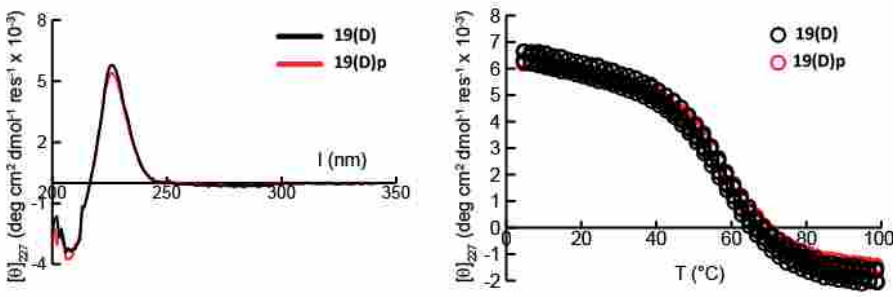
3.5.6 CD Spectra and Thermal Denaturation Plots

CD spectra and thermal denaturation plots for **26p(1), 26p(8), 26p(45), 19(D), 19(D)p, 16/19, 16p/19, 16/19p, 16p/19p, 23/26p, 23p/26p, 19/26p, 19p/26p, 26p/29, 26p/29p, 16/26p, 16p/26p, 19/23(Y→F), 19p/23(Y→F), 19/16(S→A), 19p/16(S→A), 19/32(S→A), 19p/32(S→A), 16/23(Y→F), 16p/23(Y→F), 16/32(S→A), 16p/32(S→A), 11(W→X), 26p/11(W→X), 29(T→A), 26p/29(T→A), 23(Y→F), 26p/23(Y→F), 16(S→A)/19/23(Y→F), 16(S→A)/19p/23(Y→F), 19/23(Y→FOMe), and 19p/23(Y→FOMe)** are shown in Figures 93-108, along with fitting parameters for determining T_m and ΔG_f.



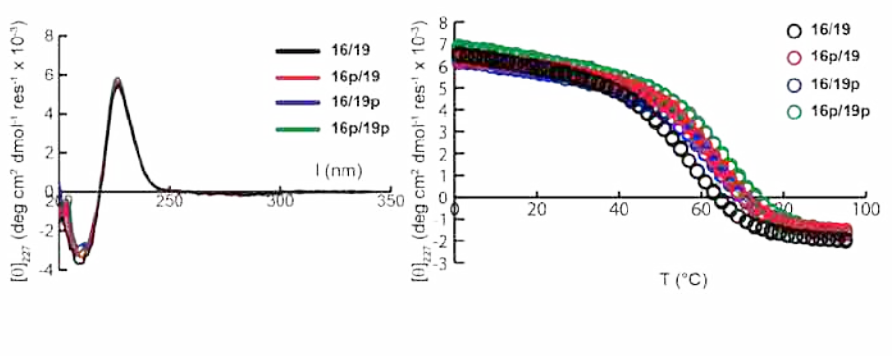
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
26p(1)	337.90 ± 0.15	0.096 ± 0.0013	0.0010 ± 0.000052	0.99989	0.032334
26p(8	335.71 ± 0.16	0.090 ± 0.0013	0.0010 ± 0.000052	0.99988	0.031806
26p(45)	333.36 ± 0.83	0.095 ± 0.0026	0.0007 ± 0.000396	0.99967	0.050379

Figure 160: CD spectra and variable temperature CD data for Pin WW domain proteins **26p(1), 26p(8), and 26p(45)** (which have Asn-linked poly(ethylene glycol) residues at position 26) in 20 mM sodium phosphate, pH 7. The length of the PEG oligomer for **26p(1)** and **26p(8)** were 1 and 8 monomer units, respectively. The PEG polymer for **26p(45)** was a polydisperse mixture 45 monomer units on average. Fit parameters appear in the table, along with parameter standard errors



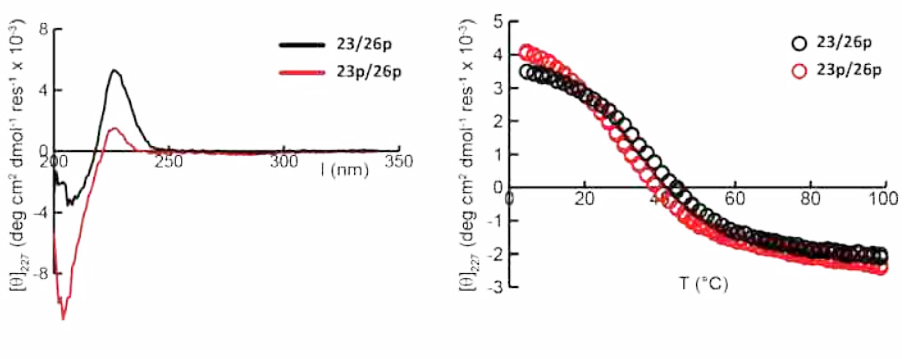
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19(D)	328.57 ± 0.29	0.093 ± 0.0011	0.0002 ± 0.000131	0.99983	0.044277
19(D)p	328.41 ± 0.28	0.090 ± 0.0009	0.0006 ± 0.000118	0.99994	0.023978

Figure 161: CD spectra and variable temperature CD data for Pin WW domain proteins **19(D)** (which has an D-Asn at position 19) and **19(D)p** (which has a D-Asn-linked poly(ethylene glycol) residue at position 19) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



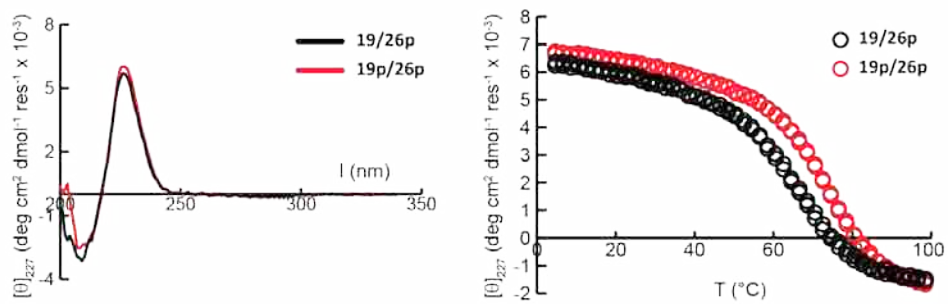
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
16/19	329.97 ± 0.16	0.091 ± 0.0009	0.0009 ± 0.000084	0.99993	0.029673
16p/19	336.27 ± 0.10	0.107 ± 0.0012	0.0011 ± 0.000061	0.99990	0.032707
16/19p	335.37 ± 0.11	0.096 ± 0.0010	0.0010 ± 0.000048	0.99992	0.028747
16p/19p	338.43 ± 0.11	0.103 ± 0.0011	0.0011 ± 0.000037	0.99992	0.028234

Figure 162: CD spectra and variable temperature CD data for Pin WW domain proteins **16/19** (which has Asn at positions 16 and 19) **16p/19** (which has an Asn-linked poly(ethylene glycol) residue at position 16 and Asn at position 19), **16/19p** (which has an Asn at position 16 and Asn-linked poly(ethylene glycol) residue at position 19) and **16p/19p** (which has Asn-linked poly(ethylene glycol) residue at positions 16 and 19) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors



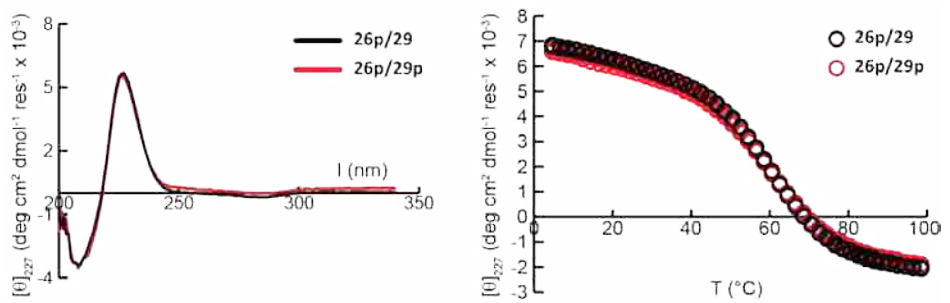
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
23/26p	308.54 ± 0.91	0.073 ± 0.0014	0.0000 ± 0.000182	0.99984	0.026656
23p/26p	305.04 ± 1.24	0.071 ± 0.0026	-0.0006 ± 0.000277	0.99948	0.052609

Figure 163: CD spectra and variable temperature CD data for Pin WW domain proteins **23/26p** (which has an Asn at position 23 and Asn-linked poly(ethylene glycol) residue at position 26) and **23p/26p** (which has Asn-linked poly(ethylene glycol) residue at positions 23 and 26) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



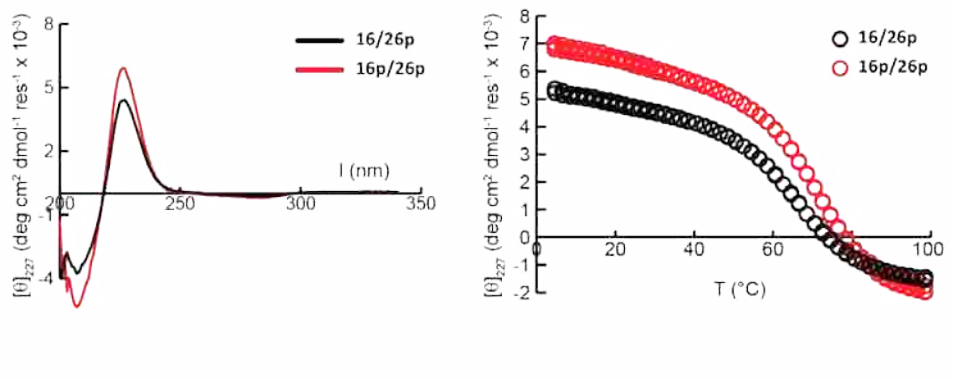
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19/26p	335.35 ± 0.12	0.093 ± 0.0012	0.0011 ± 0.000039	0.99990	0.031276
19p/26p	342.75 ± 0.14	0.105 ± 0.0011	0.0012 ± 0.000025	0.99992	0.027116

Figure 164: CD spectra and variable temperature CD data for Pin WW domain proteins **19/26p** (which has an Asn at position 19 and Asn-linked poly(ethylene glycol) residue at position 26) and **19p/26p** (which has Asn-linked poly(ethylene glycol) residue at positions 19 and 26) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



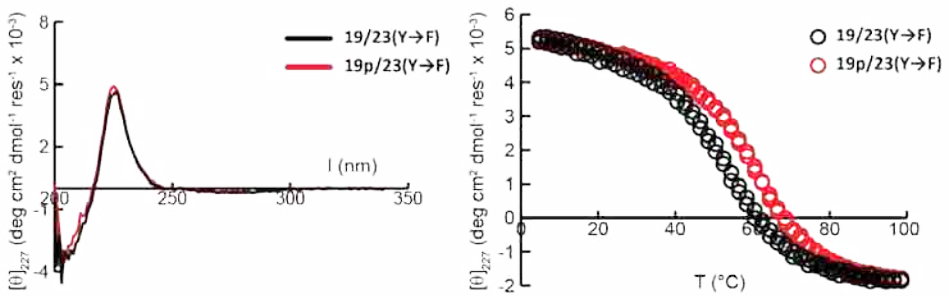
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
26p/29	330.30 ± 0.25	0.090 ± 0.0007	0.0006 ± 0.000102	0.99996	0.021485
26p/29p	329.94 ± 0.28	0.087 ± 0.0007	0.0006 ± 0.000098	0.99997	0.019442

Figure 165: CD spectra and variable temperature CD data for Pin WW domain proteins **26p/29** (which has an Asn-linked poly(ethylene glycol) residue at position 26 and Asn at position 29) and **26p/29p** (which has Asn-linked poly(ethylene glycol) residue at positions 26 and 29) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



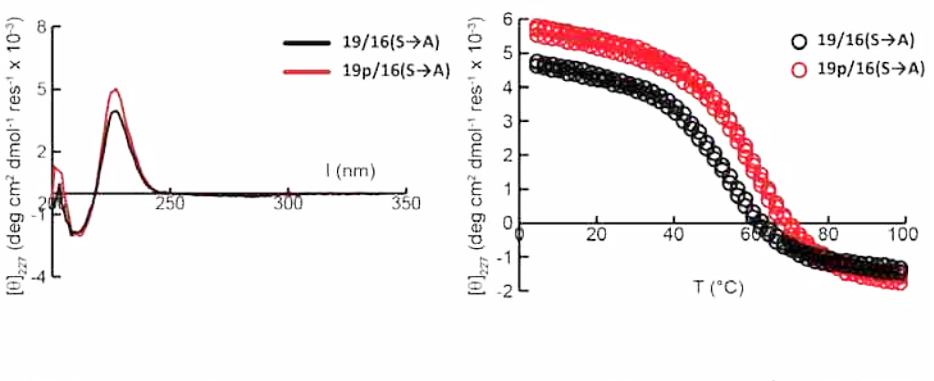
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
16/26p	335.78 ± 0.27	0.092 ± 0.0006	0.0006 ± 0.000102	0.99997	0.015115
16p/26p	340.26 ± 0.08	0.101 ± 0.0007	0.0011 ± 0.000018	0.99997	0.017028

Figure 166: CD spectra and variable temperature CD data for Pin WW domain proteins **16/26p** (which has an Asn at position 16 and Asn-linked poly(ethylene glycol) residue at position 26) and **16p/26p** (which has Asn-linked poly(ethylene glycol) residue at positions 16 and 26) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



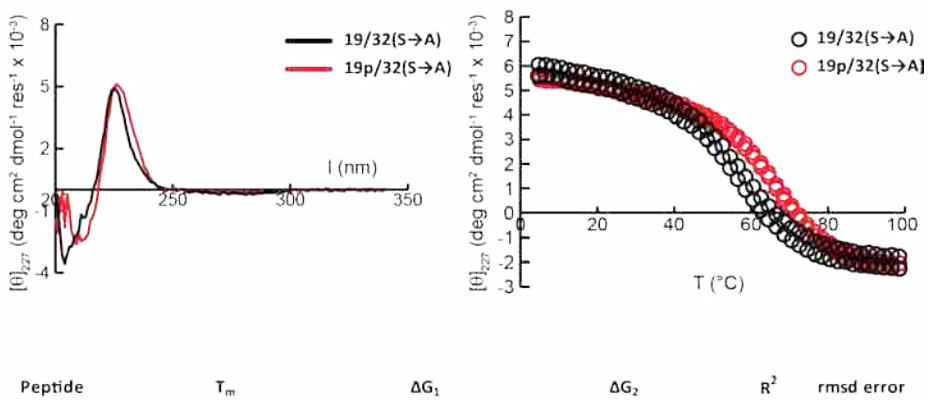
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19/23(Y→F)	324.60 ± 0.38	0.087 ± 0.0012	0.0001 ± 0.000150	0.99988	0.031440
19p/23(Y→F)	329.46 ± 0.11	0.090 ± 0.0012	0.0011 ± 0.000055	0.99988	0.031073

Figure 167: CD spectra and variable temperature CD data for Pin WW domain proteins **19/23(Y→F)** (which has an Asn at position 16 and Phe at position 23) and **19p/23(Y→F)** (which has Asn-linked poly(ethylene glycol) residue at 19 and Phe at position 23) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



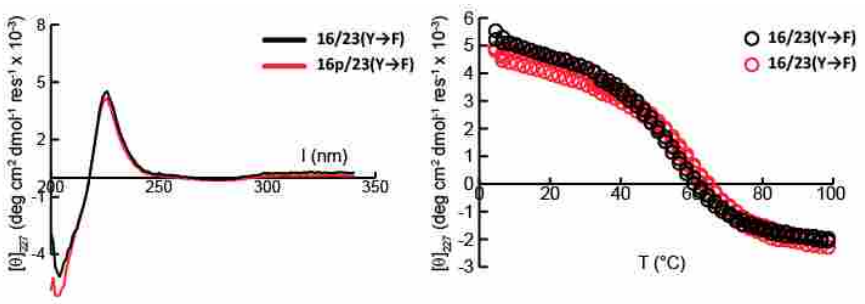
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19/16(S→A)	324.12 ± 0.23	0.092 ± 0.0008	0.0002 ± 0.000115	0.99993	0.021337
19p/16(S→A)	329.96 ± 0.20	0.093 ± 0.0008	0.0008 ± 0.000099	0.99995	0.020939

Figure 168: CD spectra and variable temperature CD data for Pin WW domain proteins **19/16(S→A)** (which has an Asn at position 19 and Ala at position 16) and **19p/16(S→A)** (which has Asn-linked poly(ethylene glycol) residue at 19 and Ala at position 16) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



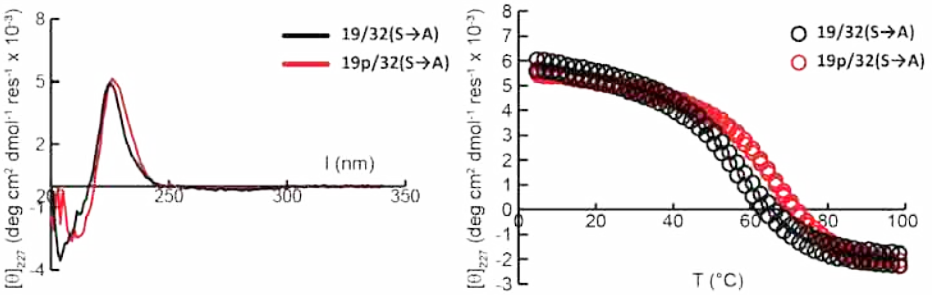
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19/32(S→A)	327.57 ± 0.27	0.087 ± 0.0007	0.0004 ± 0.000098	0.99996	0.020975
19p/32(S→A)	335.81 ± 0.11	0.092 ± 0.0008	0.0009 ± 0.000047	0.99995	0.020871

Figure 169: CD spectra and variable temperature CD data for Pin WW domain proteins **19/32(S→A)** (which has an Asn at position 19 and Ala at position 32) and **19p/32(S→A)** (which has Asn-linked poly(ethylene glycol) residue at 19 and Ala at position 32) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



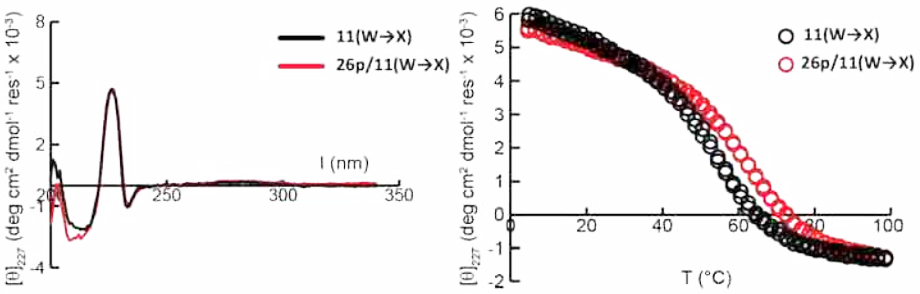
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
16/23(Y→F)	323.87 ± 0.67	0.082 ± 0.0031	0.0008 ± 0.000252	0.99958	0.059298
16p/23(Y→F)	330.21 ± 1.21	0.082 ± 0.0022	0.0004 ± 0.000340	0.99957	0.055758

Figure 170: CD spectra and variable temperature CD data for Pin WW domain proteins **16/23(Y→F)** (which has an Asn at position 16 and Phe at position 23) and **16p/23(Y→F)** (which has Asn-linked poly(ethylene glycol) residue at 16 and Phe at position 23) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



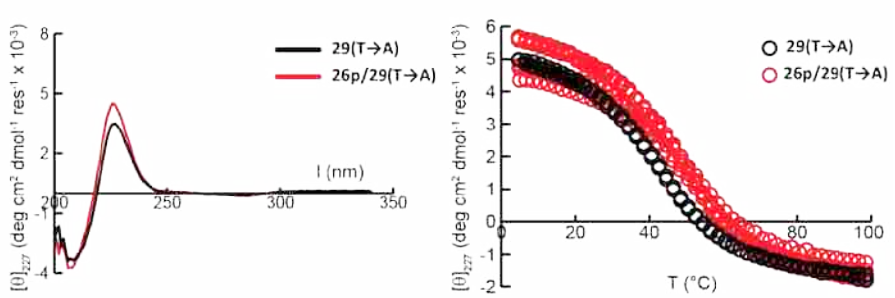
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19/32(S→A)	327.57 ± 0.27	0.087 ± 0.0007	0.0004 ± 0.000098	0.99996	0.020975
19p/32(S→A)	335.81 ± 0.11	0.092 ± 0.0008	0.0009 ± 0.000047	0.99995	0.020871

Figure 171: CD spectra and variable temperature CD data for Pin WW domain proteins **19/32(S→A)** (which has an Asn at position 19 and Ala at position 32) and **19p/32(S→A)** (which has Asn-linked poly(ethylene glycol) residue at 19 and Ala at position 32) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



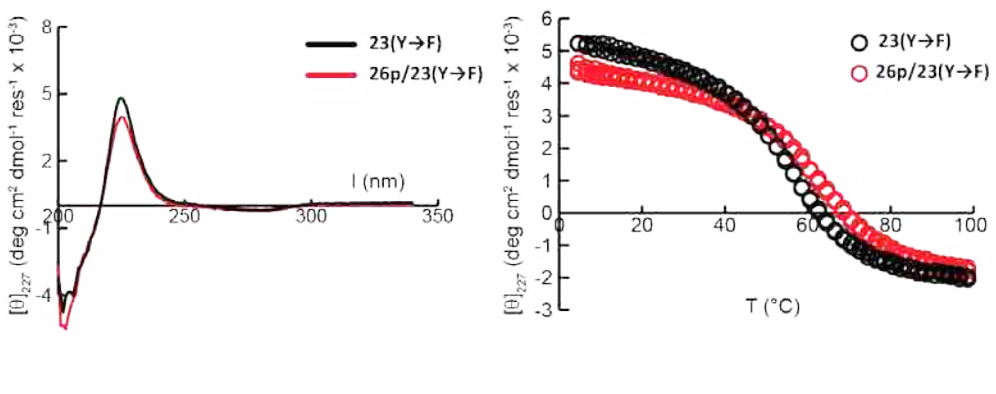
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
11(W→X)	325.19 ± 0.37	0.091 ± 0.0012	0.0005 ± 0.000166	0.99990	0.028151
26p/(W→X)	331.56 ± 0.16	0.086 ± 0.0011	0.0009 ± 0.000069	0.99991	0.024229

Figure 172: CD spectra and variable temperature CD data for Pin WW domain proteins **11(W→X)** (which has naphalalanine position 11) and **26p/11(W→X)** (which has Asn-linked poly(ethylene glycol) residue at 26 and naphalalanine at position 11) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



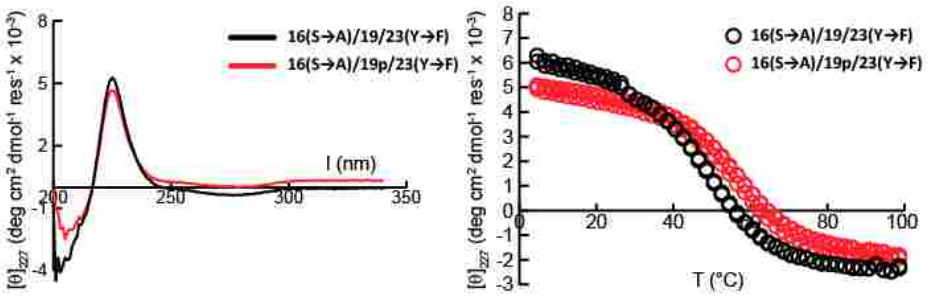
Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
26/29(T→A)	313.58 ± 0.72	0.077 ± 0.0014	0.0001 ± 0.000178	0.99986	0.030950
26/29(T→A)	318.26 ± 0.37	0.071 ± 0.0018	0.0008 ± 0.000105	0.99976	0.050668

Figure 173: CD spectra and variable temperature CD data for Pin WW domain proteins **29(T→A)** (which has Ala position 29) and **26p/29(T→A)** (which has Asn-linked poly(ethylene glycol) residue at 26 and Ala at position 29) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



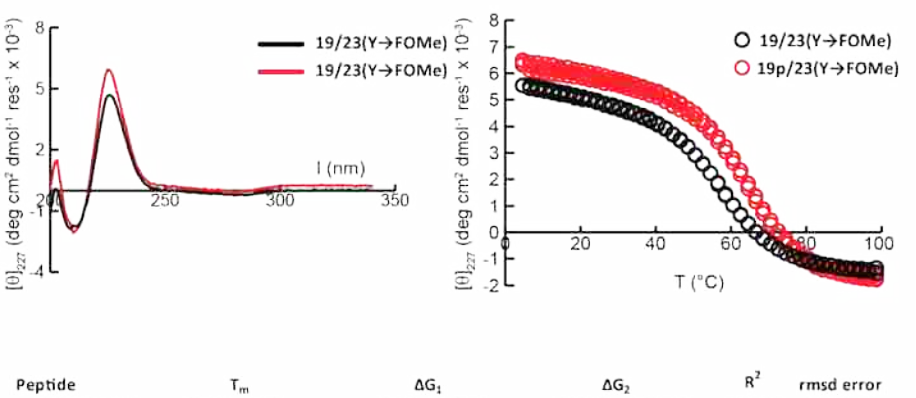
23(Y→F)	326.36 ± 0.32	0.092 ± 0.0010	0.0004 ± 0.000149	0.99990	0.028499
26p/23(Y→F)	332.97 ± 0.57	0.088 ± 0.0014	0.0004 ± 0.000197	0.99984	0.030204

Figure 174: CD spectra and variable temperature CD data for Pin WW domain proteins **23(Y→F)** (which has Phe position 23) and **26p/23(Y→F)** (which has Asn-linked poly(ethylene glycol) residue at 26 and Phe position 23) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
16(S→A)/19/23(Y→F)	318.99 ± 1.12	0.078 ± 0.0038	0.0006 ± 0.000272	0.99969	0.060946
16(S→A)/19p/23(Y→F)	326.90 ± 0.26	0.092 ± 0.0009	0.0002 ± 0.000120	0.99993	0.023550

Figure 175: CD spectra and variable temperature CD data for Pin WW domain proteins **16(S→A)/19/23(Y→F)** (which has Ala at position 16, Asn at position 19, and Phe position 23) and **16(S→A)/19/23(Y→F)** (which has Ala at position 16, Asn-linked poly(ethylene glycol) residue at 19 and Phe position 23) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors.



Peptide	T_m	ΔG_1	ΔG_2	R^2	rmsd error
19/23(Y→FOMe)	328.11 ± 0.22	0.091 ± 0.0007	0.0003 ± 0.000096	0.99995	0.018976
19p/23(Y→FOMe	333.86 ± 0.20	0.096 ± 0.0009	0.0009 ± 0.000104	0.99991	0.030227

Figure 176: CD spectra and variable temperature CD data for Pin WW domain proteins **19/23(Y→FOMe)** (which has an Asn at position 16 and methoxyphenylalanine at position 23) and **19p/23(Y→FOMe)** (which has Asn-linked poly(ethylene glycol) residue at 19 and methoxyphenylalanine at position 23) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with parameter standard errors

3.5.7 Global Fitting of Variable Temperature CD Data to Obtain Enthalpic and Entropic Components of ΔGf

It is possible to fit thermal denaturation curves to the equation

$$\Delta G_f = \frac{\Delta H(T_m) \cdot (T_m - T)}{T_m} + \Delta C_p(T - T_m - T_m\left(\frac{T}{T_m}\right)) \quad (2)$$

where T is temperature in Kelvin, ΔH is the enthalpy of folding, T_m is the melting temperature of the peptide, and ΔC_p is the change in heat capacity upon folding. These thermodynamic parameters were obtained for each peptide as parameters of a global fit, analogously to the methods described in Chapter 2. Entropy of folding can be calculated from these parameters using the equation

$$-T\Delta S = T\left(\frac{\Delta H}{\Delta C_p} + T_m \ln\frac{T}{T_m}\right) \quad (\textbf{Error! Bookmark not defined.})$$

where T is the reference temperature (in this case 60°C).

3.5.8 References

(1) Koepf, E. K.; Petrassi, H. M.; Sudol, M.; Kelly, J. W.: WW: An isolated three-stranded antiparallel [beta]-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *PRS* **1999**, *8*, 841-853.
(2) Shental-Bechor, D.; Levy, Y.: Effect of Glycosylation on Protein Folding: A Close Look at Thermodynamic Stabilization. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 8256-8261.
(3) Pandey, B. K.; Smith, M. S.; Torgerson, C.; Lawrence, P. B.; Matthews, S. S.; Watkins, E.; Groves, M. L.; Prigozhin, M. B.; Price, J. L.: Impact of Site-Specific PEGylation on the Conformational Stability and Folding Rate of the Pin WW Domain Depends Strongly on PEG Oligomer Length. *Bioconjug. Chem.* **2013**, *24*, 796-802.
(4) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M.: The folding mechanism of a β-sheet: the WW domain. *J. Mol. Biol.* **2001**, *311*, 373-393.
(5) Kaul, R.; Angeles, A. R.; Jäger, M.; Powers, E. T.; Kelly, J. W.: Incorporating β-Turns and a Turn Mimetic out of Context in Loop 1 of the WW Domain Affords Cooperatively Folded β-Sheets. *J. Am. Chem. Soc.* **2001**, *123*, 5206-5212.
(6) Rodriguez-Martinez, J. A.; Sola, R. J.; Castillo, B.; Cintron-Colon, H. R.; Rivera-Rivera, I.; Barletta, G.; Griebenow, K.: Stabilization of alpha-chymotrypsin upon PEGylation correlates with reduced structural dynamics. *Biotechnol. Bioeng.* **2008**, *101*, 1142-9.
(7) Kessel, A.; Ben-Tal, N.: Introduction to Proteins: Structure, Function, and Motion. Taylor & Francis, 2012; pp 274-285.
(8) Pokkuluri, P. R.; Raffen, R.; Dieckman, L.; Boogaard, C.; Stevens, F. J.; Schiffer, M.: Increasing protein stability by polar surface residues: domain-wide consequences of interactions within a loop. *Biophys. J.* **2002**, *82*, 391-8.
(9) Yamagata, Y.; Kubota, M.; Sumikawa, Y.; Funahashi, J.; Takano, K.; Fujii, S.; Yutani, K.: Contribution of Hydrogen Bonds to the Conformational Stability of Human Lysozyme: Calorimetry and X-ray Analysis of Six Tyrosine → Phenylalanine Mutants†,‡. *Biochemistry (Mosc).* **1998**, *37*, 9355-9362.
(10) Price, J. L.; Powers, E. T.; Kelly, J. W.: N-PEGylation of a Reverse Turn Is Stabilizing in Multiple Sequence Contexts, unlike N-GlcNAcylation. *ACS Chemical Biology* **2011**, *6*, 1188-1192.
(11) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W.: Glycosylation of the enhanced aromatic sequon is similarly stabilizing in three distinct reverse turn contexts. *Proceedings of the National Academy of Sciences* **2011**, *108*, 14127-14132.
(12) Meng, W.; Guo, X.; Qin, M.; Pan, H.; Cao, Y.; Wang, W.: Mechanistic Insights into the Stabilization of srcSH3 by PEGylation. *Langmuir* **2012**, *28*, 16133-16140.
(13) Svergun, D. I.; Ekström, F.; Vandegriff, K. D.; Malavalli, A.; Baker, D. A.; Nilsson, C.; Winslow, R. M.: Solution Structure of Poly(ethylene) Glycol-Conjugated Hemoglobin Revealed by Small-Angle X-Ray Scattering: Implications for a New Oxygen Therapeutic. *Biophys. J.* **2008**, *94*, 173-181.
(14) Murray, J. K.; Gellman, S. H.: Application of Microwave Irradiation to the Synthesis of 14-Helical β-Peptides. *Organic Letters* **2005**, *7*, 1517-1520.
(15) Edelhoch, H.: Spectroscopic Determination of Tryptophan and Tyrosine in Proteins*. *Biochemistry (Mosc).* **1967**, *6*, 1948-1954.
(16) Herzner, H.; Kunz, H.: Spacer-separated sialyl LewisX cyclopeptide conjugates as potential E-selectin ligands. *Carbohydr. Res.* **2007**, *342*, 541-57.