2016-06-01

Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic Stability

Paul B. Lawrence
Brigham Young University

Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic Stability

Paul B. Lawrence

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

Doctor of Philosophy

Joshua L. Price, Chair
Paul B. Savage
Matt A. Peterson
Merritt B. Andrus
Richard K. Watt

Department of Chemistry and Biochemistry

Brigham Young University

June 2016

ABSTRACT

Criteria for Selecting PEGylation Sites on Proteins for Higher Thermodynamic Stability

Paul B. Lawrence
Department of Chemistry and Biochemistry, BYU
Doctor of Philosophy

PEGylation of protein side-chains has been used for more than 30 years to enhance the pharmacokinetic properties of protein drugs, and has been enabled by the recent development of many chemoselective reactions for protein side-chain modification. However, there are no structure- or sequence-based guidelines for selecting sites that provide optimal PEG-based pharmacokinetic enhancement with minimal loss to biological activity.

Chapter 1 is a brief introduction to protein PEGylation. In chapter 2 we use the WW domain of the human protein Pin 1 (WW) as a model system to probe the impact of PEG on protein conformational stability. Using a combination of experimental and theoretical approaches, we develop a structure-based method for predicting which sites within WW are most likely to experience PEG-based stabilization, and show that this method correctly predicts the location of a stabilizing PEGylation site within the chicken Src SH3 domain. PEG-based stabilization in WW is associated with enhanced resistance to proteolysis, is entropic in origin, and likely involves disruption by PEG of the network of hydrogen-bound solvent molecules that surround the protein. Chapter 3 shows that PEG-based stabilization of the WW domain depends strongly on the identity of the PEG-protein linker, with the most stabilizing linkers involving conjugation of PEG to an Asn side-chain amide nitrogen. Chapter 4 investigates the interplay between structure-based guidelines for PEG-base stabilization developed in chapter 2 and the different chemistries explored in chapter 3.

Keywords: PEGylation, Therapeutic Proteins, Thermodynamic Stability, Circular Dichroism, β-sheet.

ACKNOWLEDGEMENTS

Thanks to my advisor, Joshua Price, for his overall goodness and patience throughout my graduate experience at BYU. His advice, expertise, and encouragement has made this dissertation possible.

Thanks to my wife, Jessica, and little girls, Lillyan and Lucy, for enduring an absent husband and father during these past five years. Thanks to Mason Smith, who, among other things, has so kindly trained new undergraduate and graduate students while I have been writing my dissertation. Thanks to Brijesh Pandey, our first post-doc, who was so willing to take time away from his projects to help me with mine. Thanks to Wendy, Minnie, Mason, Chad, Cameron, Jake, Andy, Mindy, Andrew, and Ryan who have helped me with synthesis, purification, and characterization of the many peptides used in my research. Thanks also to the Charles E. and Margaret P. Maw Fellowship that has given me financial support for the past year.

TABLE OF CONTENTS

1 INTRODUCTION

Peptides and proteins are attractive targets for treatment of many human diseases.[1-3] However, their benefits are limited because of fast degradation by proteases, filtration through the kidneys, aggregation, and recognition/neutralization by antibodies.[4-9] To mitigate these problems Davis and Abuchowsky[4,5] developed a technique in 1977 called PEGylation in which they covalently attached a 5000 da polyethylene glycol (PEG) molecule to the surface of bovine serum albumin. Abuschowsky et al.[5] reported that the PEGylated bovine serum albumin showed substantial changes in solubility, electrophoretic mobility in acrylamide gel, ion exchange chromatography, sedimentation, and extended protein half-life in blood. PEGylation has also been shown to reduce protein aggregation, proteolysis, recognition by antibodies, and increase protein shelf-life.[6-8,10,11] There are currently ten FDA approved PEGylated therapeutic proteins, antibody fragments, and oligonucleotides on the market, including PEGylated bovine adenosine deaminase: pegademase bovine (Adagen®); and PEGylated L-asparaginase: pegaspargase (Oncaspar®); PEGylated products of interferon-α (IFN- α): peginterferonα-2b (PegIntron®) and peginterferonα-2a (Pegasys®); PEGylated granulocyte colony stimulating factor (G-CSF): pegfilgrastim (Neulasta®); PEGylated growth hormone receptor antagonist: pegvisomant (Somavert®); a PEGylated 28-nucleotide aptamer against vascular endothelial growth factor (VEGF); PEGylated erythropoietin (Mircera®); PEGylated anti-tumornecrosis factor (TNF)-α(Cimzia®); and PEGylated recombinant porcine uricase (urate oxidase) (Puricase®).[12]

PEGylated proteins typically have longer half-lives than the parent protein due in part to the decrease in renal clearance. Hydration of the PEG polymer increases the hydrodynamic radius of the PEGylated protein 5-10 times greater than what one would calculate from the polymers molecular weight.[6] This hydration effect causes an increase to the molecular size of the PEG-protein conjugate, which reduces renal filtration and increases circulation time. PEG-protein conjugates with molecular weights of below 20 kDa are filtered through the kidneys; while molecules above 20 kDa are generally excreted via the liver.[13,14]

PEGylation also addresses the challenge of proteolytic degradation and immunogenicity of protein therapeutics. A common model that is used to explain how PEG protects proteins from proteases and antigenic determinants involves the large PEG polymer shielding the protein surface.[10] Proteases and antibodies have reduced access to the protein surface due to steric obstruction by PEG.

The FDA approved two PEGylated protein drugs (peginterferon-α2b and peginterferon-α2a) in 2000 and 2001, respectively, for the treatment of chronic hepatitis C (HCV). Prior to using peginterferon-α2b and peginterferon-α2a, hepatitis patients were treated in combination with ribavirin and interferon-α (IFN-α).[15] A problem with this treatment was the plasma half-life for IFN-α was only 4-6 hours, which required patients to receive three subcutaneous injections a week. Also, the short plasma half-life of IFN-α resulted in continual fluctuations in plasma concentration, which created periods of time for the virus to replicate and develop resistance.[10]PEGylating IFN-α with a linear 12 kDa PEG[10,16] via an amide bond to a surface exposed lysine resulted in an increased plasma half-life of 27-37 hours. IFN-α was also PEGylated with a 40 kDa branched PEG[10,17] (peginterferon-α2a) resulting in a plasma half-life

of 65 hours. PEGylation of IFN-α has resulted in approximately a ten fold increase to plasma half-life. The increased plasma half-life of peginterferon-α2a has reduced the amount of subcutaneous injections and more than doubled patients response to treatment of HCV.

1.1 Methods of PEGylation

The least complicated strategy for attaching PEG to proteins uses an excess of PEG polymers that are functionalized with succinimide[18,19], maleimide[20,21], and aldehyde[22,23] groups. The resulting PEG electrophile reacts with naturally occurring nucleophiles in the side-chains of lysine and cysteine. A major problem with this approach is that many different nucleophiles on the protein surface can react with the excess PEG reagent, resulting in a heterogeneous mixture of protein-PEG conjugates (Figure 1-1 A). Identification and purification of these protein mixtures is difficult because the proteins vary in the location of the PEGylation sites and the extent of PEGylation.[24] This method of PEGylation may result in PEG attachment near or at the active site of proteins, thereby diminishing biological activity (see Figure 1-1 A). To avoid this problem, many groups have developed site-specific PEGylation strategies that are selective for cysteine, for the N-terminal amino group, or for an unnatural amino acid.

PEGylation of the N-terminus is a common site-specific strategy (see Figure 1-2 A) that takes advantage of the different pKa values of the α-amine group of the N-terminal amino acid (pKa, 7.6-8.0) and the ε-amine group of the lys side-chain (pKa, 10.0-10.2).[25] Under mildly acidic conditions, PEG polymers that are functionalized with an aldehyde are conjugated selectively to the N-terminus via reductive alkylation.[26]

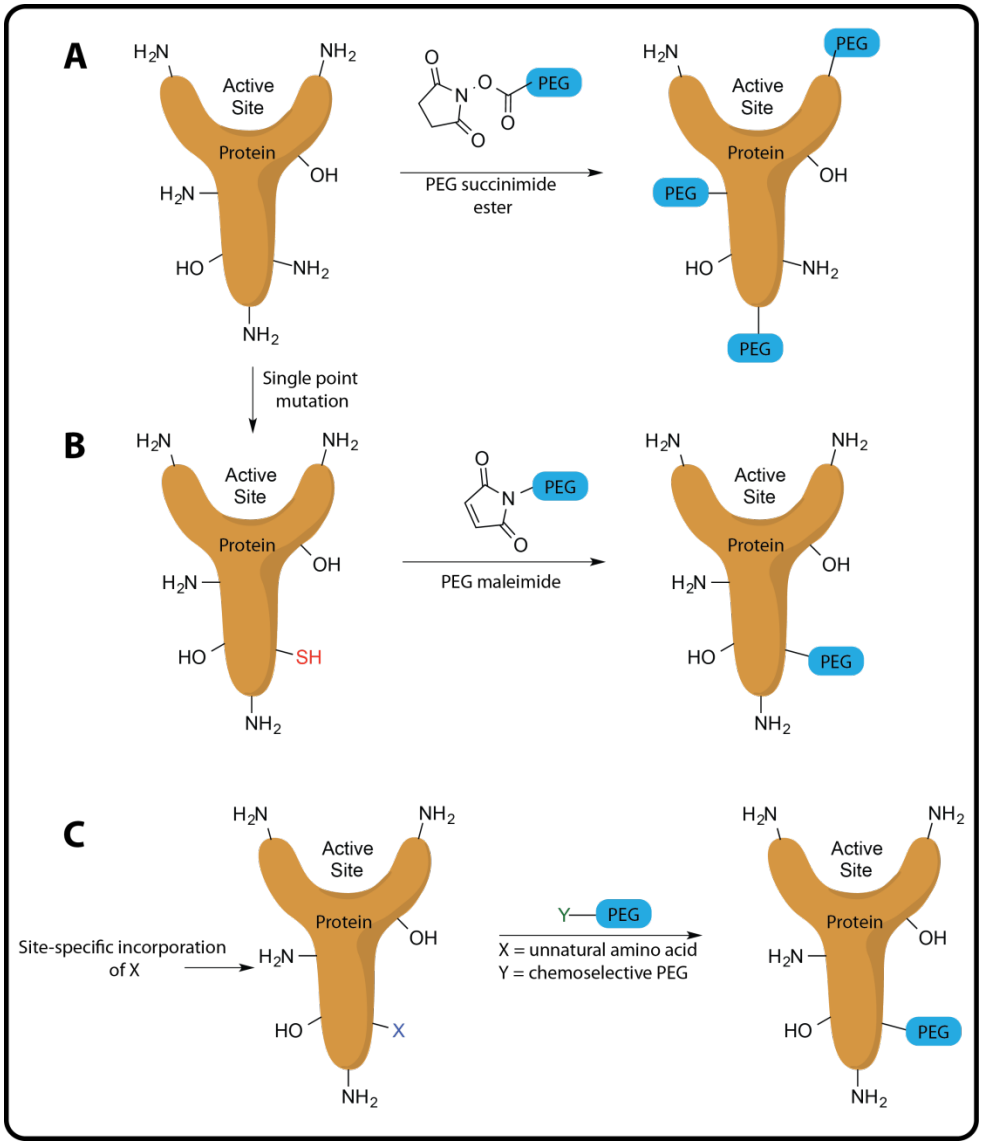


Figure 1-1 Examples of protein PEGylation. (A) Random PEGylation. (B) Cysteine PEGylation. (C) Chemoselective PEGylation of unnatural amino acids.

Site-specific PEGylation of naturally occurring or genetically introduced cysteine residues using maleimide, vinyl sulfone, ortho-pyridyl disulfide, thiol, and iodoacetamide PEG reagents is a common approach[27,28] (see Figure 1-2 c-f). Recently Davis et al.[29] have designed an elegant method for converting the thiol of a cysteine residue into dehydroalanine via bisalkylation of the thiol, followed by β-elimination to generate the unsaturated dehydroalanine containing protein.

The newly installed dehydroalanine is a Michael acceptor, which can then be treated with a thiol containing PEG polymer to form the protein-PEG conjugate (see Figure 1-4 a). Davis applied this method to a single cysteine mutant of the serine protease subtilisin *Bacillus lentus* (SBL) to installed an allyl sulfide in place of the cysteine thiol followed by cross-metathesis with an allyl-PEG to form the protein-PEG conjugate.[30] It is important to note that functionalizing dehydroalanine-containing proteins with thiols is known to result in a mixture of diastereomers, each containing one of two possible Cys-functionalized epimers at the site of modification.

Another way to incorporate PEG site-specifically into proteins is to use unnatural amino acids that have unique chemoselective handles such as azides, terminal alkynes, ketones, and alkenes. Some examples of these unnatural amino acids include: Azidohomoalanine (Aha), homopropargylglycine (Hpg), homoallylglycine (Hag), trans-crotylglycine (Tcg), azidonorleucine (Anl), 2-aminooctynoic acid (Aoa), *p*-ethynylphenylalanine (*p*–EtF) , *p*-bromophenylalanine (*p*–BrF), *p*–iodophenylalanine (*p*–IF), *p*-azidophenylalanine (*p*–AzF), *p*-acetylphenylalanine (*p*–AcF), *p*-propargyloxyphenylalanine (*p*–pPa), and oxonorvaline (Onv).[31,32] The unnatural amino acids in Figure 1-3A are all methionine analogs that use endogenous translational machinery to incorporate the amino acid site-specifically into full length proteins.[33-35] Using the amber stop codon and mutated aminoacyl tRNA synthetases (aaRSs) the unnatural amino acids in Figure 1-3B are incorporated into proteins by using E. coli[36,37], yeast[38], or cell-free methods.[31,39] After incorporation of the unnatural amino, the unique chemical handle is then used to conjugate an appropriately functionalized PEG polymer to the protein. For example, Aha (see Figure 1-4b) has an azido group that will chemoselectively react with an alkyne-functionalized PEG polymer in a bioorthogonal copper-catalyzed [3+2] cycloadditon reaction termed a “click reaction”.[40]

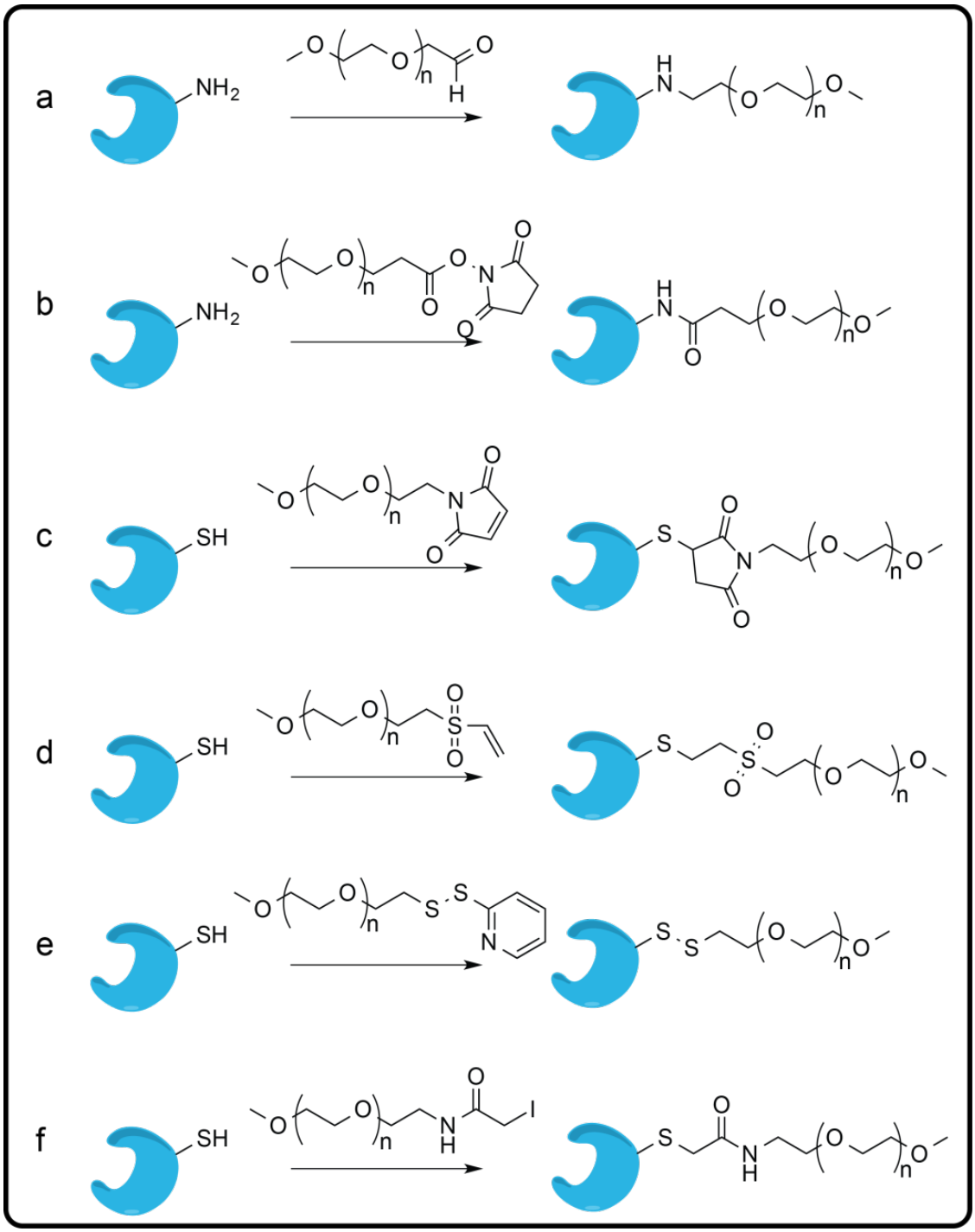


Figure 1-2. Amine and thiol reactive PEG derivatives. a) PEG aldehyde. b) PEG succinimidyl carboxyl methyl ester. c) PEG maleimide. d) PEG vinyl sulfone. e). PEG orthopyridyl disulfide. f) PEG iodoacetamide

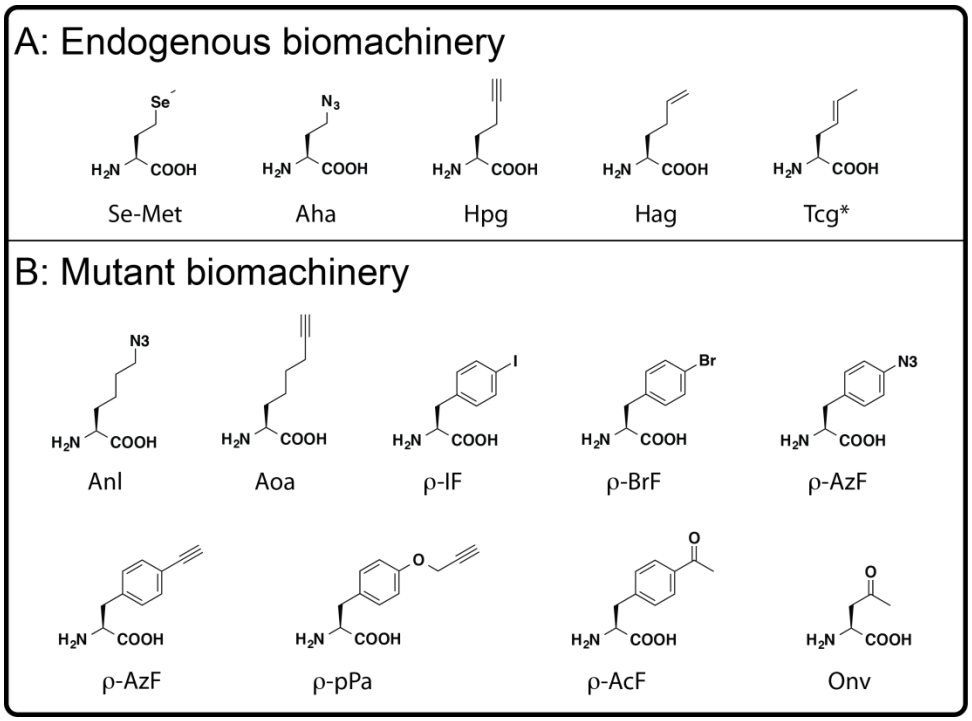


Figure 1-3. Structures of unnatural amino acids. *Tcg requires overexpression of wild type methionyl-tRNAsynthetase

Cho et al.[36] incorporated a ketone-containing amino acid called *p*-acetylphenylalanine (*p*-AcF Figure 1-3) into the human growth hormone (hGH) via amber suppression. The ketone was then reacted with a 30kDa linear PEG-oxyamine to form a very stable oxime linkage (see Figure 1-4c). Lin et al. genetically modified ubiquitin with homo-propargylglycine (Hpg see Figure 1-3) via auxotrophic expression. They then synthesized an iodoaryl-PEG polymer of 5kDa and used the Sonogashira C-C cross-coupling reaction to form the PEG-protein conjugate (see Figure 1-4d). Davis et al.[41] used the Suzuki-Miyaura cross-coupling reaction with a 2kDa functionalized PEG boronic acid and unnatural *p*-iodophenylalanine (*p*-IF Figure 1-3), which was incorporated into an all β-helix protein from *Nostopunctiforme* via amber suppression (Figure 1-4 e).

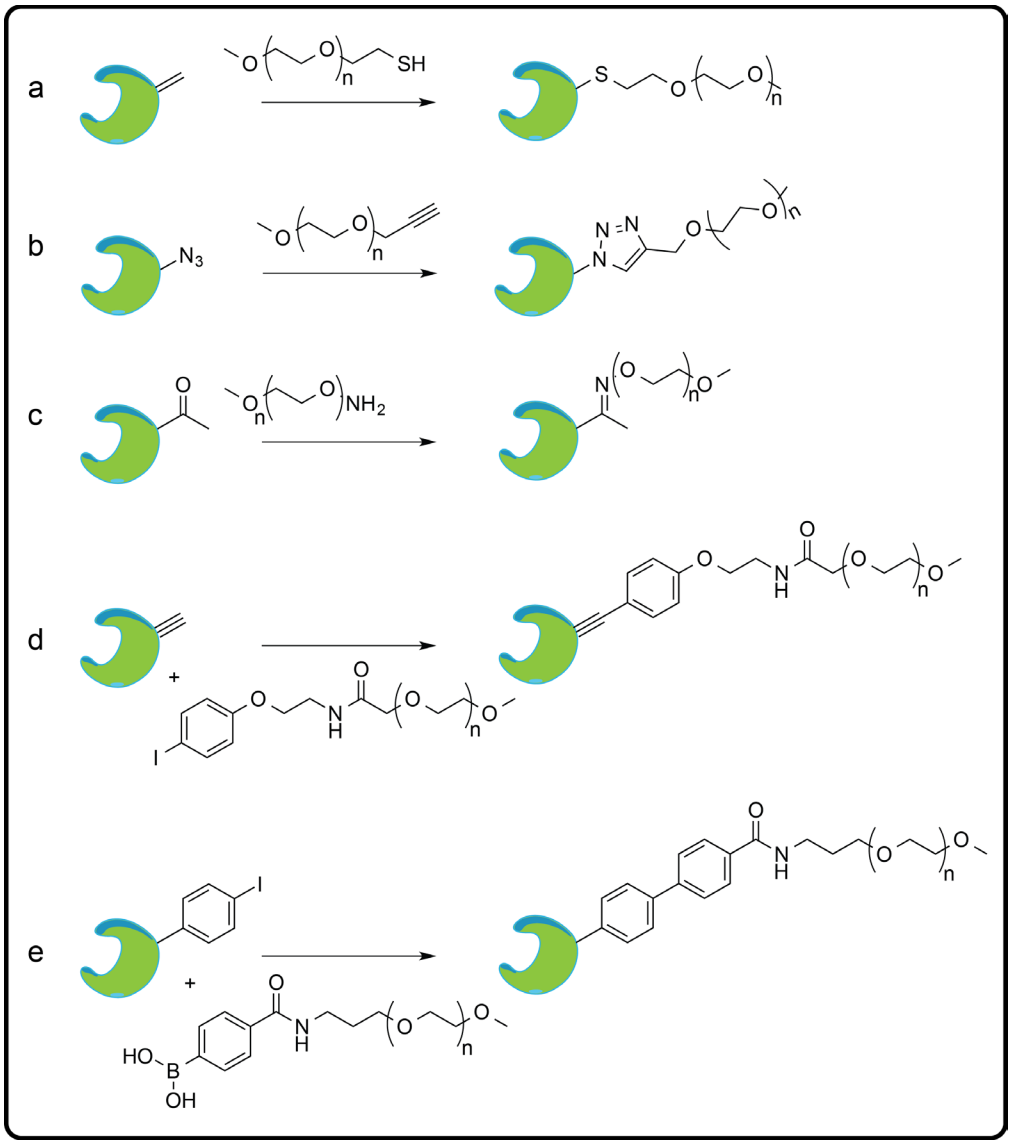


Figure 1-4. Unnatural amino acids with chemoselective handles. a) PEG thiol. b) PEG alkyne. c) PEG oxyamine. d) PEG alkyne conjugation via Sonogashira cross-coupling. e) PEG boronic acid conjugation via Suzuki-Miyaura cross-coupling.

Alternatively, efforts to incorporate PEGylated unnatural amino acids directly into proteins have experienced some recent success (see Figure 1-5). Shozen[42] and Zang[43] recently incorporated PEGylated *p*-aminophenylalanine (PEGnAF) and lysine (PEGnLys) derivatives having 4, 8, and 12 PEG units into streptavidin using a frame shift suppression method based on

a four-base codon-anticodon pair. A longer 24-unit PEG could also be incorporated, but at significantly reduced yields. A branched PEGylated variant of p-amniophenylalanine (i.e.$PEG_{12x3}AF \approx 2kDa$) was not successfully incorporated into the protein, probably due to steric bulk of the branched PEG. These early results are encouraging, but if this method is to be used more broadly, it will need to provide access to the larger PEGs (~20–40 kDa) that are more typically used in therapeutic contexts.

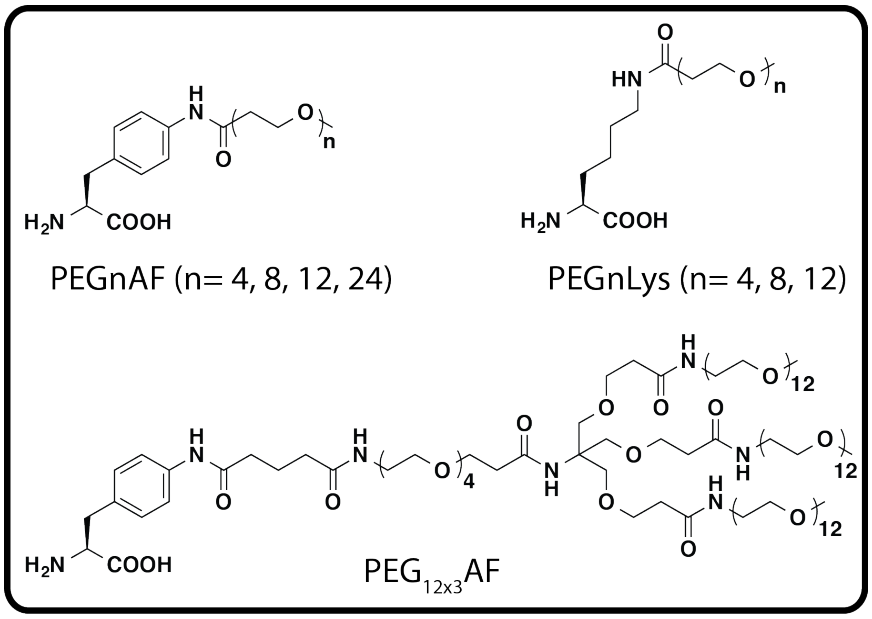


Figure 1-5. PEGylated amino acids. p-aminophenylalanine (PEGnAF) and lysine (PEGnLys)

1.2 The Effect of PEGylation on Protein Thermodynamic Stability

The pharmacokinetic benefits of PEGylated proteins are thought to derive from increased hydrodynamic volume of the PEG-protein conjugate.[11] The increased volume helps reduce the protein from being filtered out of the serum via the kidneys.[5,7,8,44,45] The large PEG polymer also shields the protein from proteases, antibodies, and aggregation.[6-8,10] However, aggregation,

proteolytic degradation, and recognition/neutralization by antibodies are also related to the protein conformational stability.[46-48] Protein aggregation[49], proteolysis[50-52], and immune responses[53,54] are accelerated when the protein folding energy landscape allows a significant population of unfolded, misfolded, or partially folded conformations. Therefore, it seems reasonable to expect PEGylated proteins that increase conformational stability upon PEGylation to have additional pharmacokinetic advantages over PEGylated proteins that did not have an increase in conformational stability.

However, the impact of PEGylation on protein conformational stability is incompletely understood. Protein PEGylation can increase,[55-80] decrease,[81,82] or have no effect on protein conformational stability,[63,83-87] and conflicting reports have indicated different results for the same protein.[60,88-92] The molecular basis for these differences is unclear, though they likely include (1) using different methods for assessing conformational stability, (2) using PEGylation strategies that generate heterogeneous mixture of protein-PEG conjugates, (3) along with differences in the location of the PEGylation site, (4) PEG length, and (5) the chemistry of the PEG-protein linkage.

Understanding how PEGylation affects the conformational stability of proteins at the molecular level is important for the rational improvement of protein therapeutics. If we could understand the mechanism of how PEG stabilizes and destabilizes proteins then scientists could develop engineering guidelines for selecting the best PEGylation sites. These newly developed guidelines would help reduce wasted time, money and effort for finding the optimal PEGylation sites in therapeutic proteins.

1.2.1 Mechanism of PEG-based Protein Stabilization

Recent studies have begun to provide some mechanistic insight to PEG-based protein stabilization. Liu et al.[93] used molecular dynamic simulations to investigate the energetic consequences of attaching PEGs of various sizes (10, 50, 100, and 200 PEG units) to LysB29 of insulin. Attachment of PEG to insulin was shown to reduce the solvent accessible surface area (SASA) for both hydrophobic and hydrophilic residues. A Lennard-Jones energy analysis of PEGylated insulin suggests that PEG-protein interactions are predominantly hydrophobic in nature. The radial distribution of water in these MD simulations indicates that water molecules near the protein surface are being released to bulk solvent upon PEGylation, a result consistent with hydrophobic PEG-protein interactions. Liu et al. find that decreasing the SASA of insulin via PEGylation is stabilizing because the backbone hydrogen bonds (which are critical for secondary structure) are protected from water molecules that would disrupt the hydrogen bonding network.[94]

Meng et al.[95] used stopped-flow fluorescence and equilibrium denaturation experiments to study the impact of PEGylation on the stability SH3 domain. Selective PEGylation at the N-terminus via reductive alkylation with a 5 kDa and 10 kDa polymer resulted in no increase to thermodynamic stability whereas non-specific modification of the N-terminal α-amine and two lysine ε amines with a 10 kDa PEG did increase the thermodynamic stability by ~0.93 kcal/mol. The data suggests that stabilization is dependent on the number of PEG polymers attached instead of the size of the polymer. They extracted the m-value from the chemical denaturation curves of the PEGylated and non-PEGylated proteins. This parameter is correlated to the amount the protein surface exposed to solvent upon folding.[96] Meng et al. concluded that PEG stabilizes

the SH3 by decreasing the SASA in the folded state more prominently than in the unfolded state. Meng et al. also investigated how PEGylation affects the folding kinetics of SH3 domain. The data show that PEG-based stabilization is in part a result of decreasing the unfolding rate significantly more than the folding rate.

Xu and coworkers[56] functionalized the trimeric α-helical coiled coil **1CW** with a 2 kDa PEG maleimide at a solvent exposed Cys located midway between the N- and C-termini of **1CW** (see figure 2). CD experiments demonstrated that PEGylation enhances the helicity of **1CW** and does not disrupt its ability to form a coiled coil. Small-angle X-ray scattering studies revealed that the PEG within PEGylated **1CW** is more compact than would be expected for a corresponding free PEG.[55,97] Molecular dynamics simulations provided evidence for stabilizing interactions of the PEG oxygen atoms with solvent exposed Lys residues on **1CW**, along with interactions between PEG and hydrophobic residues on the coiled-coil surface, with accompanying decreases in SASA.[57,98] The simulations suggest that PEG occupies the lys ε-ammonium group such that competing interactions between lysine and backbone carbonyl oxygens are weakened, thereby strengthening the backbone hydrogen bonds and increasing the helix stability.

However, careful analysis of variable temperature CD data for **1CW** and its PEGylated counterpart by Pandey et al.[99] found that PEGylation actually destabilizes the **1CW** coiled coil by a small but measurable amount. Alanine mutagenesis experiments confirmed the presence of favorable interactions between the Cys-linked PEG and Lys residues at positions 15 and 21 of **1CW**, though such interactions are not favorable enough to overcome the intrinsically destabilizing impact of PEGylation on **1CW**.[99] However, installing additional Lys residues closer to the PEGylation site (i.e., at the *i*+3 and *i*+4 positions relative to the PEGylated Cys)

results in a **1CW** variant that is stabilized by PEGylation ($\Delta\Delta G_f$ = -0.28 kcal/mol) owing to a favorable three-way interaction between the two Lys residues and the PEGylated Cys. Interestingly, this interaction can occur even when Cys is modified with a maleimide derivative that lacks PEG, indicating that the three-way interaction involves the maleimide linker and not PEG.[99]

Several researchers have hypothesized that PEGylation stabilizes proteins by reducing their internal structural dynamics.[100-103] Rodriguez-Martinez et al. probed this effect by using FTIR spectroscopy to calculate the rate of hydrogen/deuterium exchange within PEGylated α-chymotrypsin (α -CT). They observed that increasing the degree of PEGylation corresponded to slower kinetics of hydrogen/deuterium exchange. They explain that slower hydrogen/deuterium exchange demonstrates that the protein core is becoming more rigid and the structural dynamics are being reduced.

Previous simulations suggest that PEG can engage in both hydrophobic and hydrogen-bonding interactions with groups on the protein surface, though the limited number of atomic-resolution structures of PEGylated proteins make it difficult to know whether these interactions have significant structural or thermodynamic consequences. For example, a low resolution crystal structure reported by Svergun et al.[104] suggests that PEG partially covers the surface of PEGylated hemolglobin, and is presumably involved in interactions with surface residues. In contrast, the NMR and crystallography studies of Cattni et al[105] on PEGylated plastocyanin suggest that the PEG and the protein behave as independent domains without structural evidence for strong interactions.

Diverse studies agree that PEG-based changes to protein surface solvation play a key role in whether or not PEGylation results in substantial increases to protein conformational stability.

Because such increases are associated with enhanced resistance to proteolysis, structure-based mastery of these effects remains an important goal. Early work along these lines is promising and provides hope that such a goal is reasonably attainable.

1.3 Previous Work in the Price Lab

Our lab seeks to understand why PEGylation may or may not increase the thermodynamic stability of PEGylated proteins at the molecular level. We seek to develop rational structure-based guideline for the optimal design of PEGylated therapeutic protein drugs. These guidelines would assist protein chemists in determining the location, linker chemistry, and oligomer length that will provide optimal PEG-based stabilization without diminishing biological activity. We recently showed that attaching a short PEG (comprising four ethylene oxide units) to a single Asn side chain at position 19 in the WW domain of the human protein Pin 1 increases WW conformational stability by -0.70 ± 0.04 kcal mol^{-1} due to accelerated folding and slowed unfolding.[106] Shorter PEG chains impart less stability to WW than the four-unit PEG, whereas longer PEG chains provide similar stability. In this dissertation, we continue to use the WW domain as a model system for identifying structural features that are common to stabilizing PEGylation sites.

This dissertation details the progress we have made in identifying the mechanism by which PEG stabilizes the WW domain. Chapter 2 explains how the location of the PEGylation site strongly influences the ability of PEG to stabilize the WW, and suggests that this stabilizing effect comes from PEG-based partial desolvation of the protein surface. Chapter 3 shows that PEG-based stabilization of the WW domain depends strongly on the identity of the PEG-protein

linker. Chapter 4 investigates the interplay between structure-based guidelines for PEG-base stabilization developed in chapter 2 and the different chemistries explored in chapter 3.

1.4 References

(1) Caravella, J.; Lugovskoy, A. *Curr. Opin. Chem. Biol.* **2010**, *14*, 520.

(2) Leader, B.; Baca, Q. J.; Golan, D. E. *Nat. Rev. Drug. Discov.* **2008**, *7*, 21.

(3) Sood, A.; Panchagnula, R. *Chem Rev.* **2001**, *101*, 3275.

(4) Abuchowski, A.; Mccoy, J. R.; Palczuk, N. C.; Vanes, T.; Davis, F. F. *J. Biol. Chem.* **1977**, *252*, 3582.

(5) Abuchowski, A.; van Es, T.; Palczuk, N. C.; Davis, F. F. *J. Biol. Chem.* **1977**, *252*, 3578.

(6) Harris, J. M.; Chess, R. B. *Nat. Rev. Drug Discov.* **2003**, *2*, 214.

(7) Jevsevar, S.; Kunstelj, M.; Porekar, V. G. *Biotechnol. J.* **2010**, *5*, 113.

(8) Veronese, F. M.; Mero, A. *Biodrugs.* **2008**, *22*, 315.

(9) Werle, M.; Bernkop-Schnürch, A. *Amino Acids.* **2006**, *30*, 351.

(10) Fishburn, C. S. *J. Pharm. Sci.* **2008**, *97*, 4167.

(11) *PEGylated Protein Drugs: Basic Sience and Clinical Applications*; Veronese, F. M., Ed.; Birkhauser Verlag: Basel, 2009.

(12) Li, W.; Zhan, P.; Clercq, E.; Lou, H.; Liu, X. *Prog. Polym. Sci.* **2013**, *38*, 421.

(13) Yamaoka, T.; Tabata, Y.; Ikada, Y. *J Pharm Sci.* **1994**, *83*, 601.

(14) Yamaoka, T.; Tabata, Y.; Ikada, Y. *J. Pharm.Pharmacol.* **1995**, *47*, 479.

(15) Hoofnagle , J. H.; Seeff , L. B. *New Eng. J. Med.* **2006**, *355*, 2444.

(16) Bukowski, R. M.; Tendler, C.; Cutler, D.; Rose, E.; Laughlin, M. M.; Statkevich, P. *Cancer.* **2002**, *95*, 389.

(17) Zeuzem, S.; Welsch, C.; Herrmann, E. In *Pharmacokinetics of peginterferons.* **2002**; Vol. 23, p 23.

(18) Tsutsumi, Y.; Kihira, T.; Tsunoda, S.; Kanamori, T.; Nakagawa, S.; Mayumi, T. *Br. J. Cancer* **1995**, *71*, 963.

(19) Nakaoka, R.; Tabata, Y.; Yamaoka, T.; Ikada, Y. *J. Control. Rel.* **1997**, *46*, 253.

(20) Marshall, D.; Pedley, R. B.; Boden, J. A.; Boden, R.; Melton, R. G.; Begent, R. H. J. *Br J Cancer.* **1996**, *73*, 565.

(21) Kuan, C. T.; Wang, Q. C.; Pastan, I. *J Biol Chem.* **1994**, *269*, 7610.

(22) Kinstler, O.; Molineux, G.; Treuheit, M.; Ladd, D.; Gegg, C. *Adv. Drug Deliv. Rev.* **2002**, *54*, 477.

(23) Chamow, S. M.; Kogan, T. P.; Venuti, M.; Gadek, T.; Harris, R. J.; Peers, D. H.; Mordenti, J.; Shak, S.; Ashkenazi, A. *Bioconjugate Chem.* **1994**, *5*, 133.

(24) Veronese, F. M. *Biomaterials.* **2001**, *22*, 405.

(25) Kinstler, O.; Molineux, G.; Treuheit, M.; Ladd, D.; Gegg, C. *Adv. Drug Deliv. Rev.* **2002**, *54*, 477.

(26) Soderquist, R. G.; Milligan, E. D.; Sloane, E. M.; Harrison, J. A.; Douvas, K. K.; Potter, J. M.; Hughes, T. S.; Chavez, R. A.; Johnson, K.; Watkins, L. R. *J. Biomed. Mat. Res. Part A.* **2009**, *91*, 719.

(27) Jevsevar, S.; Kunstelj, M.; Porekar, V. G. *Biotechnol J.* **2010**, *5*, 113.

(28) Kunstelj, M.; Fidler, K.; Škrajnar, Š.; Kenig, M.; Smilović, V.; Kusterle, M.; Caserman, S.; Zore, I.; Porekar, V. G.; Jevševar, S. *Bioconjugate Chem.* **2013**, *24*, 889.

(29) Chalker, J. M.; Lercher, L.; Rose, N. R.; Schofield, C. J.; Davis, B. G. *Angew Chem Int Ed Engl* **2012**, *51*, 1835.

(30) Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. J.; Davis, B. G. *J Am Chem Soc.* **2008**, *130*, 9642.

(31) Bundy, B. C.; Swartz, J. R. *Bioconjugate Chem.* **2010**, *21*, 255.

(32) Ngo, J. T.; Tirrell, D. A. *Acc. Chem. Res.* **2011**, *44*, 677.

(33) van Hest, J. C.; Tirrell, D. A. *FEBS lett.* **1998**, *428*, 68.

(34) van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. *J Am Chem Soc.* **2000**, *122*, 1282.

(35) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R. *Pro. Natl. Acad. Sci. USA.* **2002**, *99*, 19.

(36) Cho, H.; Daniel, T.; Buechler, Y.; Litzinger, D. C.; Maio, Z.; Putnam, A.-M.; Kraynov, V. S.; Sim, B.-C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; Becky, G. C.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E. *Pro. Natl. Acad. Sci. USA.* **2011**, *108*, 9060.

(37) Chen, H.; Lu, Y.; Fang, Z.; Liu, J.; Tian, H.; Gao, X.; Yao, W. *Biochem. Eng. J.* **2011**, *58–59*, 25.

(38) Tada, S.; Andou, T.; Suzuki, T.; Dohmae, N.; Kobatake, E.; Ito, Y. *PLOS One.* **2012**, *7*.

(39) Goerke, A. R.; Swartz, J. R. *Biotechnol Bioeng.* **2009**, *102*, 400.

(40) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J Am Chem Soc.* **2003**, *125*, 3192.

(41) Dumas, A.; Spicer, C. D.; Gao, Z.; Takehana, T.; Lin, Y. A.; Yasukohchi, T.; Davis, B. G. *Angew. Chem. In. Ed.* **2013**, *52*, 3916.

(42) Shozen, N.; Iijima, I.; Hohsaka, T. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4909.

(43) Zang, Q.; Tada, S.; Uzawa, T.; Kiga, D.; Yamamura, M.; Ito, Y. *Chem. Commun.* **2015**, *51*, 14385.

(44) Abuchowski, A.; McCoy, J. R.; Palczuk, N. C.; Es, T.; Davis, F. F. *J Biol Chem.* **1977**, *252*, 3582.

(45) Harris, J. M.; Chess, R. B. *Nat Rev Drug Discov.* **2003**, *2*, 214.

(46) Krishnamurthy, R.; Manning, M. C. *Curr. Pharm. Biotech.* **2002**, *3*, 361.

(47) Hermeling, S.; Crommelin, D.; Schellekens, H.; Jiskoot, W. *Pharm. Res.* **2004**, *21*, 897.

(48) Frokjaer, S.; Otzen, D. E. *Nat Rev Drug Discov.* **2005**, *4*, 298.

(49) Dobson, C. M. *Nature* **2003**, *426*, 884.

(50) Daniel, R. M.; Cowan, D. A.; Morgan, H. W.; Curran, M. P. *Biochem. J.* **1982**, *207*, 641.

(51) Parsell, D. A.; Sauer, R. T. *J. Biol. Chem.* **1989**, *264*, 7590.

(52) Imoto, T.; Yamada, H.; Ueda, T. *J. Mol. Biol.* **1986**, *190*, 647.

(53) Thai, R.; Moine, G.; Desmadril, M.; Servent, D.; Tarride, J.-L.; Menez, A.; Leonetti, M. *J. Biol. Chem.* **2004**, *279*, 50257.

(54) Ohkuri, T.; Nagatomo, S.; Oda, K.; So, T.; Imoto, T.; Ueda, T. *J. Immunol.* **2010**, *185*, 4199.

(55) Shu, J. Y.; Lund, R.; Xu, T. *Biomacromolecules* **2012**, *13*, 1945.

(56) Shu, J. Y.; Tan, C.; DeGrado, W. F.; Xu, T. *Biomacromolecules* **2008**, *9*, 2111.

(57) Jain, A.; Ashbaugh, H. S. *Biomacromolecules* **2011**, *12*, 2729.

(58) Baillargeon, M.; Sonnet, P. *J. Am. Oil Chem. Soc.* **1988**, *65*, 1812.

(59) Basri, M.; Ampon, K.; Yunus, W. M. Z. W.; Razak, C. N. A.; Salleh, A. B. *J. Chem. Technol. Biotechnol.* **1995**, *64*, 10.

(60) Longo, M. A.; Combes, D. *J. Chem. Technol. Biotechnol.* **1999**, *74*, 25.

(61) Hernaiz, M. J.; Sanchez-Montero, J. M.; Sinisterra, J. V. *Enzyme Microb. Technol.* **1999**, *24*, 181.

(62) Gaertner, H. F.; Puigserver, A. J. *Enzyme Microb. Technol.* **1992**, *14*, 150.

(63) Monfardini, C.; Schiavon, O.; Caliceti, P.; Morpurgo, M.; Harris, J. M.; Veronese, F. M. *Bioconjugate Chem.* **1995**, *6*, 62.

(64) Zhang, Z.; He, Z.; Guan, G. *Biotechnol. Tech.* **1999**, *13*, 781.

(65) Treetharnmathurot, B.; Ovartlarnporn, C.; Wungsintaweekul, J.; Duncan, R.; Wiwattanapatapee, R. *Int. J. Pharm.* **2008**, *357*, 252.

(66) Chiu, K.; Agoubi, L. L.; Lee, I.; Limpar, M. T.; Lowe, J. W.; Goh, S. L. *Biomacromolecules* **2010**, *11*, 3688.

(67) Topchieva, I. N.; Efremova, N. V.; Khvorov, N. V.; Magretova, N. N. *Bioconjugate Chem.* **1995**, *6*, 380.

(68) Castellanos, I. J.; Al-Azzam, W.; Griebenow, K. *J. Pharm. Sci.* **2005**, *94*, 327.

(69) Rodriguez-Martinez, J. A.; Solá, R. J.; Castillo, B.; Cintron-Colon, H. R.; Rivera-Rivera, I.; Barletta, G.; Griebenow, K. *Biotechnol. Bioeng.* **2008**, *101*, 1142.

(70) Rodríguez-Martínez, J.; Rivera-Rivera, I.; Solá, R.; Griebenow, K. *Biotechnol. Lett.* **2009**, *31*, 883.

(71) Garcia, D.; Ortéga, F.; Marty, J.-L. *Biotechnol. Appl. Biochem.* **1998**, *27*, 49.

(72) López-Cruz, J. I.; Viniegra-Gonzalez, G.; Hernandez-Arana, A. *Bioconjugate Chem.* **2006**, *17*, 1093.

(73) Nie, Y.; Zhang, X.; Wang, X.; Chen, J. *Bioconjugate Chem.* **2006**, *17*, 995.

(74) Hinds, K. D.; Kim, S. W. *Adv. Drug Deliv. Rev.* **2002**, *54*, 505.

(75) Yang, C.; Lu, D. N.; Liu, Z. *Biochemistry* **2011**, *50*, 2585.

(76) Meng, W.; Guo, X.; Qin, M.; Pan, H.; Cao, Y.; Wang, W. *Langmuir* **2012**, *28*, 16133.

(77) Shu, J. Y.; Tan, C.; DeGrado, W. F.; Xu, T. *Biomacromolecules* **2008**, *9*, 2111.

(78) Dhalluin, C.; Ross, A.; Leuthold, L. A.; Foser, S.; Gsell, B.; Muller, F.; Senn, H. *Bioconjugate Chem.* **2005**, *16*, 504.

(79) 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. *Bioconjugate Chem.* **2006**, *17*, 618.

(80) Ramon, J.; Saez, V.; Baez, R.; Aldana, R.; Hardy, E. *Pharm. Res.* **2005**, *22*, 1374.

(81) Garcia-Arellano, H.; Valderrama, B.; Saab-Rincon, G.; Vazquez-Duhalt, R. *Bioconjugate Chem.* **2002**, *13*, 1336.

(82) Plesner, B.; Fee, C. J.; Westh, P.; Nielsen, A. D. *Eur. J. Pharm. Biopharm.* **2011**, *79*, 399.

(83) Yang, Z.; Williams, D.; Russell, A. J. *Biotechnol. Bioeng.* **1995**, *45*, 10.

(84) Yang, Z.; Domach, M.; Auger, R.; Yang, F. X.; Russell, A. J. *Enzyme Microb. Technol.* **1996**, *18*, 82.

(85) Callahan, W. J.; Narhi, L. O.; Kosky, A. A.; Treuheit, M. J. *Pharm. Res.* **2001**, *18*, 261.

(86) Plesner, B.; Westh, P.; Nielsen, A. D. *Int. J. Pharm.* **2011**, *406*, 62.

(87) Rodríguez-Martínez, J. A.; Rivera-Rivera, I.; Griebenow, K. *J. Pharm. Pharmacol.* **2011**, *63*, 800.

(88) Popp, M. W.; Dougan, S. K.; Chuang, T. Y.; Spooner, E.; Ploegh, H. L. *Proc. Natl. Acad. Sci. USA.* **2011**, *108*, 3169.

(89) Natalello, A.; Ami, D.; Collini, M.; D'Alfonso, L.; Chirico, G.; Tonon, G.; Scaramuzza, S.; Schrepfer, R.; Doglia, S. M. *PLOS One* **2012**, *7*, e42511.

(90) Nodake, Y.; Yamasaki, N. *Biosci. Biotech. Bioch.* **2000**, *64*, 767.

(91) So, T.; Ueda, T.; Abe, Y.; Nakamata, T.; Imoto, T. *J. Biochem.* **1996**, *119*, 1086.

(92) Gokarn, Y. R. Ph.D. Dissertation, University of New Hampshire, 2003.

(93) Yang, C.; Lu, D.; Liu, Z. *Biochemistry* **2011**, *50*, 2585.

(94) Hamed, E.; Xu, T.; Keten, S. *Biomacromolecules* **2013**, *14*, 4053.

(95) Wei Meng, Xinlu Guo, Meng Qin, Hai Pan, Yi Cao, and Wei Wang *Langmuir* **2012** *28* (46), 16133-16140

(96) Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein Sci.* **1995**, *4*, 2138.

(97) Lund, R.; Shu, J.; Xu, T. *Macromolecules* **2013**, *46*, 1625.

(98) Hamed, E.; Xu, T.; Keten, S. *Biomacromolecules* **2013**, *14*, 4053.

(99) Pandey, B. K.; Smith, M. S.; Price, J. L. *Biomacromolecules* **2014**, *15*, 4643.

(100) Yang, Z.; Domach, M.; Auger, R.; Yang, F. X.; Russell, A. J. *Enzyme Microb Tech.* **1996**, *18*, 82.

(101) Soares, A. L.; Guimarães, G. M.; Polakiewicz, B.; Pitombo, R. N. d. M.; Abrahão-Neto, J. *Int. J. Pharm.* **2002**, *237*, 163.

(102) López-Cruz, J. I.; Viniegra-González, G.; Hernández-Arana, A. *Bioconjugate Chem.* **2006**, *17*, 1093.

(103) García-Arellano, H.; Valderrama, B.; Saab-Rincón, G.; Vazquez-Duhalt, R. *Bioconjugate Chem.* **2002**, *13*, 1336.

(104) Svergun, D. I.; Ekström, F.; Vandegriff, K. D.; Malavalli, A.; Baker, D. A.; Nilsson, C.; Winslow, R. M. *Biophys. J.* **2008**, *94*, 173.

(105) Cattani, G.; Vogeley, L.; Crowley, P. B. *Nat Chem.* **2015**, *7*, 823.

(106) 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. *Bioconjugate Chem.* **2013**, *24*, 796.

2 CRITERIA FOR SELECTING PEGYLATION SITES ON PROTEINS FOR HIGHER THERMODYAMIC AND PROTEOLYTIC STABILITY

2.1 Introduction

Here we use a small protein, the WW domain of the human protein Pin 1 (hereafter called WW) as a model system for understanding how PEGylation generally impacts the conformational stability of β-sheet proteins (Figure 2-1). The WW domain is an extensively characterized[1-23] β-sheet protein that contains three antiparallel β-strands connected by two reverse turns.[23] The folding free energy landscape of Pin WW can be approximated as a simple two-state reaction-coordinate diagram in which the unfolded ensemble proceeds through a high-energy transition state to the folded conformation without passing through discrete intermediates.[3] The small size of WW facilitates the direct chemical synthesis of homogeneous site-specifically PEGylated variants.[24,25] WW is much smaller than many of the PEGylated proteins of pharmaceutical interest. However, recent efforts to increase WW conformational stability via glycosylation have been successfully applied in two larger proteins,[21] suggesting that insights gained from WW PEGylation will be applicable to larger therapeutically relevant proteins.

Here we identify locations within WW where PEGylation increases conformational stability and use a combination of experimental and computational approaches to probe the origins of PEG-based stabilization. We use the resulting insights, along with structural information for WW to identify features that are common to stabilizing PEGylation sites. We use these structural features to develop criteria for predicting stabilizing PEGylation sites within WW. Finally, we show that PEG-based increases to conformational stability correlate with enhanced resistance to proteolysis. These results highlight the possibility of using modern site-specific PEGylation techniques to install PEG oligomers at locations that lead to optimal increases in conformational and proteolytic stability.

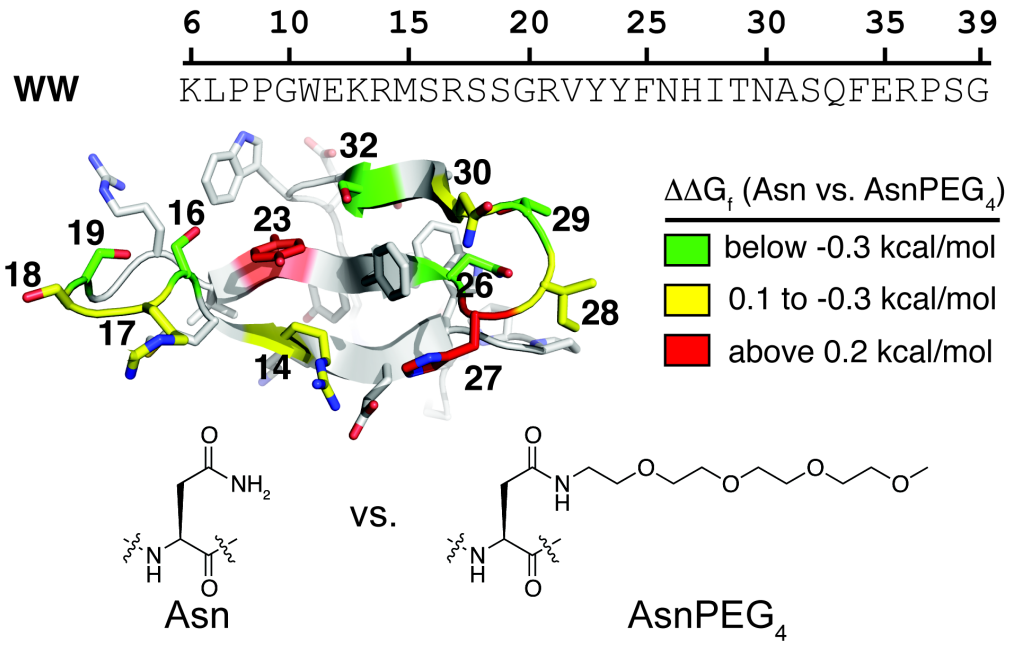


Figure 2-1. Sequence of the protein WW and ribbon diagram of WW (PDB: 1PIN), with side chains shown as sticks. Positions where we incorporated Asn vs. AsnPEG$_4$ are highlighted with color, according to the impact of PEGylation on conformational stability. Stabilizing positions are highlighted in green; neutral positions are highlighted in yellow; and destabilizing positions are highlighted in red.

2.2 Results and Discussion

2.2.1 Impact of PEGylation on WW Conformational Stability

As described previously, PEGylation of an Asn residue at position 19 with a reverse turn of WW increases WW conformational stability by accelerating folding and slowing unfolding.[25] We wondered whether this effect was unique to position 19 or whether PEGylation might similarly stabilize other positions. To address this question, we generated proteins **14**, **16**, **17**, **18**, **23**, **27**, **28**, **29** and **32**, in which wild-type residues at positions 14, 16, 17, 18, 23, 27, 28, 29 and 32, respectively, have been changed to Asn (Asn already occupies positions 26 and 30 in the unmodified protein **WW**, see Figure 2-1). We also prepared PEGylated proteins **14p**, **16p**, **17p**, **18p**, **23p**, **26p**, **27p**, **28p**, **29p**, **30p**, and **32p**, in which positions 14, 17, 18, 19, 23, 26, 30, and 32, respectively, are occupied by AsnPEG$_4$, a PEGylated Asn derivative in which a four-unit PEG oligomer has been attached to the Asn side-chain amide nitrogen (Figure 2-1). These PEGylation sites sample the various secondary structural environments present in WW, including reverse turns (positions 16, 17, 18, 26, 27, 28, 29, 30) and β-strands (positions 14, 23, 32).

Circular dichroism (CD) spectra of these variants at 25 °C (Figure 2-2) are generally very similar in shape and magnitude to that of wild-type unmodified protein **WW**, suggesting that changing wild-type residues to Asn generally does not introduce dramatic alterations to the folded conformation of the resulting Asn mutants relative to **WW**. The exceptions to this trend are easily seen in the CD spectra of proteins **14**, **23**, and their PEGylated counterparts **14p** and **23p**, which are similar in shape to that of **WW**, though substantially smaller in magnitude.

Variable temperature CD data for **14**, **14p**, **23**, and **23p** (see below) provide an explanation for this observation: **14**, **14p**, **23**, and **23p** appear to be two-state folders like **WW**, but are much less

stable. Whereas **WW** is fully folded at 25°C, **14**, **14p**, **23**, and **23p** each exist as equilibrium mixtures of fully folded and fully unfolded conformations at 25 °C. The CD spectrum of a two-state folder under equilibrium conditions is the weighted average of its fully folded and fully unfolded conformations. Therefore, the CD spectra of **14**, **14p**, **23**, and **23p** at 25 °C should be similar in shape but smaller in magnitude than the CD spectrum of **WW**. This is in fact what we observe. In contrast, variable temperature CD data indicate that **WW**, **14**, **14p**, **23**, and **23p** should be each fully folded at 2 °C. Consistent with this expectation, the CD spectra of these variants at 2°C (Figure 2-2) are much closer in magnitude to that of **WW**, suggesting that their fully folded conformations are likewise similar to that of **WW**. However, without high-resolution structural data, we cannot eliminate the possibility of substantial structural rearrangements in **14**, **14p**, **23**, and **23p**. Therefore, in the discussion below, we avoid using data from these compounds in our efforts to develop structure-based guidelines for identifying optimal PEGylation sites.

We used variable temperature CD experiments to assess the conformational stability of PEGylated proteins **14p**, **16p**, **17p**, **18p**, **23p**, **26p**, **27p**, **28p**, **29p**, **30p**, and **32p** relative to their non-PEGylated counterparts **14**, **16**, **17**, **18**, **23**, **WW, 27, 28, 29** and **32** in 20 mM aqueous sodium phosphate (pH 7.0). We also performed these same measurements on 100 μM solutions of **19p** and **19**, which were characterized previously at 10 μM and at 50 μM.[24,25] The results of this analysis appear in Figure 2-2 and Table 2-1. CD data indicate that each of these variants is a two-state folder like the wild-type **WW** protein. PEGylation substantially increases WW conformational stability at positions 16, 19, 26, 29, and 32 and moderately increases WW conformational stability at position 17. In contrast, PEGylation has essentially no impact on WW conformational stability at positions 14, 18, 28 and 30; and is substantially destabilizing at

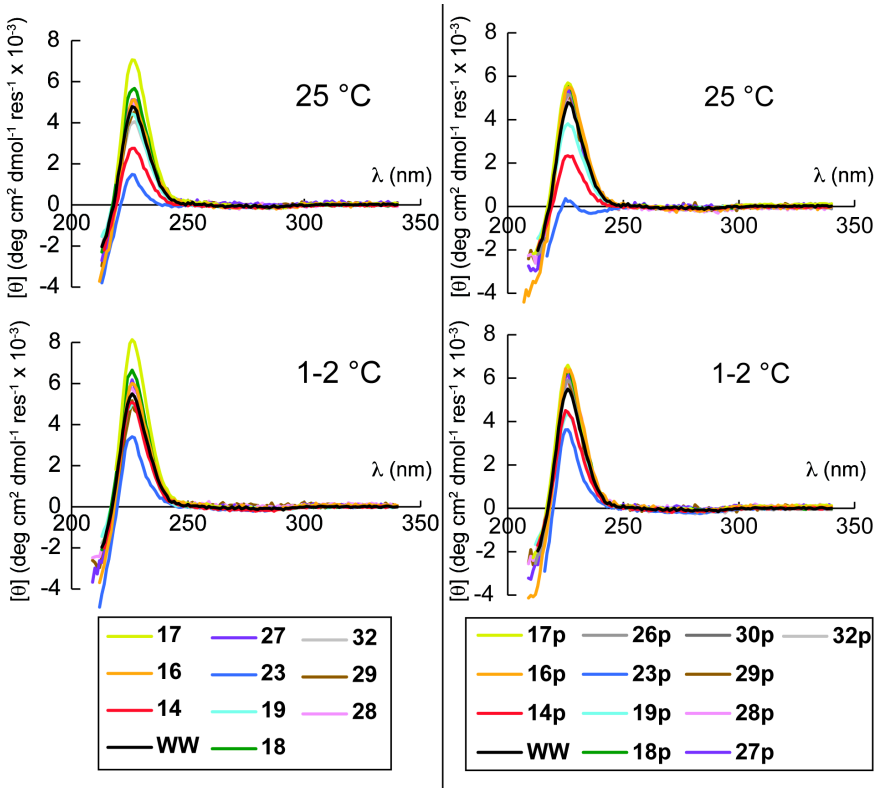


Figure 2-2. CD spectra of wild-type protein WW; non-PEGylated variants **14**, **16**, **17**, **18**, **19**, **23**, **27**, **28**, **29**, **32**; and PEGylated variants **14p**, **16p**, **17p**, **18p**, **19p**, **23p**, **26p**, **27p**, **28p**, 29**p**, **30p**, and **32p** in 20 mM sodium phosphate buffer, pH 7 at 25 °C and at low temperature (i.e. at 2°C, except for **16**, **27**, **27p**, and **29p**, which were analyzed at 1°C---variable temperature CD data for these compounds suggests that each is fully folded at 1°C and at 2°C, so these spectra are directly comparable). Spectra were obtained at 100 μM, except for **16**, **16p**, **27**, **27p**, **28**, **28p**, **29**, and **29p**, which were obtained at 50 μM.

positions at positions 23 and 27. No specific secondary structural motif appears to be generally amenable to PEG-based stabilization: stabilizing and destabilizing positions occur within both β-strands and reverse turns.

Van't Hoff analysis allows us to parse the impact of PEGylation on WW conformational stability ($\Delta\Delta G_f$) into enthalpic ($\Delta\Delta H_f$) and entropic terms ($-T\Delta\Delta S_f$). At several positions, large uncertainties in $\Delta\Delta H_f$ and in $-T\Delta\Delta S_f$ preclude further analysis. However, an interesting trend emerges from the data for stabilizing positions 16, 19, 26, and 29 (see Table 2-1). At each of these positions, $-T\Delta\Delta S_f$ is negative (i.e. favorable) whereas $\Delta\Delta H_f$ is positive (i.e. unfavorable).

This observation suggests an entropic origin for the PEG-based increase to WW conformational stability at these positions.

2.2.2 PEGylation with 2000 Da Oligomer

The PEG oligomers used in therapeutic proteins are typically much longer than the four-unit oligomer we used in the experiments described above. We previously showed that attaching a 2000 Da PEG oligomer to an Asn at position 19 continues to increase WW conformational stability, even though the 2000 Da oligomer is much longer (~45 ethylene oxide units) than the four-unit oligomer.[25] We wondered whether the energetic impact of the 45-unit PEG at the positions described above would mirror the results described in table 2-1 for the four-unit PEG.

To test this hypothesis, we prepared WW variants **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45** in which we incorporated an Asn-linked 45-unit PEG ($AsnPEG_{45}$) at positions 16, 18, 19, 26, 27, 28, and 29, respectively. We assessed the conformational stability of these variants relative to their non-PEGylated counter- parts using variable temperature CD experiments. The results of this analysis are shown in Table 2-2.

Like the four-unit PEG, the 45-unit PEG increases conformational stability at positions 16, 19, 26, and 29, and decreases stability at position 27. Whereas the four-unit PEG had no effect at positions 18 and 28, the 45-unit PEG is destabilizing at these positions. Van't Hoff analysis of these results indicates that $-T\Delta\Delta S_f$ is negative (i.e. favorable) and $\Delta\Delta H_f$ is positive (i.e. unfavorable) at stabilizing positions 16, 19, 26, and 29, suggesting that the 45-unit oligomer likewise increases WW stability via an entropic effect.

Table 2-1 . Impact of the Four-unit PEG Oligomer on WW Conformational Stability at Various Sites.

Protein	Sequence	T_m (°C)	ΔT_m (°C)	$\Delta\Delta G_f$ (kcal/mol)	$\Delta\Delta H_f$ (kcal/mol)	$-T\Delta\Delta S_f$ (kcal/mol)
14	KLPPGWEK**N**MSRSSGRVYYFNHITNASQFERPSG	34.0 ± 0.8	-0.6 ± 5.3	0.0 ± 0.4	-0.1 ± 3.4	0.1 ± 3.3
14p	KLPPGWEK**N**MSRSSGRVYYFNHITNASQFERPSG	33.4 ± 5.2				
16	KLPPGWEKRM**N**RSSGRVYYFNHITNASQFERPSG	50.6 ± 0.2	10.1 ± 0.3	-0.90 ± 0.03	3.8 ± 1.4	-4.7 ± 1.3
16p	KLPPGWEKRM**N**RSSGRVYYFNHITNASQFERPSG	60.7 ± 0.3				
17	KLPPGWEKRMS**N**SSGRVYYFNHITNASQFERPSG	53.6 ± 0.4	1.9 ± 0.6	-0.18 ± 0.05	-2.2 ± 0.8	2.0 ± 0.8
17p	KLPPGWEKRMS**N**SSGRVYYFNHITNASQFERPSG	55.5 ± 0.5				
18	KLPPGWEKRMSR**N**SGRVYYFNHITNASQFERPSG	56.9 ± 0.2	0.0 ± 0.7	0.00 ± 0.07	-3.7 ± 0.9	3.7 ± 0.9
18p	KLPPGWEKRMSR**N**SGRVYYFNHITNASQFERPSG	57.0 ± 0.7				
19	KLPPGWEKRMSRS**N**GRVYYFNHITNASQFERPSG	55.6 ± 0.2	7.7 ± 0.4	-0.70 ± 0.04	3.6 ± 1.4	-4.3 ± 1.4
19p	KLPPGWEKRMSRS**N**GRVYYFNHITNASQFERPSG	63.3 ± 0.3				
23	KLPPGWEKRMSRSSGRV**N**YFNHITNASQFERPSG	28.5 ± 0.9	-5.2 ± 1.3	0.40 ± 0.10	4.1 ± 1.4	-3.7 ± 1.3
23p	KLPPGWEKRMSRSSGRV**N**YFNHITNASQFERPSG	23.3 ± 1.0				
WW	KLPPGWEKRMSRSSGRVYYF**N**HITNASQFERPSG	58.0 ± 0.7	6.6 ± 0.7	-0.58 ± 0.06	3.4 ± 0.9	-4.0 ± 0.9
26p	KLPPGWEKRMSRSSGRVYYF**N**HITNASQFERPSG	64.6 ± 0.2				
27	KLPPGWEKRMSRSSGRVYYFN**N**ITNASQFERPSG	55.0 ± 0.1	-4.0 ± 0.4	0.38 ± 0.04	0.5 ± 0.9	-0.1 ± 0.9
27p	KLPPGWEKRMSRSSGRVYYFN**N**ITNASQFERPSG	51.0 ± 0.4				
28	KLPPGWEKRMSRSSGRVYYFNH**N**TNASQFERPSG	53.2 ± 0.5	0.0 ± 0.7	0.00 ± 0.07	0.6 ± 0.8	-0.6 ± 0.8
28p	KLPPGWEKRMSRSSGRVYYFNH**N**TNASQFERPSG	53.2 ± 0.5				
29	KLPPGWEKRMSRSSGRVYYFNHI**N**NASQFERPSG	50.0 ± 0.3	4.1 ± 0.4	-0.36 ± 0.04	3.6 ± 1.4	-4.3 ± 1.4
29p	KLPPGWEKRMSRSSGRVYYFNHI**N**NASQFERPSG	54.1 ± 0.3				
WW	KLPPGWEKRMSRSSGRVYYFNHIT**N**ASQFERPSG	58.0 ± 0.7	0.4 ± 0.7	0.00 ± 0.07	-0.5 ± 1.1	0.5 ± 1.1
30p	KLPPGWEKRMSRSSGRVYYFNHIT**N**ASQFERPSG	58.4 ± 0.2				
32	KLPPGWEKRMSRSSGRVYYFNHITNA**N**QFERPSG	45.1 ± 0.2	5.3 ± 0.3	-0.45 ± 0.02	-0.1 ± 0.6	-0.3 ± 0.6
32p	KLPPGWEKRMSRSSGRVYYFNHITNA**N**QFERPSG	50.3 ± 0.2				
16	KLPPGWEKRM**N**RSSGRVYYF**N**HITNASQFERPSG	50.6 ± 0.2	16.7 ± 0.2	-1.38 ± 0.03	8.2 ± 0.9	-9.6 ± 0.9
16p/26p	KLPPGWEKRM**N**RSSGRVYYF**N**HITNASQFERPSG	67.3 ± 0.1				
19	KLPPGWEKRMSRS**N**GRVYYF**N**HITNASQFERPSG	55.6 ± 0.2	14.2 ± 0.2	-1.26 ± 0.02	6.1 ± 0.7	-7.3 ± 0.7
19p/26p	KLPPGWEKRMSRS**N**GRVYYF**N**HITNASQFERPSG	69.8 ± 0.1				
29	KLPPGWEKRMSRSSGRVYYF**N**HI**N**NASQFERPSG	50.0 ± 0.3	6.7 ± 0.4	-0.56 ± 0.04	1.9 ± 0.6	-2.5 ± 0.6
26p/29p	KLPPGWEKRMSRSSGRVYYF**N**HI**N**NASQFERPSG	56.8 ± 0.3				
16/19	KLPPGWEKRM**N**RS**N**GRVYYFNHITNASQFERPSG	56.9 ± 0.1	8.5 ± 0.2	-0.80 ± 0.02	1.2 ± 0.5	-2.0 ± 0.5
16p/19p	KLPPGWEKRM**N**RS**N**GRVYYFNHITNASQFERPSG	65.4 ± 0.1				

The WW sequence is shown with amino acids abbreviated according to the standard one-letter code. **N** represents AsnPEG$_4$. Observed data are given ± standard error at 100 μM protein concentration in 20 mM sodium phosphate buffer, pH 7 (except for proteins **27**, **27p**, **28**, **28p**, **29**, and **29p**, which were characterized at 50 μM protein concentration). Observed values of $\Delta\Delta G_f$ were derived from variable temperature CD experiments at the melting temperature of the corresponding non-PEGylated protein.

Table 2-2. Impact of PEGylation with the 2000 Da (~45-unit) Oligomer on WW Conformational Stability at Various Sites.

Protein	T_m (°C)	ΔT_m (°C)	$\Delta\Delta G_f$	$\Delta\Delta H_f$	$-T\Delta\Delta S_f$

			(kcal/mol)	(kcal/mol)	(kcal/mol)
16	54.9 ± 0.1	4.7 ± 0.3	-0.39 ± 0.03	5.1 ± 1.2	-5.5 ± 1.2
16p45	59.6 ± 0.3				
18	55.3 ± 0.8	-0.3 ± 0.9	0.02 ± 0.08	3.3 ± 1.5	-3.3 ± 1.5
18p45	55.1 ± 0.4				
19	55.9 ± 0.2	7.4 ± 0.4	-0.67 ± 0.05	1.7 ± 1.6	-2.4 ± 1.5
19p45	63.3 ± 0.4				
WW	58.0 ± 0.7	3.3 ± 0.7	-0.27 ± 0.06	5.1 ± 0.8	-5.3 ± 0.8
26p45	61.3 ± 0.1				
27	55.0 ± 0.1	-7.0 ± 0.3	0.65 ± 0.04	-0.4 ± 1.0	1.1 ± 1.0
27p45	48.0 ± 0.3				
28	53.2 ± 0.5	-4.5 ± 0.6	0.36 ± 0.05	4.0 ± 1.0	-3.7 ± 1.0
28p45	48.7 ± 0.3				
29	48.6 ± 0.4	4.7 ± 0.5	-0.36 ± 0.04	1.6 ± 1.3	-2.0 ± 1.3
29p45	53.3 ± 0.3				

Data are given + standard error at 50 uM protein concentration in 20 mM sodium phosphate buffer, pH 7 (except for proteins **WW** and **26p45**, which were characterized at 100 uM protein concentration) at the melting temperature of the corresponding non-PEGylated protein. Values of T_m, $\Delta\Delta G_f$, $\Delta\Delta H_f$ and $-T\Delta\Delta S_f$ were derived from variable temperature CD experiments.

The close correlation between the position-dependent results for the four-unit and 45-unit oligomers in these experiments suggests that insights gained from the four-unit oligomer should be reasonably predictive for longer oligomers that more closely resemble those currently used in therapeutic proteins.

2.2.3 Mechanistic Origins of PEG-based Stabilization

We next used temperature jump kinetic experiments to assess the contribution of folding and unfolding kinetics to the PEG-based changes in conformational stability described above, with the goal of gaining insights into how PEG can stabilize proteins. At stabilizing positions 19 and 26, and to a lesser extent at position 17, PEGylation accelerates folding and slows unfolding. In contrast, at neutral positions 14, 18, and 30, PEGylation slows both folding and unfolding by similar amounts, resulting in no overall change to folding thermodynamics. These results are shown in Table 2-3. Accelerated folding and slowed unfolding could be consistent with

simultaneous stabilization of the native state and the transition state, with the native state experiencing greater stabilization. Alternatively, these observations are also consistent with simultaneous destabilization of the unfolded ensemble and transition state, with the unfolded ensemble experiencing greater destabilization.

Table 2-3. Predicted and Experimentally Observed Impact of PEGylation with the Four-Unit Oligomer on WW Folding Thermodynamics and Kinetics at Selected Sites.

		Experimental Observations				CG Model Predictions		
Protein	**$\Delta\Delta G_f$ (kcal/mol)**	**k_f (s^{-1})**	**k_f ratio**	**k_u (s^{-1})**	**k_u ratio**	**$\Delta\Delta G_f$ (kcal/mol)**	**k_f ratio**	**k_u ratio**
14	0.0 ± 0.4	1.9 ± 0.1	0.7 ± 0.2	1.9 ± 0.2	0.7 ± 0.5	0.39	0.96	0.92
14p		1.2 ± 0.3		1.3 ± 0.8				
16	-0.90 ± 0.03	---	---	---	---	0.28	---	---
16p		---		---				
17	-0.18 ± 0.05	5.7 ± 0.2	1.08 ± 0.05	5.7 ± 0.3	0.82 ± 0.08	0.05	0.85	0.95
17p		6.2 ± 0.3		4.7 ± 0.4				
18	0.00 ± 0.07	7.1 ± 0.2	0.78 ± 0.05	7.1 ± 0.3	0.78 ± 0.09	0.06	1.07	1.07
18p		5.6 ± 0.3		5.5 ± 0.6				
19	-0.70 ± 0.04	6.2 ± 0.2	1.3 ± 0.1	6.2 ± 0.2	0.44 ± 0.04	0.11	1.04	1.33
19p		8.0 ± 0.5		2.7 ± 0.2				
23	0.40 ± 0.10	---	---	---	---	0.68	---	---
23p		---		---				
WW	-0.58 ± 0.06	9.0 ± 0.7	2.3 ± 0.2	9.0 ± 1.1	0.9 ± 0.1	0.86	0.50	1.90
26p		20.5 ± 0.6		8.4 ± 0.4				
27	0.38 ± 0.04	---	---	---	---	0.24	---	---
27p		---		---				
28	0.00 ± 0.07	---	---	---	---	0.16	---	---
28p		---		---				
29	-0.36 ± 0.04	---	---	---	---	0.03	---	---
29p		---		---				
WW	0.00 ± 0.07	9.0 ± 0.7	0.86 ± 0.07	9.0 ± 1.1	0.9 ± 0.1	0.02	0.78	1.07
30p		7.7 ± 0.2		7.7 ± 0.3				

Observed data are given ± standard error at 100 μM protein concentration in 20 mM sodium phosphate buffer, pH 7 (except for protein **27**, **27p**, **28**, **28p**, **29**, and **29p**, which were characterized at 50 μM protein concentration). Observed values of $\Delta\Delta G_f$were derived from variable temperature CD experiments; k_f and k_u were derived from temperature jump kinetic experiments. Predicted values of $\Delta\Delta G_f$, and k_f and k_u rations are from coarse-grained native topology simulations. Both are given at the melting temperature of the corresponding non-PEGylated protein.

To help discern between these two mechanistic possibilities, we studied PEGylated proteins **14p**, **16p**, **17p**, **18p**, **19p**, **23p**, **26p**, and **30p** and their non-PEGylated counterparts using

a coarse-grained native-topology-based (CG) model in which only the heavy atoms of the protein and the PEG conjugate are included. We have used similar models in the past to study the effect of glycosylation, ubiquitination, and myristoylation on protein folding.[26-29]

The CG model includes native interactions only within the protein, but may provide quantitative predictions regarding the impact of PEGylation on WW folding kinetics and thermodynamics. In this CG approach, the PEG is modeled as an excluded-volume polymer that is exposed to the solvent and cannot form favorable interactions with protein side-chain or backbone groups. Therefore, PEG-based changes to WW conformational stability in this model are assumed to come from changes in the free energy of (1) the unfolded state ensemble, which might not be as compact as the folded state and might therefore be more affected by an excluded volume PEG oligomer; or (2) the folded state due to unfavorable steric interactions between the PEG oligomer and the protein, which may prevent complete folding of the protein. Both effects (i.e. (1) and (2)) are expected to depend on the position of the PEGylation site.

The CG model captures the observed destabilization of **23p** and **27p** relative to **23** and **27**, respectively **(**Table 2-3). For the variants where PEGylation has no substantial observed impact on conformational stability (**14p**, **18p**, **28p**, **30p**), the CG model simulations also predict a minimal effect, with the exception of position 14, where the CG model predicts strong destabilization. The small effect of PEG on stability for these variants is also reflected by the kinetic rates predicted from the CG simulations (Table 2-3). A more substantial disagreement between the CG model and experimental observations is seen at stabilizing positions 16, 19, 26, and 29. The CG model predicts that PEGylation will strongly destabilize **16p** and **26p** relative to **16** and **WW**, respectively, and have a minimal effect on **19p** and **29p** relative to **19** and **29**, respectively. In contrast, we observe strong stabilization at each of these positions and faster

folding rate. The limited predictive ability of the coarse-grained native topology model suggest that PEG does not influence WW stability via a primarily excluded-volume mechanism.

An alternative to this mechanistic hypothesis is that PEG-based increase to conformational stability comes from stabilization of the transition state and native state relative to the unfolded ensemble. This scenario could potentially involve favorable PEG-protein interactions in the transition state and in the native state. In the crystal structure of the parent WW domain, the side chain at position 19 appears to be oriented toward several nearby OH-containing side-chains, including Ser16, Tyr23, and Ser32 (Figure 2-1). We wondered whether interactions between PEG and nearby OH groups contribute to the observed PEG-based stabilization. If so, the orientation of the side-chain at position 19 should also be an important factor.

To test this hypothesis, we prepared proteins **D-19** and **D-19p**, in which D-Asn or D-AsnPEG$_4$ occupy position 19, respectively (D-AsnPEG$_4$ is the enantiomer of AsnPEG$_4$ shown in Figure 2-1). Incorporating D-Asn or D-AsnPEG$_4$ should invert the orientation of the side-chain at this position. Previous work by Kelly and coworkers indicates that WW can tolerate D-amino acids within this reverse turn without substantial disruption of secondary and tertiary structure.[4] The observed similarity of the CD spectra of **D-19** and **D-19p** to those of their counterparts **19** and **19p** (see Figure 2-136) is consistent with this assertion, as are the nearly identical melting temperatures of **19** and **D-19** (55.6 ± 0.3 °C and 55.4 ± 0.3 °C, respectively. Whereas PEGylation of Asn at position 19 increases WW conformational stability by -0.70 ± 0.04 kcal/mol, PEGylation of D-Asn at position 19 has no effect ($\Delta\Delta G_f$ = 0.01 ± 0.04 kcal/mol), suggesting that side-chain orientation is an important feature of stabilizing PEGylation sites.

We recently probed the extent to which the Asn-linked PEG-oligomer at position 19 engages in favorable interactions with nearby Ser16 and Tyr23 side-chains.[30] For convenience, these previously reported data are also shown in Table 2-4. Removing the OH group at position 16 by replacing Ser with Ala reduces the stabilizing impact of PEGylation from -0.70 ± 0.04 kcal mol^{-1} (compare **19p** vs. **19**) to -0.51 ± 0.02 kcal mol^{-1} (compare **19p-16A** vs. **19-16A** in Table 2-4). Similarly, replacing Tyr23 with Phe reduces the stabilizing impact of PEGylation to -0.43 ± 0.03 (compare **19p-23F** vs. **19-23F** in Table 2-4). In contrast, we observed here that replacing Ser32 with Ala has no significant effect (compare **19p-32A** vs. **19-32A** in Table 2-4). In principle, these results could be interpreted in terms of direct favorable interactions between PEG at position 19 and the OH groups at positions 16 and 23, but not at position 32, presumably because of its distance from position 19.

However, direct PEG-OH interactions are absent from previously reported MD simulations of **19p**,[30] suggesting that such interactions are not a significant component of PEG-based stabilization. Instead, the simulations show the PEG oligomer at position 19 extending predominantly into the solvent, with a high degree of conformational entropy. The flexible PEG oligomer also appears to increase the conformational entropy of amino acid residues within **19p** relative to **19** (as measured by root-mean-square deviations of the simulated structures for **19p** and **19** vs. the crystal structure of the parent WW protein), but without substantially disrupting the native-state interactions present in the reverse turns and β-strands of **19p**.[30] These simulations are consistent with our observations that PEG-based stabilization at position 19 is associated with an unfavorable increase in enthalpy, which is offset by a favorable increase in entropy (Table 2-1). Moreover, the simulations imply that the influence of nearby OH groups on PEG-

based stabilization at position 19 must occur via an indirect mechanism rather than via direct PEG-OH contacts.

We wondered whether OH groups near other "stabilizing" PEGylation sites might be similarly (though indirectly) important to the observed PEG-based stabilization. To address this question, we identified one or more OH-containing side chains (Ser, Thr, or Tyr) near positions 16, 26, 29, and 32, and replaced these residues individually with Ala or Phe. The results of this analysis are shown in Table 2-4. Replacing Tyr23 or at Ser32 with Phe or Ala, respectively, decreases the stabilizing impact of PEGylation at position 16 (Table 2-4; compare **16p-23F** vs. **16-23F**; and **16p-32A** vs. **16-32A**).

Similarly, replacing Thr29 with Ala decreases the stabilizing impact of PEGylation at position 26 (Table 2-4; compare **26p-29A** vs. **26-29A**). In contrast, PEG-based stabilization actually increases at positions 29 and 32 upon removal of OH groups at Ser32 and Tyr23, respectively (Table 2-4; compare **29p-32A** vs. **29-32A**; **32p-23F** vs. **32-23F**). Interpretation of these last two results is complicated by the strong destabilizing impact of the Ser32Ala and Tyr23Phe mutations in these variants. In any case, these mutagenesis experiments are difficult to rationalize on the basis of direct favorable PEG-OH interactions and hint at a more indirect influence.

In agreement with this conclusion, fully atomistic simulations of PEGylated proteins **16p**, **19p**, **26p,** and **29p** provide no evidence for strong direct PEG-OH interactions. Figure 2-3 shows the results of these simulations. For each variant, we calculated the average interaction energy between PEG and every residue within WW (Figure 2-3, large graphs), along with the total energy of PEG-protein interface during the simulation (Figure 2-3, insets). Snapshots from the

simulation of each variant are also shown in Figure 2-3, though these do not indicate the lifetime of individual interactions, which in some cases are relatively transient.

Table 2-4. Effect of mutagenesis near selected PEGylation sites within WW.

Protein	Sequence	T_m (°C)	ΔT_m (°C)	$\Delta\Delta G_f$ (kcal/mol)	$\Delta\Delta H_f$ (kcal/mol)	$-T\Delta\Delta S_f$ (kcal/mol)
19	KLPPGWEKRMSRS**N**GRVYYFNHITNASQFERPSG	55.6 ± 0.2	7.7 ± 0.4	-0.70 ± 0.04	3.6 ± 1.4	-4.3 ± 1.4
19p	KLPPGWEKRMSRS**N**GRVYYFNHITNASQFERPSG	63.3 ± 0.3				
D-19	KLPPGWEKRMSRS**n**GRVYYFNHITNASQFERPSG	55.4 ± 0.3	-0.2 ± 0.4	0.01 ± 0.04	0.7 ± 0.5	-0.7 ± 0.5
D-19p	KLPPGWEKRMSRS**n**GRVYYFNHITNASQFERPSG	55.2 ± 0.3				
19-16A	KLPPGWEKRM**A**RS**N**GRVYYFNHITNASQFERPSG	51.0 ± 0.2	5.8 ± 0.3	-0.51 ± 0.02	2.5 ± 0.5	-3.0 ± 0.5
19p-16A	KLPPGWEKRM**A**RS**N**GRVYYFNHITNASQFERPSG	56.8 ± 0.1				
19-23F	KLPPGWEKRMSRS**N**GRV**F**YFNHITNASQFERPSG	51.4 ± 0.4	5.0 ± 0.5	-0.43 ± 0.03	1.8 ± 0.6	-2.2 ± 0.6
19p-23F	KLPPGWEKRMSRS**N**GRV**F**YFNHITNASQFERPSG	56.5 ± 0.1				
19-32A	KLPPGWEKRMSRS**N**GRVYYFNHITNA**A**QFERPSG	54.5 ± 0.3	8.4 ± 0.3	-0.71 ± 0.03	2.2 ± 0.4	-2.9 ± 0.4
19p-32A	KLPPGWEKRMSRS**N**GRVYYFNHITNA**A**QFERPSG	62.8 ± 0.1				
19-16A,23F	KLPPGWEKRM**A**RS**N**GRV**F**YFNHITNASQFERPSG	45.8 ± 1.0	8.0 ± 1.1	-0.72 ± 0.08	-4.2 ± 1.5	3.5 ± 1.5
19p-16A,23F	KLPPGWEKRM**A**RS**N**GAV**F**YFNHITNASQFERPSG	53.7 ± 0.3				
16	KLPPGWEKRM**N**RSSGRVYYFNHITNASQFERPSG	50.6 ± 0.2	10.1 ± 0.3	-0.90 ± 0.04	3.8 ± 1.4	-4.7 ± 1.3
16p	KLPPGWEKRM**N**RSSGRVFYFNHITNASQFERPSG	60.7 ± 0.3				
16-Y23F	KLPPGWEKRM**N**RSSGRV**F**YFNHITNASQFERPSG	50.9 ± 0.6	5.5 ± 0.7	-0.45 ± 0.06	1.7 ± 1.5	-2.1 ± 1.5
16p-Y23F	KLPPGWEKRM**N**RSSGRV**F**YFNHITNASQFERPSG	56.4 ± 0.3				
16-S32A	KLPPGWEKRM**N**RSSGRVYYFNHITNA**A**QFERPSG	55.0 ± 0.3	6.6 ± 0.3	-0.58 ± 0.03	2.3 ± 0.5	-2.9 ± 0.5
16p-S32A	KLPPGWEKRM**N**RSSGRVYYFNHITNA**A**QFERPSG	61.6 ± 0.1				
WW	KLPPGWEKRMSRSSGRVYYF**N**HITNASQFERPSG	58.0 ± 0.7	6.6 ± 0.7	-0.58 ± 0.06	3.4 ± 0.9	-4.0 ± 0.9
26p	KLPPGWEKRMSRSSGRVYYF**N**HITNASQFERPSG	64.6 ± 0.2				
WW-T29A	KLPPGWEKRMSRSSGRVYYF**N**HI**A**NASQFERPSG	40.4 ± 0.7	4.8 ± 0.8	-0.32 ± 0.06	4.0 ± 0.7	-4.4 ± 0.7
26p-T29A	KLPPGWEKRMSRSSGRVYYF**N**HI**A**NASQFERPSG	45.2 ± 0.2				
29	KLPPGWEKRMSRSSGRVYYFNHI**N**NASQFERPSG	50.0 ± 0.3	4.1 ± 0.4	-0.36 ± 0.04	0.3 ± 0.5	-0.6 ± 0.5
29p	KLPPGWEKRMSRSSGRVYYFNHI**N**NASQFERPSG	54.1 ± 0.3				
29-S32A	KLPPGWEKRMSRSSGRVYYFNHI**N**NA**A**QFERPSG	40.4 ± 0.5	11.6 ± 0.8	-0.88 ± 0.06	-4.1 ± 1.7	3.3 ± 1.7
29p-S32A	KLPPGWEKRMSRSSGRVYYFNHI**N**NA**A**QFERPSG	52.0 ± 0.7				
32	KLPPGWEKRMSRSSGRVYYFNHITNA**N**QFERPSG	45.1 ± 0.2	5.3 ± 0.3	-0.45 ± 0.02	-0.1 ± 0.6	-0.3 ± 0.6
32p	KLPPGWEKRMSRSSGRVYYFNHITNA**N**QFERPSG	50.3 ± 0.2				
32-Y23F	KLPPGWEKRMSRSSGRV**F**YFNHITNA**N**QFERPSG	30.0 ± 1.3	10.3 ± 1.5	-0.61 ± 0.11	7.7 ± 2.9	-8.3 ± 2.9
32p-Y32F	KLPPGWEKRMSRSSGRV**F**YFNHITNA**N**QFERPSG	40.3 ± 0.7				

Data are given ± standard error at 100 μM protein concentration in 20 mM sodium phosphate buffer, pH at the melting temperature of the corresponding non-PEGylated protein. **N** represents AsnPEG$_4$.**n**and **n** represent D-Asn and D-AsnPEG$_4$ respectively. Values of T_m, $\Delta\Delta G_f$, $\Delta\Delta H_f$,and $-T\Delta\Delta S_f$ were derived from variable temperature CD experiments.

Simulation results for neutral positions **14**, **17**, **18**, and **28** (Figure 2-3B) generally show weak PEG-protein interactions while simulations for stabilizing positions **16p**, and **26p** show a

strong PEG-protein interaction (Figure 2-3A). For **16p**, the simulations suggest the presence of strong interactions between PEG and Arg17 and Arg21. Similarly, the PEG in **26p** appears to interact strongly with Trp11 and Gln3. However, strong tight PEG-protein interfaces even occur in simulations of destabilized variants **23p** and **27p** (Figure 2-3C), suggesting that favorable PEG-protein interactions are not sufficient for increasing the overall conformational stability of WW. Moreover, simulations of **19p** and **29p** show that PEG-based stabilization can occur even in the absence of strong PEG-protein interactions (Figure 2-3A), indicating that direct PEG-OH interactions are not responsible for the observed impact of OH groups on PEG-based stabilization. Other factors, including the conformational entropy of PEG as well the solvation of WW surface residues must also make important contributions.

However, the ability of a highly flexible PEG to stabilize some WW variants even in the absence of strong PEG-protein interactions, together with the observed impact of nearby OH groups described above suggest the possibility that changes in WW solvation may also play a role in PEG-based stabilization. One possibility is that differential solvation of these nearby OH groups in the presence or absence of PEG affects protein conformational stability.

We investigated this possibility in the simulations of **16p** and **26p** by analyzing the organization of water near residues that interact strongly with PEG. Figure 2-4 show plots of the radial distribution function of water about Tyr23 or Phe34 in **16p** vs. **16** and about Trp11 or Thr29 in **26p** vs. **26**. These radial distribution function plots show the density of water molecules as a function of the distance from the indicated residues in **16p** and **26p** vs. **16** and **WW**, respectively. For proteins **16p** vs. **16**, PEGylation results in lower water density (and therefore

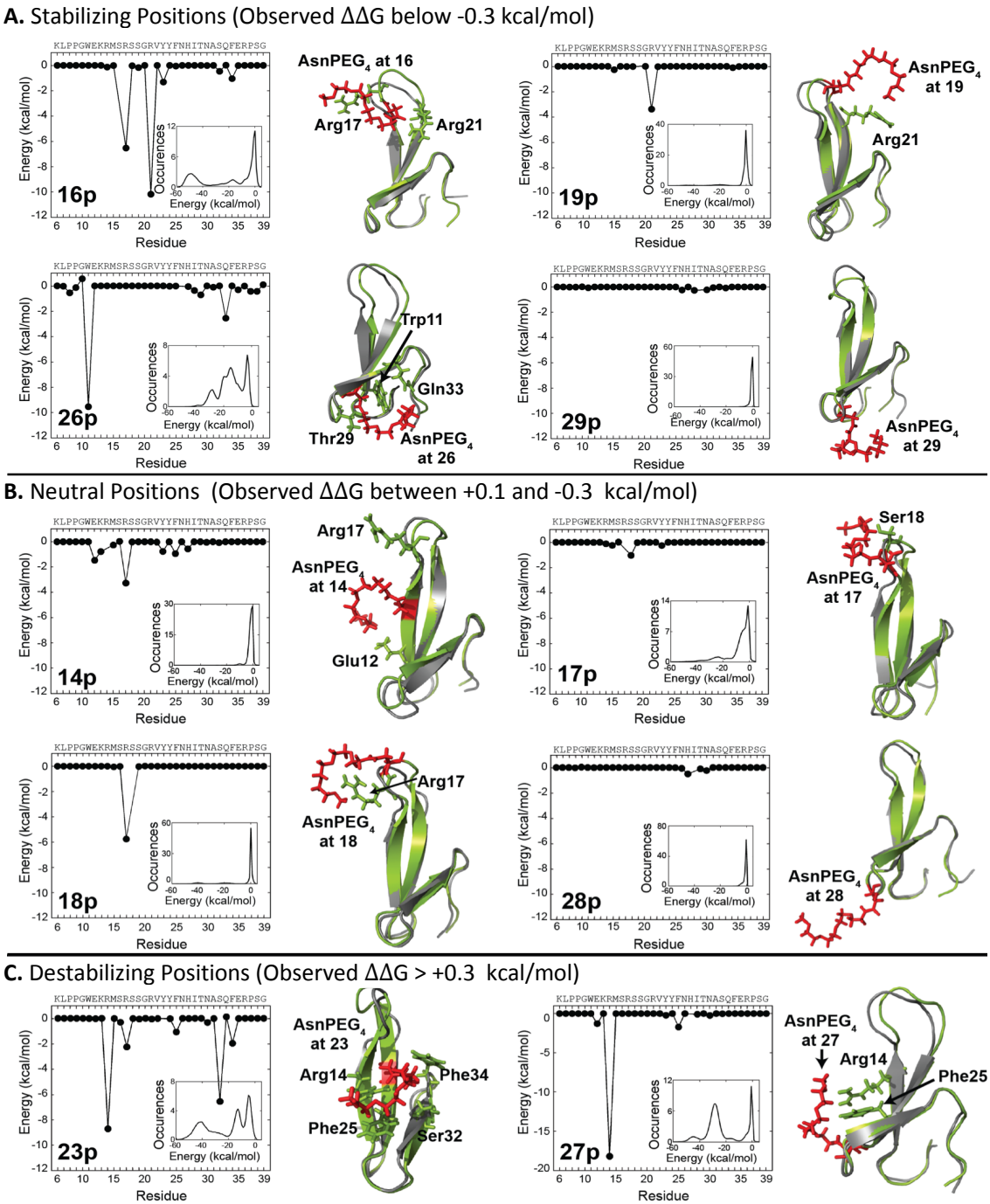


Figure 2-3. Results of atomistic simulations of WW variants PEGylated at (A) stabilizing, (B) neutral, and (C) destabilizing positions. Plots show the average interaction energy between PEG at a given position and every other residue within WW. Insets show histograms of these interactions. A snapshots from each simulation of each variant (green) is overlaid with the crystal structure of the unmodified protein WW (grey). AsnPEG4 is highlighted in red. Side chains that appear to engage in interactions with the PEG are shown as green sticks. In some cases, the PEG-protein interactions are relatively transient; therefore snapshots are not always representative of the entire simulation.

higher water disorder) around Tyr23 and Phe34. Interestingly, this change in water molecule organization is long-range and can extend out to 10 Å from the protein, indicating that PEG not only affects the first hydration shell but also more distant shells. We observe similar effects in the water around Trp11 and Thr29 in proteins **26p** vs. **WW**. The insets in each panel of Figure 2-4 show histograms of the number of water molecules observed in the simulations at a distance < 3 Å from the indicated side chain (i.e., the first hydration shell). In **26p**, PEG results in a decrease in the number of water molecules (i.e. dehydration) in the first solvation shell around Trp11. A smaller amount of dehydration occurs about Phe34 in **16p**.

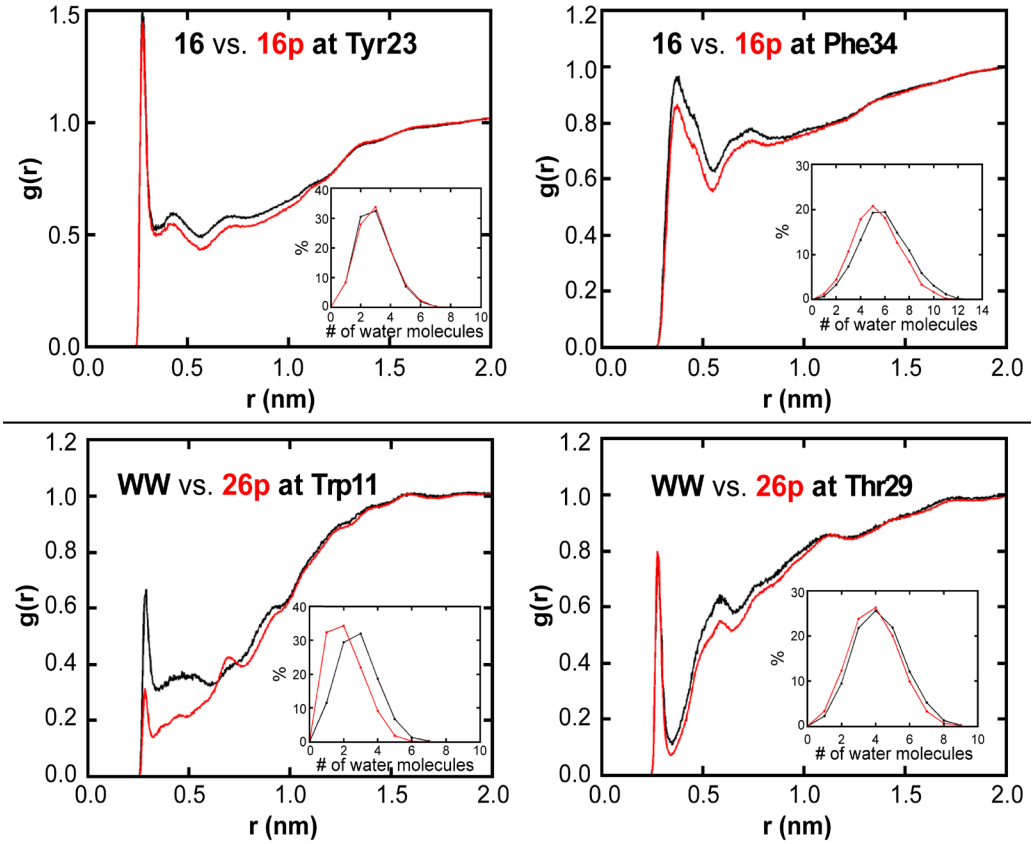


Figure 2-4. Simulated radial distribution function of water about Tyr23 or Phe34 acids in 16 vs. 16p (top panel) or about Trp11 or Thr29 in WW vs. 26p (bottom panel). Insets show histograms of the number of water molecules at a distance of < 3 Å from the indicated side chains (i.e. the first hydration shell).

We would expect this PEG-based dehydration (i.e. release of water from the protein surface to bulk solvent) to be entropically favorable, offset by a smaller increase in enthalpy due to the loss of protein-water hydrogen bonds, an expectation consistent with our earlier observations that PEG-based stabilization is entropic in origin (Table 2-1, compare **19p** vs **19**: $\Delta\Delta H_f = 3.6 \pm 1.4$ kcal mol^{-1}; $-T\Delta\Delta S_f = -4.3 \pm 1.4$ kcal mol^{-1}). We speculate that this dehydration effect is more pronounced near water-binding OH groups and is the origin of the observed impact of OH groups on the PEG-based stabilization of WW.

We explored this possibility experimentally by assessing the impact of increasing amounts of heavy water (D_2O) on the conformational stability of **19p** vs. **19**. The results of this analysis are shown in Figure 2-5A. Non-PEGylated **19** is -0.47 ± 0.05 kcal mol^{-1} more stable in buffer containing 98% D_2O than in buffer containing no D_2O. In contrast, a similar increase in D_2O only increases the stability of **19p** by -0.25 ± 0.04 kcal mol^{-1}.

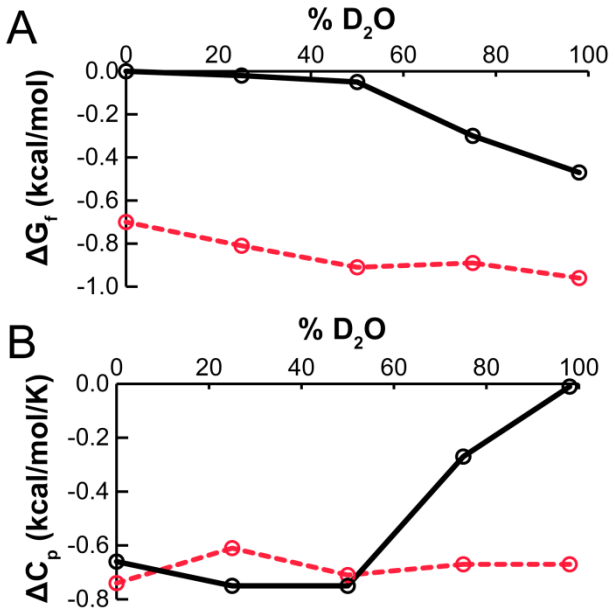


Figure 2-5. The change in (A) conformational stability (ΔG_f) or (B) heat capacity (ΔC_p) associated with folding of **19** (black solid line) or **19p** (red dashed line) in the presence of increasing amounts of D_2O in 20 mM phosphate (pH 7) at 25 °C and at a concentration of 100 μM.

Previous studies indicate that D_2O decreases the internal flexibility and increases the conformational stability of proteins, and suggests that the origin of this effect is in the increased strength of the non-covalent O–D···X interaction (i.e., a deuterium bond) relative to the non-covalent O–H···X interaction (i.e., a hydrogen bond).[31] This difference in strength provides an energetic incentive for the oxygen atoms within D_2O to engage in more solvent-solvent deuterium bonds and fewer solvent-protein hydrogen bonds. This effect increases the compactness of the folded protein and makes unfolding less favorable. We hypothesize that increasing D_2O concentration to 98% stabilizes non-PEGylated **19** more profoundly than PEGylated **19p** because **19p** is less solvated than **19**, with fewer solvent-protein hydrogen bonds to replace with stronger solvent-solvent deuterium bonds.

Increasing the D_2O concentration also affects the heat capacity change due to folding (ΔC_p) for **19** and **19p** (Figure 2-5B). For **19**, ΔC_p increases from -0.66 ± 0.07 kcal mol^{-1} K^{-1} (no D_2O) to 0.01 ± 0.12 kcal mol^{-1} K^{-1} (98% D_2O). In contrast, the ΔC_p for **19p** (-0.74 ± 0.05 kcal mol^{-1} K^{-1} in H_2O) is not substantially affected by increasing amounts of D_2O. In the context of protein folding, negative values of ΔC_p are associated with folding processes that decrease the amount of solvent-accessible surface area by burying non-polar side chains (or, alternatively, with unfolding processes that increase solvent-accessible surface area by exposing non-polar side chains to solvent).[32]We hypothesize that increasing D_2O concentration makes ΔC_p of **19** less negative because the unfolded conformation of **19** in D_2O is more compact, with less exposed non-polar surface area than the unfolded conformation of **19** in H_2O (i.e., the stronger network deuterium bonds in D_2O more effectively constrains the unfolded conformation of **19** than does the weaker network hydrogen bonds in H_2O). In contrast, we hypothesize that the ΔC_p of **19p** is

independent of D_2O concentration because PEG disrupts the strong network of deuterium bonds surrounding the protein, thereby attenuating the penalty for unfolding in D_2O.

2.2.4 PEGylation at Two Stabilizing Sites

We next wondered whether simultaneously PEGylating two of the identified "stabilizing" positions would result in more substantial increases to WW conformational stability. To address this question, we prepared PEGylated proteins **16p/26p**, **19p/26p**, **26p/29p**, and **16p/19p** and their mono- and non-PEGylated counterparts (**16/26p**, **16p**, and **16**; **19/26p**, **19p**, and **19**;**16/19p**, **16p/19**, and **16/19**; and **26p/29**, **29p**, and **29**, respectively, see Table 2-5), in which Asn or $AsnPEG_4$ has been incorporated at each of two stabilizing positions as indicated. We assessed the conformational stability of these variants using variable temperature CD experiments; results are shown in Table 2-5.

Doubly PEGylated compound **19p/26p** is -1.26 ± 0.03 kcal mol^{-1} more stable than its non-PEGylated counterpart **19**, a larger increment than we observed above for mono-PEGylation at position 19 or at position 26. We used double mutant cycle analysis to determine whether the PEG oligomers at positions 19 and 26 stabilize **19p/26p** independently or whether they interact with each other (either favorably or unfavorably). When Asn occupies position 19, PEGylation at position 26 stabilizes WW by -0.59 ± 0.02 (compare **19/26p** with **19**). When $AsnPEG_4$ occupies position 19, PEGylation at position 26 stabilizes WW by -0.55 ± 0.04 (compare **19p/26p** with **19p**), a nearly identical amount. These results suggest that the PEG oligomers at these positions do not interfere with each other and contribute independently and additively to the overall stabilization of **19p/26p** relative to **19**.

Table 2-5. Simultaneous PEGylation with the Four-unit Oligomer at Two Positions within WW.

Protein	T_m (°C)	ΔT_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)	ΔH_f (kcal/mol)	$\Delta\Delta H_f$ (kcal/mol)	$-T\Delta S_f$ (kcal/mol)	$-T\Delta\Delta S_f$ (kcal/mol)
16	50.6 ± 0.2	12.1 ± 0.3	0.00 ± 0.02	-1.02 ± 0.03	-30.2 ± 0.8	4.6 ± 1.2	30.5 ± 0.8	-5.6 ± 1.2
16/26p	62.6 ± 0.3		-1.02 ± 0.02		-25.9 ± 0.8		24.9 ± 0.8	
16p	60.7 ± 0.3	6.6 ± 0.3	-0.90 ± 0.03	-0.49 ± 0.03	-26.6 ± 1.1	4.3 ± 1.1	25.8 ± 1.1	-4.8 ± 1.1
16p/26p	67.3 ± 0.1		-1.38 ± 0.01		-22.3 ± 0.3		20.9 ± 0.3	
19	55.6 ± 0.2	6.8 ± 0.2	0.00 ± 0.02	-0.59 ± 0.02	-31.4 ± 0.6	4.4 ± 0.7	31.4 ± 0.6	-5.0 ± 0.7
19/26p	62.4 ± 0.1		-0.59 ± 0.02		-26.9 ± 0.4		26.3 ± 0.4	
19p	63.3 ± 0.3	6.4 ± 0.3	-0.70 ± 0.04	-0.55 ± 0.04	-27.8 ± 1.2	2.5 ± 1.3	27.1 ± 1.2	-3.0 ± 1.3
19p/26p	69.8 ± 0.1		-1.26 ± 0.02		-25.3 ± 0.4		24.0 ± 0.4	
16/19	56.9 ± 0.1	5.4 ± 0.2	0.00 ± 0.01	-0.49 ± 0.02	-30.2 ± 0.3	1.4 ± 0.5	30.2 ± 0.3	-1.9 ± 0.5
16/19p	62.4 ± 0.1		-0.49 ± 0.01		-28.8 ± 0.4		28.3 ± 0.4	
16p/19	63.2 ± 0.1	2.2 ± 0.1	-0.63 ± 0.01	-0.17 ± 0.02	-31.6 ± 0.5	2.6 ± 0.6	30.9 ± 0.4	-2.7 ± 0.6
16p/19p	65.4 ± 0.1		-0.80 ± 0.01		-29.0 ± 0.4		28.2 ± 0.4	
29	50.0 ± 0.3	7.1 ± 0.4	0.00 ± 0.03	-0.60 ± 0.04	-28.1 ± 0.4	1.3 ± 0.6	28.1 ± 0.4	-1.9 ± 0.6
26p/29	57.1 ± 0.2		-0.60 ± 0.02		-26.8 ± 0.5		26.2 ± 0.5	
29p	54.1 ± 0.3	2.6 ± 0.4	-0.36 ± 0.02	-0.20 ± 0.03	-27.8 ± 0.4	1.7 ± 0.6	27.5 ± 0.4	-1.9 ± 0.6
26p/29p	56.8 ± 0.3		-0.56 ± 002		-26.1 ± 0.5		25.6 ± 0.5	

Tabulated data are given ± standard error at 100 μM protein concentration in 20 mM sodium phosphate buffer, pH 7. Values of $\Delta\Delta G_f$were derived from variable temperature CD experiments and are given at the melting temperature of the corresponding non-PEGylated protein (i.e. **16**, **19**, **16/19**, **23**, and **29**, respectively).

In the other doubly PEGylated variants shown in Table 2-5, the PEG oligomers appear to interfere with each other unfavorably such that full additive stabilization is not observed. For example, mono-PEGylated **16p/19** and **16/19p** are -0.63 ± 0.02 and -0.49 ± 0.02 kcal mol^{-1} more stable, respectively, than their non-PEGylated counterpart **16/19**. If the PEG oligomers at these positions contributed independently to WW stability in doubly PEGylated **16p/19p**, we would expect **16p/19p** to be -1.13 ± 0.02 kcal/mol more stable than non-PEGylated **16/19**. In contrast, we observe that **16p/19p** is only -0.80 ± 0.02 kcal mol^{-1} more stable than **16/19**, indicating that the PEG oligomers at positions 16 and 19 interfere unfavorably with each other in **16p/19p**. We observe analogous results for **29, 26p/29, 29p, and 26p/29p**.

In both of these situations, the doubly PEGylated compounds contain two PEG oligomers within a single reverse turn. It is possible that the oligomers interfere sterically with each other in such close proximity. Alternatively, it is possible that PEG oligomers at these two neighboring positions tend to interact with the same partners on the WW surface such that when two oligomers are present, full additive stabilization is not possible.

More difficult to understand is the unfavorable interference between PEG oligomers at positions 16 and 26 in doubly PEGylated **16p/26p**. Positions 16 and 26 are not close to each other in primary sequence, are more than 15 Å apart in the WW structure, and actually occupy opposite faces of WW. The unfavorable interference between PEG oligomers at positions 16 and 26 must not involve direct contact between the PEG oligomers or even competing interactions with the same side-chains, but might involve more subtle effects.

Fully atomistic simulations of doubly PEGylated compounds **16p/19p** and **19p/26p** provide support for our hypothesis that PEG oligomers in close proximity can interfere with each other. For example, in **16p/19p**, the PEG oligomers at positions 16 and 19 are unable to engage in the same PEG-protein interactions that were available to them in mono-PEGylated **16p** and **19p** (i.e. with Arg21), respectively, presumably because of steric interference (Figure 2-6A). In contrast, the PEG-protein interactions at positions 19 and 26 in **19p/26p** are the same that were available in mono-PEGylated **19p** and **26p**, respectively, suggesting that distance can prevent PEG oligomers from interfering with each other (Figure 2-6B).

However, PEG-based changes to the network of hydrogen-bound solvent molecules around might also be able to explain the non-additive interference observed between the PEG oligomers in **16p/19p**. It is possible that the PEG oligomers at 16 and 19 perturb this network of

solvent molecules in the same general area such that introducing a second nearby PEG doesn't have a large additional impact.

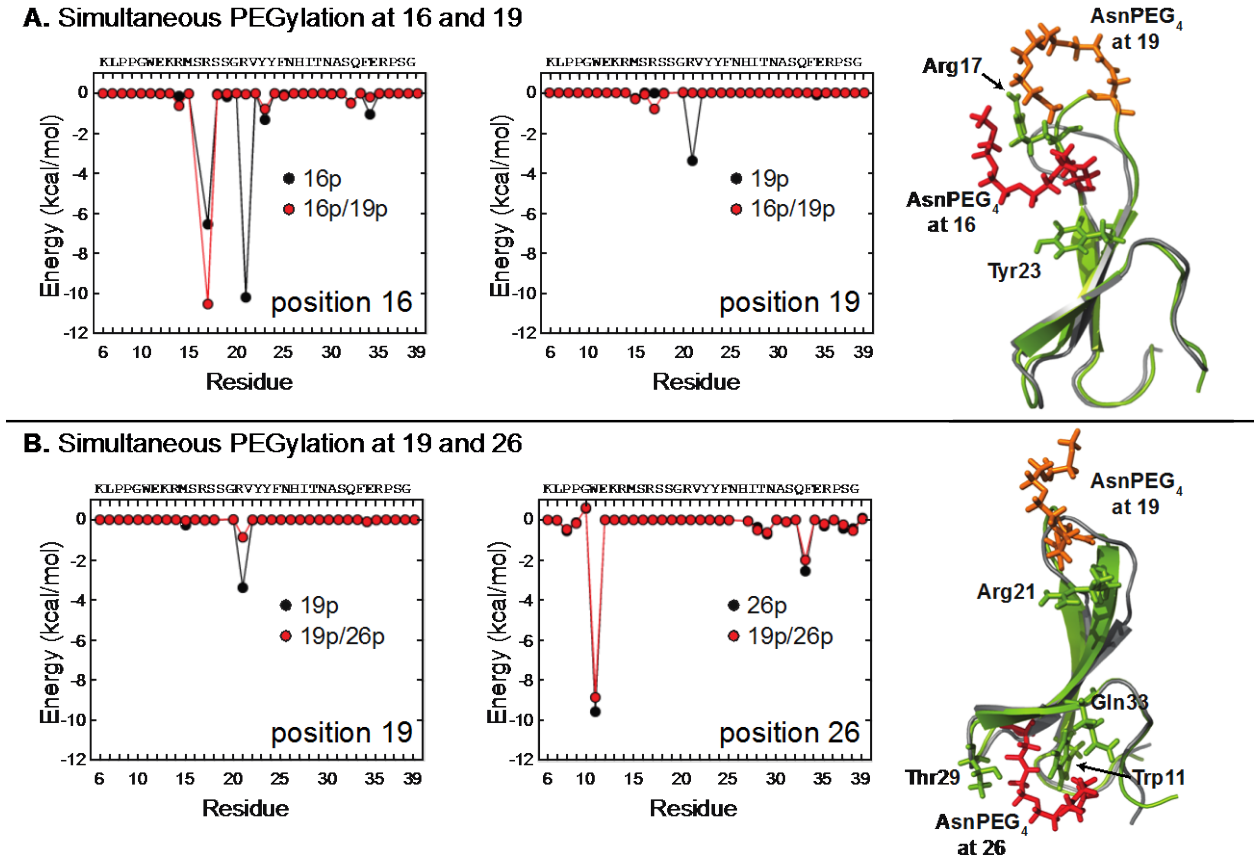


Figure 2-6. Comparison between the PEG-protein interactions in (A) doubly PEGylated 16p/19p vs. mono-PEGylated 16p and 19p and in (B) doubly PEGylated 19p/26p vs. mono-PEGylated 19p and 26p. Graphs show the average PEG-protein interaction energy between $AsnPEG_4$ at the indicated positions and every other residue within WW. Results from mono-PEGylated variants are shown in black, whereas results from doubly PEGylated variants are shown in red. Also shown are snapshot from the simulations of the doubly PEGylated variants (green), overlaid with the crystal structure of the unmodified WW (grey). In each snapshot, $AsnPEG_4$ is highlighted in orange at position 19 and in red at positions 16 or 26. Side chains that appear to engage in interactions with PEG in each snapshot are shown as green sticks.

This is consistent with what we observe in the double mutant cycle for **16/19**, **16/19p**, **16p/19**, and **16p/19p**, as described above: the second introduced PEG has a smaller impact on WW conformational stability than the first. In contrast, when the attached PEGs are sufficiently far apart, the first and second PEGs impact WW conformational stability independently and additively.

2.2.5 Structure-Based Selection of Stabilizing PEGylation Sites

Based on these mechanistic insights, we wondered whether (1) side chain orientation and (2) the presence of nearby OH groups could be used as structure-based criteria to identify positions most likely to experience substantial entropic PEG-based stabilization. To this end, we analyzed each of the PEGylation sites discussed above in x-ray crystal structure of the parent WW domain from which **16p**, **17p**, **18p**, **19p**, **26p**, **27p**, **28p**, **29p**, **30p**, **32p** and their non-PEGylated counterparts were derived. We limited this analysis to these variants because their CD spectra indicate close structural similarity to the parent WW domain.

At each PEGylation site, we defined vectors **a** and **b** (Figure 2-7): vector **a** begins with the backbone alpha carbon and ends at the side-chain center-of-mass (determined by averaging the *x,y,z* coordinates of each side-chain atom); vector b begins with the side-chain center-of-mass and ends at side-chain oxygen atom of the nearest Ser, Thr, or Tyr residue. We then measured the angle θ between vectors **a** and **b** at each position using the following relationship: $\cos\theta = \mathbf{a}\cdot\mathbf{b} / (|\mathbf{a}|\cdot|\mathbf{b}|)$. Small values of θ indicate that a side chain is oriented toward the nearest Ser, Thr, Tyr residue, whereas large values of θ indicate orientation away from the nearest Ser, Thr, or Tyr residue. Values of θ for all the positions investigated are shown in Table 2-6.

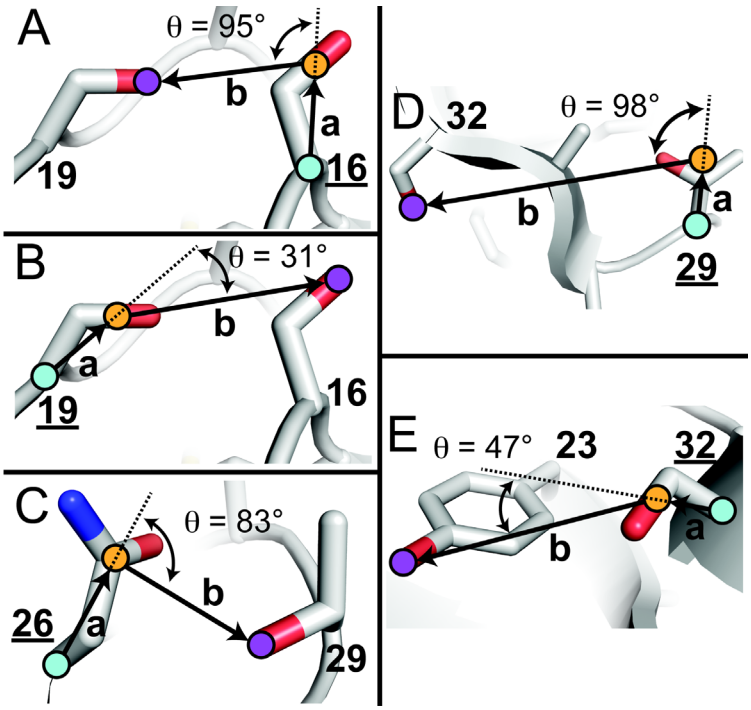


Figure 2-6. The angle θ between vectors a and b at positions (A) 16, (B) 19, (C) 26, (D) 29, and (E) 32. Alpha carbons (C□), side-chain centers-of-mass (COM), and oxygens of nearest OH-containing side-chain are highlighted with blue-, orange-, and purple-filled circles, respectively.

Table 2-6. The angle θ at various PEGylation sites within WW

PEGylation Site	Native residue	ΔΔG$_f$ (kcal/mol)	θ (°)
16	Ser	-0.90 ± 0.03	95
17	Arg	-0.18 ± 0.05	121
18	Ser	0.00 ± 0.07	145
19	Ser	-0.70 ± 0.04	31
26	Asn	-0.58 ± 0.06	83
27	His	0.38 ± 0.04	155
28	Ile	0.00 ± 0.07	128
29	Thr	-0.36 ± 0.04	98
30	Asn	0.00 ± 0.07	150
32	Ser	-0.45 ± 0.02	47

ΔΔG$_f$ values associated with PEGylation at each position are from table 2-1.

Next, we examined the relationship between the angle θ and the PEG-based stabilization ($\Delta\Delta G_f$) of **16p**, **17p**, **18p**, **19p**, **26p**, **27p**, **28p**, **29p**, **30p**, and **32p**, relative to their non-PEGylated counterparts. Figure 2-8 indicates that PEGylation tends to be most stabilizing at positions with smaller values of θ (i.e. at positions that are oriented toward nearby Ser, Thr, or Tyr side chains). Most importantly, had we used the correlation line shown in Figure 2-8 prospectively, we would have correctly predicted that PEGylation at positions 16, 19, 26, 29 and 32 would result in substantial (< -0.30 kcal mol^{-1}) increases to conformational stability.

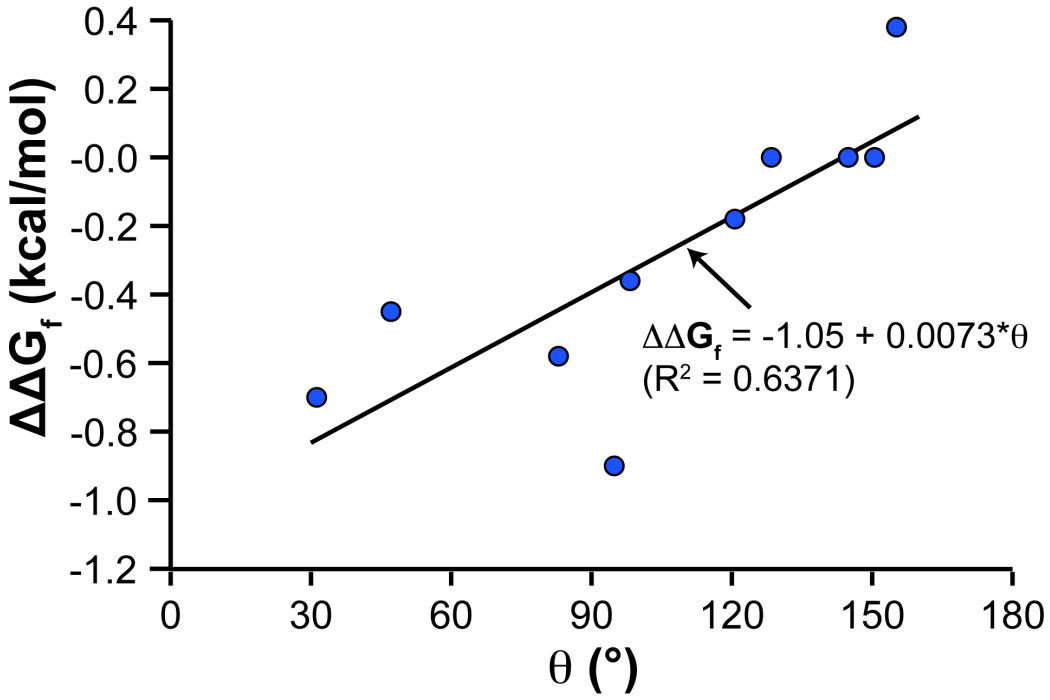


Figure 2-7. Relationship between the angle θ and PEG-based stabilization at a given site ($\Delta\Delta G_f$)

We tested the utility of the angle θ as a predictor of PEG-based stabilization within other proteins by calculating θ for each residue within the chicken Src SH3 domain (hereafter called SH3, Figure 2-9A).[33] Thr20 and Thr22 are within the same loop near the N-terminus of SH3 and

are only 3.7 Å apart from each other. More importantly, Thr20 is oriented toward Thr22, with θ = 77 °C (Figure 2-9A). The correlation observed between θ and PEG-based stabilization in the context of WW (Figure 2-8) led us to predict that PEGylation of an Asn residue at position 20 of SH3 would enhance conformational stability. To test this hypothesis, we used solid-phase peptide synthesis to prepare **SH3 T20N** and **SH3 T20NPEG**, in which Thr20 has been replaced by Asn or by $Asn(PEG)_4$, respectively. We assessed the conformational stability of these variants using variable temperature CD experiments, run in triplicate for each variant (Figure 2-9B).

Data from these experiments are readily fit to equations derived from a two-state folding-unfolding model. The melting temperature T_m of **SH3 T20NPEG** (74.0 ± 0.8 °C) is 13.0 ± 0.9 °C higher than that of non-PEGylated **SH3 T20N** (T_m = 61.1 ± 0.3 °C), corresponding to an increase in stability of -1.2 ± 0.1 kcal mol^{-1} at 61.1 °C. Atomistic simulations suggest that PEGylation of **SH3 T20NPEG** is associated with lower water density (and higher water disorder) around nearby polar side chains, including Thr22, suggesting a similar origin for PEG-based stabilization in SH3[34] and in WW (see Figure 2-10). It is remarkable that a four-unit PEG oligomer can have such a large effect on the stability of the 56-residue SH3 domain. This substantial increase in SH3 conformational stability is consistent, suggesting that the angle θ (i.e. the orientation of a side-chain with respect to nearby OH groups) is a reasonable predictor of PEG-based stabilization in β-sheet- and reverse-turn-containing proteins.

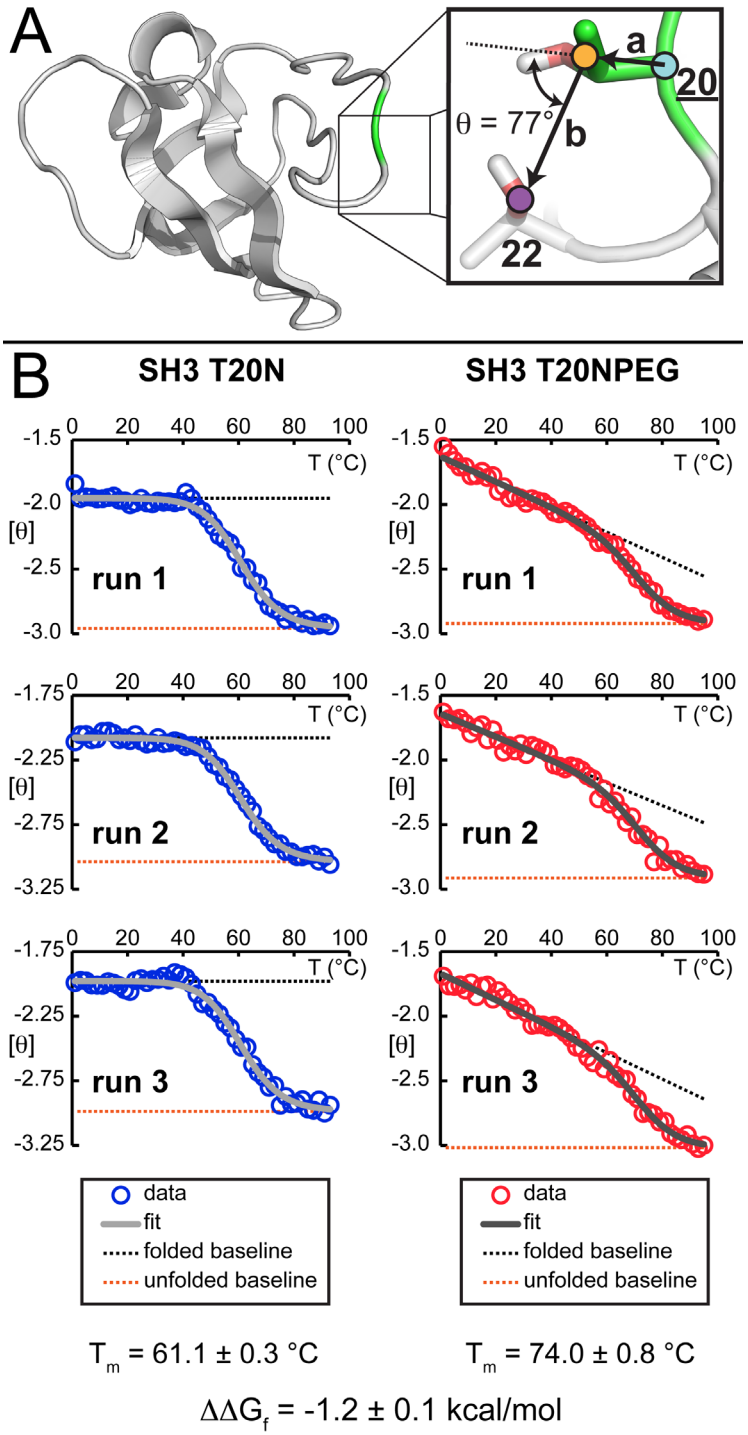


Figure 2-8. (A) Ribbon diagram of chicken Src SH3 (PDB ID: 1SRL), with Thr20 highlighted in green. Inset shows the angle θ between vectors a and b at position 20. Alpha carbons (Cα), side-chain centers-of-mass (COM), and oxygens of nearest OH-containing side-chain are highlighted with blue-, orange-, and purple-filled circles, respectively. (B) Variable temperature CD data for SH3 T20N and SH3 T20NPEG at 50 μM in 20 mM sodium phosphate buffer, pH 7, run in triplicate for each variant.

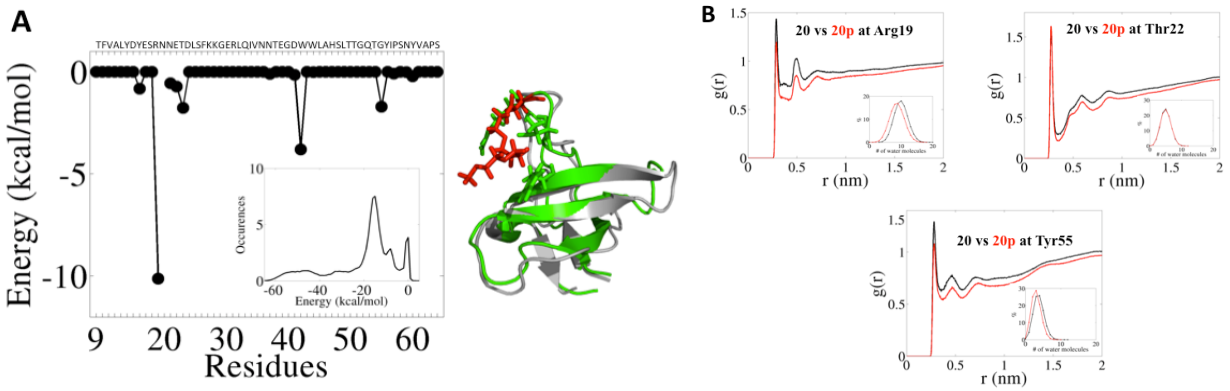


Figure 2-9. (A) Results of atomistic simulations of SH3 T20NPEG. The plot shows the average interaction energy between PEG at a given position and every other residue within SH3 T20N. Inset shows histograms of these interactions. A snapshot from the simulation of SH3 T20NPEG (green) is overlaid with the crystal structure of the unmodified protein SH3. $AsnPEG_4$ is highlighted in red. Side chains that appear to engage in interactions with the PEG are shown as green sticks. In some cases, the PEG-protein interactions are relatively transient; therefore snapshots are not always representative of the entire simulation. (B) Simulated radial distribution function of water about the indicated nearby polar side chains in SH3 T20N vs. SH3 T20NPEG. Insets show histograms of the number of water molecules at a distance of < 3 Å from the indicated side chains (i.e. the first hydration shell).

We wondered whether the PEG-based stabilization observed at selected positions in WW would be associated with enhanced protection from proteolytic degradation. To address this question, we assessed the resistance of the WW variants described above to degradation by pronase and by proteinase K.[35] We previously showed that PEGylation at stabilizing position 19 protected WW from proteolysis.[25] Recall that PEGylation at position 19 is stabilizing ($\Delta\Delta G_f$ = -0.70 ± 0.04 kcal mol^{-1}. The half-life of PEGylated **19p** in pronase is 3.6 ± 0.3 times longer than that of non-PEGylated **19**; we observed a similar effect in proteinase K (Table 2-7**)**. At neutral position 18 (where PEGylation does not change conformational stability), the same four-unit PEG oligomer provides much less protection against proteolysis: the half-lives of **18p** in pronase and in proteinase K are only1.23 ± 0.06 and 1.7 ± 0.3 longer, respectively, than those of non-PEGylated **18** (Table 2-7).

Proteolysis experiments with the other PEGylated WW variants and their non-PEGylated counterparts described above follow a similar trend (see Table 2-7), with increased resistance to

proteolysis observed in cases where PEG is most strongly stabilizing. For example, the half-lives of doubly PEGylated **19p**/**26p** in pronase and in proteinase K are 5.9 ± 0.9 and 6.9 ± 1.9 times longer, respectively, than those of non-PEGylated **19/26**. This trend is illustrated in Figure 2-11, which shows the ratio of the half-lives of each matched pair of PEGylated vs. non-PEGylated WW variants plotted against the conformational stability of each PEGylated variant relative to its non-PEGylated counterpart. The plot suggests that beyond a certain basal level, the proteolytic protection imparted by the four-unit oligomer is substantially enhanced at positions where PEG also increases WW conformational stability.

Table 2-7. Impact of PEGylation with PEG4 at various sites on resistance of WW variants to proteolysis

Site	$\Delta\Delta G_f$ (kcal/mol)	Pronase $t_{1/2}$ ratio	Proteinase K $t_{1/2}$ ratio
16	-0.90 ± 0.03	2.0 ± 0.2	2.3 ± 0.3
17	-0.18 ± 0.05	1.19 ± 0.07	---
18	0.00 ± 0.07	1.23 ± 0.06	1.7 ± 0.3
19	-0.70 ± 0.04	3.6 ± 0.3	3.4 ± 0.4
26	-0.58 ± 0.06	1.7 ± 0.2	2.7 ± 0.3
27	0.38 ± 0.04	0.86 ± 0.05	0.9 ± 0.1
28	0.00 ± 0.07	1.1 ± 0.1	1.0 ± 0.1
29	-0.36 ± 0.04	1.7 ± 0.2	2.0 ± 0.3
16/26	-1.38 ± 0.01	2.3 ± 0.2	4.1 ± 0.5
19/26	-1.26 ± 0.02	5.9 ± 0.9	6.9 ± 1.9
16/19	-0.80 ± 0.01	3.1 ± 0.3	---
26/29	-0.56 ± 0.02	3.2 ± 0.4	---

Tabulated data are given ± standard error. Values of $\Delta\Delta G_f$ are presented as given in table 2-1. The $t_{1/2}$ ratio for a given site in pronase or proteinase K is the ratio of the half-life of the PEGylated WW variant to the half-life of the corresponding non-PEGylated WW variant in the indicated protease. Proteolysis experiments were performed at 50 μM protein concentration in 20 mM sodium phosphate, pH 7.

These data provide support for our hypothesis that globally optimal PEGylation sites are characterized by the ability of the PEG oligomer to increase protein conformational stability.

However, one could argue that this observed dependence of proteolytic resistance on PEG-based conformational stabilization (Table 2-7, Figure 2-11) is a result of local effects that are important for four-unit PEGs, but which are insignificant for longer PEGs. Indeed, one might expect steric hindrance to be the dominant contributor to the proteolytic resistance associated with longer PEGs, independent of conformational stabilization. To test this hypothesis, we assessed the ability of a 45-unit PEG to protect WW variants **16p45**, **19p45**, **26p45, 28p45,** and **29p45** from proteolysis. This 45-unit PEG is clearly much shorter than the 20–40 kDa PEGs typically found in PEGylated protein drugs. However, WW is a small protein (~4 kDa); the 45-unit PEG (~2 kDa) comprises ~33% of the total masses of these PEGylated WW variants, a PEG/protein composition approaching that of many PEGylated protein drugs (Pegfilgrastim, for example, is 50% PEG: a ~20 kDa PEG attached to a ~20 kDa protein).

If steric hindrance were the only significant contributor to the proteolytic resistance associated with longer PEGs, one would expect PEG-based changes in protein conformational stabilization to matter less and less with increasing PEG/protein ratios. For example, one would expect proteolytic resistance in the 33:67 PEG/protein conjugates (e.g., the 45-unit PEG WW variants) to be less dependent on conformational stability than in the 5:95 PEG/protein conjugates (e.g., the 4-unit PEG WW variants). Instead, we find that a PEG/protein ratio of 33:67, PEG-based increases to proteolytic resistance remain strongly correlated with the impact of the 45-unit PEG on conformational stability. At "stabilizing" positions 16, 19, 26, and 29, the increases in conformational stability associated with the 45-unit PEG oligomer are accompanied by 2.3-, 5.6-, 2.8-, and 2.8-fold increases in half-life, respectively, in the presence of pronase (see Table 2-8). However, at "destabilizing" position 28 ($\Delta\Delta G_f = 0.36 \pm 0.05$ kcal mol^{-1}), the 45-unit

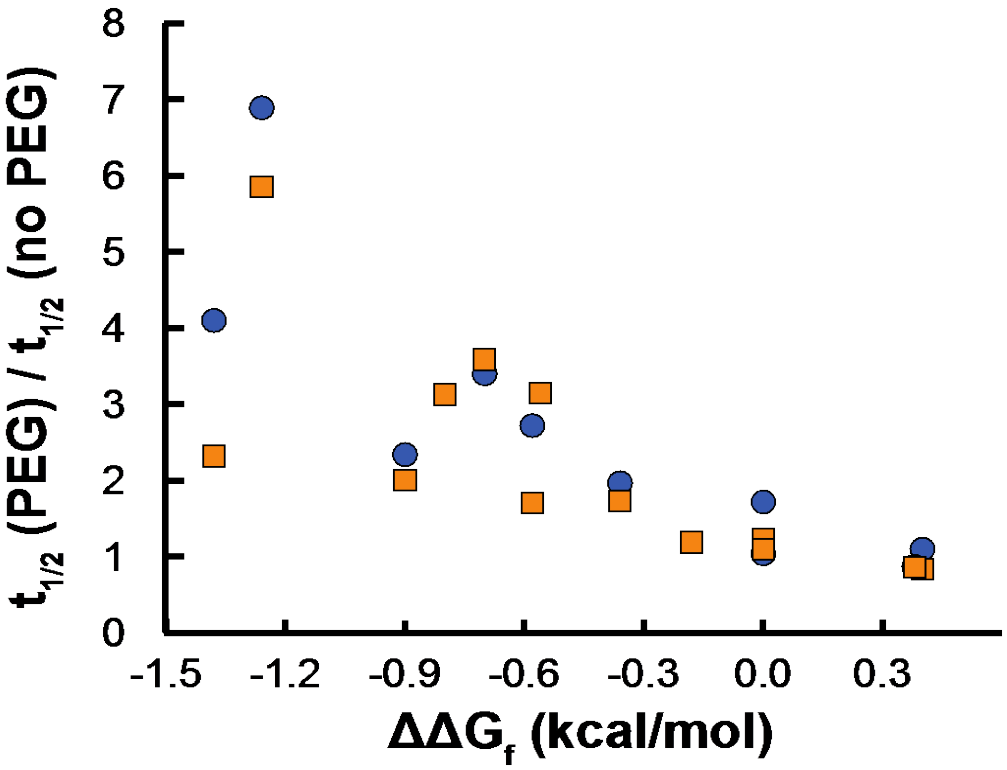


Figure 2-10. Plot of PEG-based proteolytic stability (expressed as the ratio of half-life of a given PEGylated WW variant to the half-life of its sequence-matched non-PEGylated counterpart) in the presence of pronase (blue circles) or proteinase K (orange squares) vs. PEG-based conformational stability ($\Delta\Delta G_f$).

PEG oligomer has no substantial impact on proteolytic stability. These observations are not consistent with the hypothesis that steric hindrance is the only significant factor contributing to PEG-based proteolytic resistance, and suggest that PEG-based changes to conformational stability also play an important role for PEG/protein conjugates approaching the compositions typical of PEGylated protein drugs.

Table 2-8. Impact of PEGylation with PEG45 at various sites on resistance of WW variants to proteolysis

Proteins	$\Delta\Delta G_f$ (kcal/mol)	Pronase $t_{1/2}$ ratio
16p45 vs. **16**	-0.39 ± 0.03	2.3 ± 0.5
19p45 vs. **19**	-0.67 ± 0.05	5.6 ± 1.2
26p45 vs. **WW**	-0.27 ± 0.06	2.8 ± 0.5
28p45 vs. **28**	0.36 ± 0.05	1.1 ± 0.3
29p45 vs. **29**	-0.59 ± 0.08	2.8 ± 0.5

Tabulated data given ± standard error. Values of $\Delta\Delta G_f$ are presented in table 2-2. The $t_{1/2}$ ratio for a given site in pronase is the ratio of the half-life of the PEGylated WW variant to the half –life of the corresponding non-PEGylated WW variant. Proteolysis experiments were performed at 50 μM protein concentration in 20 mM sodium phosphate, pH 7.

2.3 Conclusion

Advances in protein chemistry now allow site-specific PEGylation of any arbitrary position on the protein surface. Why pursue predictive tools for identifying optimal PEGylation sites when one can simply scan a PEGylated side chain through a list of potential sites and pick the one(s) that provide the best balance between enhanced pharmacokinetic properties and biological function?[36,37] Such a trial-and-error approach is unsatisfying from a scientific point of view; is both time- and resource-intensive (site-specific side-chain modification is much more challenging to carry out than alanine-scanning mutagenesis, for example) and may therefore be limited by practical considerations to a subset of potential surface sites; and must be repeated for each new protein of interest. In contrast, rational structure-based guidelines for identifying optimal PEGylation sites have the potential to circumvent this time-consuming step in PEGylated protein drug development.

We have developed a structure-based method for predicting which sites within the WW domain are most likely to experience PEG-based stabilization, and have shown that PEG-based stabilization is associated with enhanced resistance to proteolysis. We developed this method

based on mutagenesis experiments, which showed that side-chain orientation and the presence of nearby OH groups can modulate PEG-based stabilization at a given site. MD simulations suggest that stabilization cannot always be explained by favorable PEG-protein interactions because the formation of a tight PEG-protein interface is coupled by an entropic loss in many cases. While direct PEG-OH interactions cannot explain the increased thermodynamic stability, it is likely that nearby OH groups may instead exert a more indirect influence, involving the network of hydrogen-bound solvent molecules surrounding the protein. The simulations indicate that PEG can increase the disorder of water molecules around nearby residues. Solvent isotope experiments are consistent with this possibility, as are our observations that PEG-based stabilization is entropic in origin, with beneficial increases in entropy compensating for unfavorable increases in enthalpy.

We find that 45-and four-unit PEGs have a similar impact on WW conformational and proteolytic stability, suggesting that the structure-based model developed using the four-unit PEG will apply in the context of the larger oligomers typically used in therapeutically relevant proteins. Most importantly, we have also shown that our structure-based method can correctly predict a location within the Src SH3 domain (another β-sheet protein) where PEGylation enhances conformational stability.

2.4 Supporting Information

2.4.1 Atomistic Simulations of PEGylated WW variants

To study the effect of PEGylation on the folded state of WW Prof. Yaakov Levy's lab used all-atom modeling. Using Coot7 software,[38] we built the models of the modified version of WW with Asn or AsnPEG$_4$ at the various positions (14, 16, 17, 18, 19, 23, 26, 27, 28, 29, 30).

We also built the models of the variant with Asn or Asn-PEG at the position 19 with the additional mutations (16F, 21T, both of them); the impact of PEGylation on these variants was assessed previously by variable temperature circular dichroism experiments.[24] For all these variants, simulations reveal a change in solvation of the residues in sites 17 and 23 is observed, which is consistent with the observed PEG-based stabilization.

To build models for the Asn-linked PEG, we constructed a model $AsnPEG_4$ amino acid, which was blocked with an acetyl group on the main-chain amino group, and an amide on the main chain carbonyl group. Using Gaussian 09 Rev. B.01[39] at the Hartree-Fock level of theory and the 6-13G(d) basis set,[40-48] we derived the Mulliken partial charges[49] for this system. The programs *antechamber*, *parmchck,* and *tleap*[50] followed by acpype[51] were used to prepare GROMACS-format force field files based on the AMBER99SB-ILDN force field. These parameters were then incorporated into the AMBER99SB-ILDN files.

The molecular dynamics (MD) simulations were performed using GROMACS Version 4.5.4.[52] We used the modified AMBER99SB-ILDN force field[53] with the incorporated Asn-PEG residue. The various WW variants were solvated in a box with periodic boundary conditions containing pre-equilibrated SPC/E[54] water molecules. The size of the box was selected to assure distance of at least 1 nm between the solute and the box walls. The average sizes of the box for the non-PEGylated and PEGylated variants are 5x5x5 and 6x6x6 nm^3, respectively (with ~14,000 and 17,000 water molecules, respectively, in the box). Sodium and chloride ions were added to maintain overall system neutrality. The LINCS algorithm[55] was used to control bonds during the simulation. Leapfrog algorithm was employed with step of 2 fs. All the simulations were performed at constant pressure (1 atm) and temperature (300 K); the latter was controlled using a modified scheme of Berendsen thermostat.[56]

Before the runs, the proteins were relaxed using steepest descent method of energy minimization. The minimization was folled by the equilibration of the system in two phases. The first phase was conducted under an NVT ensemble; second under NPT enxemble (100 ps each phase). For all modified WW variants as well as for unmodified WW variants we performed 3 simulations for 100 ns each. The radial distribution analysis for water molecules was calculated as the probability to find the water oxygen atom next to specific atoms (Oγ for Ser, Nε for Arg, OH for Tyr, Oγ1 for Thr, and Cz for Phe).

For each variant, we calculated the average interaction energy between PEG and every residue within WW (Figure 2-3, large graphs). We also calculated the total energy of the PEG-protein interface during the simulation. The insets in Figure 2-3 show the histogram of this interface for each variant, providing insight into the strength and diversity of the observed PEG-protein interactions. For example, at a given PEGylation site, a histogram with sharp peak near 0 kcal mol^{-1} indicates that the PEG oligomer at this position is mostly solvent-exposed and does not engage extensively in PEG-protein interactions. In contrast, a histogram with lower energies suggests a more stable, more extended PEG-protein interaction interface. The width of the histogram suggests a diverse interface, namely, low specificity. Snapshots from the simulation of each variant are also shown in Figure 2-3, though these do not indicate the lifetime of specific interactions, which in some cases are relatively transient.

2.4.2 WW variant Synthesis

Proteins**14**, **14p**, **16**, **16p**,**16-Y23F**, **16p-Y23F**, **16-32A**, **16p-32A**,**17**, **17p**, **18**, **18p**, **19**, **19p**,**D-19**,**D-19p**, **19-32A**, **19p-32A**, **23**, **23p**,**WW**, **26p**,**WW-T29A**, **26p-T29A**,**27**, **27p**, **28**, **28p**, **29**, **29p**,**29-S32A**, **29p-S32A**,**30p**, **32**, **32p**,**32-Y23F**, **32p-Y23F**,**16/26p**, **16p/26p**, **19/26p**,

19p/26p, **16/19**, **16/19p**, **16p/19**, **16p/19p**, **26p/29**, **26p/29p**, **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45** Figure 2-12 were synthesized as C-terminal acids, by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc Nα protection strategy as described previously.[24,25]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) and Fmoc-Asn(PEG45)-OH, which were synthesized as described previously,[24] and Fmoc-D-Asn(PEG4)-OH (N^2-fluorenylmethyoxycarbonyl-N^4-[11-methoxy-3,6,9-trioxaundecyl]-D-asparagine), which was synthesized as described below.

WW variants were synthesized on a 25μmol scale. A general protocol for manual solid-phase peptide synthesis follows: Fmoc-Gly-loaded NovaSyn Wang resin (69.4mg, 25μmol 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). 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 1.25 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.

Figure 2-11. Sequences for PEGylated and non-PEGylated WW variants.

For coupling of an activated amino acid, we prepared a stock coupling solution of 100 mL NMP, 3.17 g HBTU (0.01mol, 0.1M) and 1.53 g HOBt (0.01 mol, 0.1M) for a final concentration of 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 vacuum 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.

2.4.3 WW Variant 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 spectrometry (ESI-TOF, spectra appear below in Figures 2-13 through 2-64), and purity was analyzed by Analytical HPLC shown in Figures 2-65 through 2-118).

2.4.4 ESI-TOF data

ESI-TOF spectra for proteins **WW, 14**, **14p**, **16**, **16p**,**16-Y23F**, **16p-Y23F**, **16-32A**, **16p-32A**,**17**, **17p**, **18**, **18p**, **19**, **19p**,**D-19**,**D-19p**, **19-32A**, **19p-32A**, **23**, **23p**,**26p**,**WW-T29A**, **26p-T29A**,**27**, **27p**, **28**, **28p**, **29**, **29p**,**29-S32A**, **29p-S32A**,**30p**, **32**, **32p**,**32-Y23F**, **32p-Y23F**,**16/26p**, **16p/26p**, **19/26p**, **19p/26p**, **16/19**, **16/19p**, **16p/19**, **16p/19p**, **26p/29**, **26p/29p**, **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45** are shown in Figures 2-13 through 2-64.

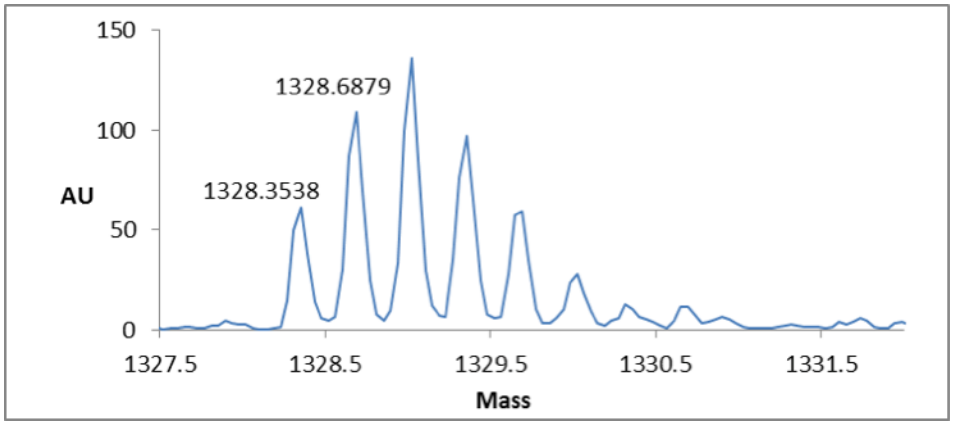


Figure 2-13. ESI TOF spectrum for parent protein **WW**. Expected $[M+3H]^{3+}/3$ = 1328.3351 Da. Observed $[M+3H]^{3+}/3$ = 1328.3538 Da.

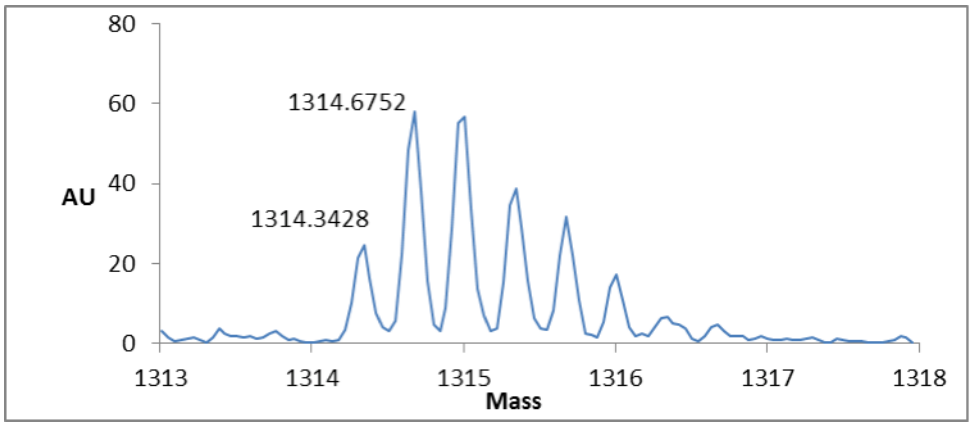


Figure 2-14. ESI TOF spectrum for the WW variant **14**. Expected $[M+3H]^{3+}/3$ = 1314.3157 Da. Observed $[M+3H]^{3+}/3$ = 1314.3428 Da

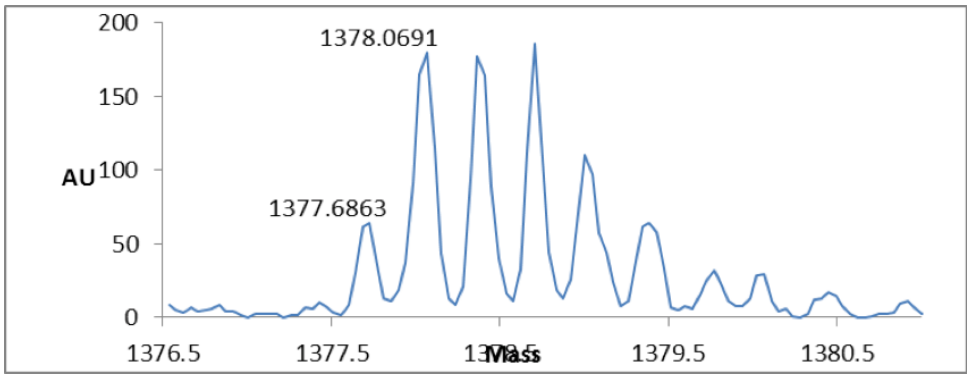


Figure 2-15. ESI TOF spectrum for WW variant **14p**. Expected $[M+3H]^{3+}/3$ = 1377.6892 Da. Observed $[M+3H]^{3+}/3$ = 1377.6863 Da.

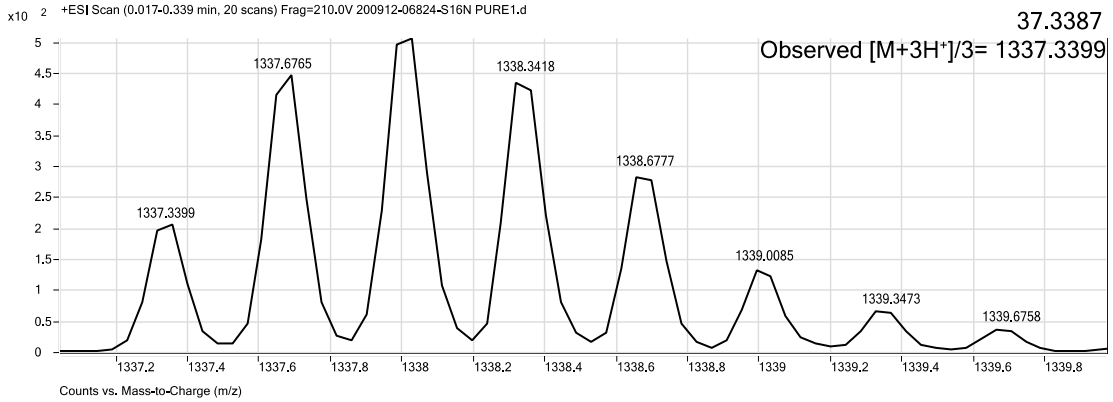


Figure 2-16. ESI TOF spectrum for WW variant **16.** Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3389 Da.

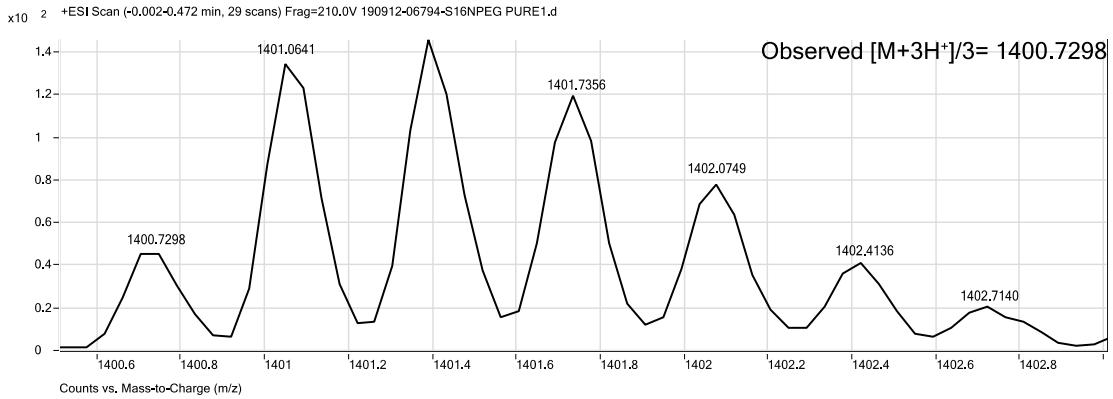


Figure 2-17. ESI TOF spectrum for WW variant **16p**. Expected $[M+3H]^{3+}/3$ =1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.7298 Da.

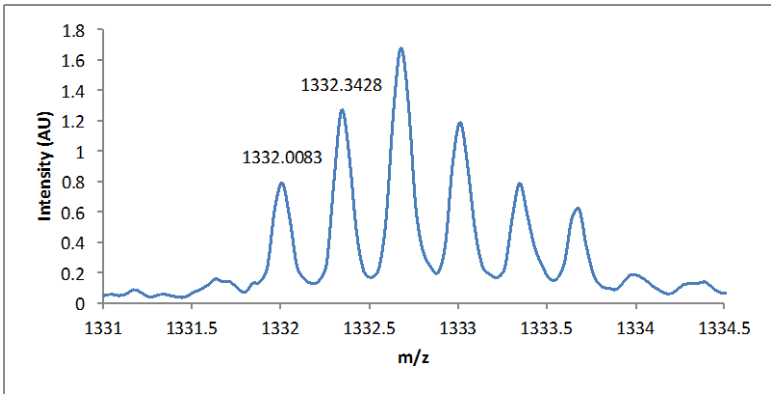


Figure 2-18. ESI TOF spectrum for WW variant **16-Y23F**. Expected $[M+3H]^{3+}/3$ =1332.007 Da. Observed $[M+3H]^{3+}/3$ = 1332.008 Da.

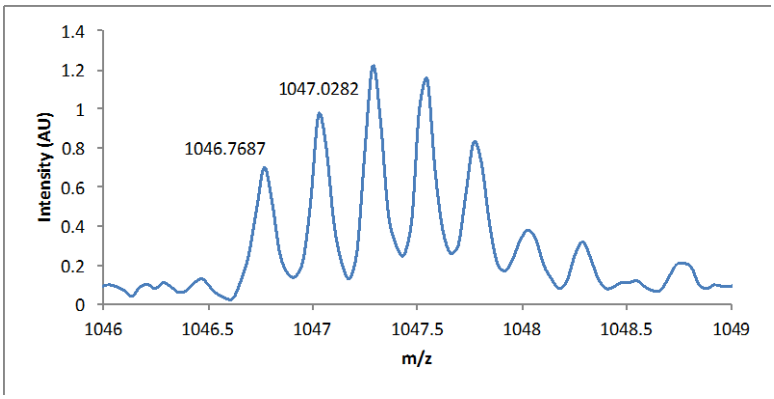


Figure 2-19. ESI TOF spectrum for WW variant **16p-Y23F**. Expected $[M+4H]^{4+}/4$ = 1046.787 Da. Observed $[M+4H]^{4+}/4$ = 1046.769 Da.

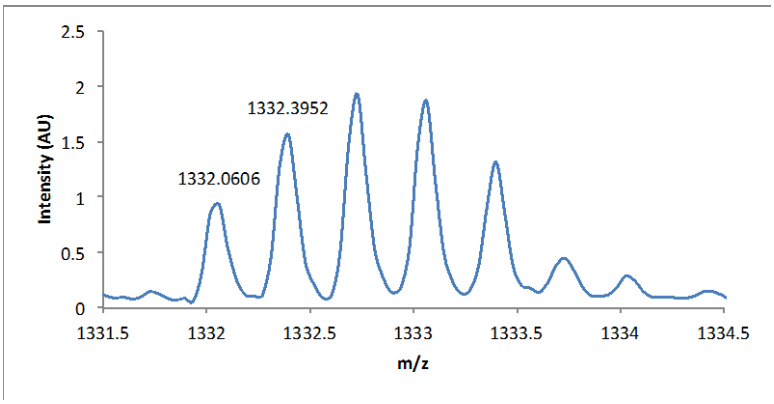


Figure 2-20. ESI TOF spectrum for WW variant **16-S32A**. Expected $[M+3H]^{3+}/3$ = 1332.007 Da. Observed $[M+3H]^{3+}/3$ = 1332.061 Da.

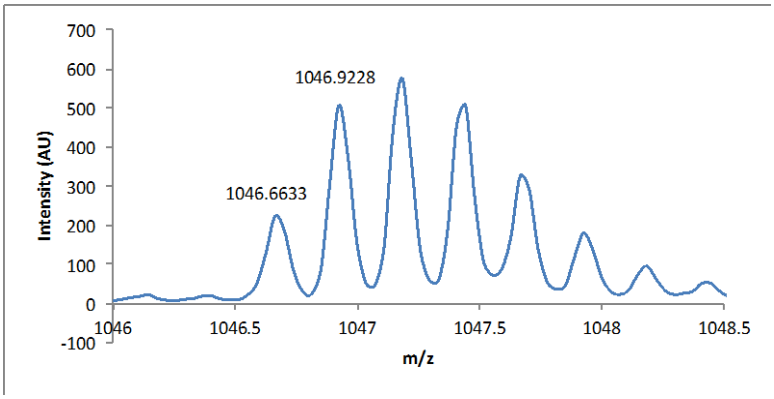


Figure 2-21. ESI TOF spectrum for WW variant **16p-S32A**. Expected $[M+4H]^{4+}/4$ = 1046.787 Da. Observed $[M+4H]^{4+}/4$ = 1046.663 Da.

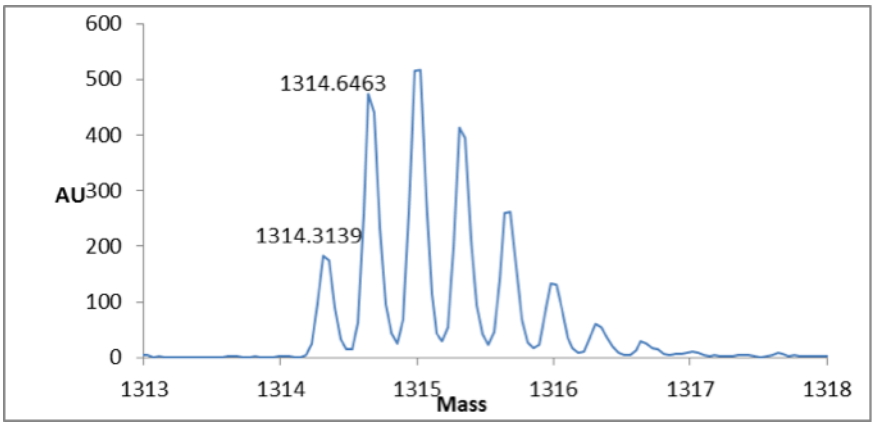


Figure 2-22. ESI TOF spectrum for WW variant **17**. Expected [M+3H]3+/3 = 1314.3157 Da. Observed [M+3H]3+/3 = 1314.3139 Da

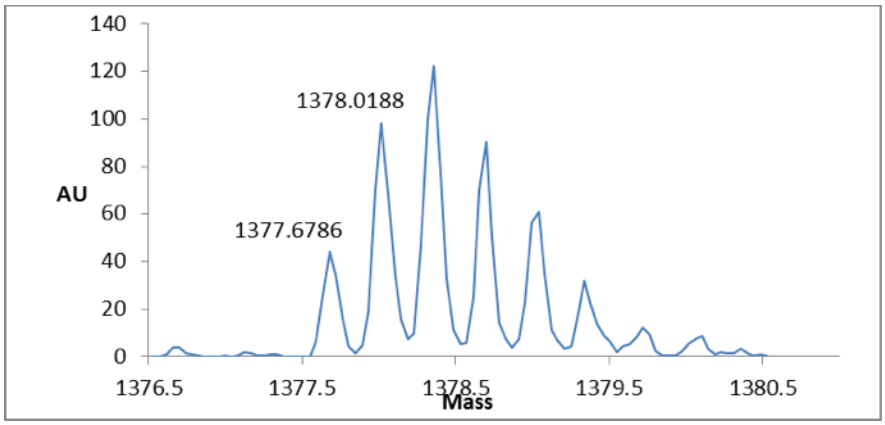


Figure 2-23. ESI TOF spectrum for WW variant **17p**. Expected $[M+3H]^{3+}/3$ = 1377.6892 Da. Observed $[M+3H]^{3+}/3$ = 1377.6786 Da.

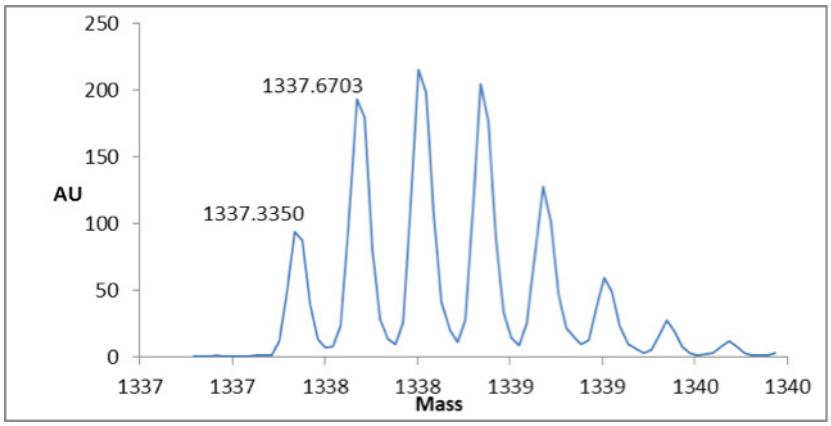


Figure 2-24. ESI TOF spectrum for WW variant **18**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3350 Da.

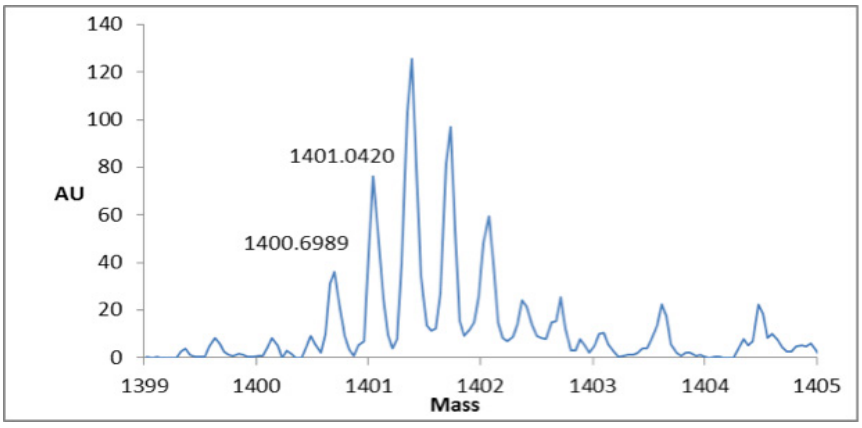


Figure 2-25. ESI TOF spectrum for WW variant **18p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.6989 Da.

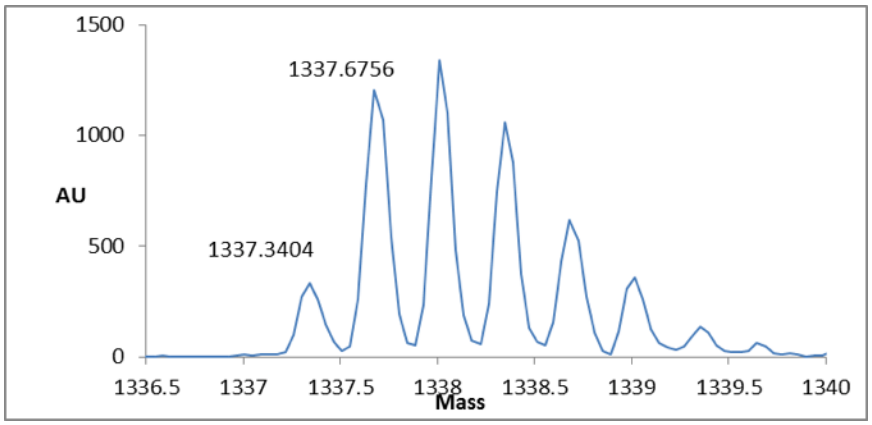


Figure 2-26. ESI TOF spectrum for WW variant **19**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1377.3404 Da.

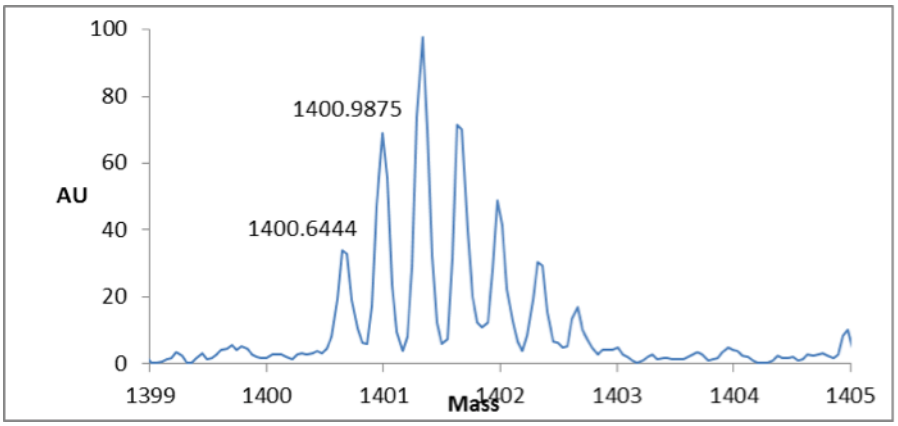


Figure 2-27. ESI TOF spectrum for WW variant **19p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.6444 Da.

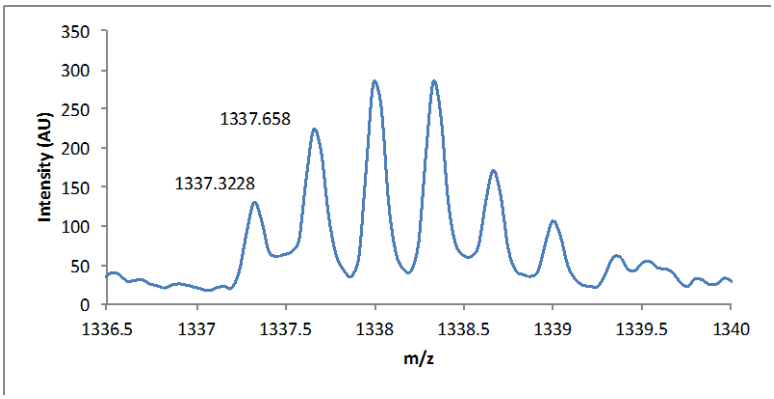


Figure 2-28. ESI TOF spectrum for WW variant **D-19**. Expected $[M+3H]^{3+}/3$ = 1337.339 Da. Observed $[M+3H]^{3+}/3$ = 1337.323 Da.

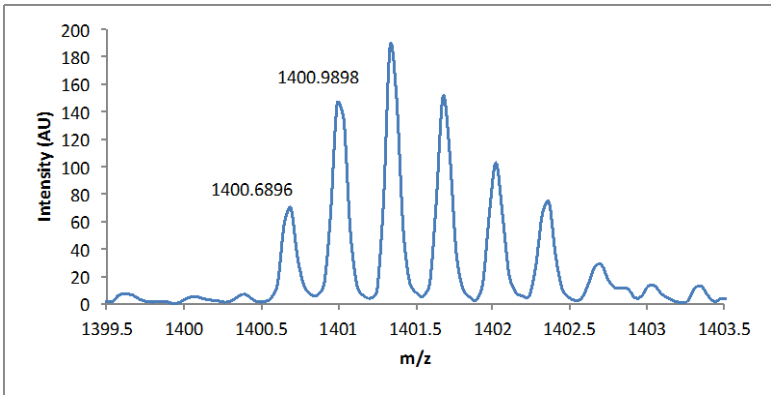


Figure 2-29. ESI TOF spectrum for WW variant **D-19p**. Expected $[M+3H]^{3+}/3$ = 1400.712 Da. Observed $[M+3H]^{3+}/3$ = 1400.690 Da.

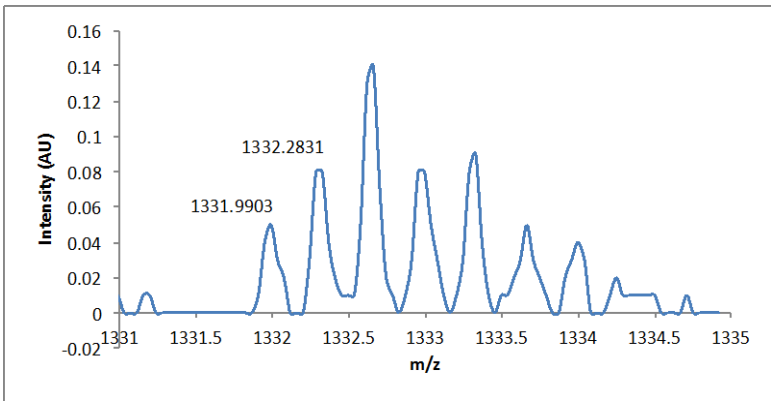


Figure 2-30. ESI TOF spectrum for WW variant **19-32A**. Expected $[M+3H]^{3+}/3$ = 1332.007 Da. Observed $[M+3H]^{3+}/3$ = 1331.990 Da.

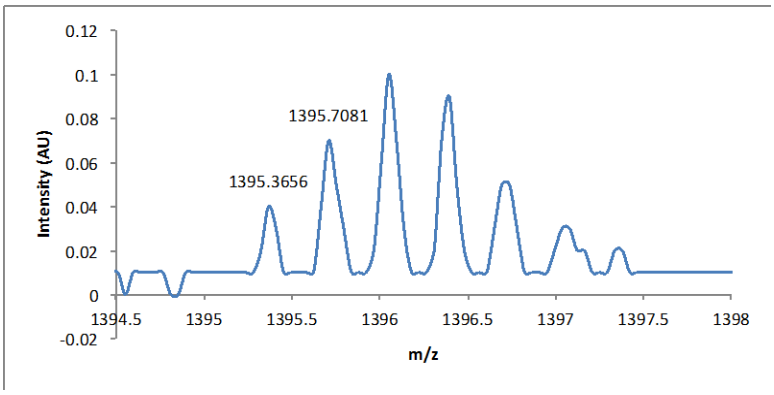


Figure 2-31. ESI TOF spectrum for WW variant **19p-32A**. Expected $[M+3H]^{3+}/3$ = 1395.381 Da. Observed $[M+3H]^{3+}/3$ = 1395.366.

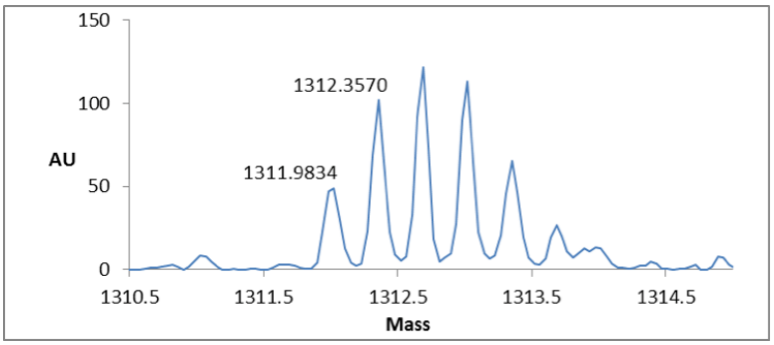


Figure 2-32. ESI TOF spectrum for WW variant **23**. Expected $[M+3H]^{3+}/3$ = 1311.9950 Da. Observed $[M+3H]^{3+}/3$ = 1311.9834 Da.

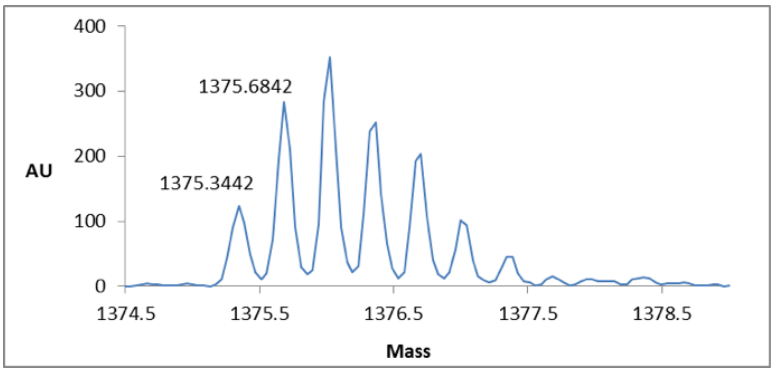


Figure 2-33. ESI TOF spectrum for WW variant **23p**. Expected $[M+3H]^{3+}/3$ = 1375.3685 Da. Observed $[M+3H]^{3+}/3$ = 1375.3442 Da.

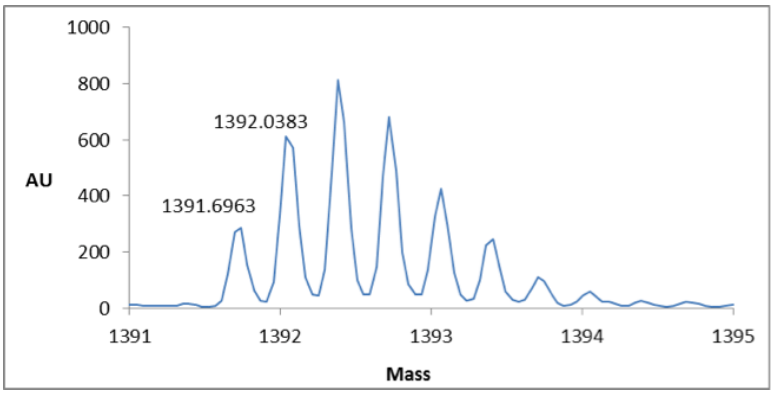


Figure 2-34. ESI TOF spectrum for WW variant **26p**. Expected $[M+3H]^{3+}/3$ = 1391.7086 Da. Observed $[M+3H]^{3+}/3$ = 1391.6963 Da.

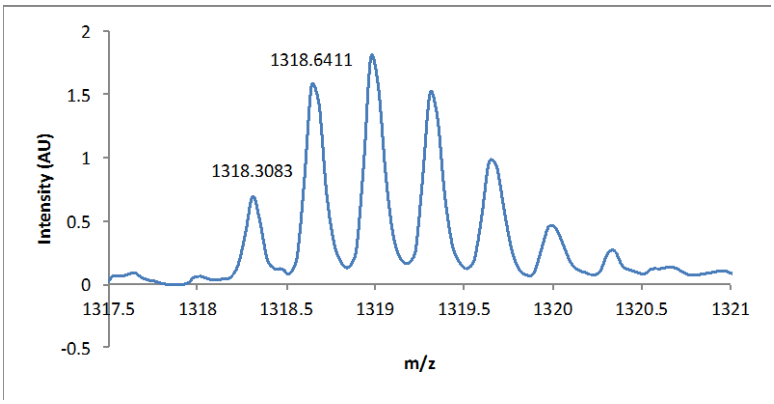


Figure 2-35. ESI TOF spectrum for WW variant **WW-T29A**. Expected $[M+3H]^{3+}/3$ = 1318.332 Da. Observed $[M+3H]^{3+}/3$ = 1318.308 Da.

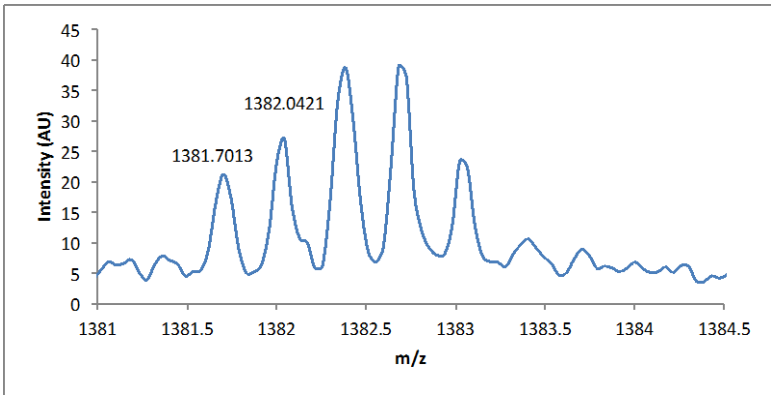


Figure 2-36. ESI TOF spectrum for WW variant **26p-T29A**. Expected $[M+3H]^{3+}/3$ = 1381.705 Da. Observed $[M+3H]^{3+}/3$ = 1381.701 Da.

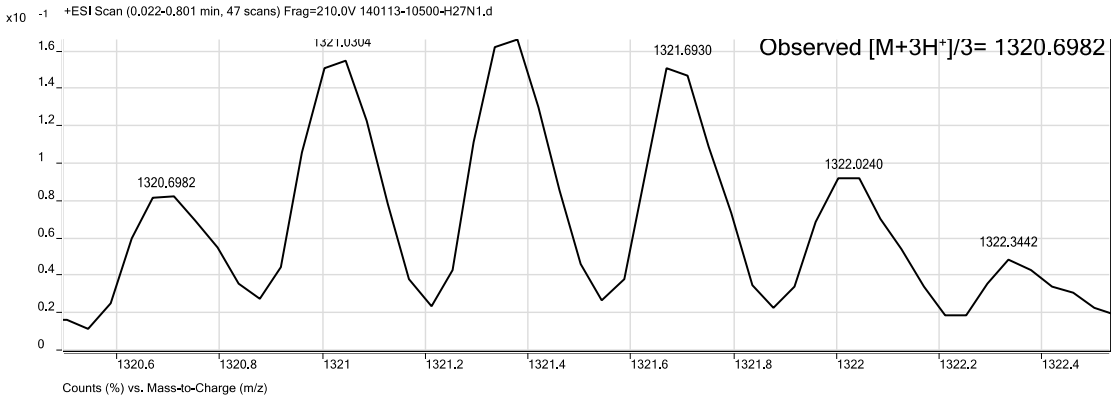


Figure 2-37. ESI TOF spectrum for WW variant **27**. Expected $[M+3H]^{3+}/3$ = 1320.6631 Da. Observed $[M+3H]^{3+}/3$ = 1320.6982 Da

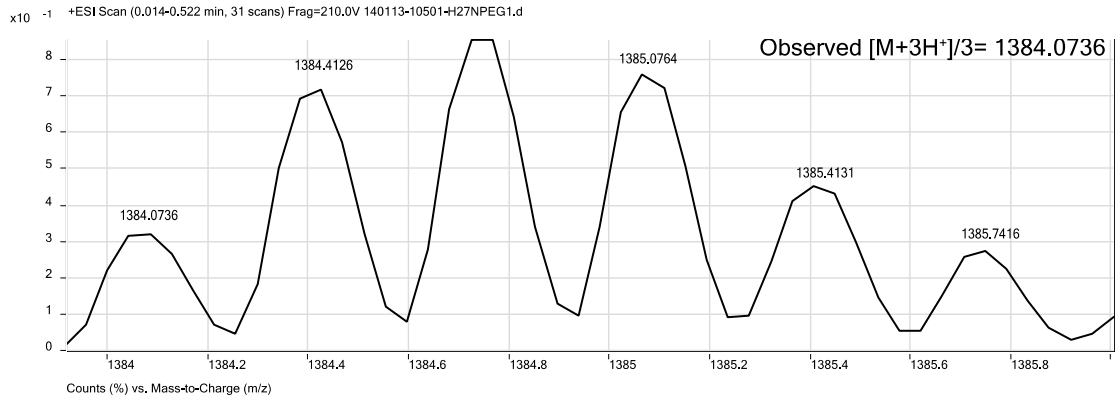


Figure 2-38. ESI TOF spectrum for WW variant **27p**. Expected $[M+3H]^{3+}/3$ = 1384.0366 Da. Observed $[M+3H]^{3+}/3$ = 1384.0736 Da.

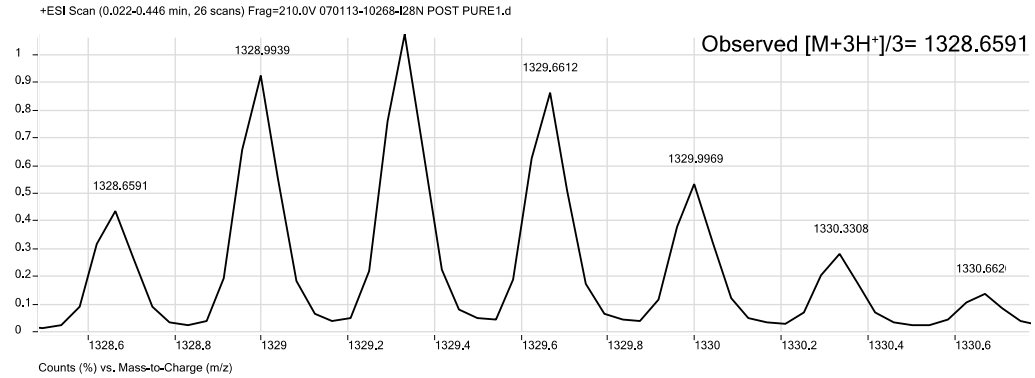


Figure 2-39. ESI TOF spectrum for WW variant **28**. Expected $[M+3H]^{3+}/3$ = 1328.6547 Da. Observed $[M+3H]^{3+}/3$ = 1328.6591 Da.

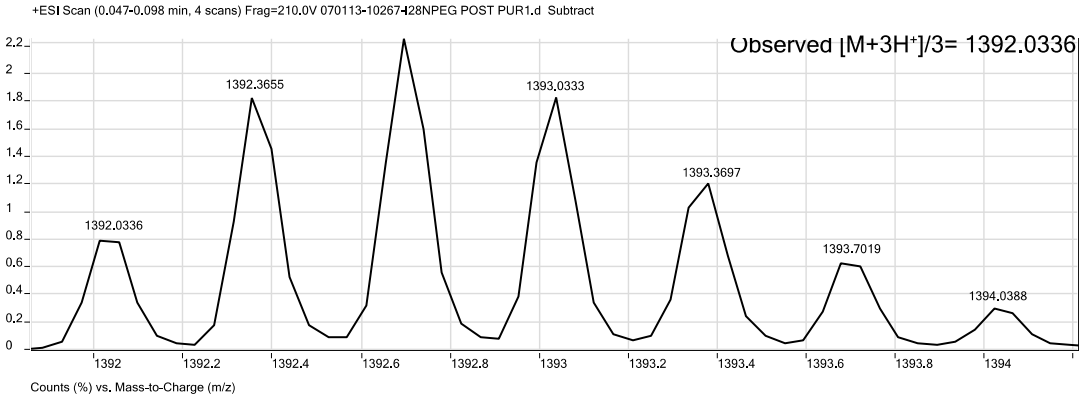


Figure 2-40. ESI TOF spectrum for WW variant **28p**. Expected $[M+3H]^{3+}/3$ = 1392.0282 Da. Observed $[M+3H]^{3+}/3$ = 1392.0336 Da.

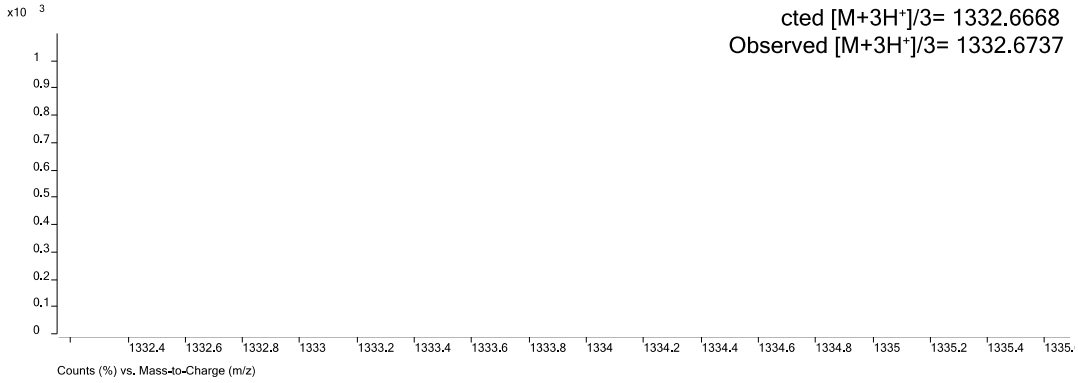


Figure 2-41. ESI TOF spectrum for WW variant **29**. Expected $[M+3H]^{3+}/3$ = 1332.6668 Da. Observed $[M+3H]^{3+}/3$ = 1332.6737 Da.

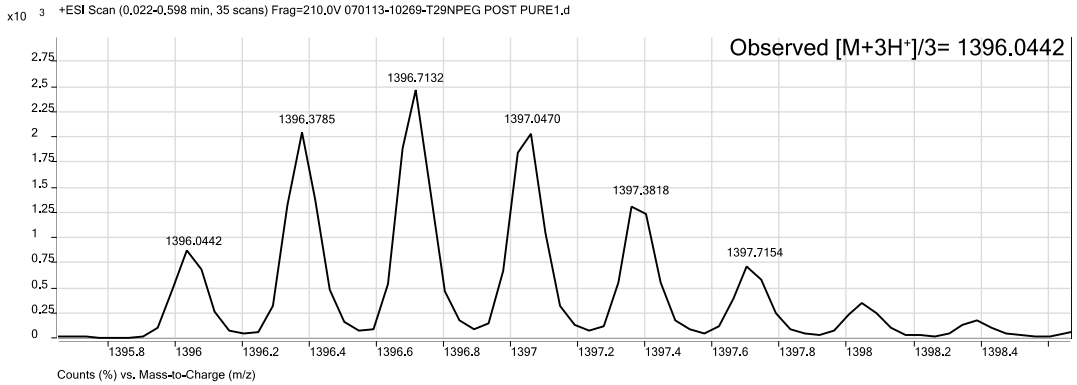


Figure 2-42. ESI TOF spectrum for WW variant **29p**. Expected $[M+3H]^{3+}/3$ = 1396.0404 Da. Observed $[M+3H]^{3+}/3$ = 1396.0446 Da.

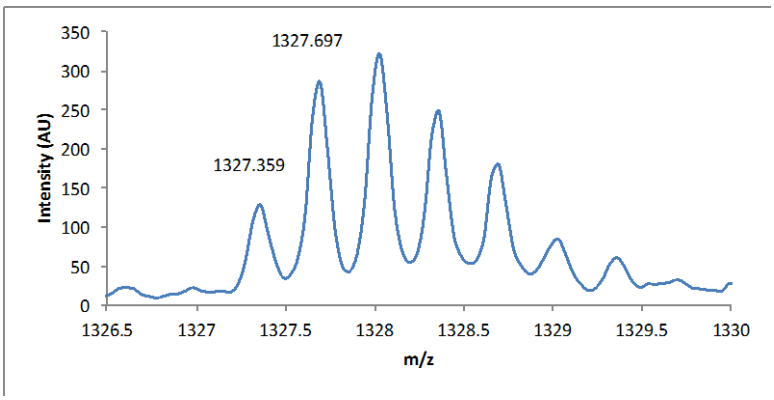


Figure 2-43.ESI TOF spectrum for WW variant **29-S32A**. Expected $[M+3H]^{3+}/3$ = 1327.355 Da. Observed $[M+3H]^{3+}/3$ = 1327.359 Da.

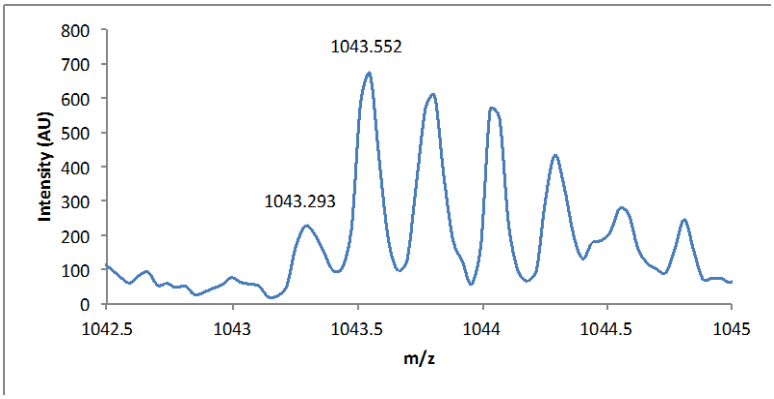


Figure 2-44. ESI TOF spectrum for WW variant **29p-S32A**. Expected $[M+4H]^{4+}/4$ = 1043.284 Da. Observed $[M+4H]^{4+}/4$ = 1043.293 Da.

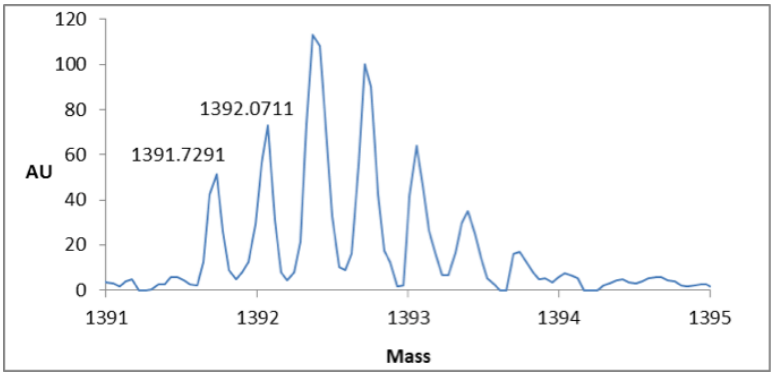


Figure 2-45. ESI TOF spectrum for WW variant **30p**. Expected $[M+3H]^{3+}/3$ = 1391.7086 Da. Observed $[M+3H]^{3+}/3$ = 1391.7291 Da.

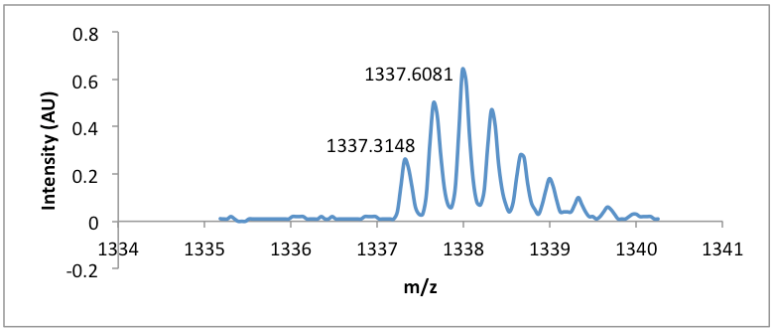


Figure 2-46. ESI TOF spectrum for WW variant **32**. Expected $[M+3H]^{3+}/3$ = 1337.3387 Da. Observed $[M+3H]^{3+}/3$ = 1337.3148 Da.

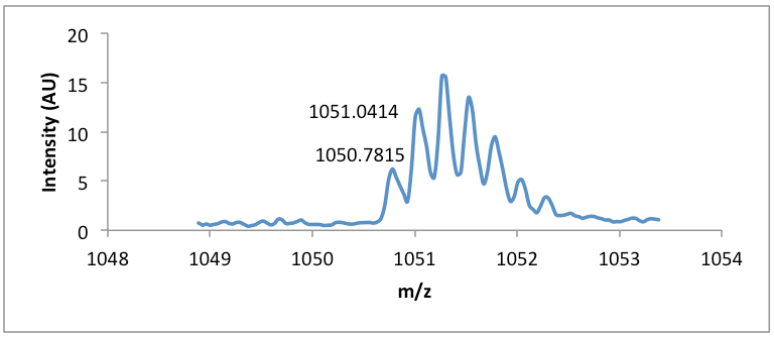


Figure 2-47. ESI TOF spectrum for WW variant **32p**. Expected $[M+4H]^{4+}/4$ = 1050.7861 Da. Observed $[M+4H]^{4+}/4$ = 1050.7815 Da.

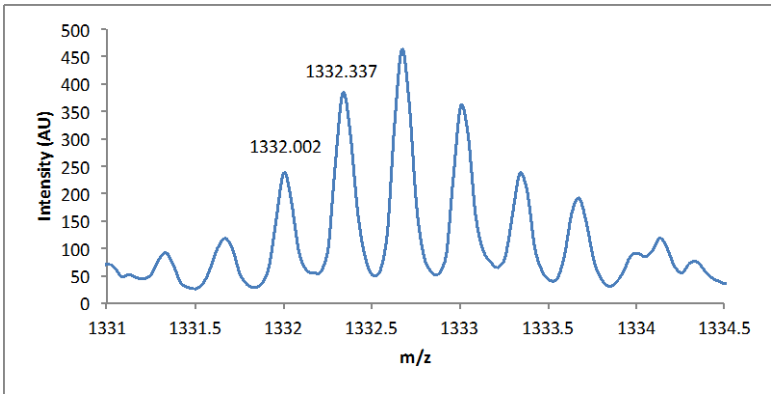


Figure 2-48. ESI TOF spectrum for WW variant **32-Y23F**. Expected $[M+3H]^{3+}/3$ = 1332.007 Da. Observed $[M+3H]^{3+}/3$ = 1332.002 Da

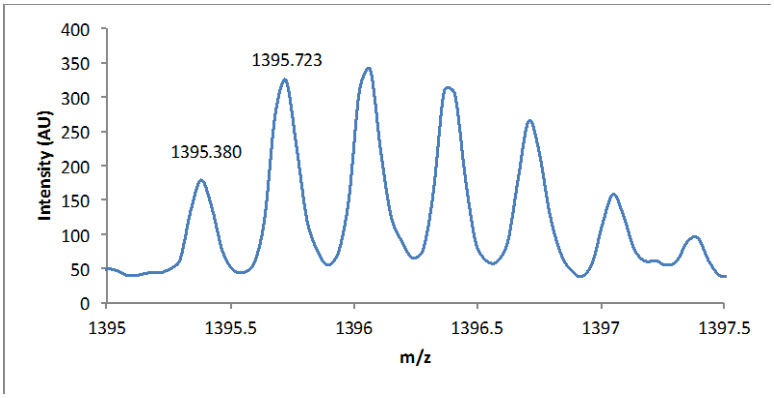


Figure 2-49. ESI TOF spectrum for WW variant **32p-Y23F**. Expected $[M+3H]^{3+}/3$ = 1395.381 Da. Observed $[M+3H]^{3+}/3$ = 1395.380 Da.

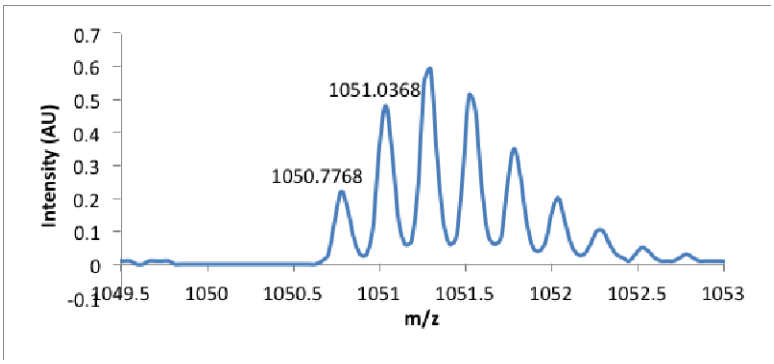


Figure 2-50. ESI-TOF spectrum for WW variant **16/26p**. Expected $[M+4H]^{4+}/4$ = 1050.7861 Da. Observed $[M+4H]^{4+}/4$ = 1050.7768 Da.

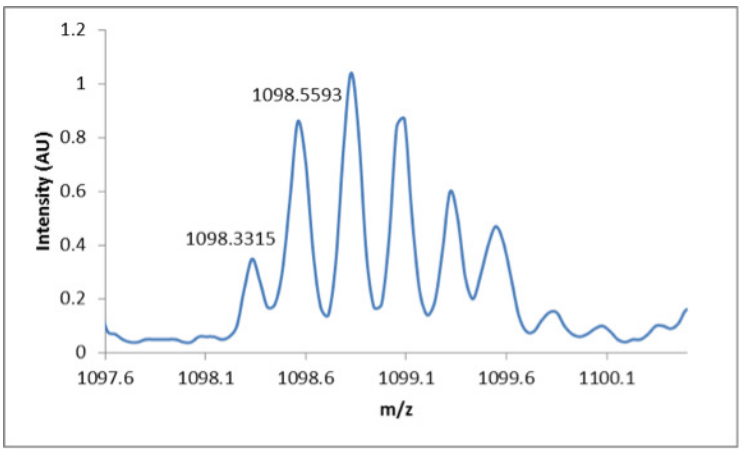


Figure 2-51. ESI TOF spectrum for doubly PEGylated WW variant **16p/26p**. Expected $[M+4H]^{4+}/4$ = 1098.3163 Da. Observed $[M+4H]^{4+}/4$ = 1098.3315 Da.

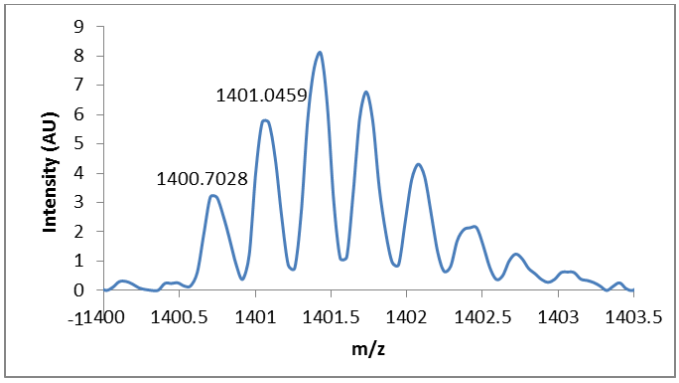


Figure 2-52. ESI-TOF spectrum for WW variant **19/26p**. Expected $[M+3H]^{3+}/3$ = 1400.7123 Da. Observed $[M+3H]^{3+}/3$ = 1400.7028 Da.

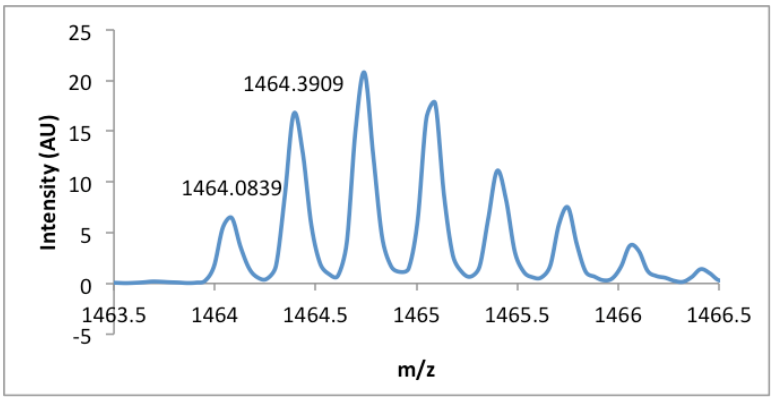


Figure 2-53. ESI TOF spectrum for WW variant **19p/26p**. Expected $[M+3H]^{3+}/3$ = 1464.0858 Da. Observed $[M+3H]^{3+}/3$ = 1464.0839 Da.

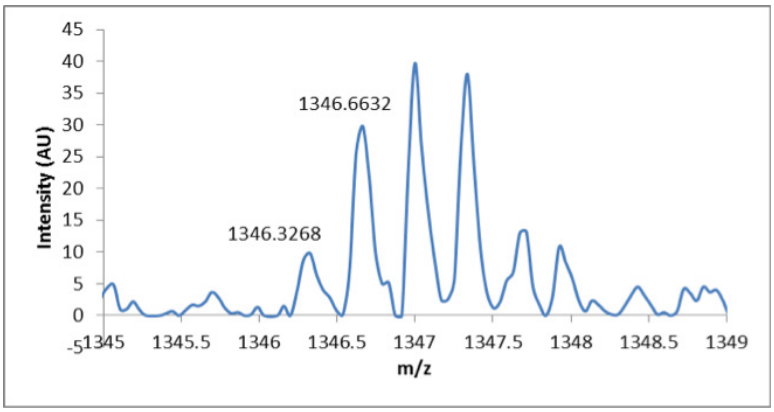


Figure 2-54. ESI TOF spectrum for WW variant **16/19**. Expected $[M+3H]^{3+}/3$ = 1346.3424 Da. Observed $[M+3H]^{3+}/3$ = 1346.3268 Da.

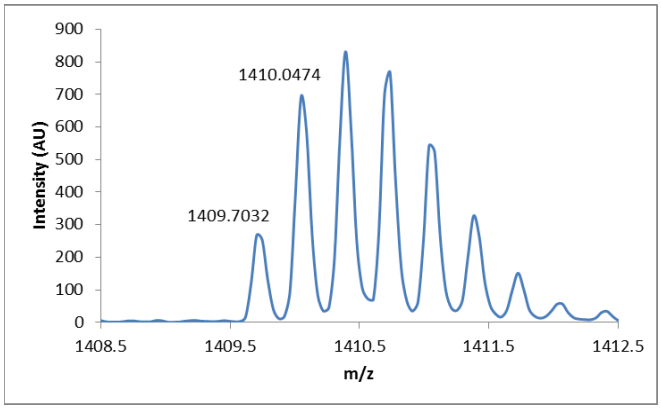


Figure 2-55. ESI-TOF spectrum for WW variant **16p/19**. Expected $[M+3H]^{3+}/3$ = 1409.7159 Da. Observed $[M+3H]^{3+}/3$ = 1409.7032 Da.

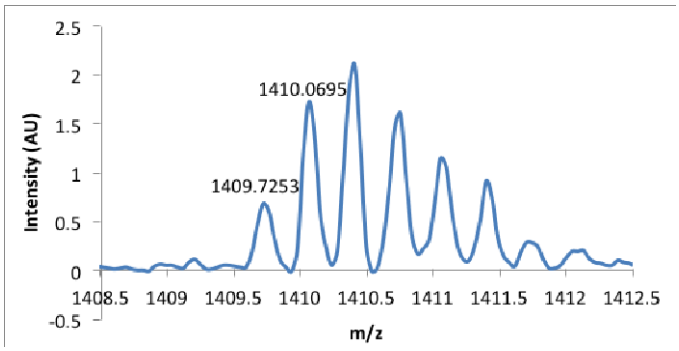


Figure 2-56. ESI-TOF spectrum for WW variant **16/19p**. Expected $[M+3H]^{3+}/3$ = 1409.7159 Da. Observed $[M+3H]^{3+}/3$ = 1409.7253 Da.

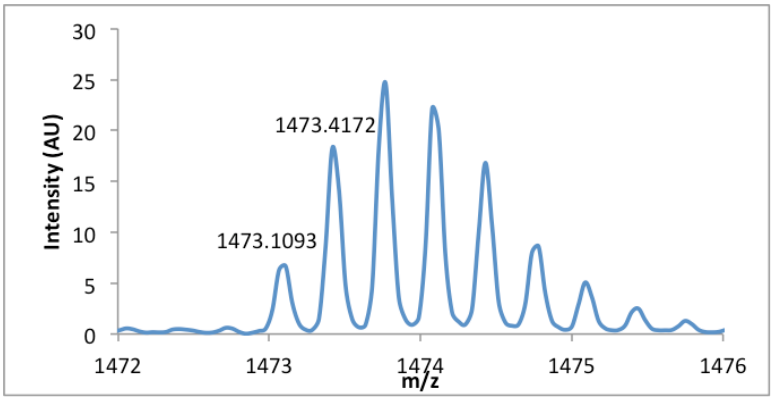


Figure 2-57. ESI TOF spectrum for doubly PEGylated WW variant **16p/19p**. Expected $[M+3H]^{3+}/3$ = 1473.0894 Da. Observed $[M+3H]^{3+}/3$ = 1473.1093 Da.

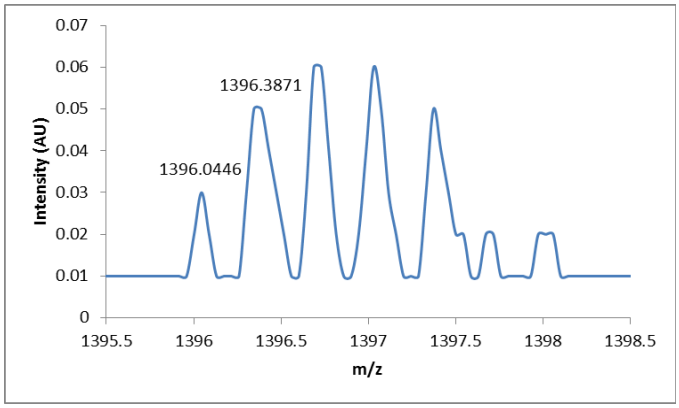


Figure 2-58. ESI-TOF spectrum for WW variant **26p/29**. Expected $[M+3H]^{3+}/3$ = 1396.0404 Da. Observed $[M+3H]^{3+}/3$ = 1396.0446 Da.

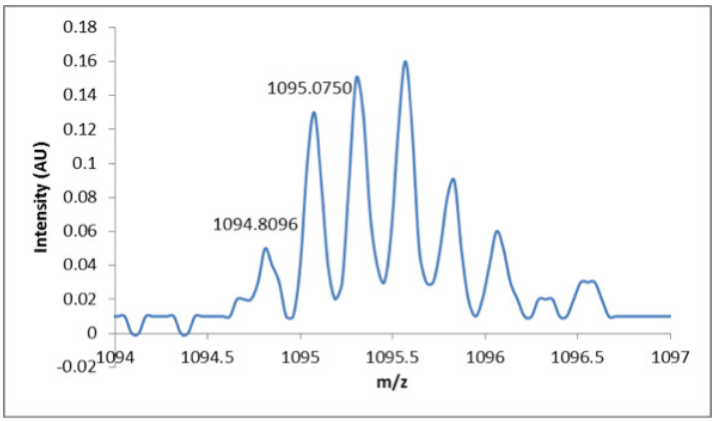


Figure 2-59. ESI TOF spectrum for doubly PEGylated WW variant **26p/29p**. Expected $[M+4H]^{4+}/4$ = 1094.8124Da. Observed $[M+4H]^{4+}/4$ = 1094.8096.

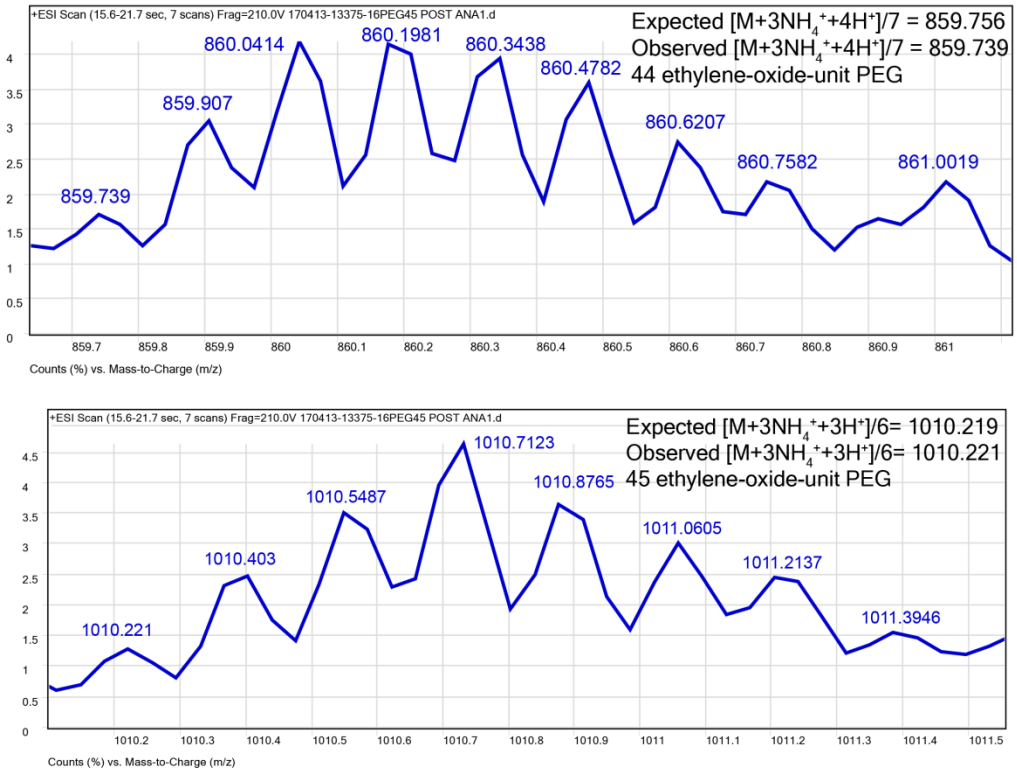


Figure 2-60. ESI TOF spectra for WW variant **16p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide units.

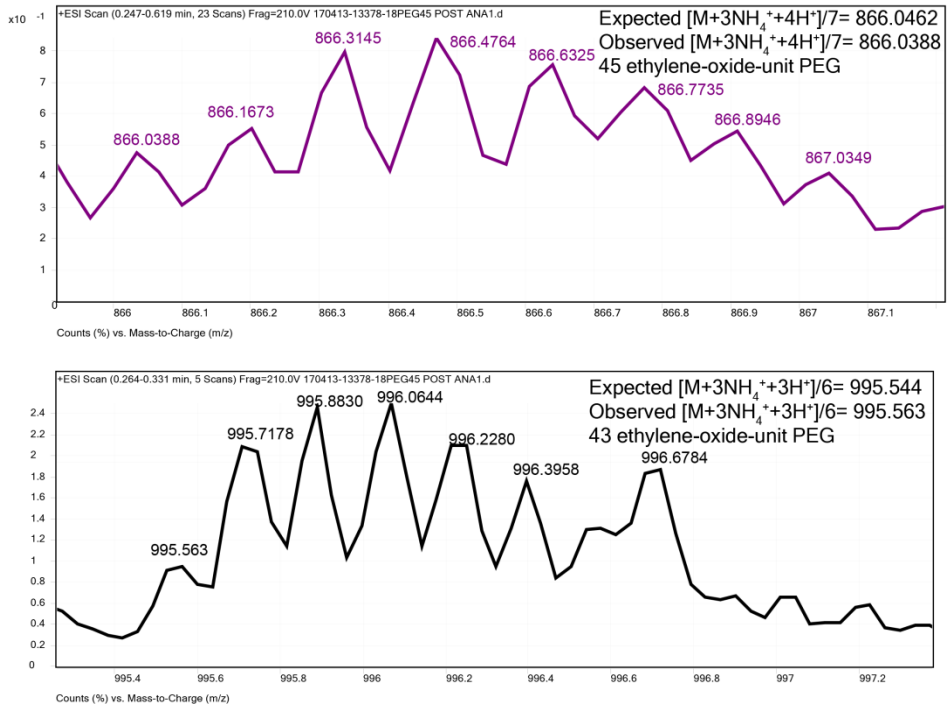


Figure 2-61. ESI TOF spectra for WW variant **18p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide unit

x

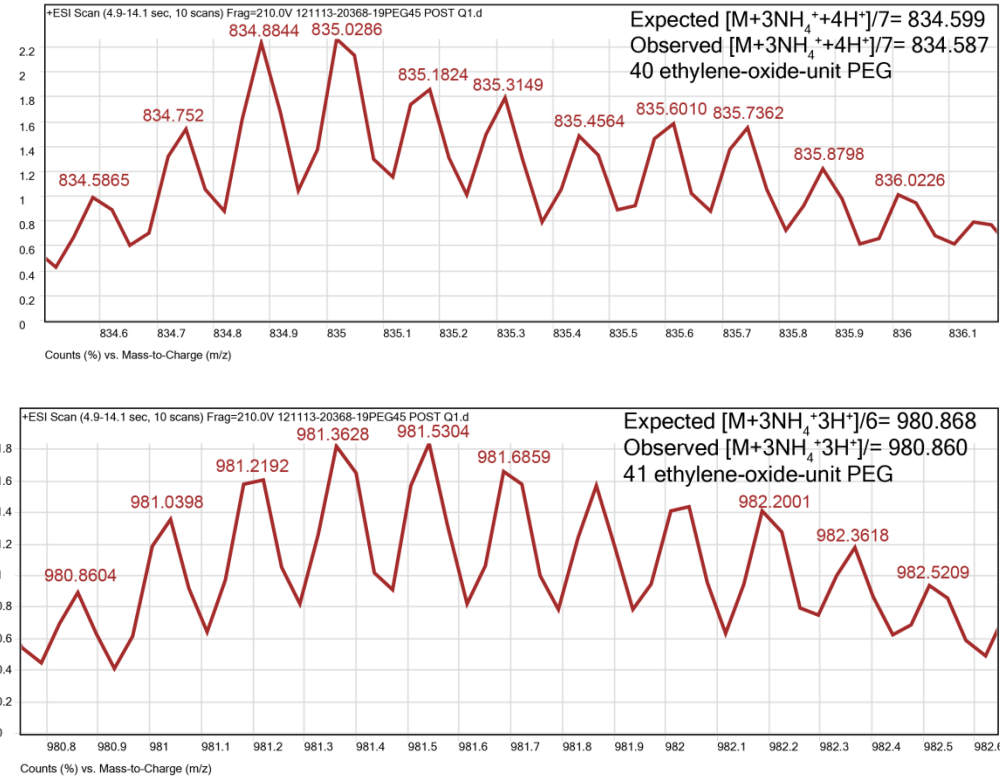


Figure 2-62. ESI TOF spectra for WW variant **19p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide units.

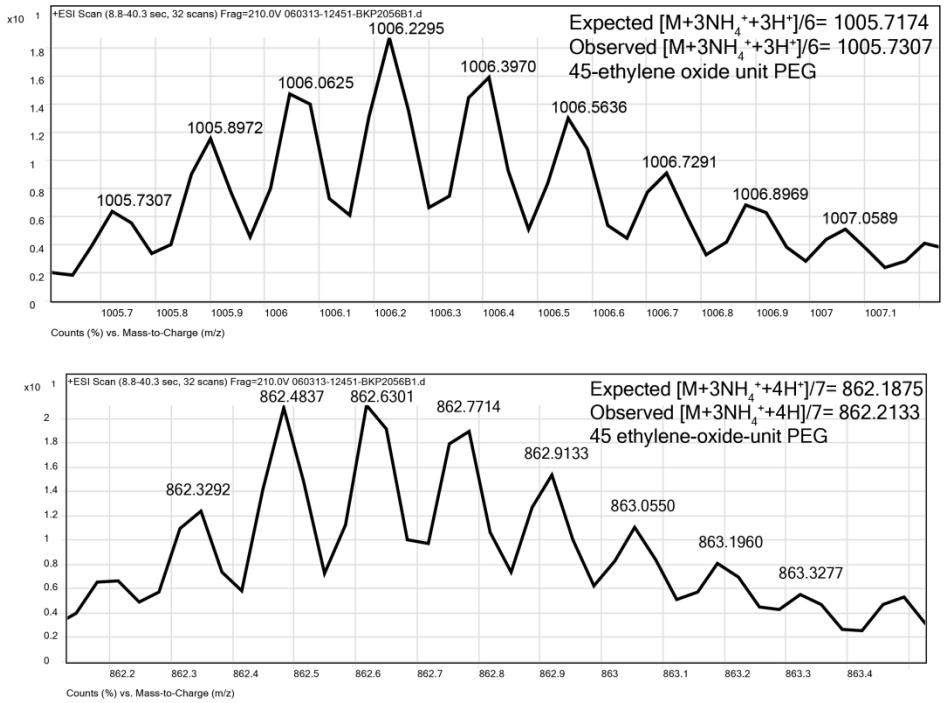


Figure 2-63. ESI TOF spectra for WW variant **26p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide units.

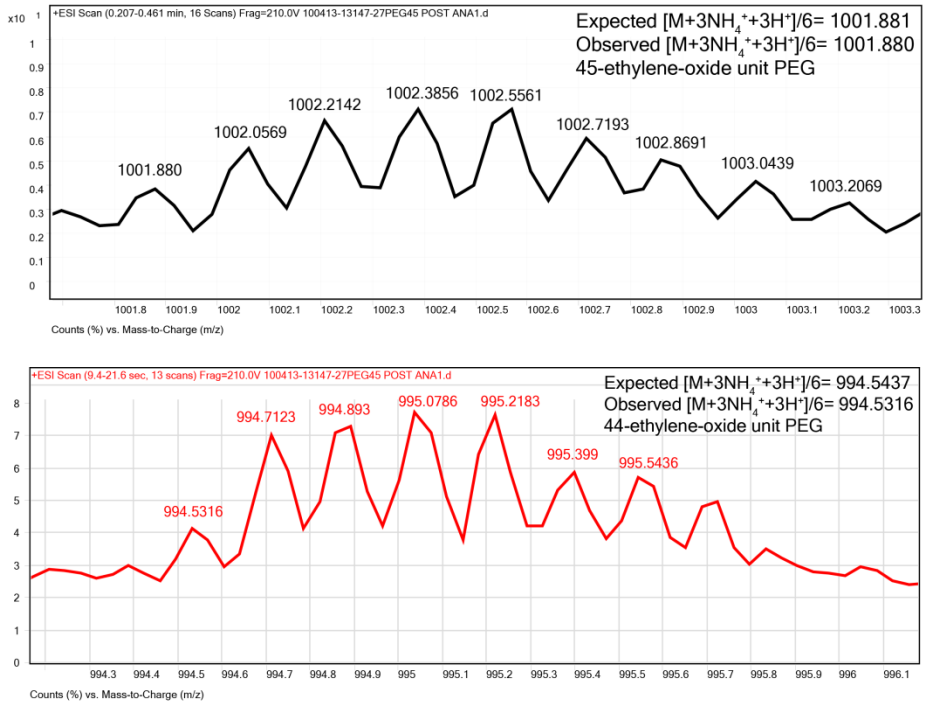


Figure 2-64. ESI TOF spectra for WW variant **27p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide units.

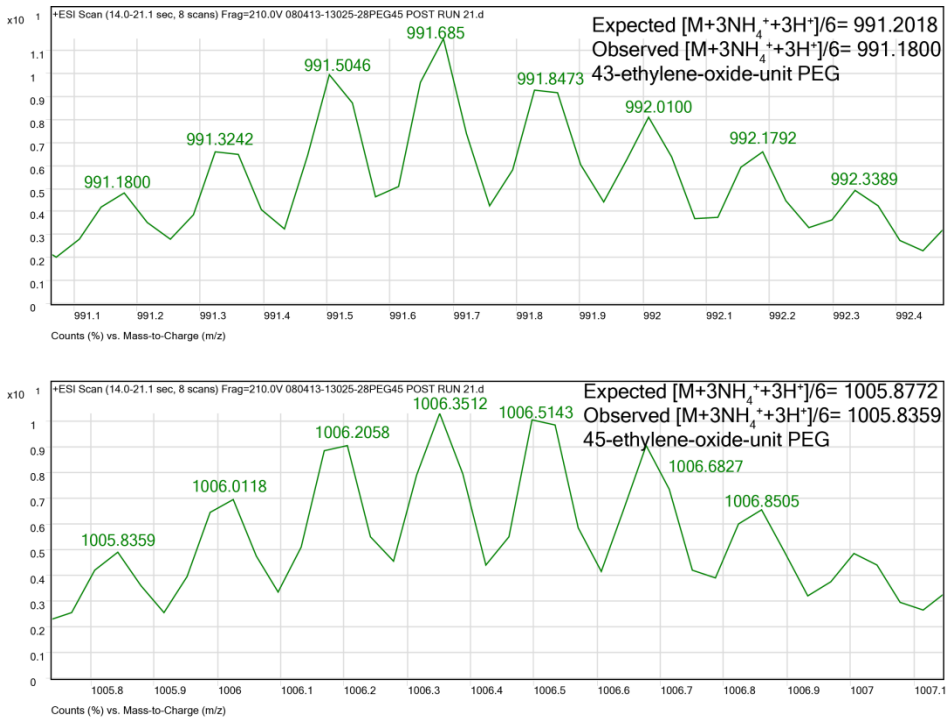


Figure 2-65. ESI TOF spectra for WW variant **28p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide units.

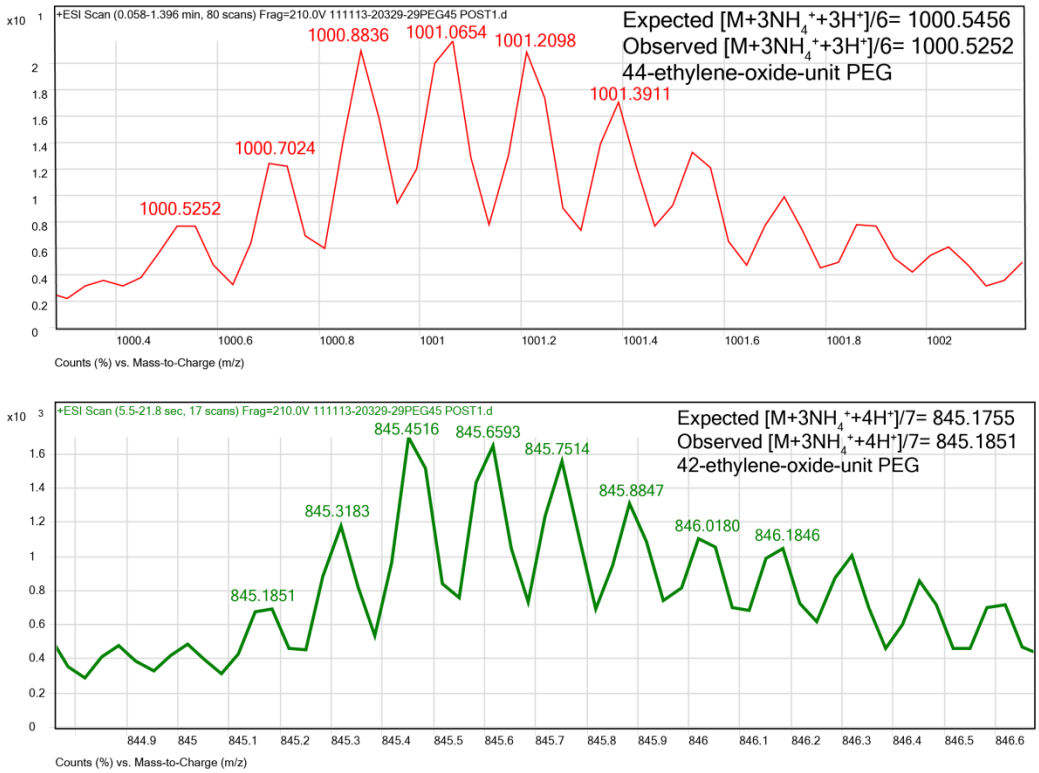


Figure 2-64. ESI TOF spectra for WW variant **29p45**, which contains a polydisperse PEG oligomer with an average of ~45 ethylene oxide units.

2.4.5 Analytical HPLC data

HPLC traces for proteins**WW, 14**, **14p**, **16**, **16p**,**16-Y23F**, **16p-Y23F**, **16-32A**, **16p-32A**,**17**, **17p**, **18**, **18p**, **19**, **19p**,**D-19**,**D-19p**, **19-32A**, **19p-32A**, **23**, **23p**,**26p**,**WW-T29A**, **26p-T29A**,**27**, **27p**, **28**, **28p**, **29**, **29p**,**29-S32A**, **29p-S32A**,**30p**, **32**, **32p**,**32-Y23F**, **32p-Y23F**,**16/26p**, **16p/26p**, **19/26p**, **19p/26p**, **16/19**, **16/19p**, **16p/19**, **16p/19p**, **26p/29**, **26p/29p**, **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45**are shown in Figures2-65 through 2-118.

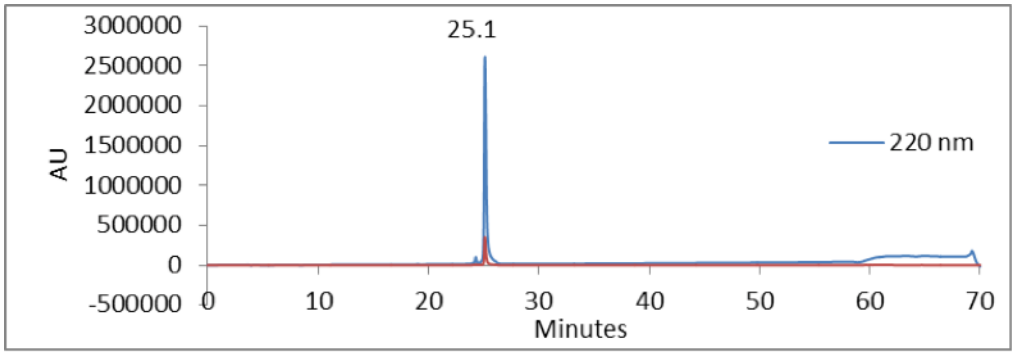


Figure 2-65. Analytical HPLC Data for parent protein **WW**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration

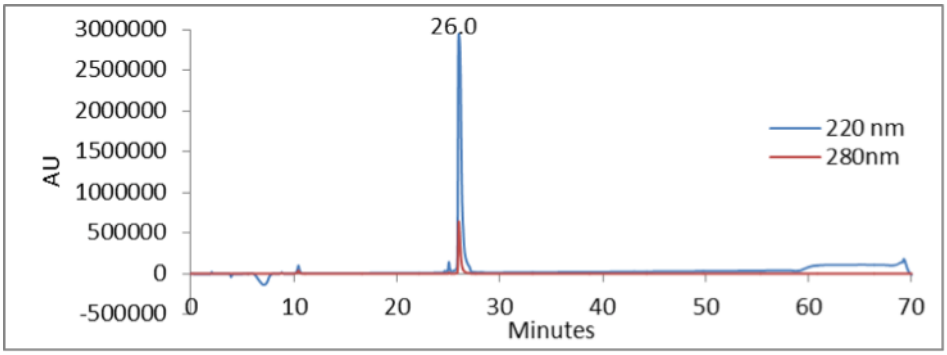


Figure 2-66. Analytical HPLC Data for WW variant **14**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

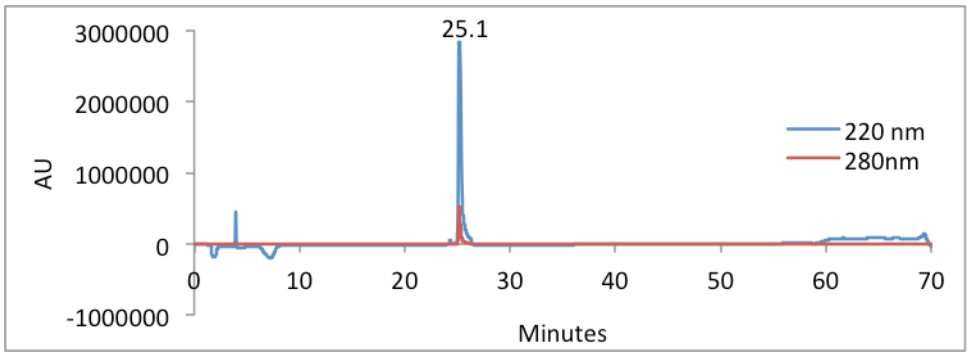


Figure 2-67. Analytical HPLC Data for WW variant **14p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

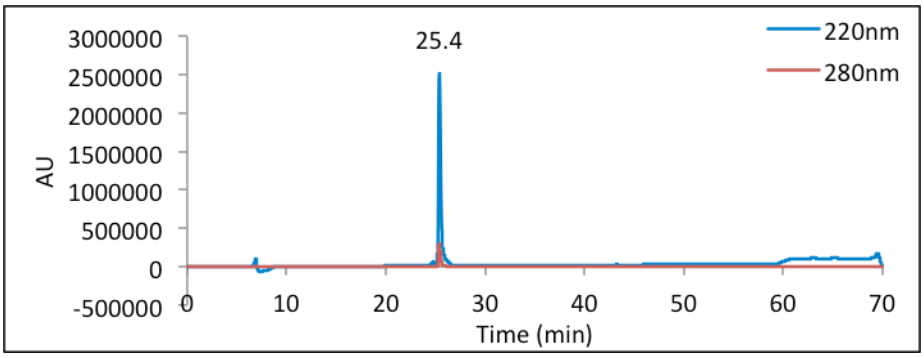


Figure 2-68. Analytical HPLC Data for WW variant **16**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

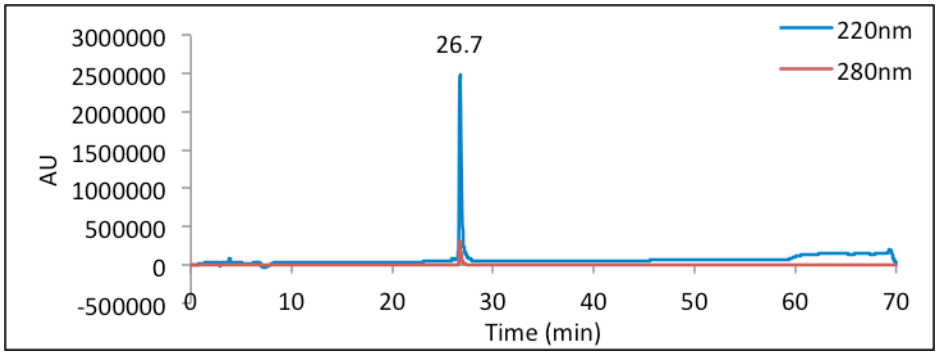


Figure 2-69. Analytical HPLC Data for WW variant **16p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

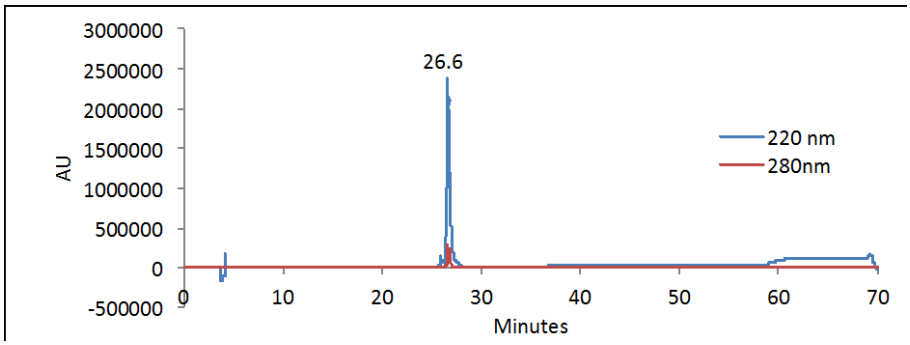


Figure 2-70. Analytical HPLC Data for WW variant **16-Y23F**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

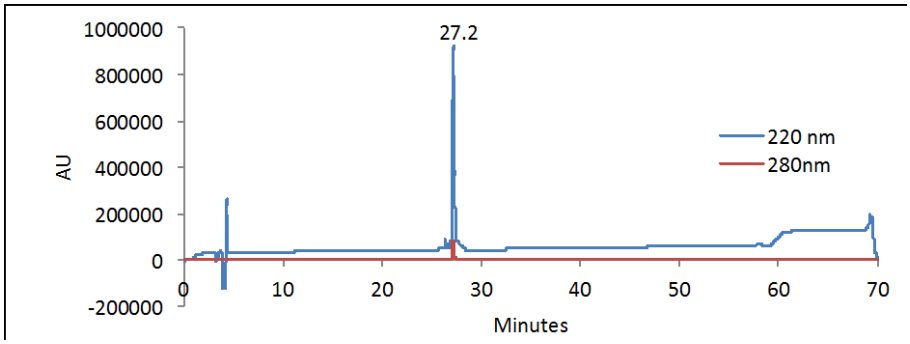


Figure 2-71. Analytical HPLC Data for WW variant **16p-Y23F**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

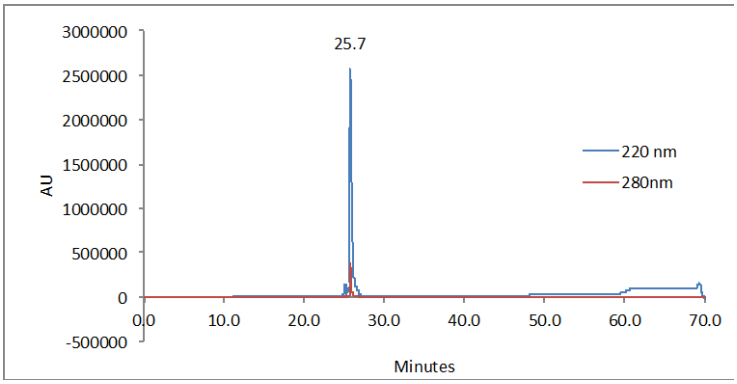


Figure 2-72. Analytical HPLC Data for WW variant **16-S32A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

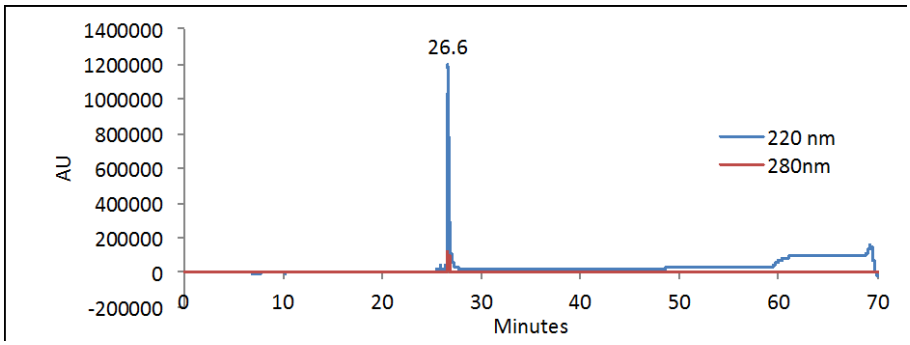


Figure 2-73. Analytical HPLC Data for WW variant **16p-S32A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

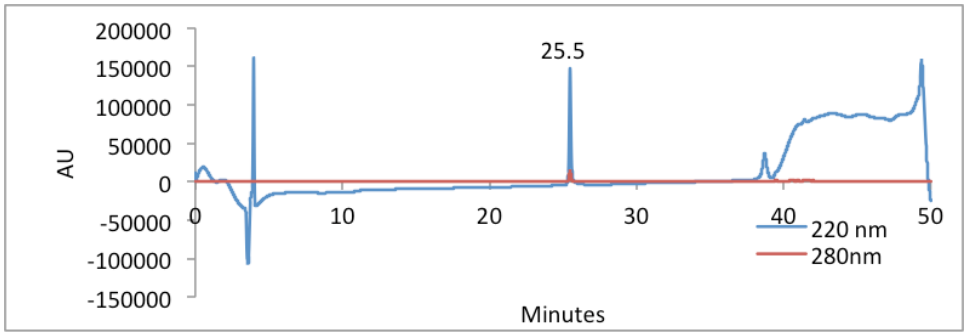


Figure 2-74. Analytical HPLC Data for WW variant **17**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration

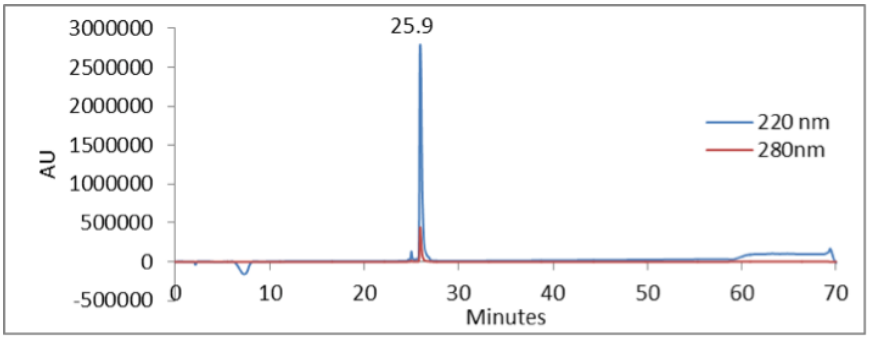


Figure 2-75. Analytical HPLC Data for WW variant **17p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

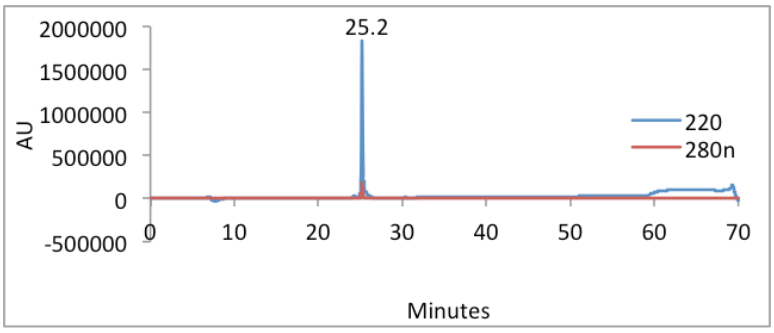


Figure 2-76Analytical HPLC Data for WW variant **18**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

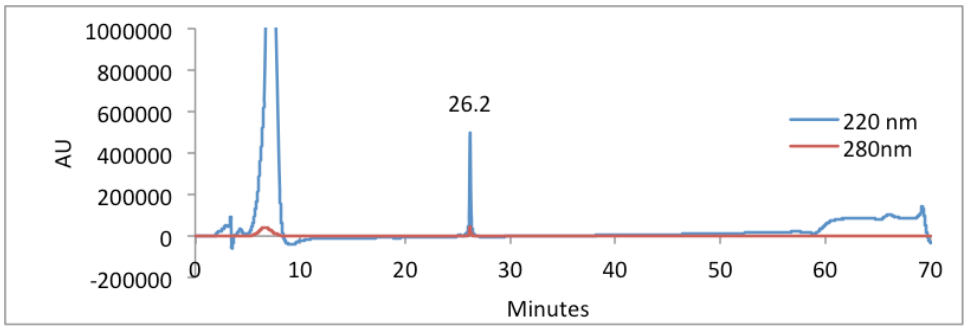


Figure 2-77. Analytical HPLC Data for WW variant **18p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

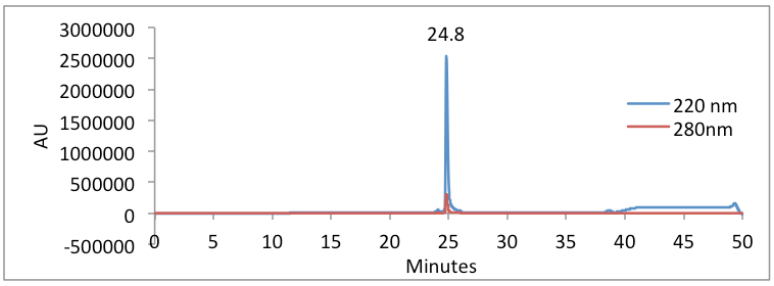


Figure 2-78. Analytical HPLC Data for WW variant **19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

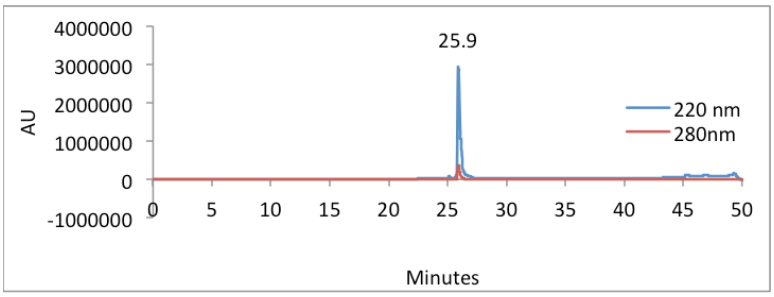


Figure 2-79.Analytical HPLC Data for WW variant **19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H9XO, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated early at 50 min.

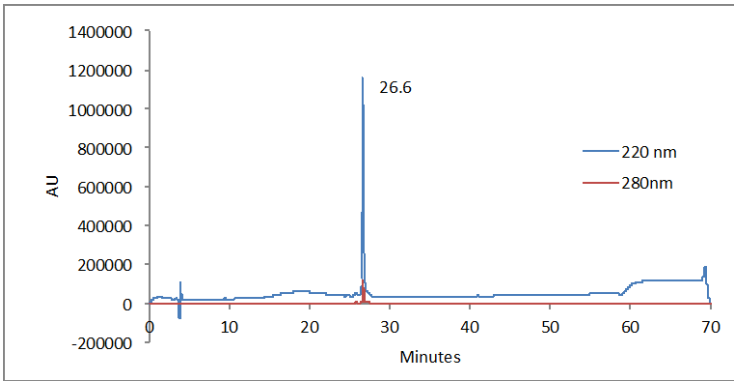


Figure 2-80. Analytical HPLC Data for WW variant **D-19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

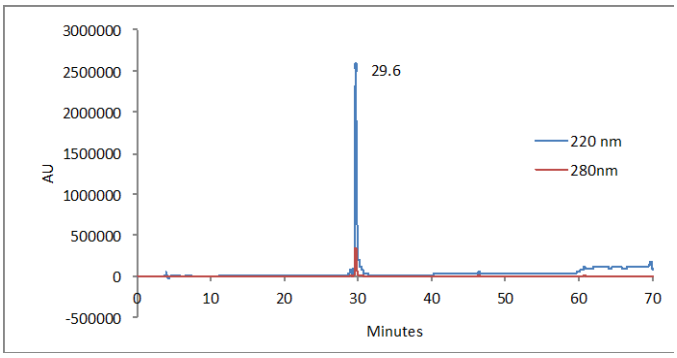


Figure 2-81. Analytical HPLC Data for WW variant **D-19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

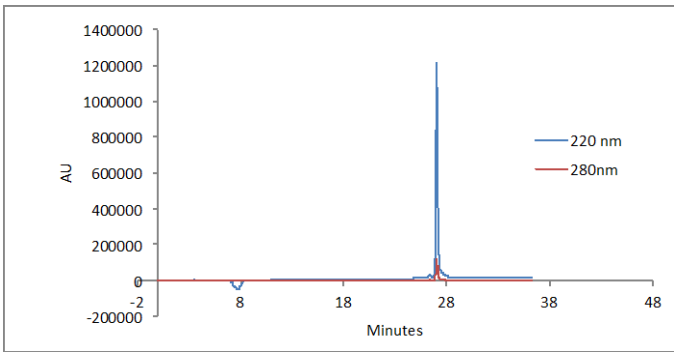


Figure 2-82.Analytical HPLC Data for WW variant **19-S32A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient intended to run from 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated several minutes after the desired peak eluted.

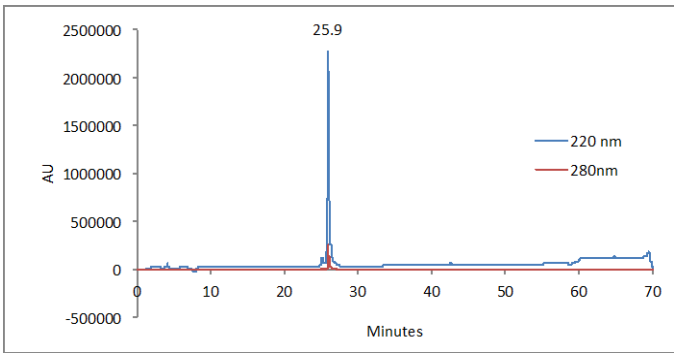


Figure 2-83. Analytical HPLC Data for WW variant **19p-S32A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min

.

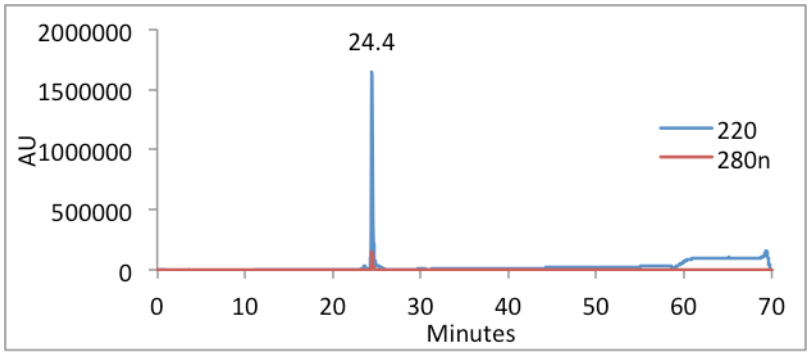


Figure 2-84. Analytical HPLC Data for WW variant **23**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

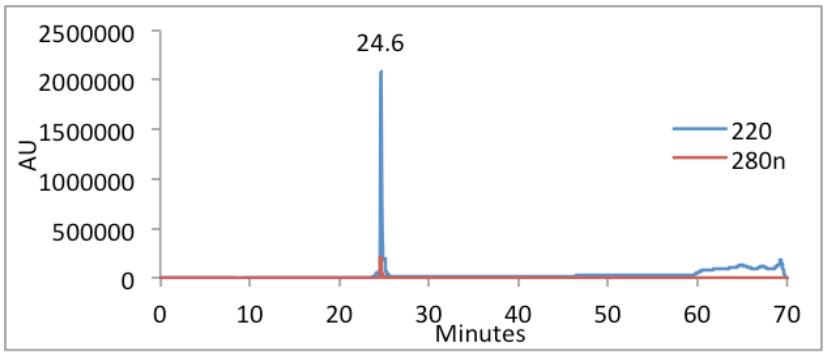


Figure 2-85. Analytical HPLC Data for WW variant **23p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

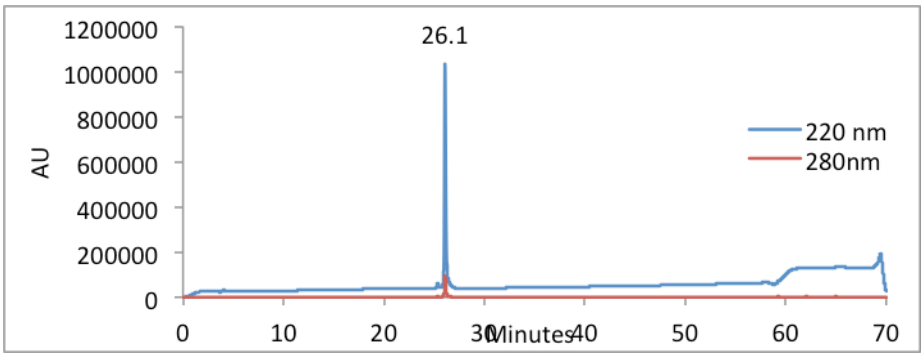


Figure 2-86. Analytical HPLC Data for WW variant **26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

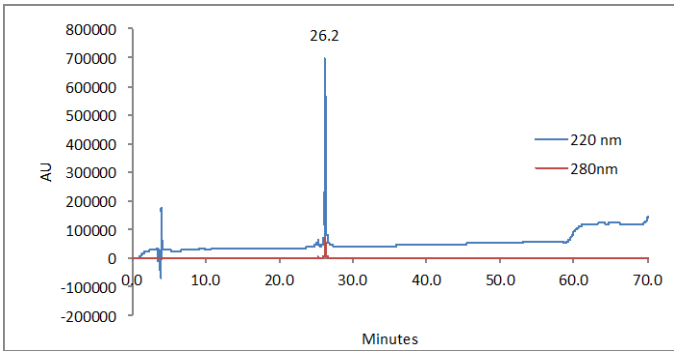


Figure 2-87. Analytical HPLC Data for WW variant **WW-T29A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

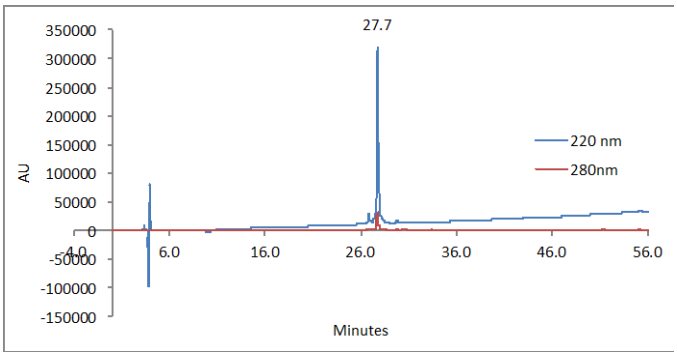


Figure 2-88. Analytical HPLC Data for WW variant **26p-T29A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

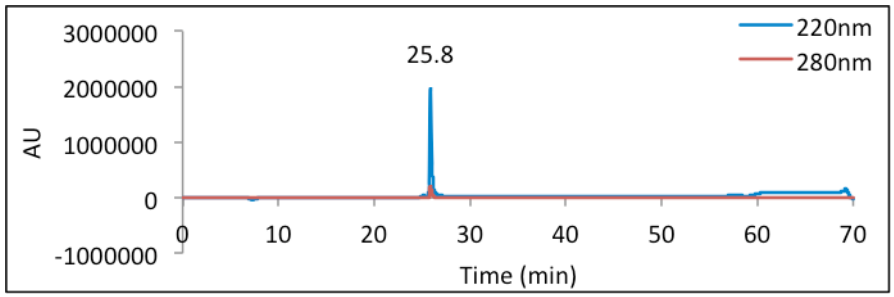


Figure 2-89.Analytical HPLC Data for WW variant **27**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

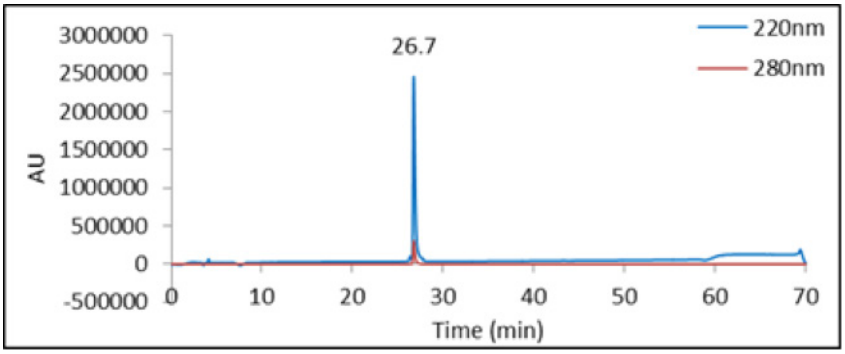


Figure 2-90. Analytical HPLC Data for WW variant **27p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

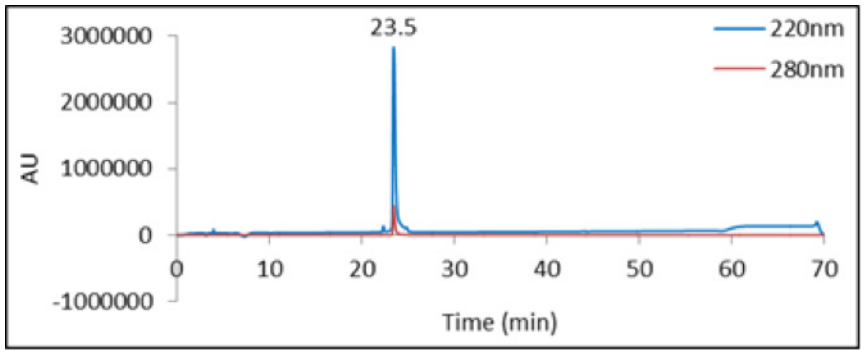


Figure 2-91. Analytical HPLC Data for WW variant **28**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

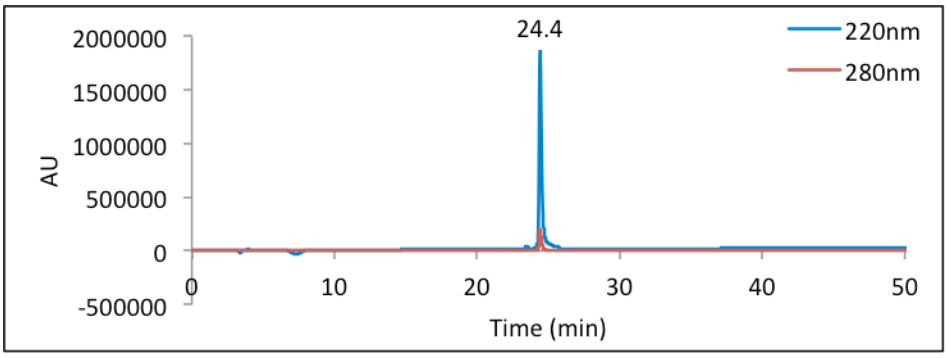


Figure 2-92.Analytical HPLC Data for WW variant **28p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated early at 50 min.

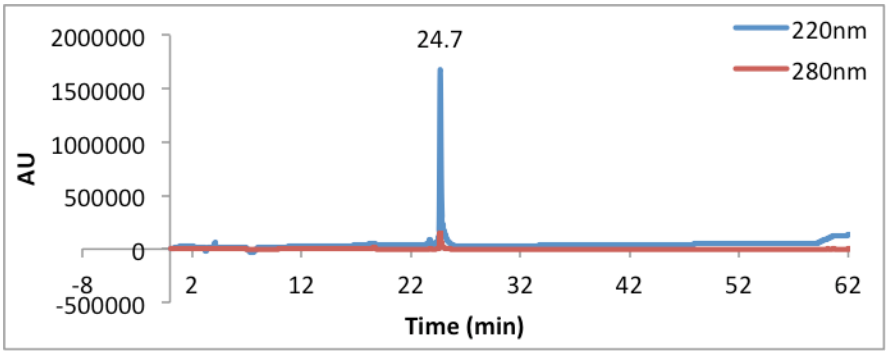


Figure 2-93. Analytical HPLC Data for WW variant **29**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated early at ~62 min.

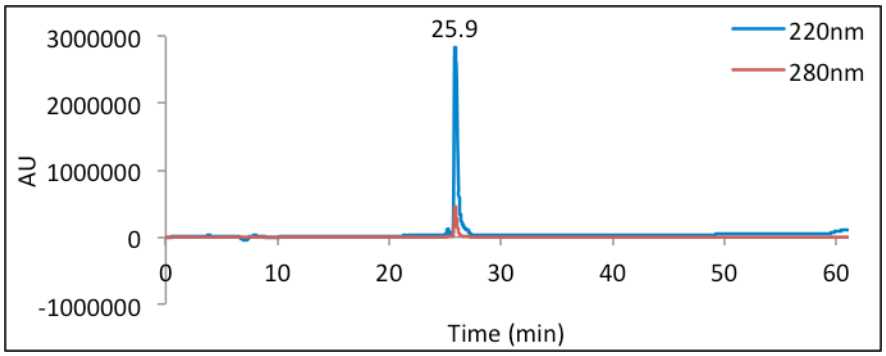


Figure 2-94.Analytical HPLC Data for WW variant **29p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated early at ~61 min.

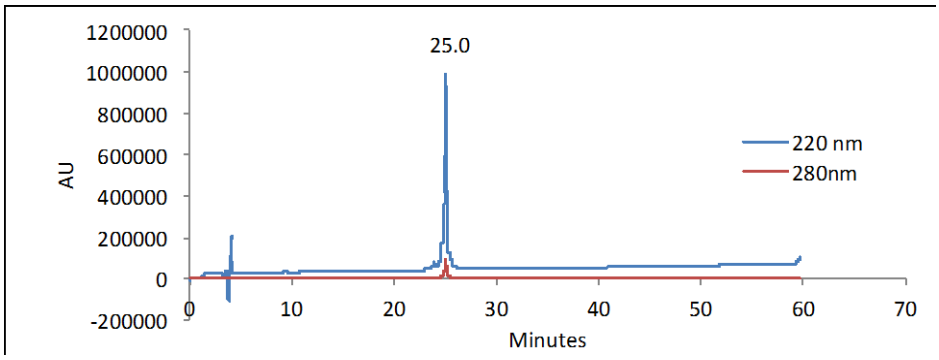


Figure 2-95. Analytical HPLC Data for WW variant **29-S32A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated prematurely at 60 min.

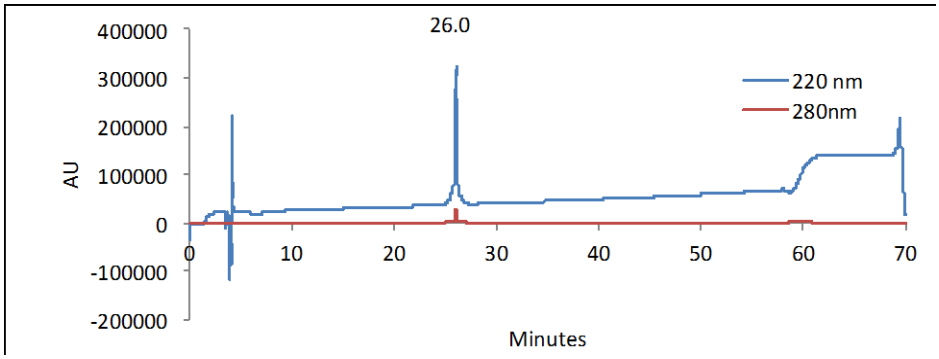


Figure 2-96. Analytical HPLC Data for WW variant **29p-S32A**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

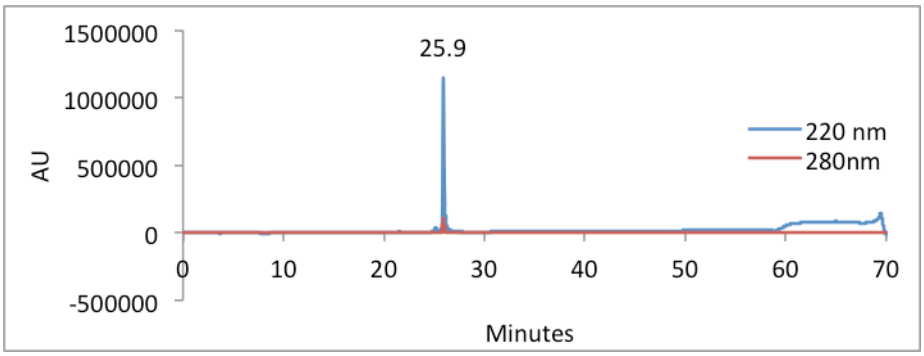


Figure 2-97. Analytical HPLC Data for WW variant **30p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

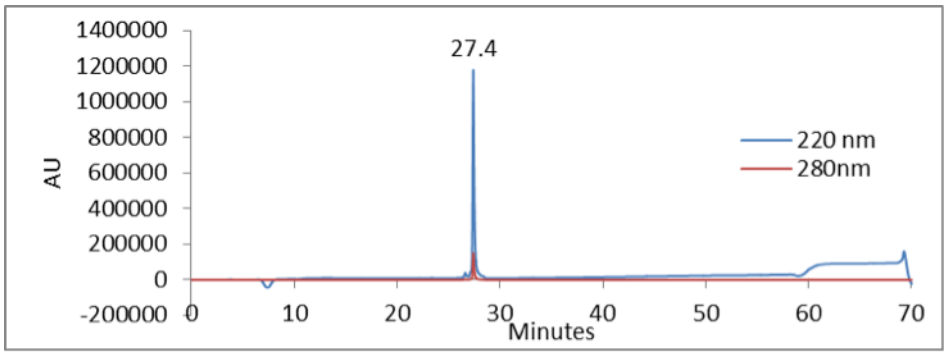


Figure 2-98. Analytical HPLC Data for WW variant **32**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

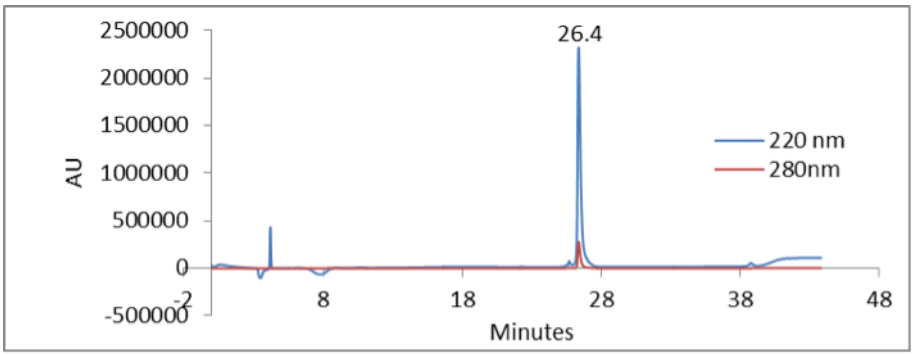


Figure 2-99.Analytical HPLC Data for WW variant **32p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

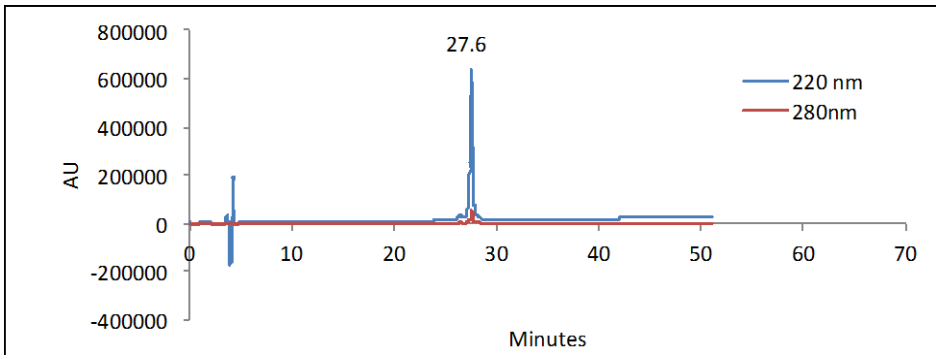


Figure 2-100.Analytical HPLC Data for WW variant **32-Y23F**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. This run was terminated prematurely at 51 min

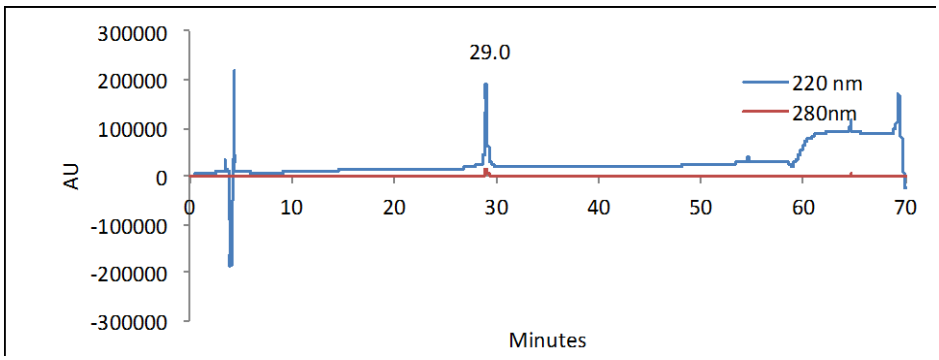


Figure 2-101.Analytical HPLC Data for WW variant **32p-Y23F**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

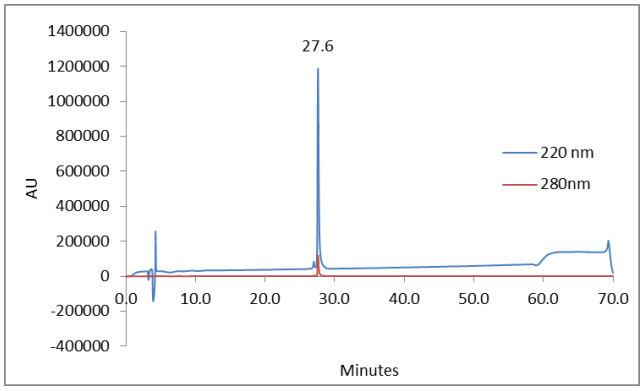


Figure 2-102. Analytical HPLC Data for Pin WW domain protein **16/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

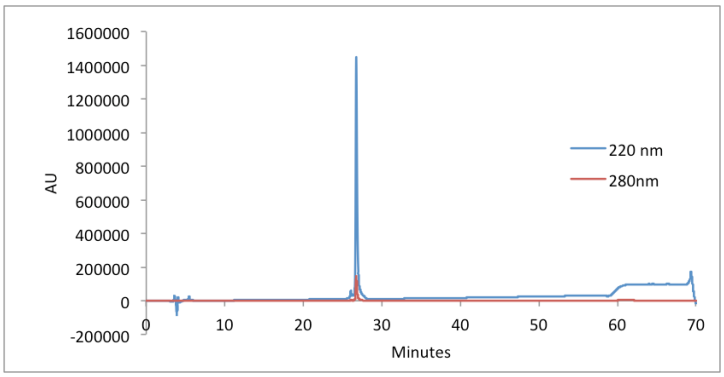


Figure 2-103.Analytical HPLC Data for WW variant **16p/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

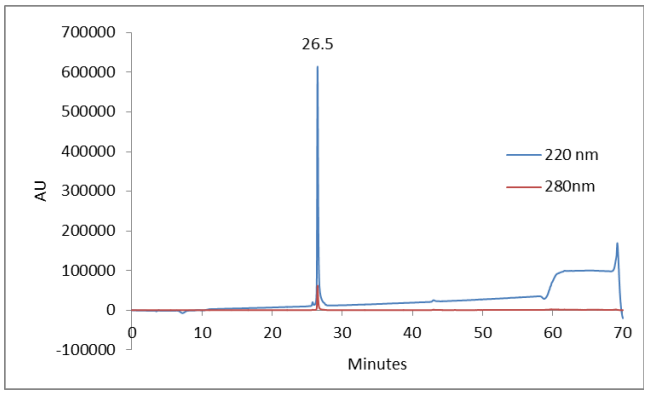


Figure 2-104. Analytical HPLC Data for WW variant **19/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

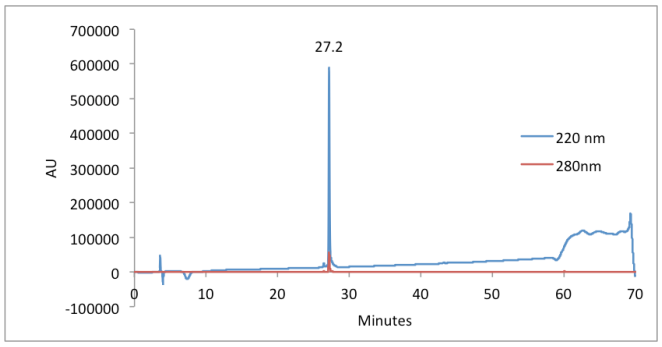


Figure 2-105. Analytical HPLC Data for WW variant **19p/26p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

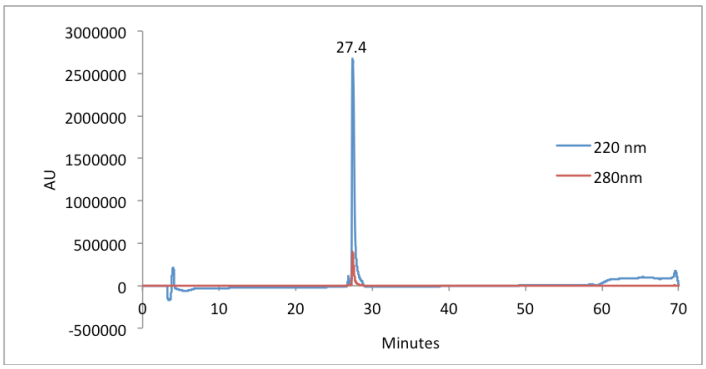


Figure 2-106. Analytical HPLC Data WW variant **16/19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

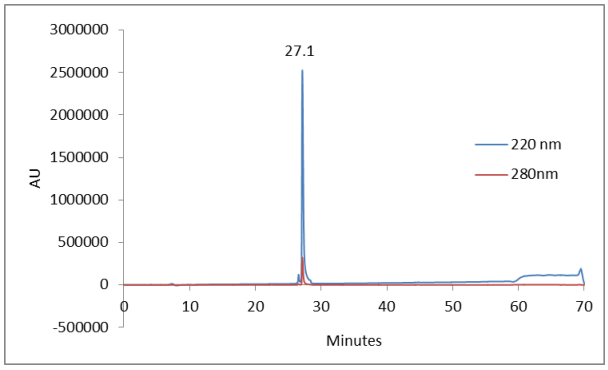


Figure 2-107. Analytical HPLC Data for WW variant **16p/19**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

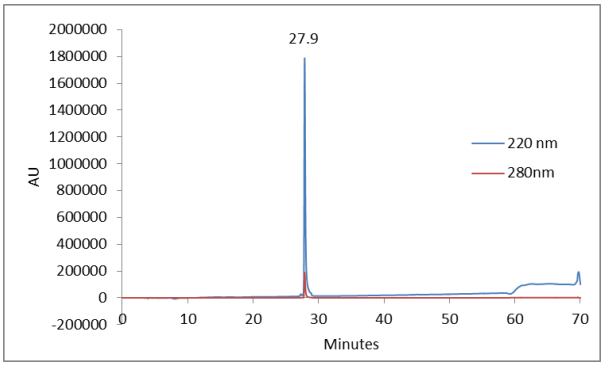


Figure 2-108.Analytical HPLC Data for WW variant **16/19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

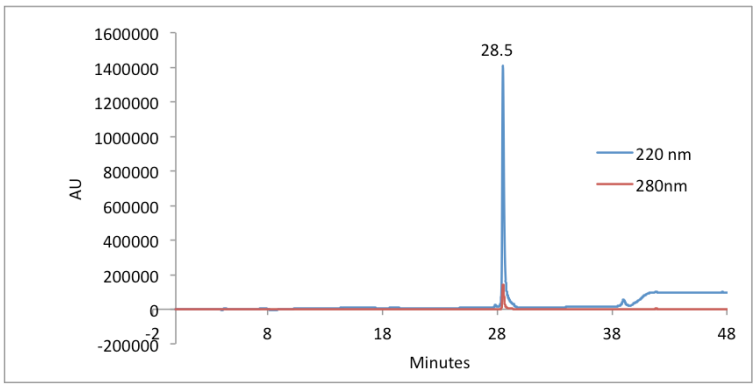


Figure 2-109.Analytical HPLC Data doubly PEGylated WW variant **16p/19p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-40% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 30 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

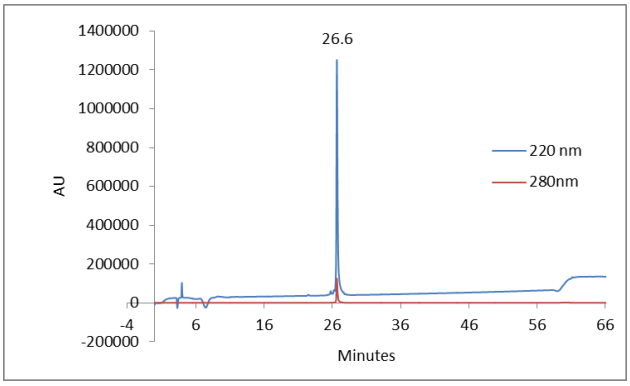


Figure 2-110.Analytical HPLC Data for WW variant **26p/29**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min. Analysis was truncated after 66 minutes (during column re-equilibration).

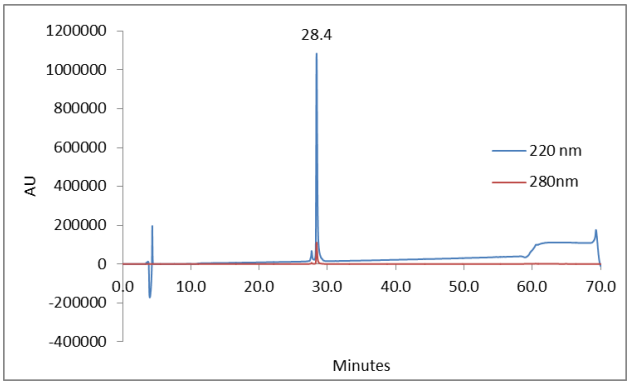


Figure 2-111. Analytical HPLC Data doubly PEGylated WW variant **26p/29p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

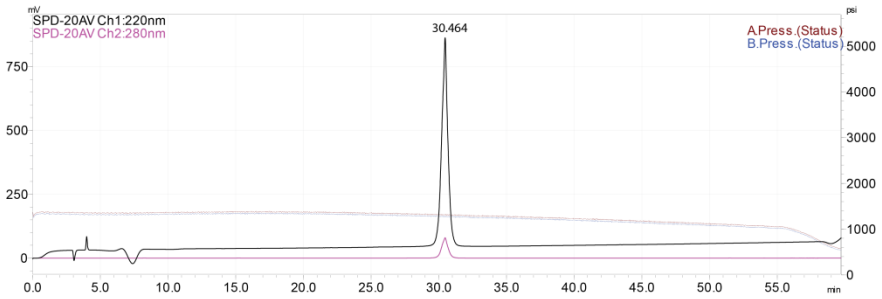


Figure 2-112. Analytical HPLC Data for WW variant **16p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

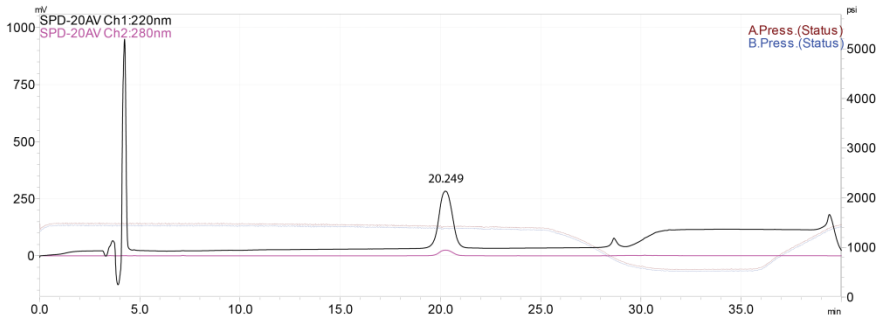


Figure 2-113.Analytical HPLC Data for WW variant **18p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

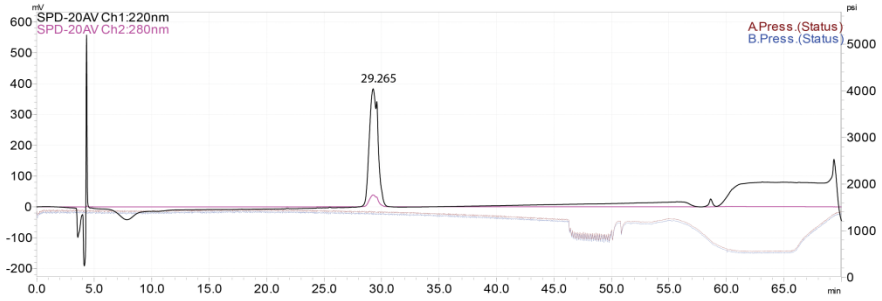


Figure 2-114. Analytical HPLC Data for WW variant **19p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

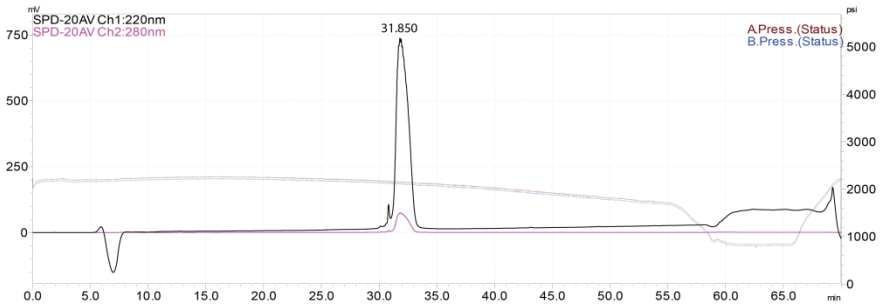


Figure 2-115. Analytical HPLC Data for WW variant **26p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

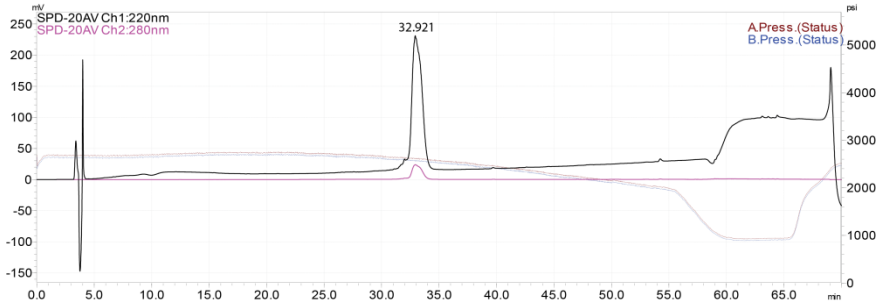


Figure 2-116.Analytical HPLC Data for WW variant **27p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

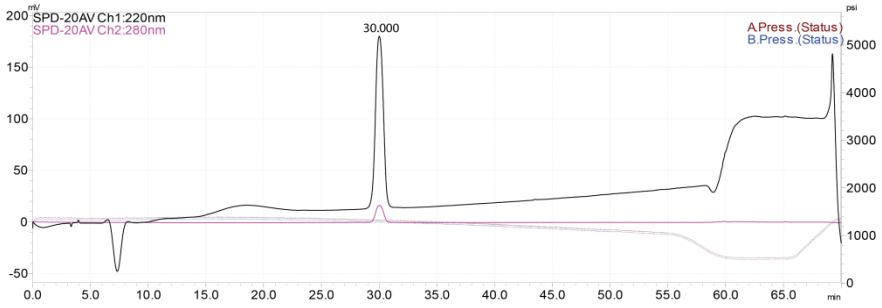


Figure 2-117.Analytical HPLC Data for WW variant **28p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

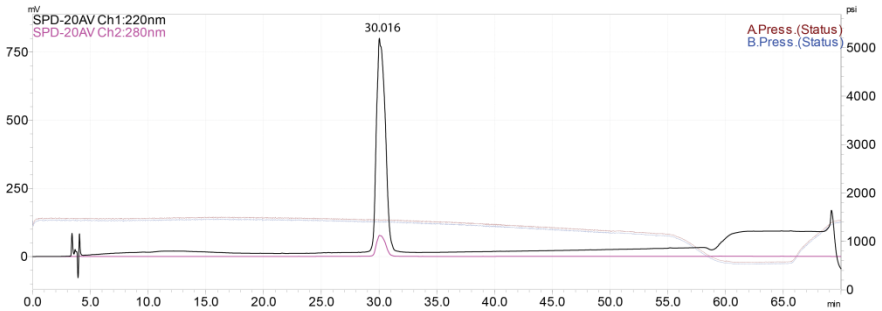


Figure 2-118.Analytical HPLC Data for WW variant **29p45**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10-minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 ml/min.

2.4.6 Organic Synthesis

2.4.7 Fmoc-D-Asn(PEG)-OtBu

Fmoc-D-Asn(PEG)-OtBu was synthesized following a procedure analogous to that of Herzner and Kunz[57]: 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).

^{1}H 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**H$_{(a)}$**H$_{(b)}$-O-);4.30 (H, apparent t, Fmoc Ar2CH-CH$_{(a)}$**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**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O--C**H$_2$**-C**H$_2$**-O-); 3.45 (2H, m, -CONH-C**H$_2$**-CH_2-O-);3.36 (3H, s, -O-C**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**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 Fmoc-D-Asn(PEG)-OtBu is shown in Figure 2-119.

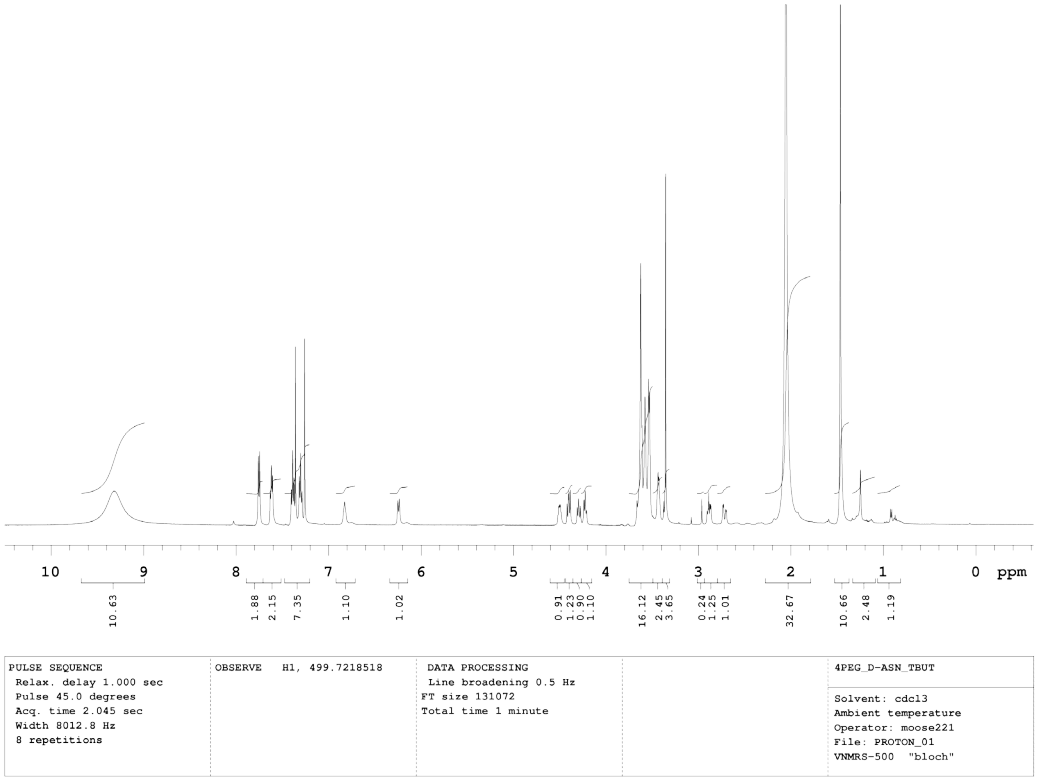


Figure 2-119.[1]H NMR spectrum for Fmoc-D-Asn(PEG)-OtBu.

^{13}C NMR (126 MHz, $CDCl_3$): 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_2CH-**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 Fmoc-D-Asn(PEG)-OtBu is shown in Figure 2-120.

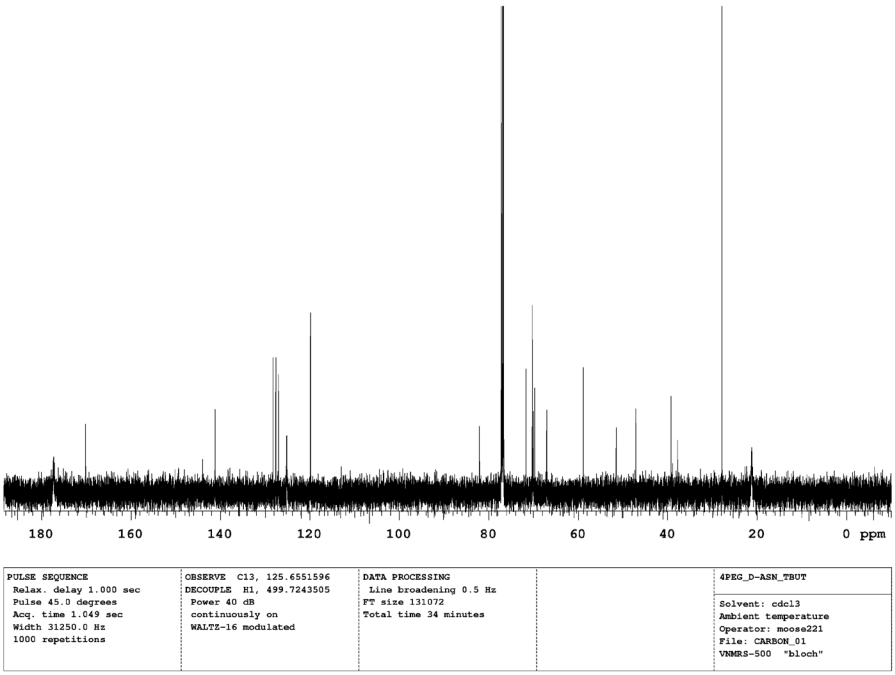


Figure 2-120. ^{13}C NMR spectrum for Fmoc-D-Asn(PEG)-OtBu.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-D-Asn(PEG)-OtBu were made by analogy with published spectral data for related compounds,[58,59] and with the assistance of a 2D HSQC experiment (Figure 2-121), using the one-bond C-H correlations shown in Table 2-9.

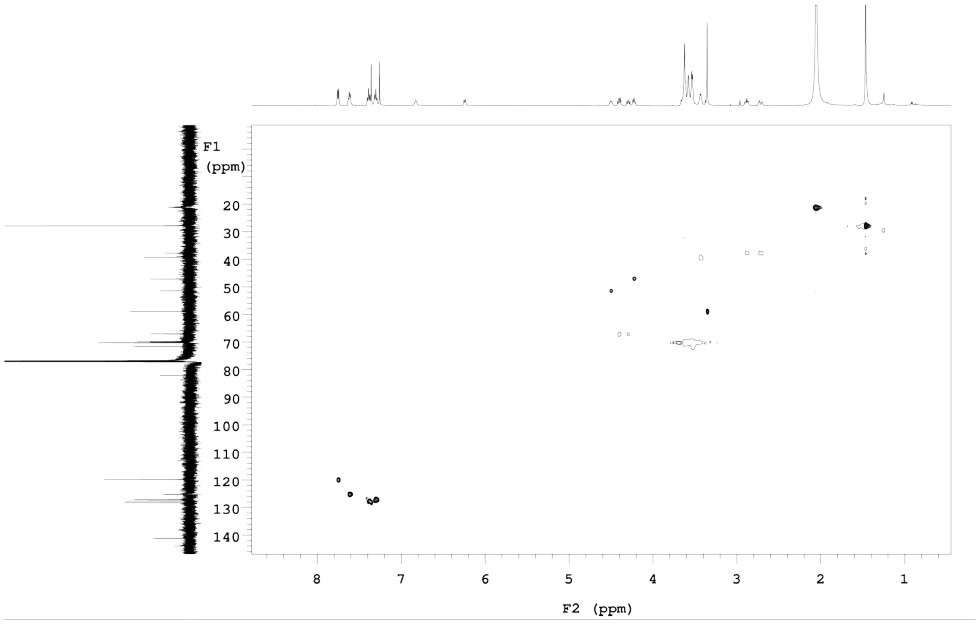


Figure 2-121. 2D HSQC spectrum of Fmoc-D-Asn(PEG)-OtBu.

Table 2-9.One-Bond C-H correlations identified from HSQC data for Fmoc-D-Asn-PEG-OtBu.

^{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 2-122.

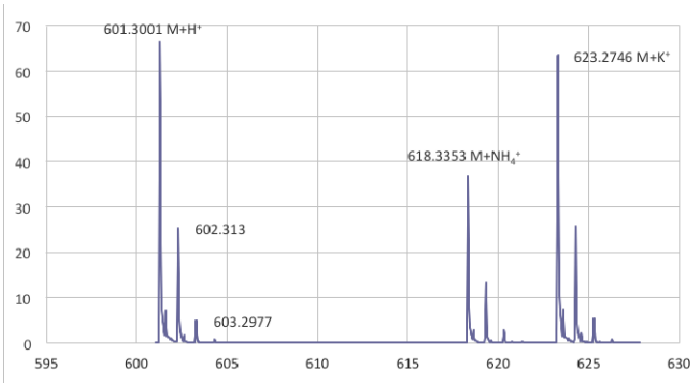


Figure 2-122.ESI-TOF MS data for Fmoc-D-Asn(PEG)-OtBu. Calculated m/z for $C_{32}H_{44}N_2O_9$ ($M+H^+$) is 601.31, found 601.30.

2.4.8 Fmoc-D-Asn(PEG4)-OH

To a solution of TFA (95% in water, 50ml) was added 0.89g Fmoc-D-Asn(PEG4)-OtBu, and the solution was stirred for 4 h under an argon atmosphere. The product was concentrated by rotary evaporation, and used without further purification.

^{1}H 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**H$_{(a)}$**H$_{(b)}$-O-); 4.31 (H, apparent t, Fmoc Ar2CH-CH$_{(a)}$**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**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O--C**H$_2$**-C**H$_2$**-O-); 3.45 (2H, m, -CONH-C**H$_2$**-CH_2-O-); 3.32 (3H, s, -O-C**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 Fmoc-D-Asn(PEG4)-OH is shown in Figure 2-123.

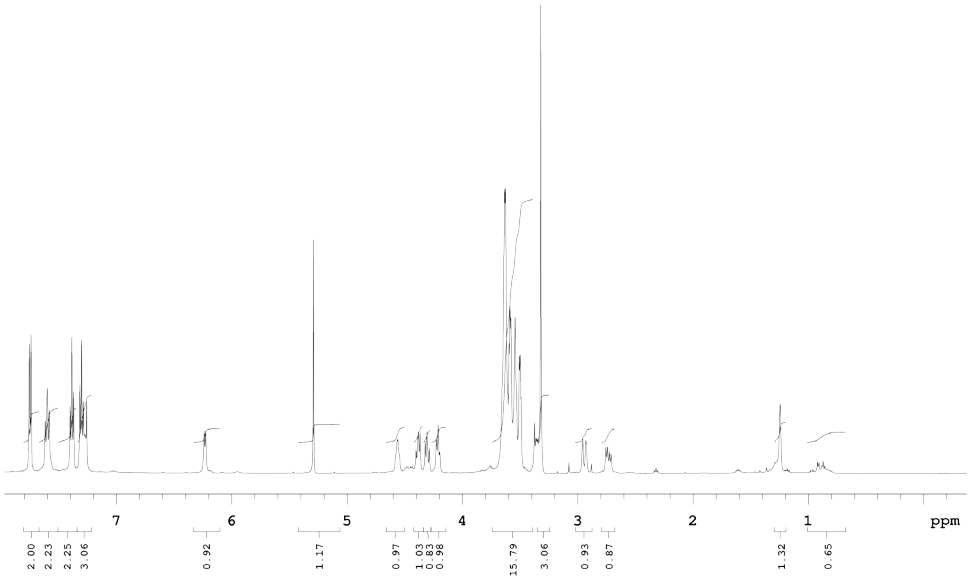


Figure 2-123.[1]H NMR spectrum of Fmoc-D-Asn(PEG4)-OH.

[13]C NMR (126 MHz, $CDCl_3$): 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 Fmoc-D-Asn(PEG4)-OH is shown in Figure 2-124.

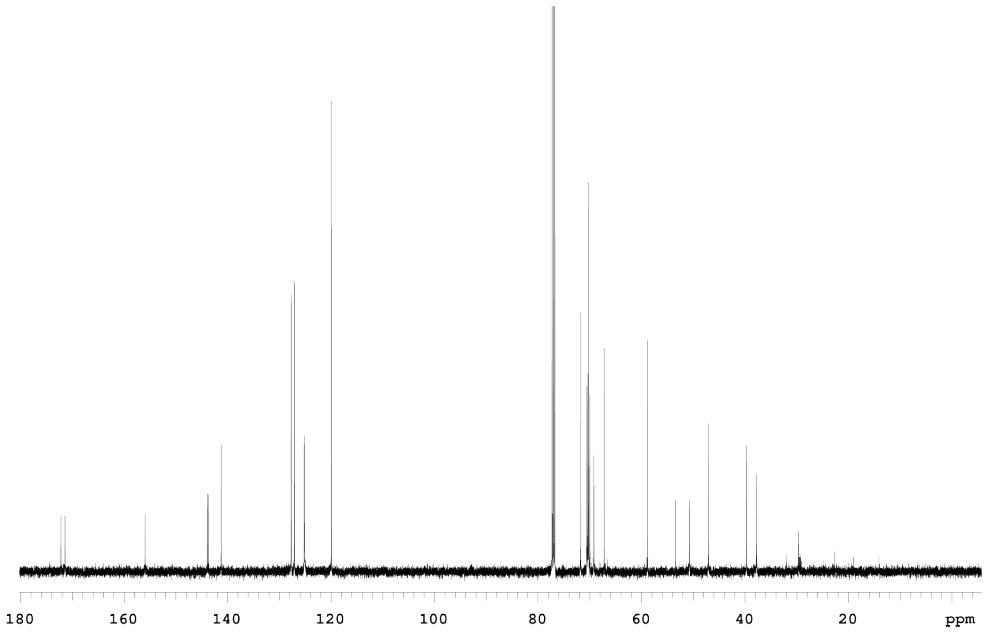


Figure 2-124.[13]C NMR spectrum of Fmoc-D-Asn(PEG4)-OtBu.

Assignments of the [1]H and [13]C NMR spectra for the Fmoc-D-Asn(PEG)-OH were made by analogy with published spectral data for related compounds, [58,59] and with the assistance of a 2D HSQC experiment (Figure 2-125), using the one-bond C-H correlations shown in Table 2-10.

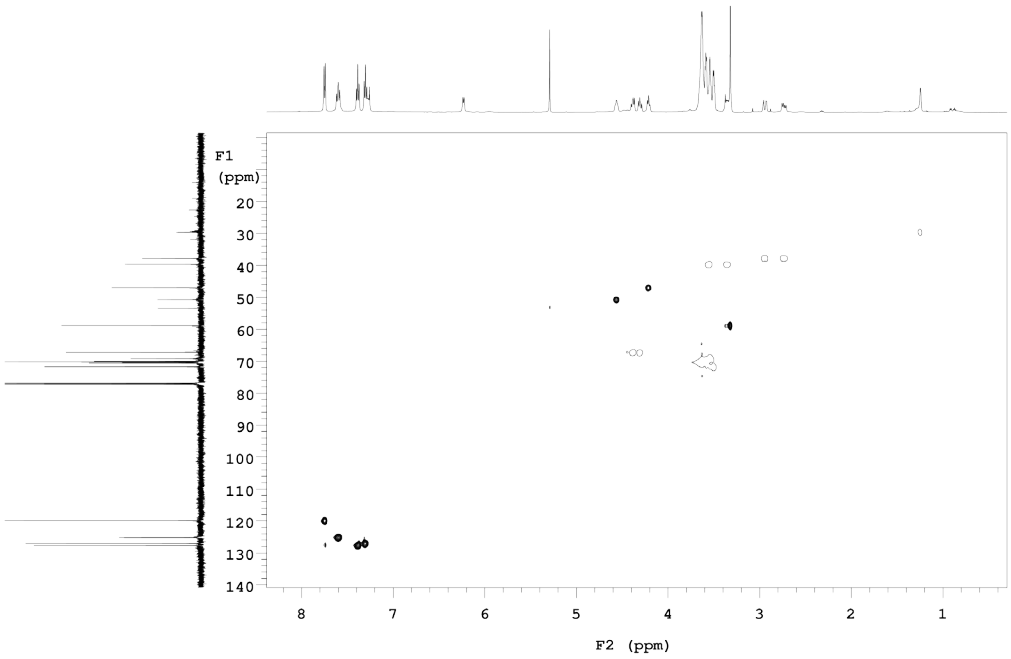


Figure 2-125.2D HSQC spectrum of Fmoc-D-Asn(PEG)-OH.

Table 2-10. One-Bond C-H correlations identified from HSQC experiment on Fmoc-D-Asn(PEG)-OH

^{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 2-126.

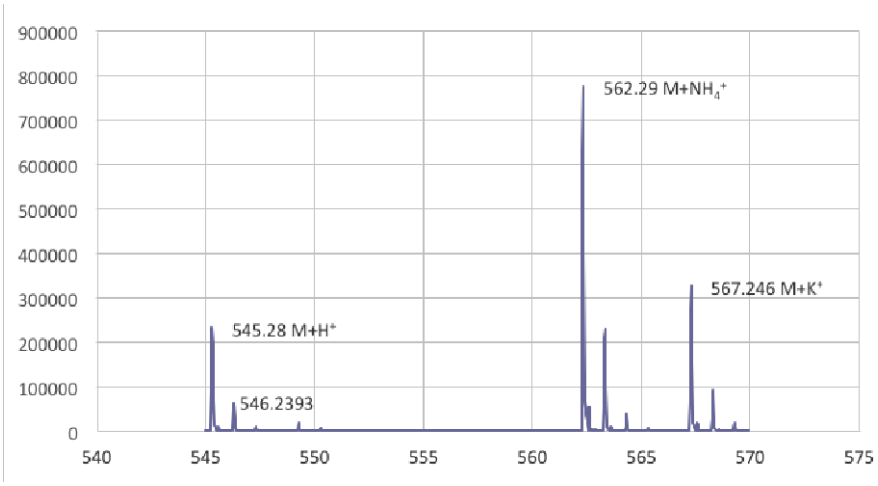


Figure 2-126.ESI-TOF MS data for (Fmoc-D-Asn(PEG)-OH) 2. Calculated m/z for $C_{28}H_{36}N_2O_9$ ($M+H^+$) 545.25, found 545.28.

2.4.9 Analysis of Folding Thermodynamic and Kinetics of WW Variants

2.4.10 Circular Dichroism Spectroscopy

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 (ε_{Trp} = 5690 $M^{-1}cm^{-1}$, ε_{Tyr} = 1280 $M^{-1}cm^{-1}$).[60] CD spectra of 50 or 100 μM solutions were obtained from 340 to 200 nm at 25ºC and in some cases at 1°C or 2°C. Variable temperature CD data were obtained at least in triplicate for 50 or 100 μM solutions of **WW, 14**, **14p**, **16**, **16p**, **16-Y23F**, **16p-Y23F**, **16-32A**, **16p-32A**, **17**, **17p**, **18**, **18p**, **19**, **19p**, **D-19**,

D-19p, **19-32A**, **19p-32A**, **23**, **23p**, **26p**, **WW-T29A**, **26p-T29A**, **27**, **27p**, **28**, **28p**, **29**, **29p**, **29-S32A**, **29p-S32A**, **30p**, **32**, **32p**, **32-Y23F**, **32p-Y23F**, **16/26p**, **16p/26p**, **19/26p**, **19p/26p**, **16/19**, **16/19p**, **16p/19**, **16p/19p**, **26p/29**, **26p/29p**, **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45** 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.

2.4.11 Laser Temperature Jump Experiments

PEGylated peptides **14p, 17p, 18p, 19p**,**26p,** and **30p** as well as corresponding non-PEGylated peptides **14, 17, 18, 19, 26,** and **30** (100 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,[3,61-63] 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 below 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:

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

Equation 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 $\chi_1(t)$:

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

Equation 2

which is plotted as a function of time for each protein at several temperatures.

2.4.12 Global Fitting of Variable Temperature CD Data

For **16**, **16p**,**16-Y23F**, **16p-Y23F**, **16-32A**, **16p-32A**,**D-19**,**D-19p**, **19-32A**, **19p-32A**, **23**, **23p**,**WW-T29A**, **26p-T29A**,**27**, **27p**, **28**, **28p**, **29**, **29p**,**29-S32A**, **29p-S32A**,**32**, **32p**,**32-Y23F**, **32p-Y23F**,**16/26p**, **16p/26p**, **19/26p**, **19p/26p**, **16/19**, **16/19p**, **16p/19**, **16p/19p**, **26p/29**, **26p/29p**, **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45**,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:

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

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

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

Equation 4

where R is the universal gas constant (0.0019872 kcal/mol/K). $\Delta G_f(T)$ was fit to the following equation:

$$\Delta G_f = \frac{\Delta H(T_m) \cdot (T_m - T)}{T_m} + \Delta C_p \cdot (T - T_m - T \cdot \ln\left[\frac{T}{T_m}\right])$$

Equation 5

where the fit parameters are T_m(the midpoint of the unfolding transition; the temperature at which $\Delta G_f = 0$); $\Delta H(T_m)$, the change in enthalpy upon folding at T_m; and ΔC_p, the change in heat capacity upon folding. The parameters for equations 3-5 were used to calculate the values of the folding free energy ΔG_f for WW variants in the main text and in tables below.

2.4.13 Global Fitting of Variable Temperature CD and Laser Temperature Jump Experiments

For proteins **WW**, **14**, **14p**, **17**, **17p**, **18**, **18p**, **19**, **19p**, **26p**, and **30p**, 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 ΔG_f, the folding and unfolding activation energies ($\Delta G^{\ddagger}_f$ and $\Delta G^{\ddagger}_u$, respectively), and folding and unfolding rates (k_f and k_u, respectively). As before, variable temperature CD data were fit to the following equation:

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

Equation 6

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.

Relaxation data from temperature jump experiments were simultaneously fit to the following equations:

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

Equation 7

where t is time, A_0 is the initial value of x_1 at $t = 0$, y_0 is the value of c_1 at $t = \infty$, and k_f is the folding rate. Each relaxation trace derived from several replicated temperature jump experiments had distinct A_0 and y_0 values. In contrast, temperature-dependent K_f(see equations 8 and 9) and k_f(see equation 10) were constrained to be the same across all variable temperature CD and temperature jump experiments for a given protein.

Equilibrium folding constant K_f was defined by the following equations:

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

Equation 8

$$\Delta G_f = \frac{\Delta H(T_m) \cdot (T_m - T)}{T_m} + \Delta C_p \cdot (T - T_m - T \cdot \ln\left[\frac{T}{T_m}\right])$$

Equation 9

where R is the universal gas constant (0.0019872 kcal/mol/K), $\Delta G_f$is the folding free energy, T_m is the midpoint of the variable temperature thermal unfolding transition (i.e., the temperature at which $\Delta G_f = 0$), $\Delta H(T_m)$ is the change in enthalpy upon folding at T_m, and ΔC_p is the change in heat capacity upon folding.

Folding rate k_f was defined by the following Kramers[64-66] model equation:

$$k_f(T) = \nu(59\ ^\circ C) \cdot \frac{\eta(59\ ^\circ C)}{\eta(T)} \exp\left[-\frac{\Delta G_0^\dagger + \Delta G_1^\dagger \cdot (T - T_m) + \Delta G_2^\dagger \cdot (T - T_m)^2}{RT}\right]$$

Equation 10

in which $\Delta G^\ddagger_f(T)$, the folding activation energy 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 10 and 11). The pre-exponential term in Equation 10 represents the viscosity-corrected frequency ν of the characteristic diffusional folding motion at the barrier[67,68] (at 59 °C, $\nu = 5 \times 10^5\ s^{-1}$).[17] η(59 °C) is the solvent viscosity at 59 °C and η(T) is the solvent viscosity at temperature T, both calculated with equation 11:

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

Equation 11

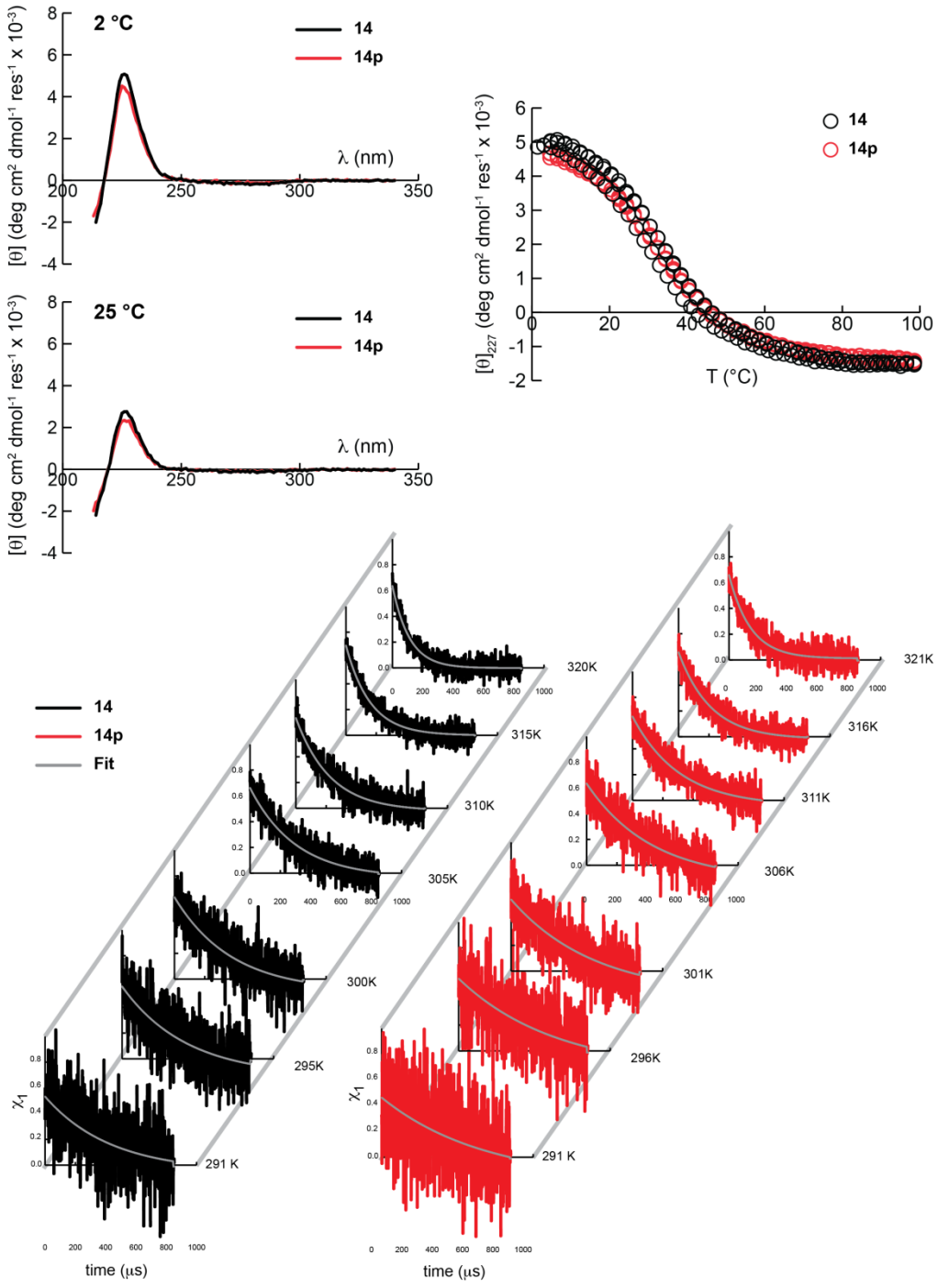
where $A = 2.41 \times 10^5$ Pa·s, B = 247.8 K, and C = 140 K.[69]

The parameters for equations 6–11 were used to calculate the values of the folding free energy ΔG_f, folding rate k_f and unfolding rate k_u ($k_u = k_f / K_f$) for proteins **WW**, **14**, **14p**, **17**, **17p**, **18**, **18p**, **19**, **19p**, **26p**, and **30p**that are presented in **Table 2-3**.

2.4.14 Plots of CD Spectra, Variable Temperature CD Data, and Temperature Jump Kinetic Data

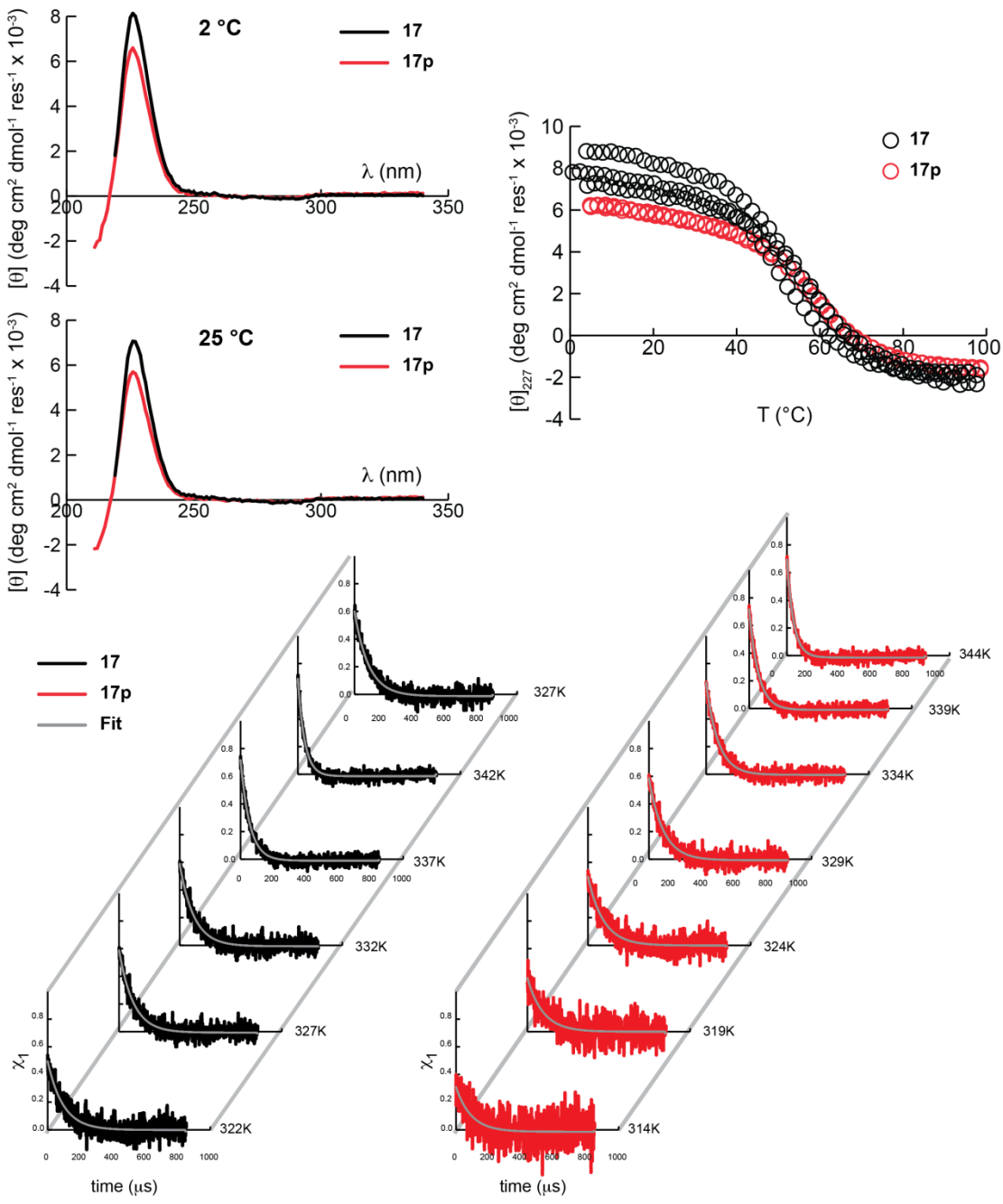
CD spectra, variable temperature CD data, and temperature jump kinetic data for proteins **WW**, **14**, **14p**, **17**, **17p**, **18**, **18p**, **19**, **19p**, **26p**, **30p**are shown in **Figures 2-127** through **2-132**, along with the parameters of equations 6-11 that were used to generate the global fits for each compound. The standard error for each fitted parameter is also shown.

Similarly, CD spectra and variable temperature CD data for **16**, **16p**,**16-Y23F**, **16p-Y23F**, **16-32A**, **16p-32A**,**D-19**,**D-19p**, **19-32A**, **19p-32A**, **23**, **23p**,**WW-T29A**, **26p-T29A**,**27**, **27p**, **28**, **28p**, **29**, **29p**,**29-S32A**, **29p-S32A**,**32**, **32p**,**32-Y23F**, **32p-Y23F**,**16/26p**, **16p/26p**, **19/26p**, **19p/26p**, **16/19**, **16/19p**, **16p/19**, **16p/19p**, **26p/29**, **26p/29p**, **16p45**, **18p45**, **19p45**, **26p45**, **27p45**, **28p45**, and **29p45**, along with data for **19** and **19p** in the presence of various amounts of D_2O, are shown in **Figures 2-127** through **2-157**, 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.



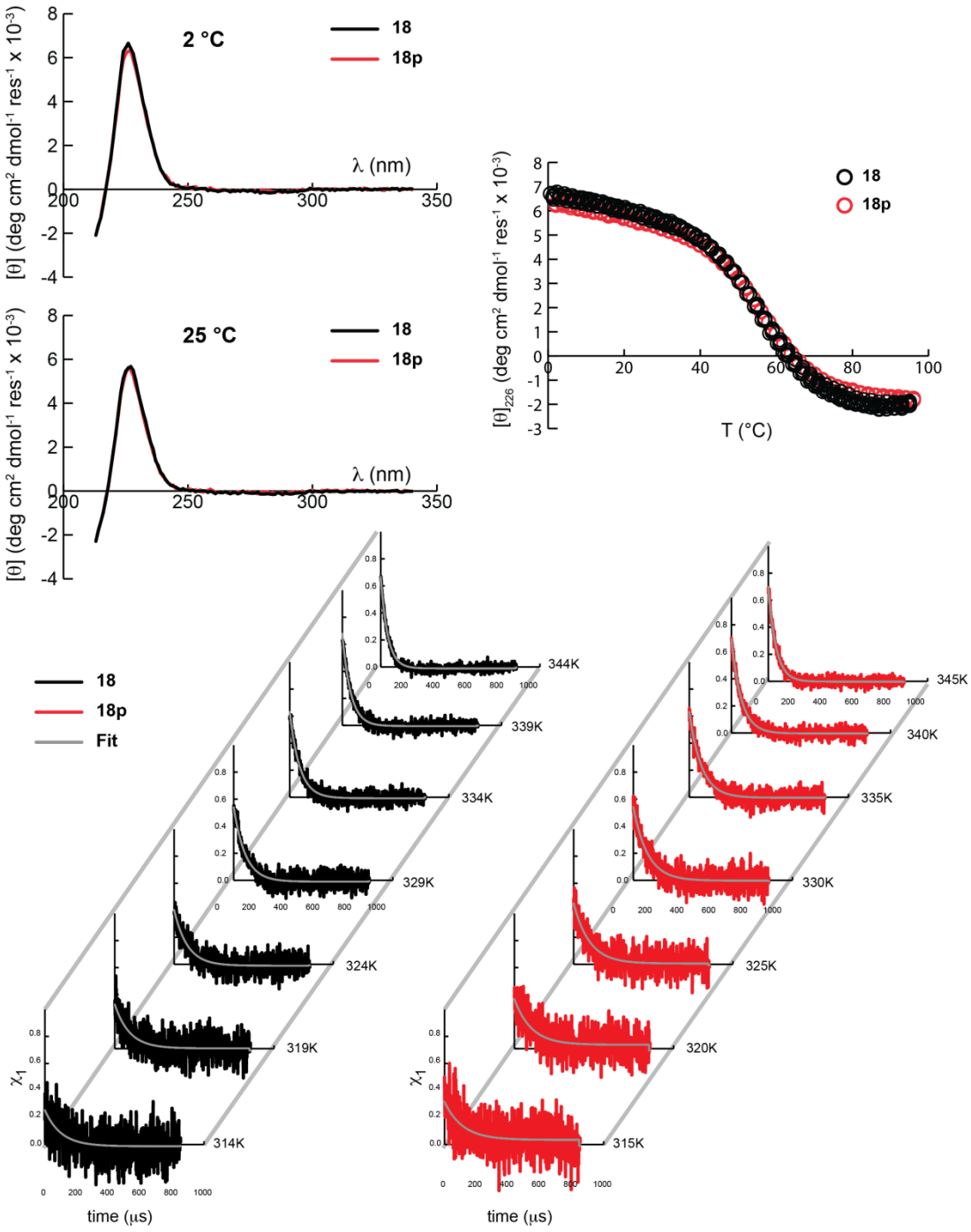
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error	$\Delta G^{\dagger}_0$	$\Delta G^{\dagger}_1$	$\Delta G^{\dagger}_2$
14	307.13 ± 0.85	-21.282 ± 1.479	0.5420 ± 0.1170	0.92985	0.11781	3.1480 ± 0.0263	0.0217 ± 0.0032	-0.0004 ± 0.0002
14p	306.51 ± 5.22	-21.528 ± 2.690	0.2450 ± 0.6330	0.88262	0.14266	3.3992 ± 0.0695	0.0233 ± 0.0180	-0.0003 ± 0.0007

Figure 2-127. CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for 100 μM solutions of proteins **14** (black) and **14p** (red) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data for each compound to equations 6–11.



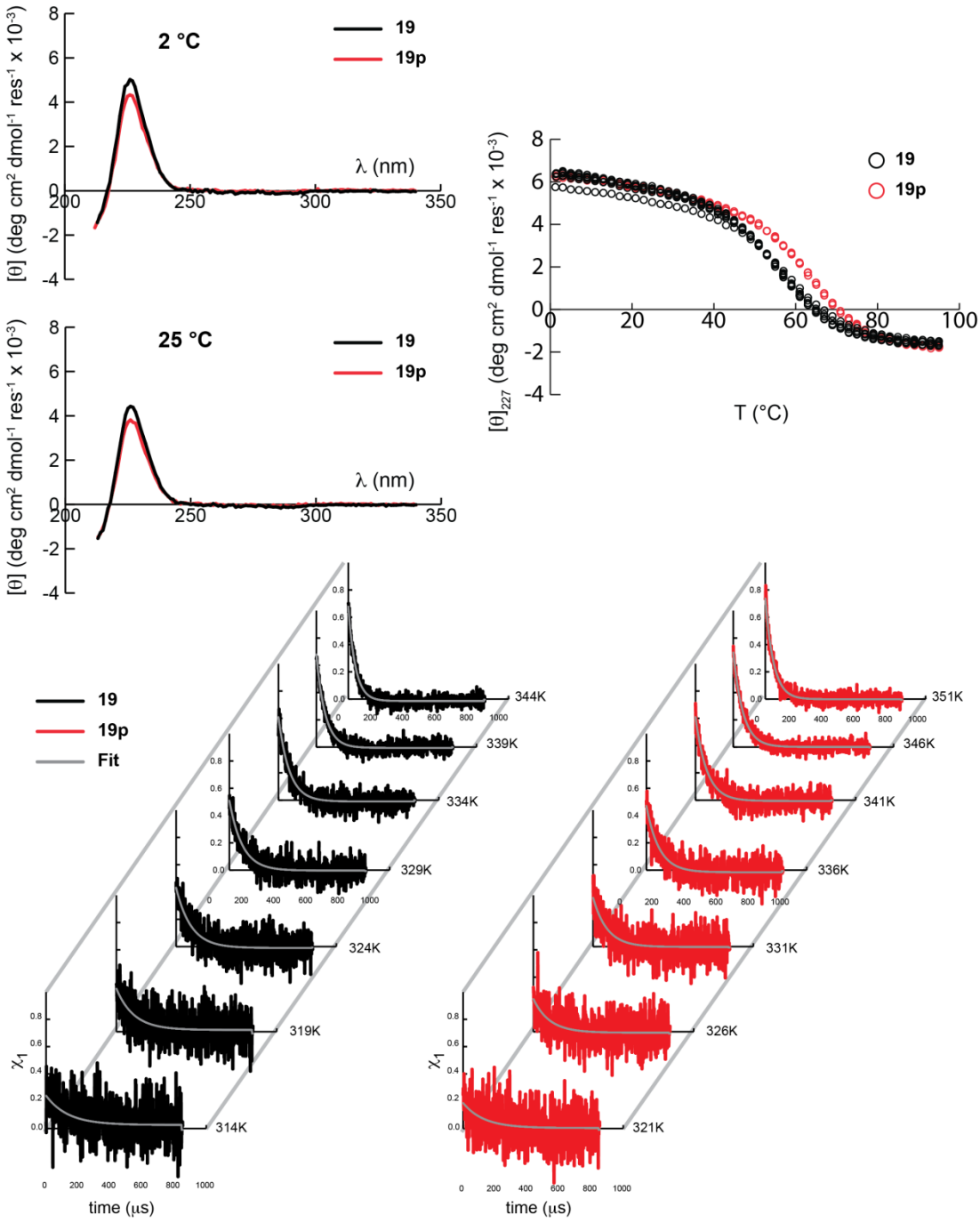
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error	$\Delta G^{\dagger}_0$	$\Delta G^{\dagger}_1$	$\Delta G^{\dagger}_2$
17	326.79 ± 0.35	-28.424 ± 0.431	0.0390 ± 0.0820	0.99825	0.03941	2.8513 ± 0.0089	0.0461 ± 0.0029	0.0001 ± 0.0002
17p	328.68 ± 0.48	-31.176 ± 0.547	-0.3040 ± 0.1580	0.99326	0.05281	2.8942 ± 0.0120	0.0510 ± 0.0024	0.0006 ± 0.0002

Figure 2-128.CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for 100 μM solutions of proteins **17** (black) and **17p** (red) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data for each compound to equations 6–11.



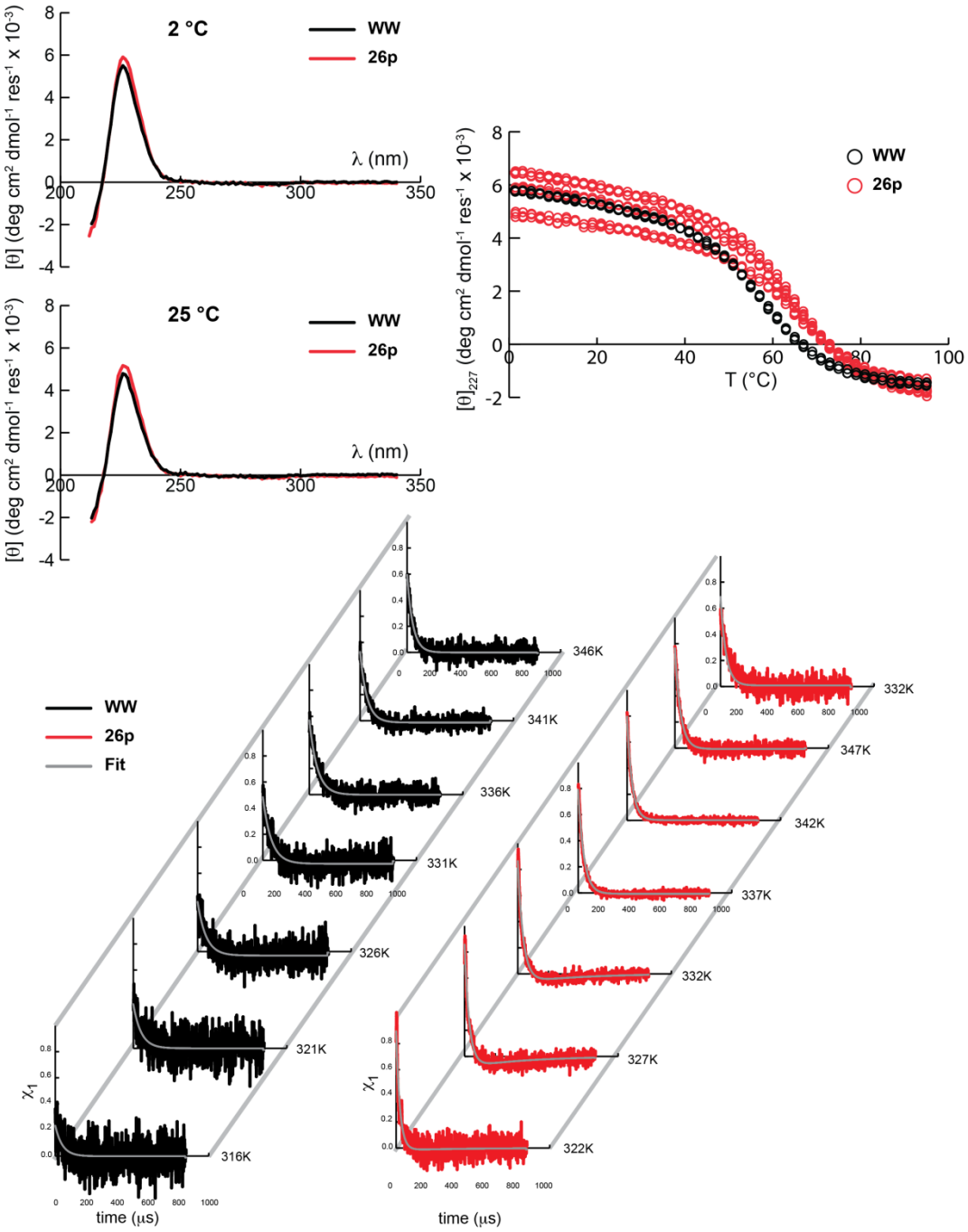
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error	$\Delta G^{\ddagger}_0$	$\Delta G^{\ddagger}_1$	$\Delta G^{\ddagger}_2$
18	330.08 ± 0.23	-27.166 ± 0.353	-0.4880 ± 0.0590	0.99597	0.06058	2.7705 ± 0.0155	0.0435 ± 0.0016	0.0012 ± 0.0002
18p	330.12 ± 0.67	-30.846 ± 0.745	-0.3790 ± 0.2070	0.98890	0.06799	2.9325 ± 0.0165	0.0496 ± 0.0026	0.0011 ± 0.0002

Figure 2-129.CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for 100 μM solutions of proteins **18** (black) and **18p** (red) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data for each compound to equations 6–11.



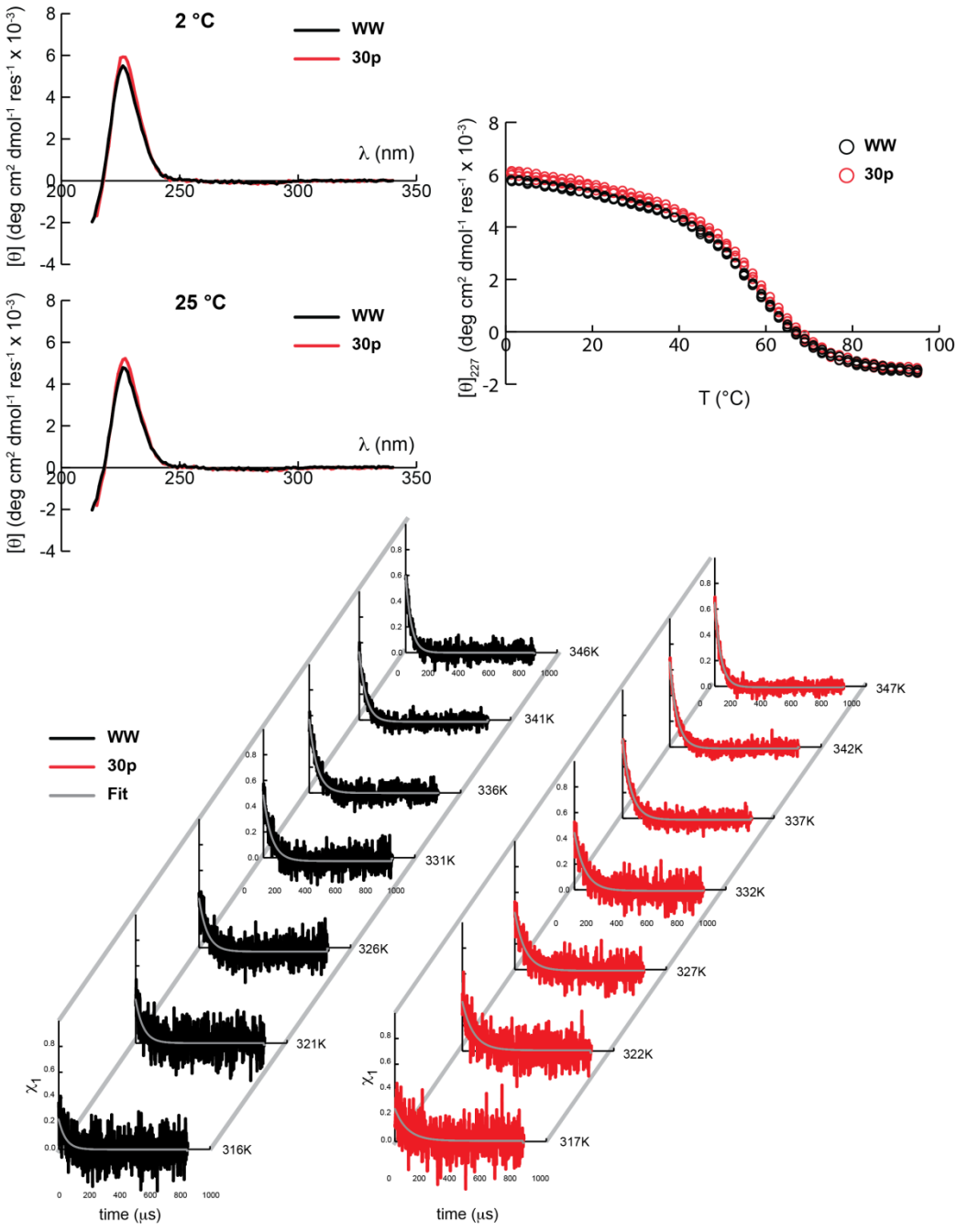
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error	$\Delta G^{\dagger}_0$	$\Delta G^{\dagger}_1$	$\Delta G^{\dagger}_2$
19	329.02 ± 0.16	-29.516 ± 0.261	-0.5120 ± 0.0520	0.99863	0.02985	2.9132 ± 0.0152	0.0619 ± 0.0024	0.0011 ± 0.0001
19p	337.00 ± 0.10	-33.039 ± 0.324	-0.7190 ± 0.0150	0.99881	0.02409	2.8973 ± 0.0066	0.0628 ± 0.0011	0.0014 ± 0.0001

Figure 2-130. CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for 100 μM solutions of proteins **19** (black) and **19p** (red) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data for each compound to equations 6–11.



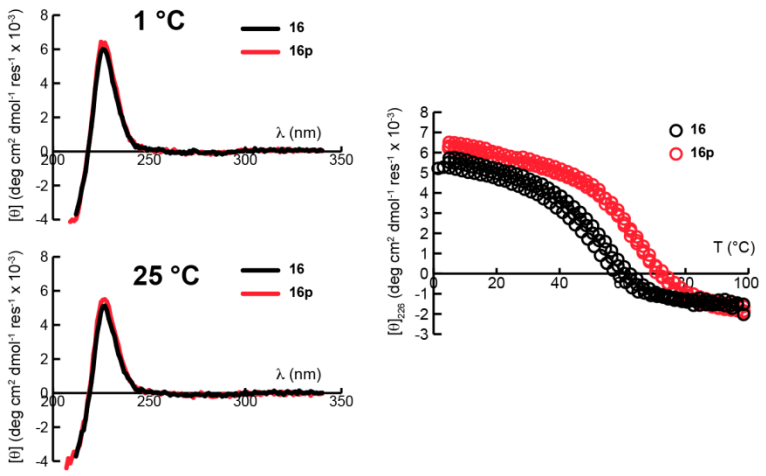
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error	$\Delta G^{\dagger}_0$	$\Delta G^{\dagger}_1$	$\Delta G^{\dagger}_2$
WW	331.11 ± 0.66	-31.116 ± 0.658	-0.3540 ± 0.2040	0.99445	0.05606	2.6299 ± 0.0265	0.0704 ± 0.0054	0.0002 ± 0.0005
26p	337.75 ± 0.20	-31.831 ± 0.505	-0.6160 ± 0.0410	0.99700	0.03896	2.4642 ± 0.0088	0.0558 ± 0.0019	-0.0001 ± 0.0001

Figure 2-131.CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for 100 μM solutions of proteins **WW** (black) and **26p** (red) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data for each compound to equations 6–11.



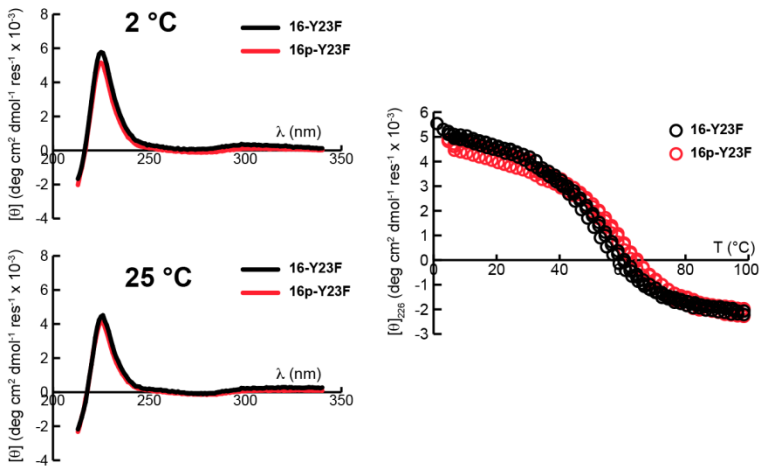
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error	$\Delta G^{\ddagger}_0$	$\Delta G^{\ddagger}_1$	$\Delta G^{\ddagger}_2$
WW	331.11 ± 0.66	-31.116 ± 0.658	-0.3540 ± 0.2040	0.99445	0.05606	2.6299 ± 0.0265	0.0704 ± 0.0054	0.0002 ± 0.0005
30p	331.5 ± 0.22	-31.606 ± 0.812	-0.6655 ± 0.0713	0.98980	0.06426	2.7407 ± 0.0172	0.0553 ± 0.0024	0.0016 ± 0.0002

Figure 2-132. CD spectra (lines, top right), variable temperature CD data (circles, top left) and laser temperature jump relaxation data (lines, bottom) for 100 μM solutions of proteins **WW** (black) and **30p** (red) in 20 mM sodium phosphate, pH 7. Grey lines show the global fit of the kinetic data for each compound to equations 6–11.



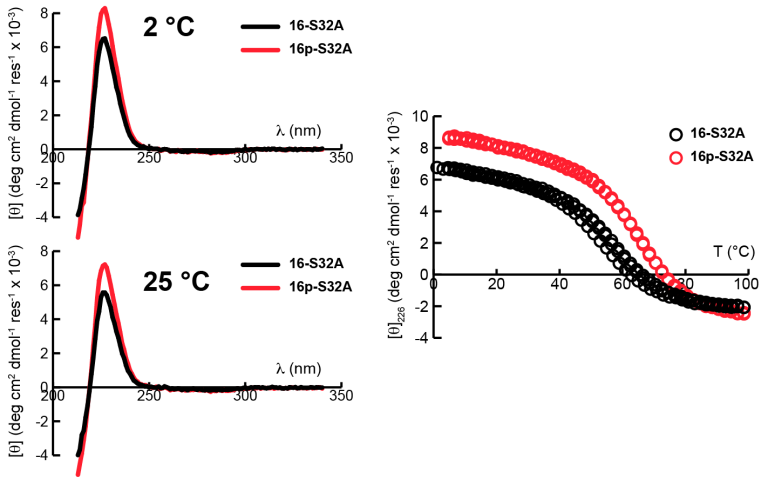
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16	323.7 ± 0.2	-30.5 ± 0.8	-0.83 ± 0.04	0.9998	0.051
16p	333.9 ± 0.3	-32.4 ± 0.5	-0.56 ± 0.10	0.9999	0.041

Figure 2-133. CD spectra (50 µM) and variable temperature CD data (100 µM) for WW variants **16** (black) and **16p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



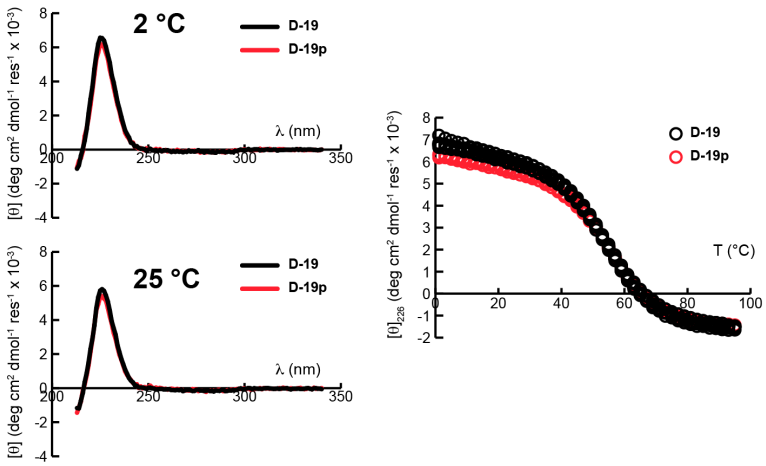
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16-Y23F	324.0 ± 0.6	-26.5 ± 0.9	-0.5 ± 0.2	0.9995	0.059
16p-Y23F	329.5 ± 0.3	-28.8 ± 1.1	-0.72 ± 0.05	0.9995	0.059

Figure 2-134.CD spectra (100 µM) and variable temperature CD data (100 µM) for WW variants **16-Y23F** (black) and **16p-Y23F** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



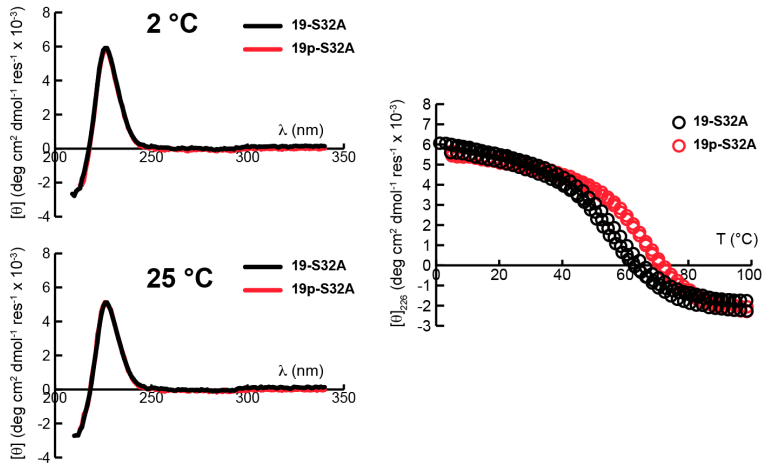
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16-S32A	328.1 ± 0.3	-29.3 ± 0.3	-0.2 ± 0.1	0.9999	0.031
16p-S32A	334.7 ± 0.1	-31.5 ± 0.3	-0.68 ± 0.02	0.9999	0.032

Figure 2-135. Figure S128. CD spectra (100 µM) and variable temperature CD data (100 µM) for WW variants **16-S32A** (black) and **16p-S32A** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



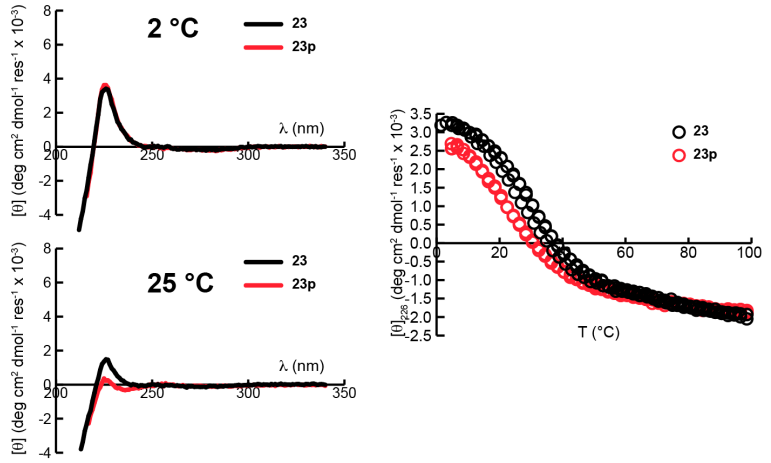
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
D-19	328.6 ± 0.3	-30.4 ± 0.3	-0.12 ± 0.09	0.9998	0.044
D-19p	328.4 ± 0.3	-29.6 ± 0.3	-0.42 ± 0.08	0.9999	0.024

Figure 2-136.Variable temperature CD data (100 µM) for WW variants **D-19** (black) and **D-19p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



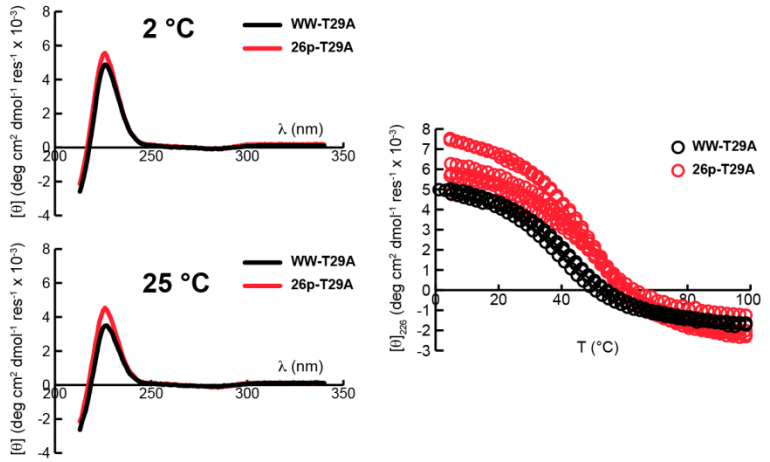
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
19-S32A	327.5 ± 0.3	-28.3 ± 0.2	-0.29 ± 0.07	0.99995	0.021
19p-S32A	335.9 ± 0.1	-31.1 ± 0.3	-0.59 ± 0.03	0.99995	0.021

Figure 2-137.Variable temperature CD data (100 µM) for WW variants **19-S32A** (black) and **19p-S32A** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



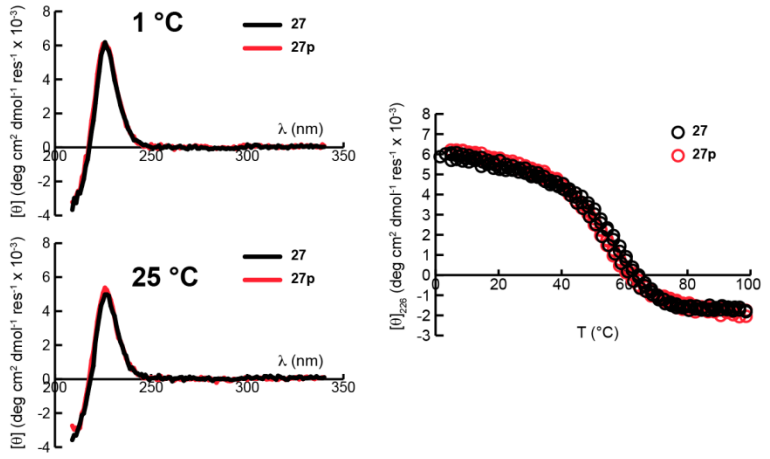
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
23	301.7 ± 0.9	-25.2 ± 0.5	-0.3 ± 0.2	0.9997	0.035
23p	296.4 ± 1.0	-24.1 ± 1.0	-0.57 ± 0.09	0.9991	0.047

Figure 2-138.CD spectra and variable temperature CD data for 100 µM solutions of WW variants **23** (black) and **23p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



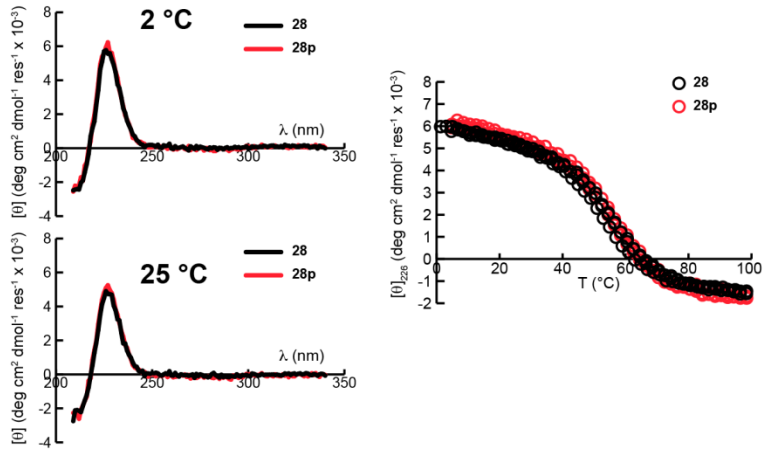
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
WW-T29A	313.6 ± 0.7	-24.1 ± 0.5	-0.1 ± 0.1	0.9998	0.031
26p-T29A	318.4 ± 0.2	-22.7 ± 0.4	-0.54 ± 0.04	0.9997	0.051

Figure 2-139.CD spectra and variable temperature CD data for 100 µM solutions of WW variants **WW-T29A** (black) and **26p-T29A** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



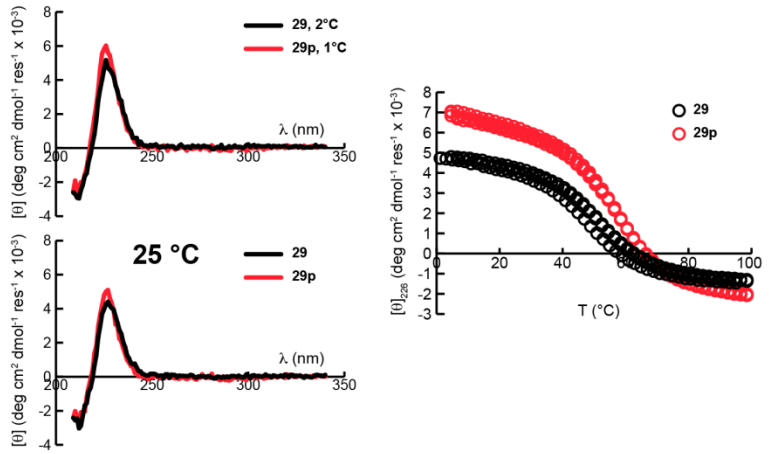
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
27	328.2 ± 0.1	-31.8 ± 0.5	-0.66 ± 0.07	0.9997	0.054
27p	324.2 ± 0.4	-30.9 ± 0.5	-0.1 ± 0.1	0.9998	0.052

Figure 2-140.CD spectra and variable temperature CD data for 50 µM solutions of WW variants **27** (black) and **27p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



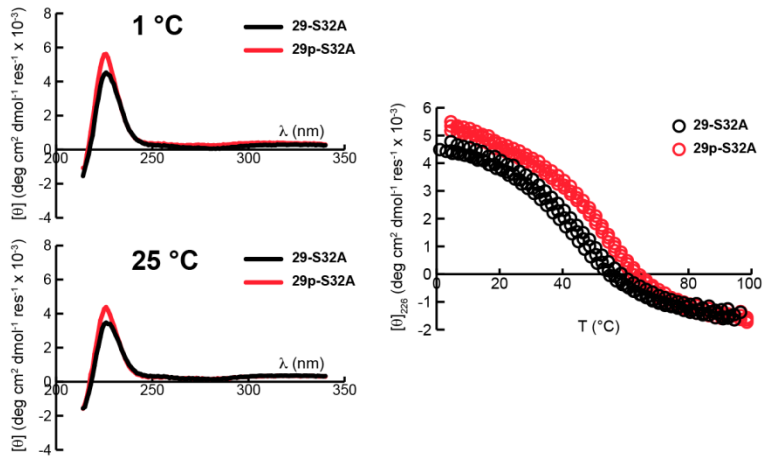
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
28	326.4 ± 0.5	-30.4 ± 0.5	-0.3 ± 0.2	0.9998	0.046
28p	326.4 ± 0.5	-29.8 ± 0.5	-0.3 ± 0.2	0.9997	0.053

Figure 2-141.CD spectra and variable temperature CD data for 50 µM solutions of WW variants **28** (black) and **28p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



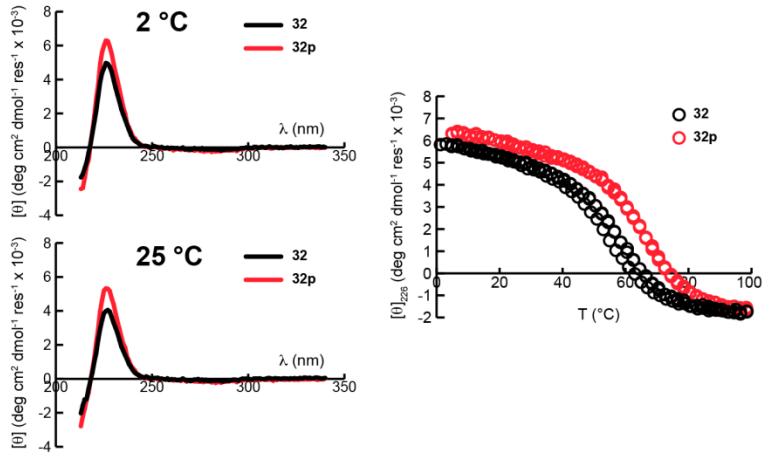
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
29	323.2 ± 0.3	-28.1 ± 0.4	-0.01 ± 0.09	0.9999	0.025
29p	327.3 ± 0.3	-29.3 ± 0.3	-0.36 ± 0.07	0.99996	0.024

Figure 2-142.CD spectra (50 µM) and variable temperature CD data (100 µM) for WW variants **29** (black) and **29p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



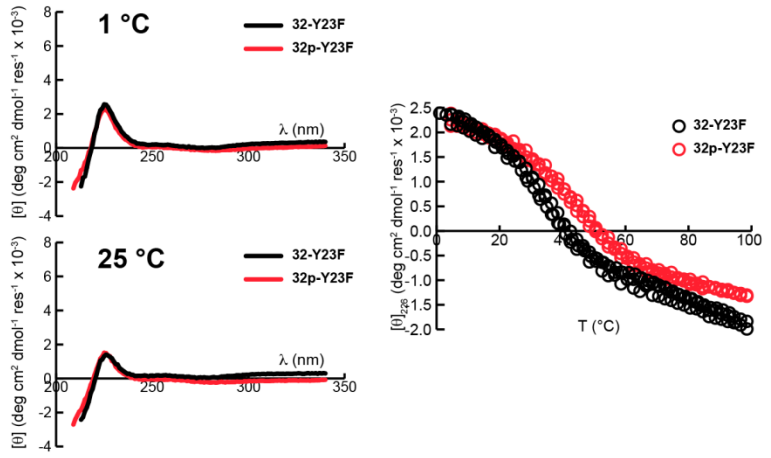
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
29-S32A	313.6 ± 0.5	-18.9 ± 0.6	-0.52 ± 0.05	0.9999	0.036
29p-S32A	325.1 ± 0.7	-26.6 ± 0.6	-0.3 ± 0.1	0.9999	0.028

Figure 2-143.CD spectra (100 µM) and variable temperature CD data (100 µM) for WW variants **29-S32A** (black) and **29p-S32A** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



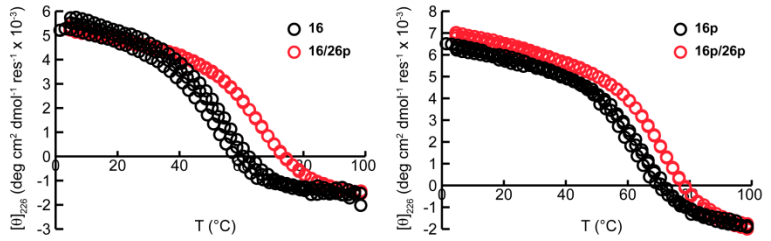
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
32	318.2 ± 0.2	-26.7 ± 0.4	-0.61 ± 0.05	0.9999	0.024
32p	323.5 ± 0.2	-28.9 ± 0.3	-0.41 ± 0.06	0.99996	0.021

Figure 2-144.CD spectra and variable temperature CD data for 100 µM solutions of WW variants **32** (black) and **32p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



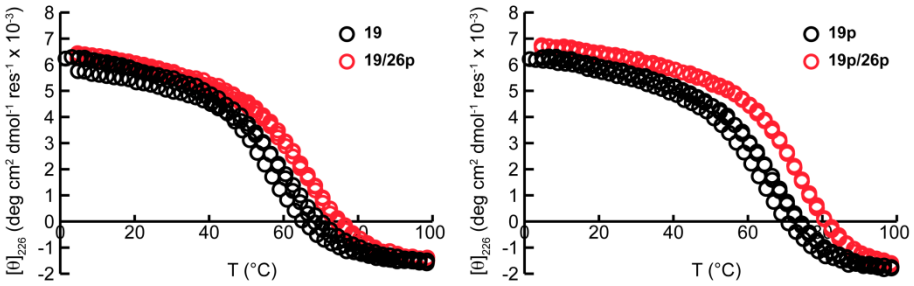
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
32-Y23F	303.1 ± 1.3	-23.2 ± 2.0	-1.0 ± 0.2	0.9991	0.047
32p-Y23F	313.5 ± 0.7	-21.9 ± 1.1	-0.6 ± 0.1	0.9994	0.033

Figure 2-145. CD spectra (100 µM) and variable temperature CD data (100 µM) for WW variants **32-Y23F** (black) and **32p-Y23F** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



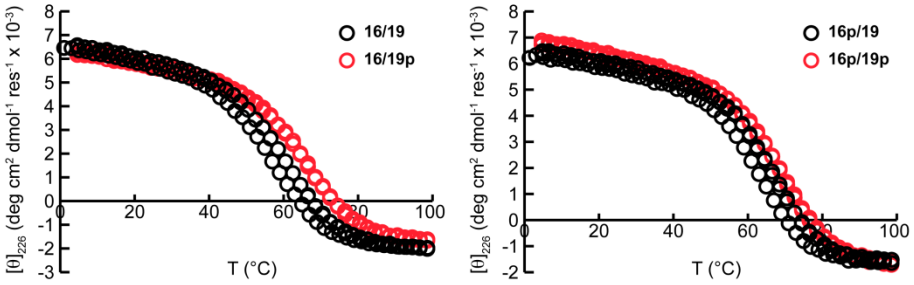
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16	323.7 ± 0.2	-30.5 ± 0.8	-0.83 ± 0.04	0.9998	0.051
16/26p	335.8 ± 0.3	-30.8 ± 0.2	-0.4 ± 0.1	0.99997	0.015
16p	333.9 ± 0.3	-32.4 ± 0.5	-0.6 ± 0.1	0.9999	0.041
16p/26p	340.4 ± 0.1	-34.2 ± 0.2	-0.71 ± 0.01	0.99997	0.01

Figure 2-146.Variable temperature CD data for 100 µM solutions of WW variants **16**, **16/26p**, **16p**, and **16p/26p** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



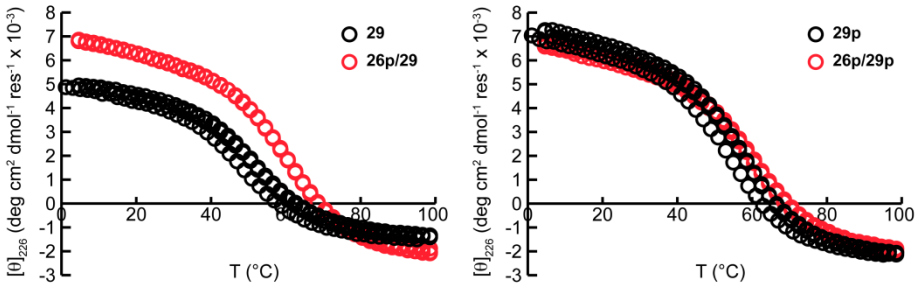
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
19	328.8 ± 0.2	-31.4 ± 0.4	-0.66 ± 0.07	0.9927	0.077
19/26p	335.5 ± 0.1	-31.4 ± 0.4	-0.66 ± 0.02	0.9999	0.031
19p	336.5 ± 0.3	-33.5 ± 1.1	-0.74 ± 0.05	0.9842	0.086
19p/26p	342.9 ± 0.1	-35.8 ± 0.4	-0.75 ± 0.02	0.9999	0.027

Figure 2-147.Variable temperature CD data for 100 µM solutions of WW variants **19**, **19/26p**, **19p**, **19p/26p** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



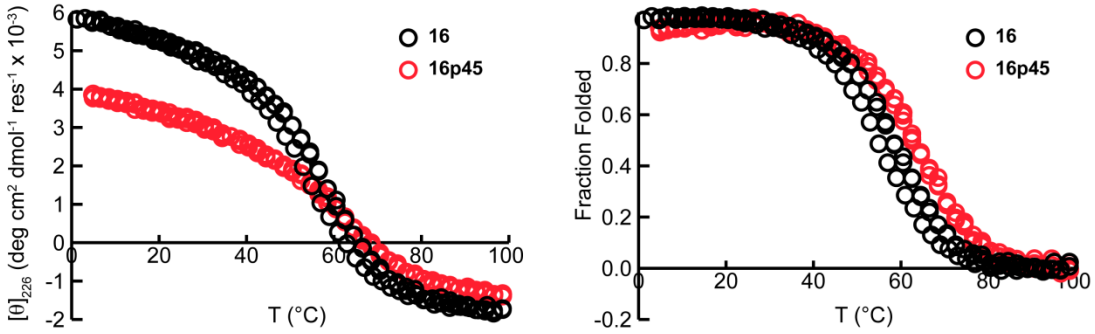
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16/19	330.1 ± 0.1	-30.2 ± 0.3	-0.56 ± 0.05	0.9999	0.030
16/19p	335.5 ± 0.1	-32.3 ± 0.3	-0.64 ± 0.03	0.9999	0.029
16p/19	336.4 ± 0.1	-36.1 ± 0.4	-0.72 ± 0.03	0.9999	0.032
16p/19p	338.6 ± 0.1	-35.0 ± 0.4	-0.70 ± 0.02	0.9999	0.028

Figure 2-148.Variable temperature CD data for 100 µM solutions of WW variants **16/19**, **16/19p**, **16p/19**, and **16p/19p** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



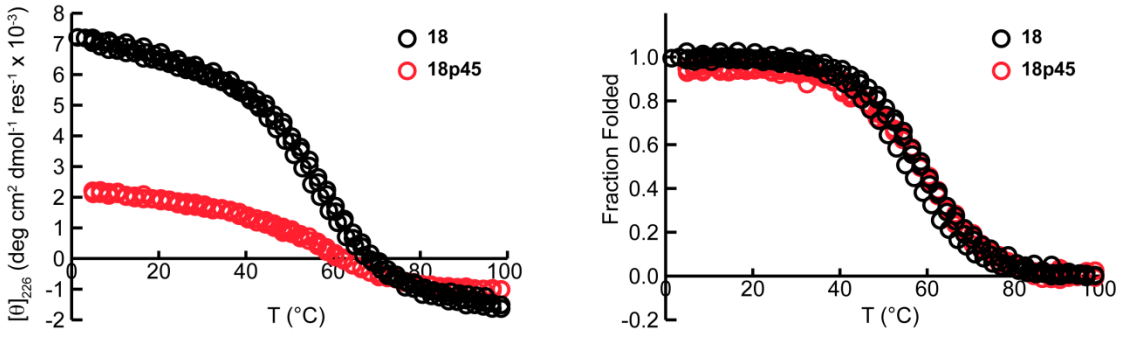
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
29	323.2 ± 0.3	-28.1 ± 0.4	0.0 ± 0.1	0.9999	0.024
26p/29	330.2 ± 0.2	-29.6 ± 0.2	-0.4 ± 0.1	0.99996	0.021
29p	327.3 ± 0.3	-29.3 ± 0.3	-0.4 ± 0.1	0.99995	0.024
26p/29p	329.9 ± 0.3	-28.6 ± 0.2	-0.4 ± 0.1	0.99996	0.019

Figure 2-149. Variable temperature CD data for 100 µM solutions of WW variants **29**, **26p/29**, **29p, 26p/29p** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



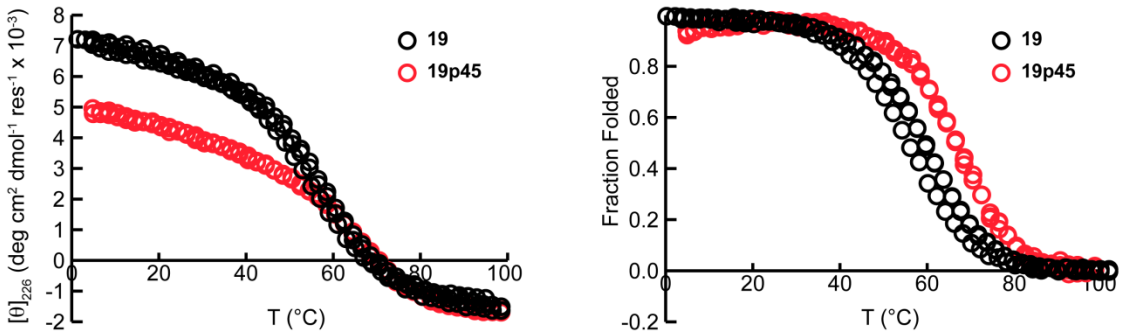
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16	328.0 ± 0.1	-31.1 ± 0.5	-0.65 + 0.06	0.9998	0.040
16p45	332.8 ± 0.3	-29.3 ± 1.0	-0.69 ± 0.04	0.9996	0.037

Figure 2-150.Variable temperature CD data for 50 µM solutions of WW variants **16** and **16p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



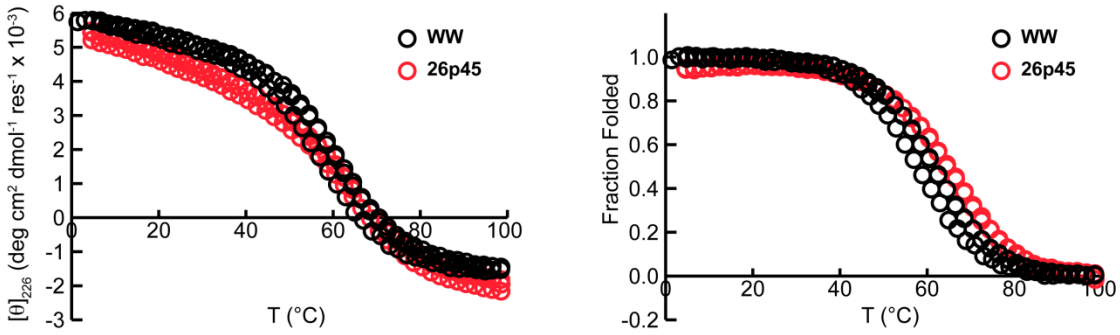
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
18	328.5 ± 0.8	-29.6 ± 0.8	-0.2 ± 0.2	0.9994	0.082
18p45	328.2 ± 0.4	-26.1 ± 1.2	-0.59 ± 0.07	0.9992	0.033

Figure 2-151. Variable temperature CD data for 50 µM solutions of WW variants **18** and **18p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



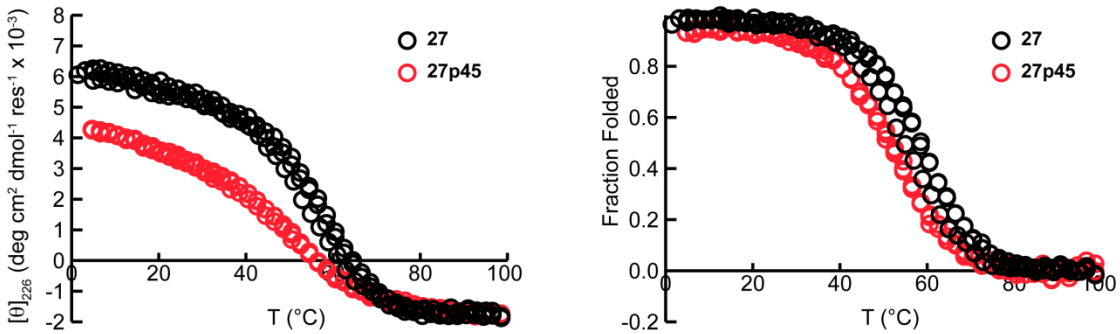
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
19	329.0 ± 0.2	-29.5 ± 0.3	-0.51 ± 0.05	0.9986	0.030
19p45	336.4 ± 0.4	-33.6 ± 1.5	-0.79 ± 0.05	0.9875	0.068

Figure 2-152.Variable temperature CD data for 50 µM solutions of WW variants **19** and **19p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations S3–S5 appear in the table, along with standard errors. Data for 50 µM 19 are from ref [25]. Data for 50 µM 19p45 are from ref [25]. These data were refit using equations 3-5 according to the methods described above.



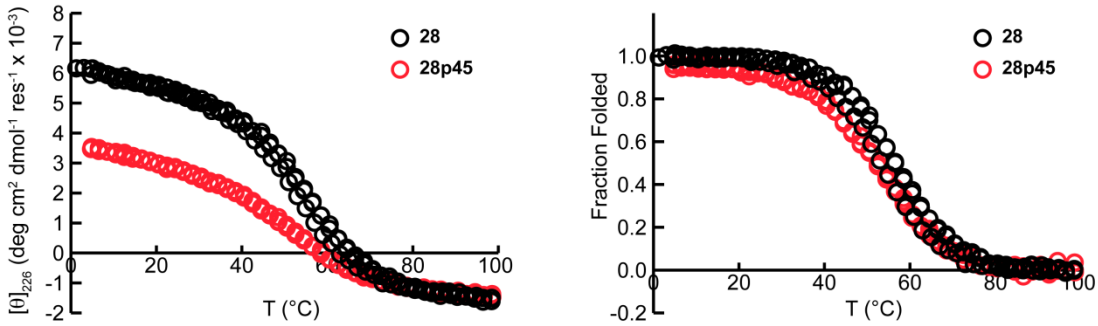
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
WW	331.1 ± 0.7	-31.1 ± 0.7	-0.4 ± 0.2	0.9944	0.056
26p45	334.4 ± 0.1	-28.1 ± 0.4	-0.61 ± 0.02	0.9999	0.020

Figure 2-153.Variable temperature CD data for 100 µM solutions of WW variants **26** and **26p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



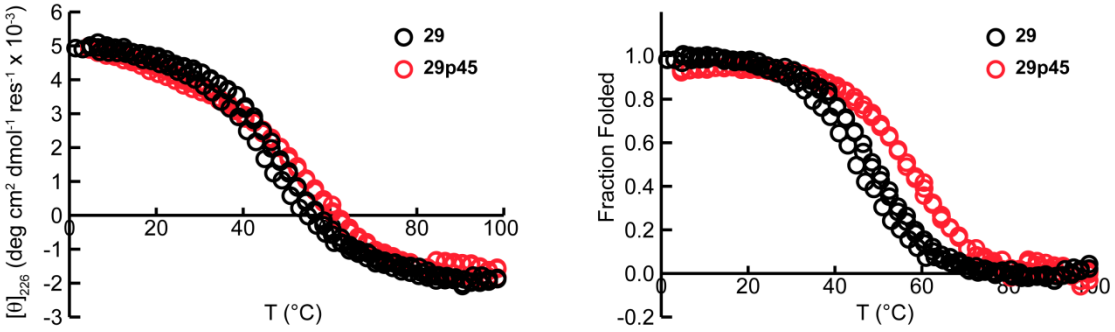
Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
27	328.2 ± 0.1	-31.8 ± 0.5	-0.66 ± 0.07	0.9997	0.054
27p45	321.2 ± 0.3	-27.4 ± 0.8	-0.70 ± 0.05	0.9997	0.036

Figure 2-154.Variable temperature CD data for 50 µM solutions of WW variants **27** and **27p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
28	326.4 ± 0.5	-30.4 ± 0.5	-0.3 ± 0.2	0.9997	0.046
28p45	321.8 ± 0.3	-24.0 ± 0.7	-0.51 ± 0.09	0.9996	0.034

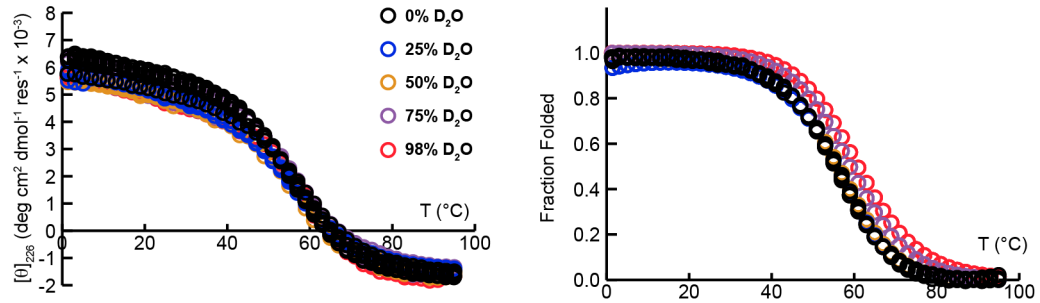
Figure 2-155.Variable temperature CD data for 50 µM solutions of WW variants **28** and **28p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.



Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
29	318.5 ± 0.8	-28.9 ± 1.4	-0.4 ± 0.2	0.9993	0.070
29p45	326.4 ± 0.3	-27.1 ± 1.0	-0.66 ± 0.05	0.9996	0.052

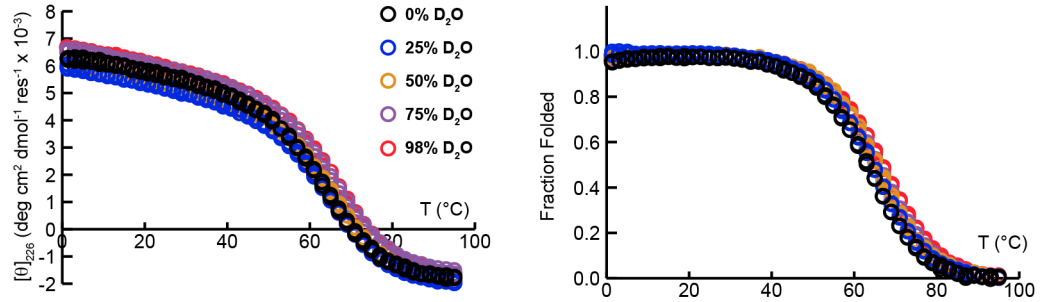
Figure 2-156.Variable temperature CD data for 50 µM solutions of WW variants **29** and **29p45** in 20 mM sodium phosphate, pH 7. Fit parameters from equations 3–5 appear in the table, along with standard errors.

Non-PEGylated Protein **19**



%D_2O	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
0	328.8 ± 0.2	-31.4 ± 0.6	-0.66 ± 0.07	0.9928	0.077
25	328.9 ± 0.2	-30.4 ± 0.7	-0.75 ± 0.04	0.9999	0.045
50	329.3 ± 0.3	-30.4 ± 1.3	-0.75 ± 0.06	0.9997	0.084
75	332.1 ± 0.5	-30.8 ± 0.4	-0.27 ± 0.13	0.9999	0.040
98	334.0 ± 0.5	-30.1 ± 0.7	-0.01 ± 0.12	0.9999	0.038

PEGylated Protein **19p**



%D_2O	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
0	336.5 ± 0.3	-33.5 ± 1.1	-0.74 ± 0.05	0.9843	0.086
25	337.6 ± 0.1	-33.6 ± 0.3	-0.61 ± 0.03	0.99995	0.021
50	338.5 ± 0.1	-35.0 ± 0.5	-0.71 ± 0.03	0.9999	0.032
75	338.9 ± 0.1	-33.1 ± 0.3	-0.67 ± 0.01	0.99997	0.019
98	339.8 ± 0.1	-33.1 ± 0.3	-0.67 ± 0.02	0.99995	0.025

Figure 2-157. Variable temperature CD data for 100 µM solutions of WW variants **19** and **19p** in 20 mM sodium phosphate, pH 7, in the presence of increasing amounts of D_2O. Fit parameters from equations 3–5 appear in the table, along with standard errors.

2.4.15 Proteoylysis

50 µM protein solutions in 20 mM sodium phosphate buffer (pH7) were incubated at ambient temperature with 5µg/mL pronase or with 10 µg/mL proteinase K respectively for up to

10 hours. At each of the several time points, the proteolysis reaction was quenched by adding 310 μL of aqueous trifluoroacetic acid (1% v/v) to 150 uL of the reaction mixture. The quenched mixture was then analyzed in triplicate by reverse phase HPLC analytical column, monitored by a UV-Vis detector at 220 nm. The degradation of the proteins was assessed using the integrated HPLC peak area to account for how much of the full-length protein remained at each time point. The protein half-lives were calculated by fitting the integrated peak areas as a function of time to a monoexponential decay equation:

$$\text{Area}(t) = A_0 \bullet \exp\,[-t/\tau],$$

Equation 12

where t is time in minutes, A_0 is a constant corresponding to relative integrated peak area at t = 0, and τ is the decay time, which is related to the protein half-life $t_{1/2}$ ($t_{1/2} = \tau \ln 2$). Decay 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**, **32p**, **16p/26p**, **19p/26p**, **26p/29p**, **16/19**, and **16p/19p, 16p45, 19p45, 26p45, 28p45, 29p45** are shown in Figures 2-158 through 2-174.

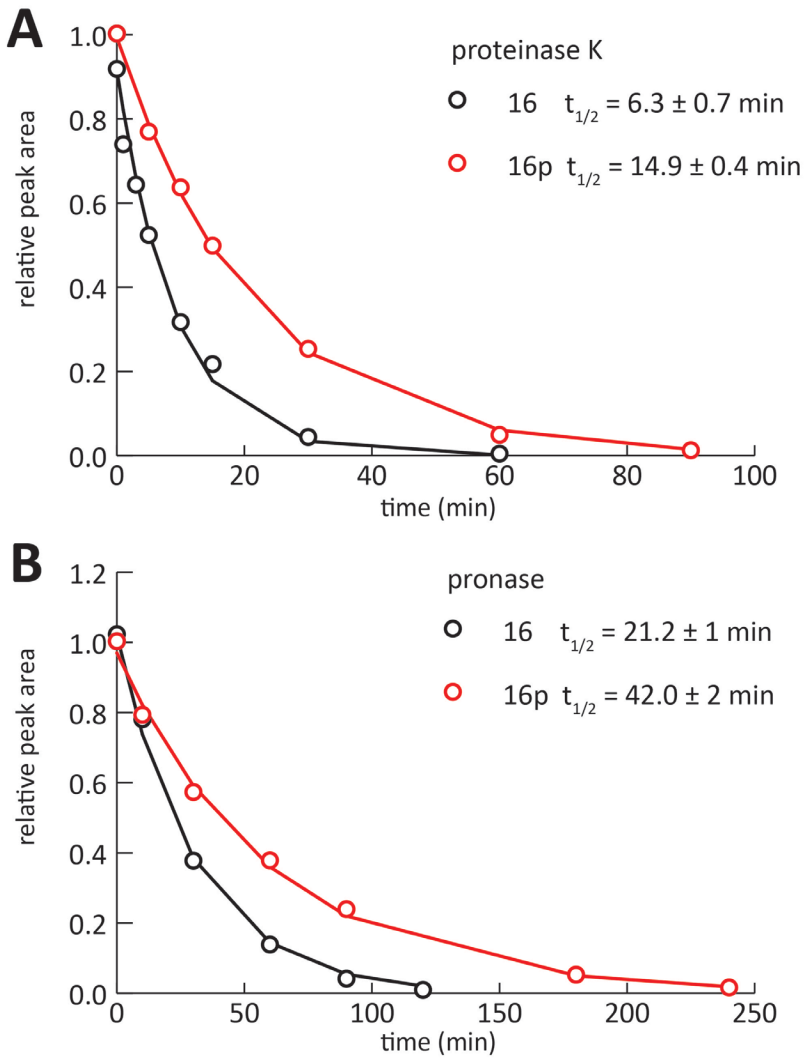


Figure 2-158. Proteolysis of **16** or **16p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **16p** and **16** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

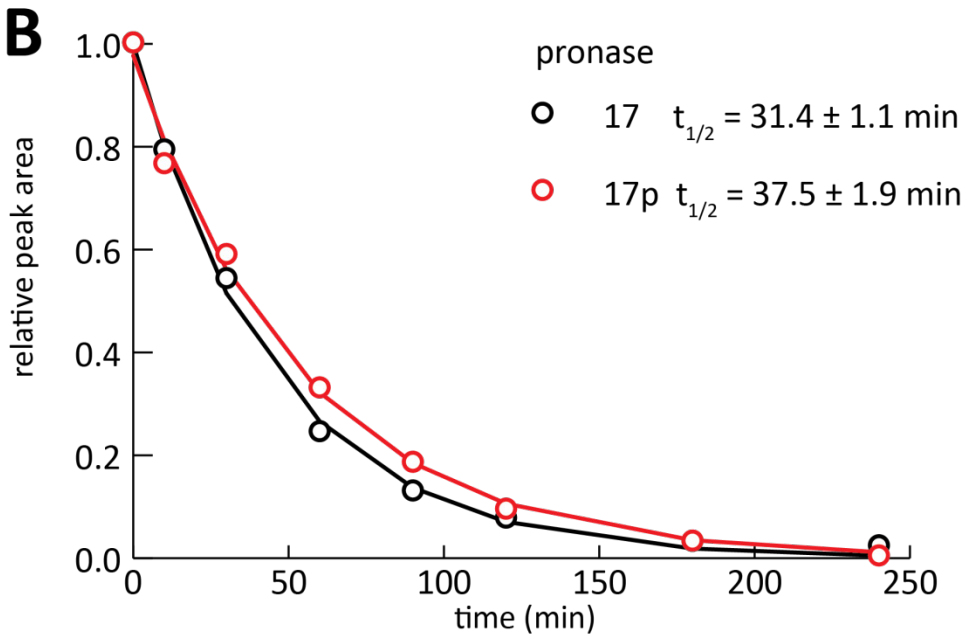


Figure 2-159. Proteolysis of **17** or **17p** (50 μM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **17p** and **17** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

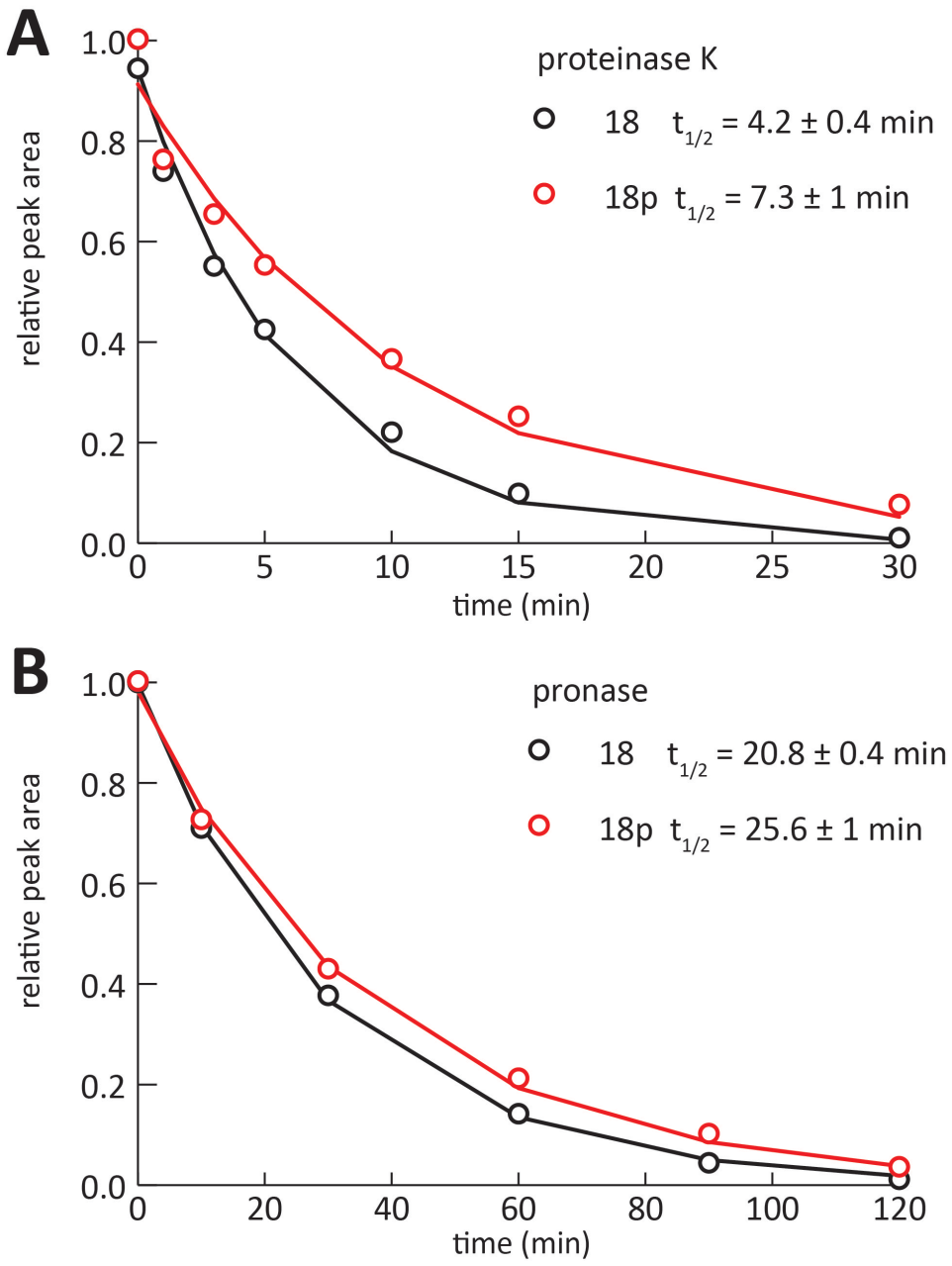


Figure 2-160. Proteolysis of **18** or **18p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **18p** and **18** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

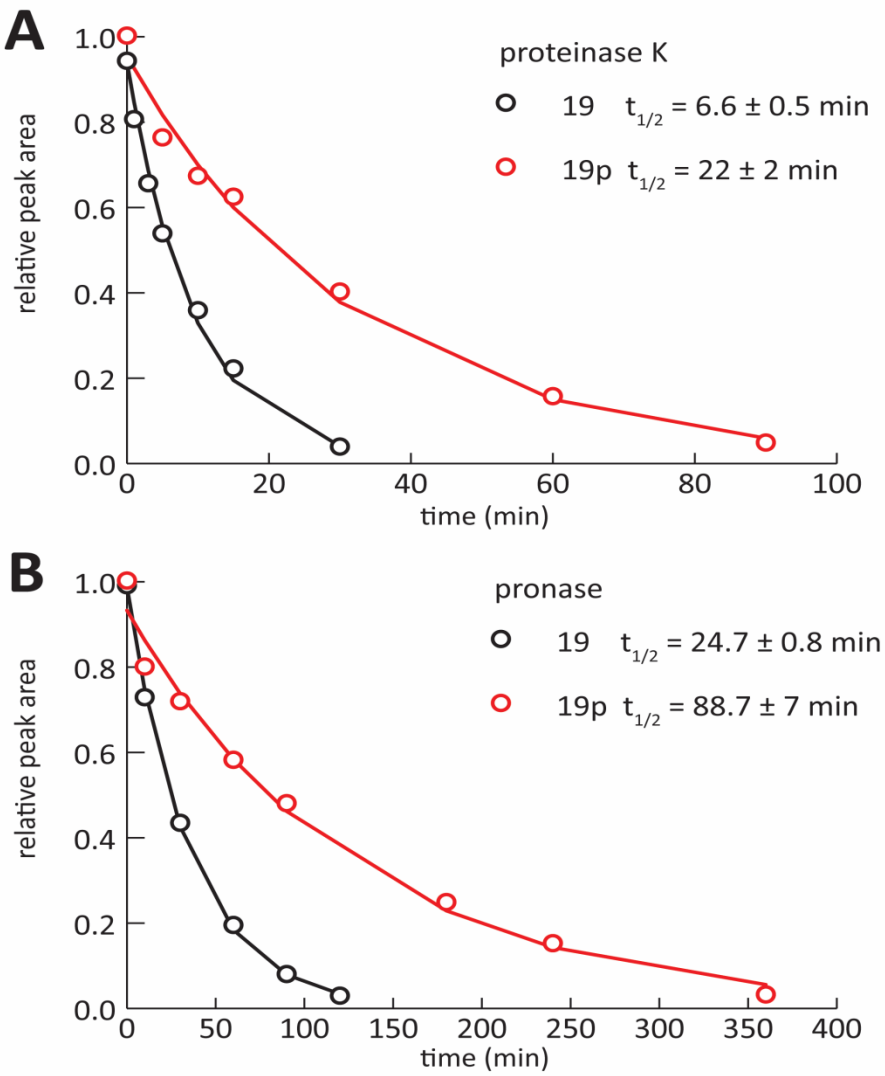


Figure 2-161.Proteolysis of **19** or **19p** (50 μM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **19p** and **19** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

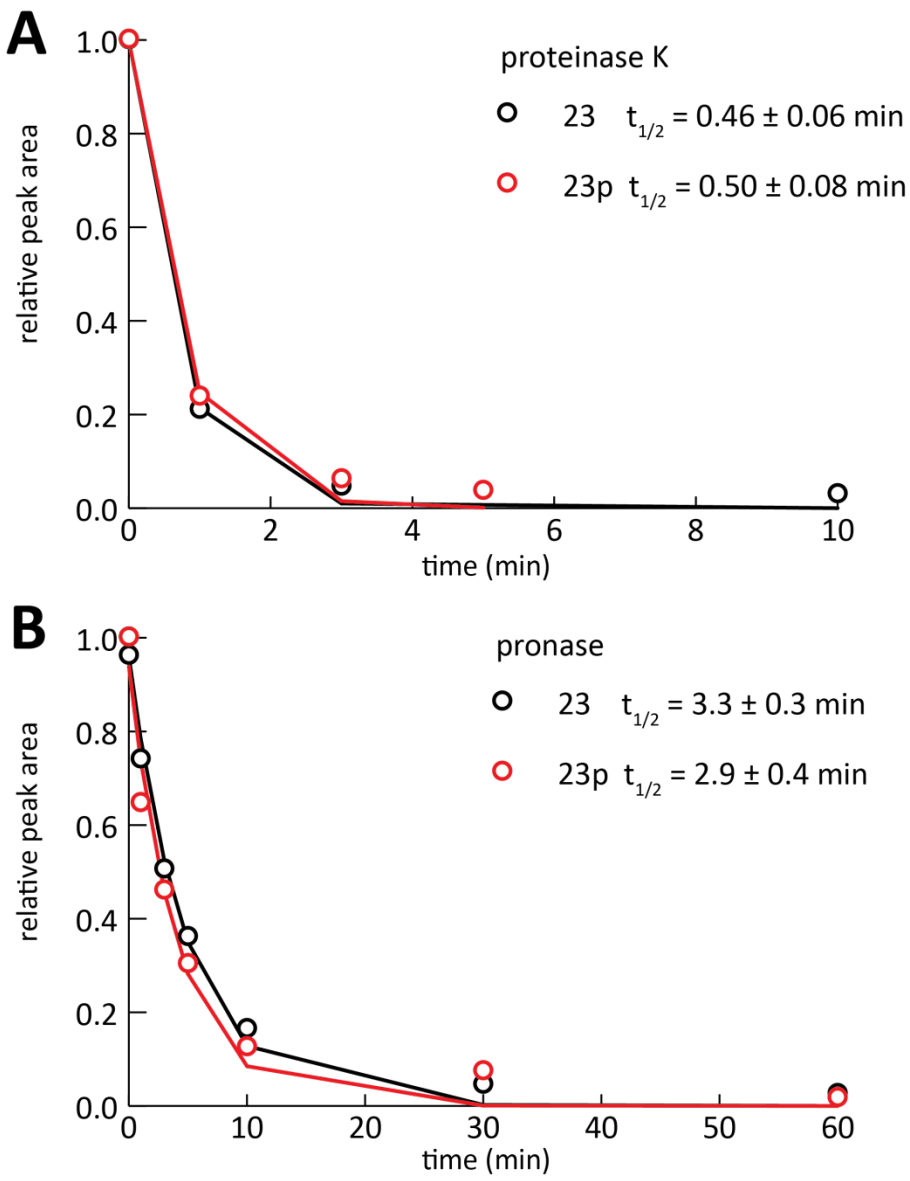


Figure 2-162.Proteolysis of **23** or **23p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **23p** and **23** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

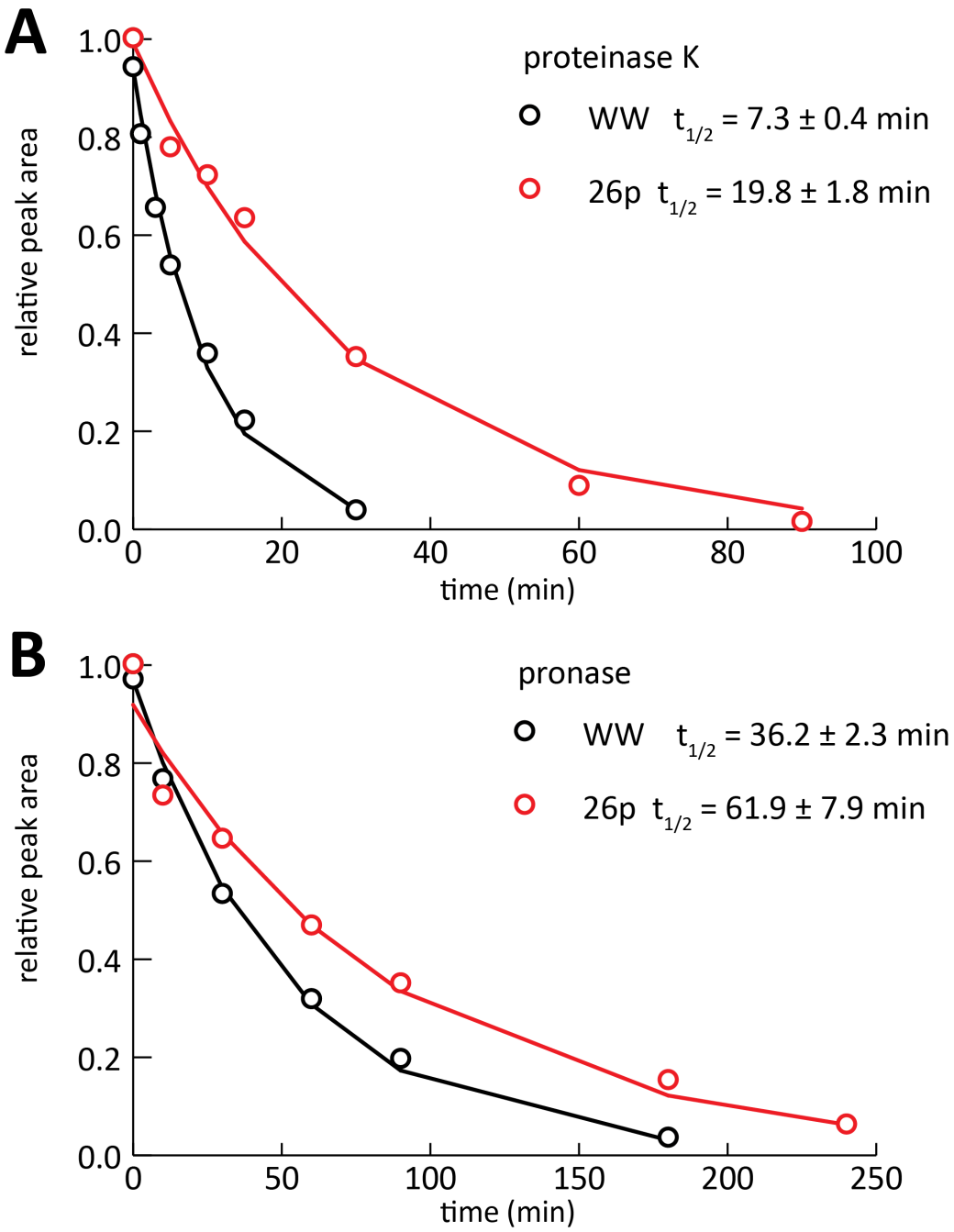


Figure 2-163. Proteolysis of **WW** or **26p** (50 μM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **26p** and **WW** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

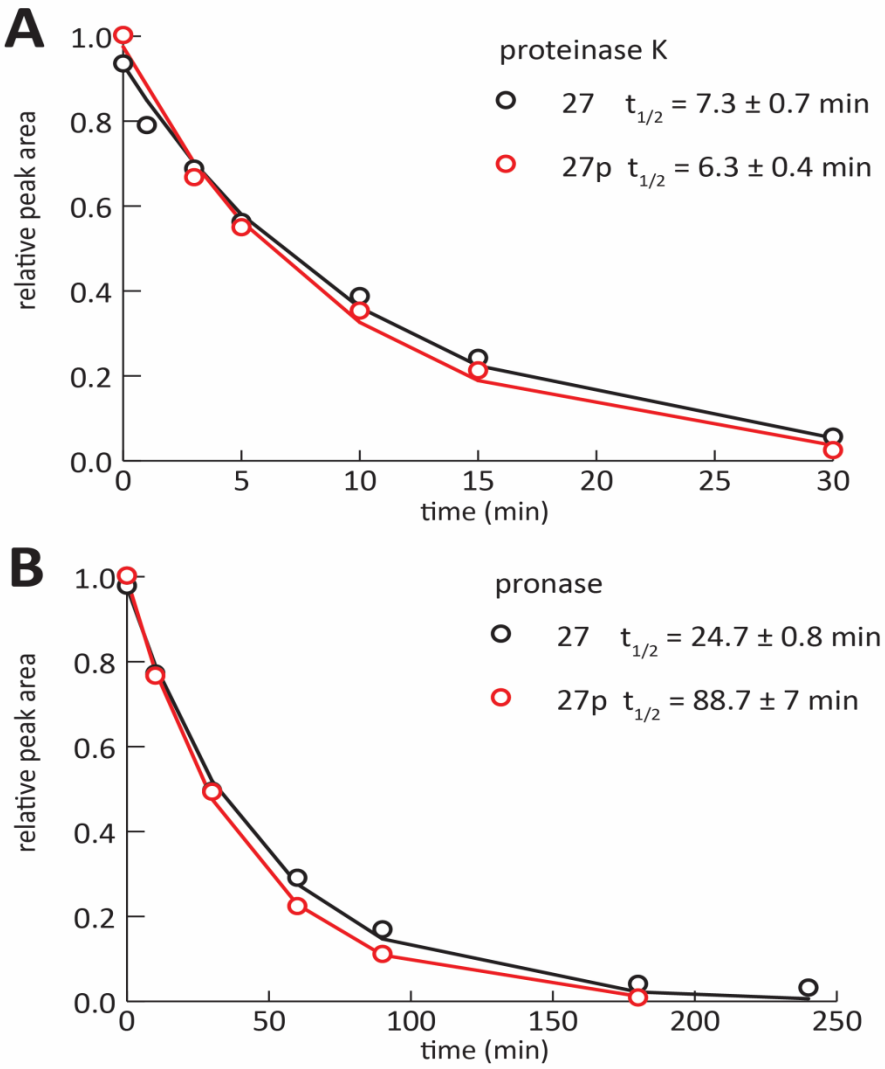


Figure 2-164. Proteolysis of **27** or **27p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **27p** and **27** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

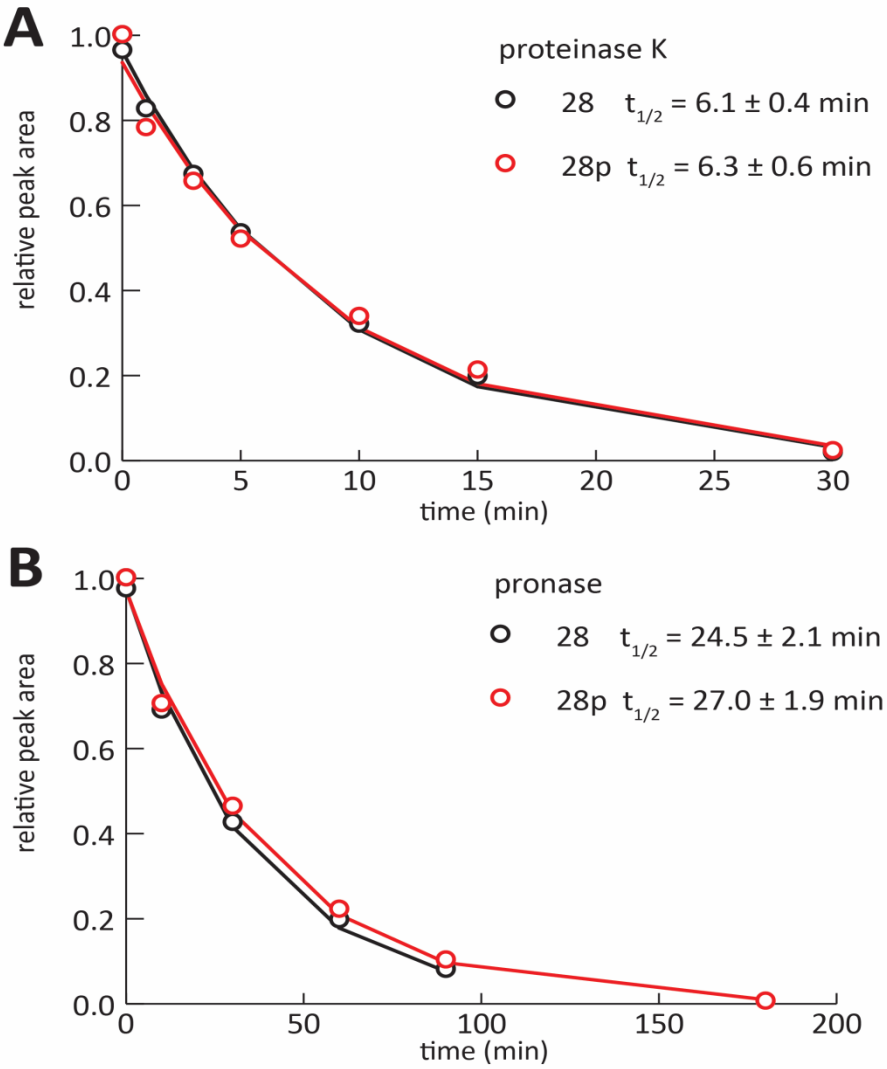


Figure 2-165. Proteolysis of **28** or **28p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **28p** and **28** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

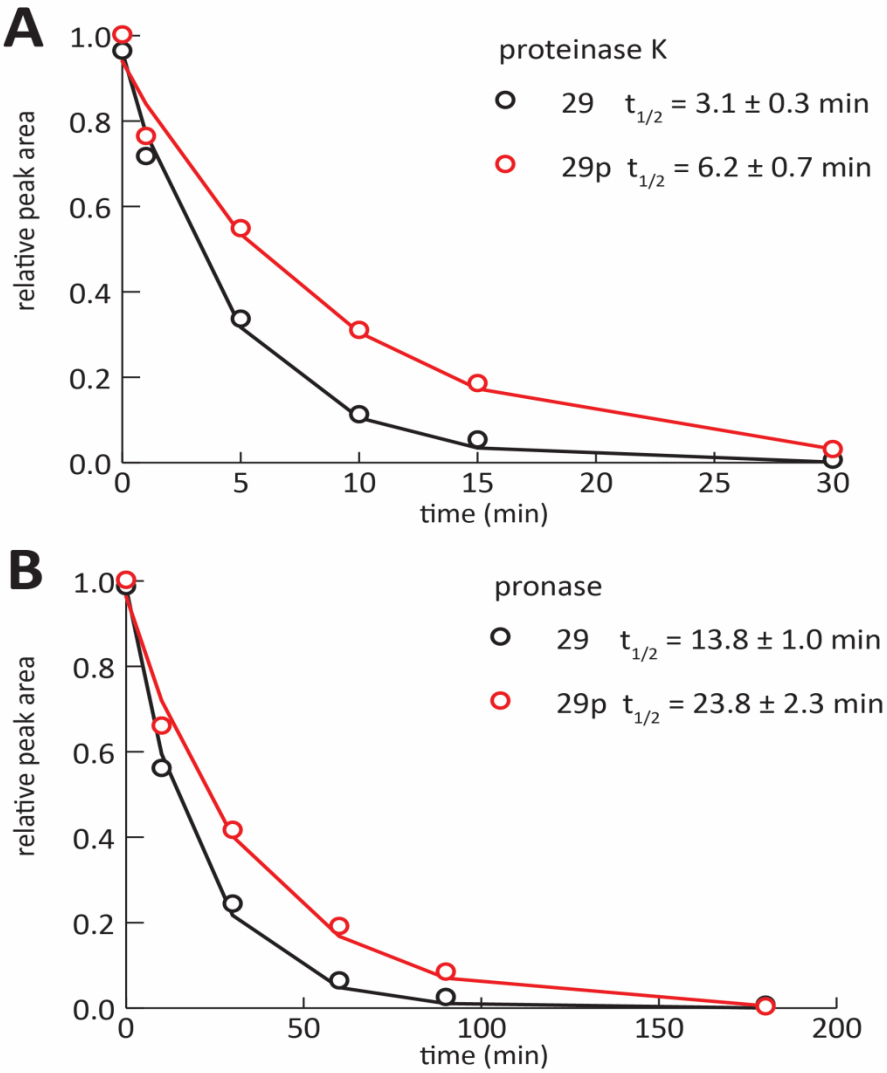


Figure 2-166. Proteolysis of **29** or **29p** (50 μM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **29p** and **29** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

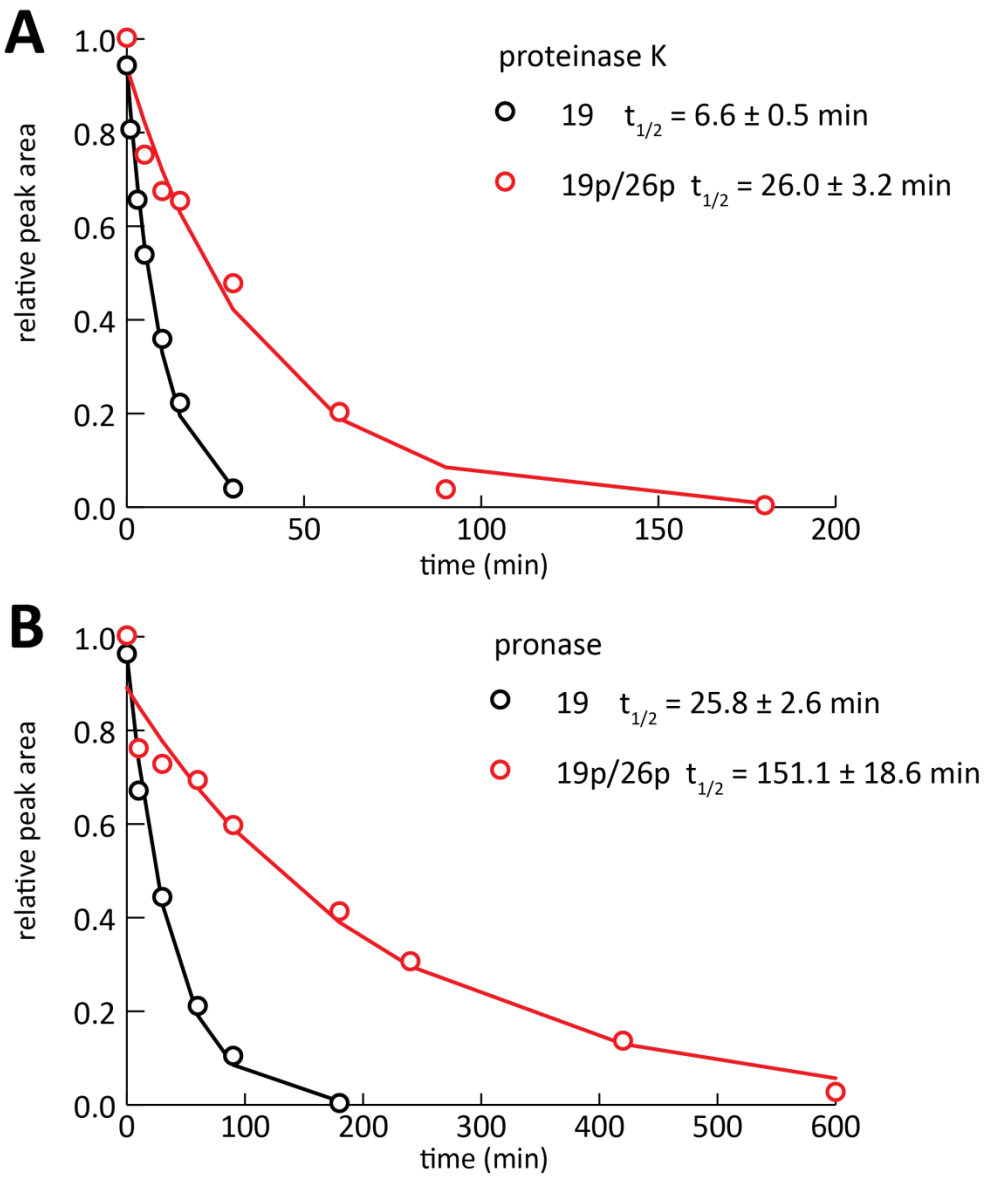


Figure 2-167.Proteolysis of **19** or **19p/26p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by (A) proteinase K (10 mg/mL) or (B) Pronase (5 mg/mL) as monitored by HPLC. Data points for **19p/26p** and **19** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

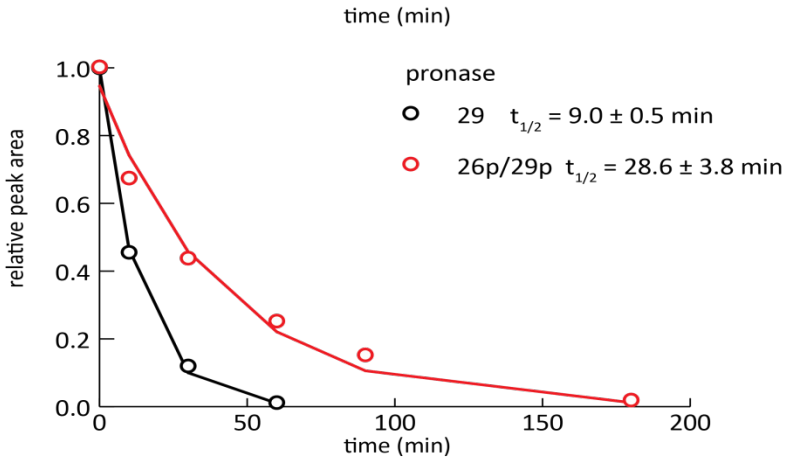


Figure 2-168. Proteolysis of **26/29** and **26p/29p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **26p/29p** and **26/29** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

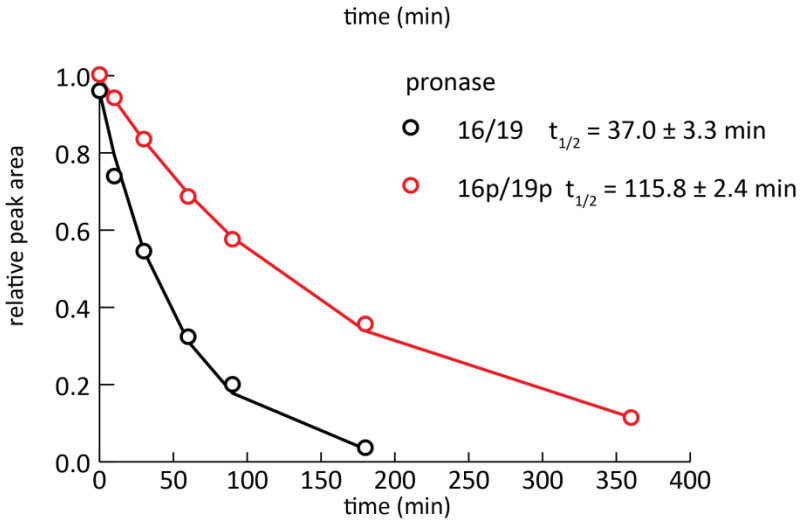


Figure 2-169.Proteolysis of **16/19** and **16p/19p** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **16p/19p** and **16/19** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

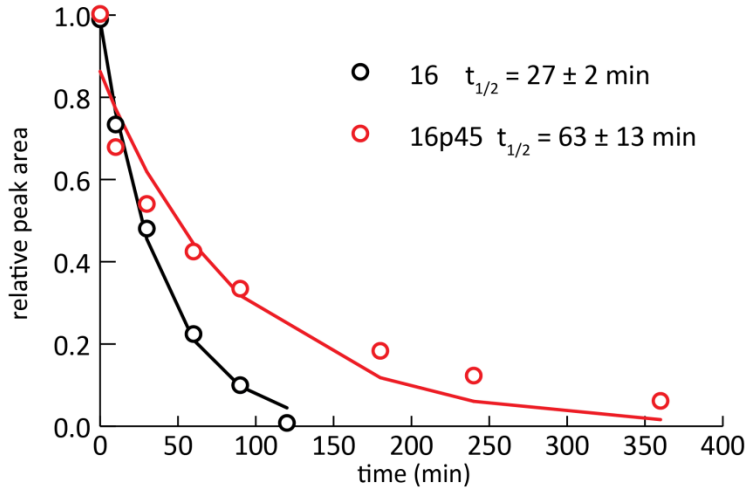


Figure 2-170.Proteolysis of **16** and **16p45** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **16p45** and **16** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

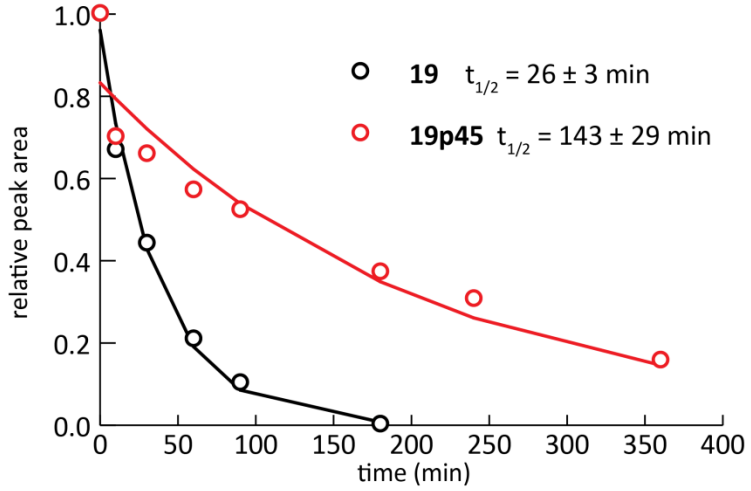


Figure 2-171.Proteolysis of **19** and **19p45** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **19p45** and **19** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

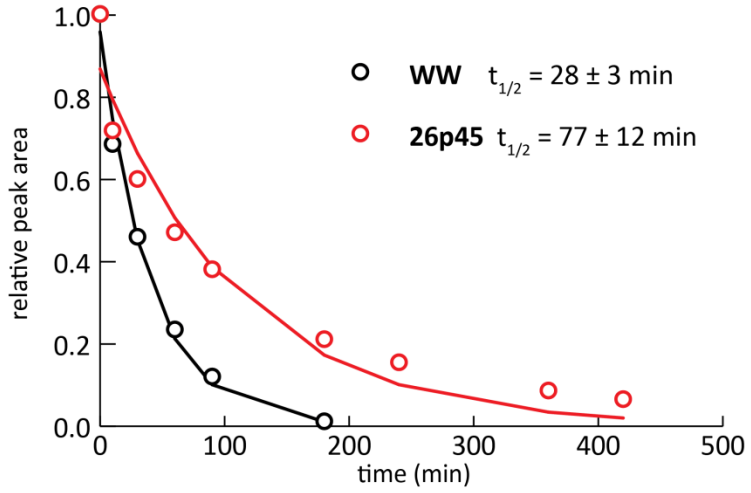


Figure 2-172. Proteolysis of **WW** and **26p45** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **26p45** and **WW** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

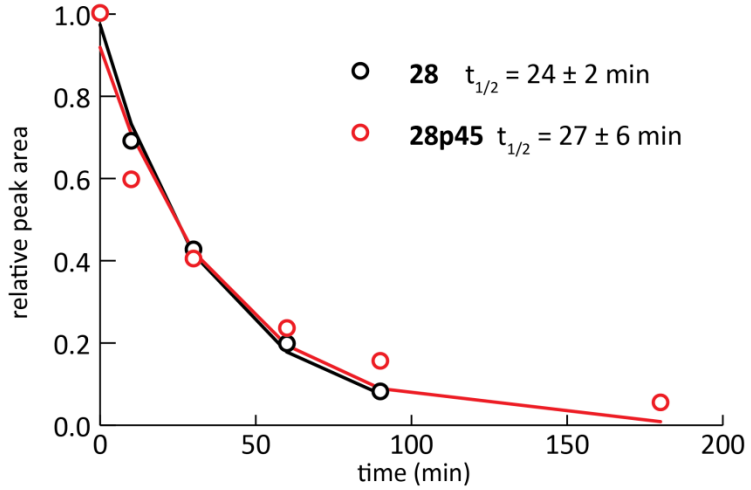


Figure 2-173. Proteolysis of **28** and **28p45** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **28p45** and **28** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

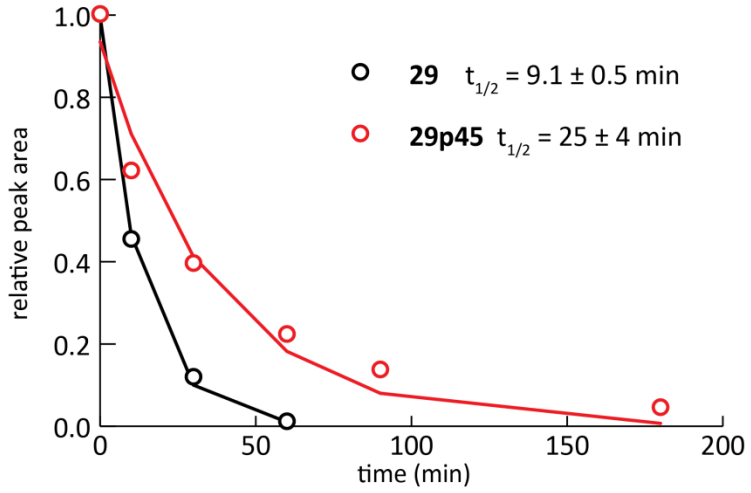


Figure 2-174.Proteolysis of **29** and **29p45** (50 µM protein concentration in 20 mM sodium phosphate buffer, pH 7) by Pronase (5 mg/mL) as monitored by HPLC. Data points for **29p45** and **29** are shown as red and black circles, respectively, and each represents the average of three replicate experiments. Solid lines represent fits of the data to a monoexponential decay function, which was used to calculate the indicated half-lives.

2.4.16 Synthesis, Characterization, and Thermodynamic Analysis of a PEGylated Src SH3 Variant

Protein **SH3 T20N** is a variant of the chicken Src SH3 domain (hereafter called SH3) in which Thr20 has been replaced by Asn. PEGylated protein **SH3 T20NPEG** is a variant of SH3 in which Thr20 has been replaced by AsnPEG$_4$. The sequences of these proteins are shown below:

```
                    9      15    20    25    30    35    40    45    50    55    60
                    |      |     |     |     |     |     |     |     |     |     |
SH3 T20N       H2N-TFVALYDYESRNETDLSFKKGERLQIVNNTEGDWWLAHSLTTGRTGYIPSNYVAPS-CONH2
SH3 T20NPEG    H2N-TFVALYDYESRNETDLSFKKGERLQIVNNTEGDWWLAHSLTTGRTGYIPSNYVAPS-CONH2
```

The bold underlined **N** in **SH3 T20NPEG** represents AsnPEG$_4$. These SH3 domain variants were synthesized as C-terminal amides by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc Nα protection strategy as described above. 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). Rink Amide MBHA resin LL, Fmoc-protected psuedoproline dipeptides (incorporated at the positions highlighted above in bold blue font; used to facilitate successful synthesis of this lengthy peptide), and 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.[24]

The Rink Amide MBHA LL resin has a nominal loading capacity of 0.38 mmol/g. We found that our synthetic efforts were more successful when we reduced the loading capacity of the resin prior to synthesis. To do this, we coupled a mixture of Fmoc-protected serine (the C-terminal residue of **SH3 T20N** and **SH3 T20NPEG**) and Boc-protected valine to the free amino groups on the resin. We reasoned that this approach would reduce the loading capacity of the resin by a factor of two because Boc is stable to peptide coupling and piperidine-mediated Fmoc deprotection conditions. Spectroscopic characterization of the amount of dibenzofulvene released upon the Fmoc deprotection of a defined amount of this modified resin indicated that its loading capacity was 0.16 mmol/g.

We aliquotted reduced-capacity modified resin into two fritted polypropylene syringes (one for each of the SH3 variants described above; 134 mg per syringe, 22 μmol at ~0.16 mmol/g resin loading) and allowed the resin to swell first in CH_2Cl_2, and then in dimethylformamide (DMF). Solvent was then drained from the resin using a vacuum manifold. A general synthetic procedure follows:

To remove the Fmoc protecting group on a resin-linked amino acid, 1.1 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 1.1 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. After incorporation of Asp41 we used 5% piperazine 0.1M HOBT in DMF for each subsequent deprotection reaction to prevent piperidine-mediated aspartimide formation.

For coupling of an activated amino acid to an amino group on resin, we prepared a stock coupling solution of 100 mL NMP, 3.17 g HBTU (0.01 mol, 0.1 M) and 1.53 g HOBt (0.01 mol, 0.1 M) for a final concentration of 0.1 M HBTU and 0.1 M HOBt. The desired Fmoc-protected amino acid (110 μmol, 5 eq) was dissolved by vortexing in 1.1 mL coupling solution. To the dissolved amino acid solution was added 37 μL DIEA (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 10 minutes. Following the coupling reaction, the activated amino acid solution was drained from the resin with a vacuum 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.

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 (Figure 2-175 through 2-176). Proteins were then re-purified by semi-preparative reverse-phase HPLC on a C18 column using a linear gradient of water in acetonitrile with 0.1% v/v TFA (Figure 2-177 through 2-178). HPLC fractions containing the desired protein product were pooled, frozen, and lyophilized. Protein purity was assessed by analytical HPLC (Figures 2-179 through 2-180), and proteins were identified by electrospray ionization time of flight mass spectrometry (ESI-TOF, Figures 2-181 through 2-182).

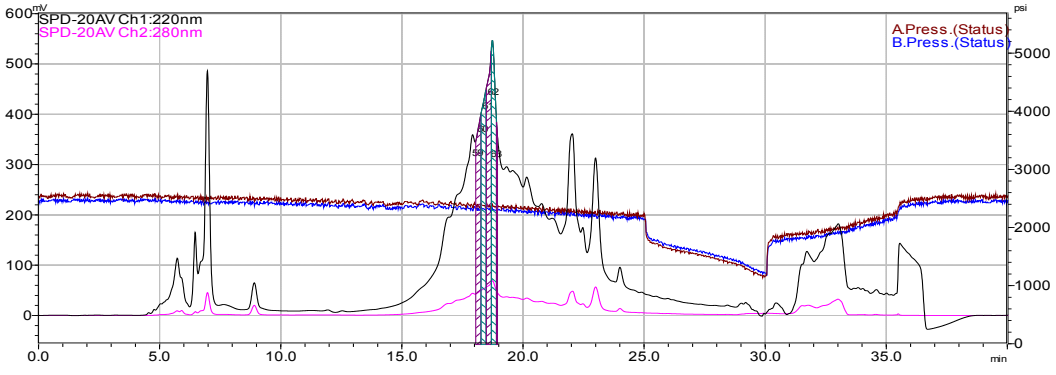


Figure 2-175. Preparative reverse-phase HPLC purification data for **SH3 T20N**. Protein solution was injected onto a C18 preparative column and eluted using a linear gradient of 25-45% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 25 minutes, followed by a 5 minute rinse (95% B), and a 10 minute column re-equilibration (25% B) with a flow rate of 10 mL/min.

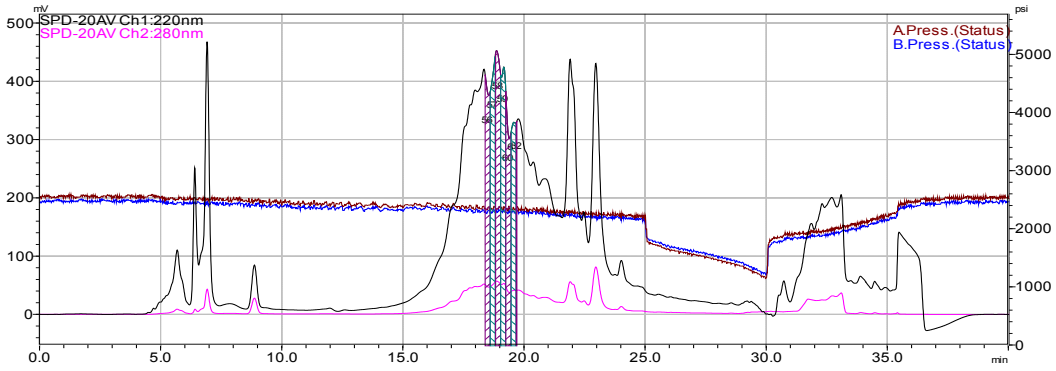


Figure 2-176.Preparative reverse-phase HPLC purification data for **SH3 T20NPEG**. Protein solution was injected onto a C18 preparative column and eluted using a linear gradient of 25-45% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 25 minutes, followed by a 5 minute rinse (95% B), and a 10 minute column re-equilibration (25% B) with a flow rate of 10 mL/min.

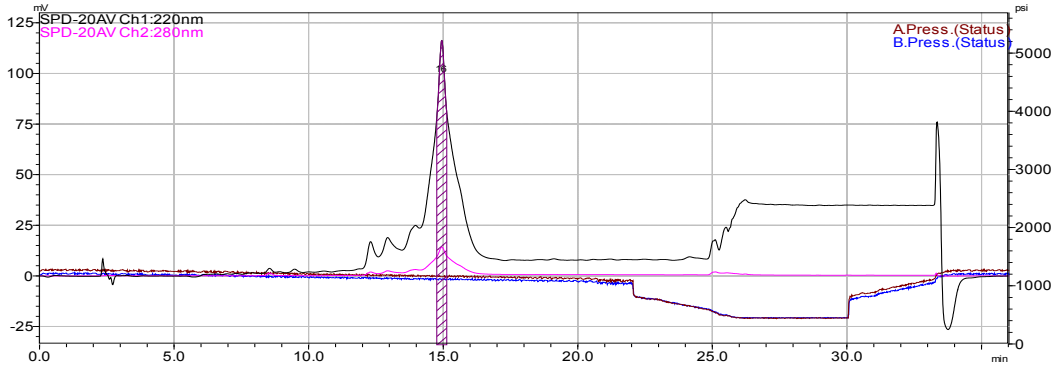


Figure 2-177. Semi-preparative reverse-phase HPLC repurification data for **SH3 T20N**. Protein solution was injected onto a C18 semi-preparative column and eluted using a linear gradient of 23-43% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 22 minutes, followed by a 7 minute rinse (95% B), and a 6 minute column re-equilibration (23% B) with a flow rate of 6 mL/min.

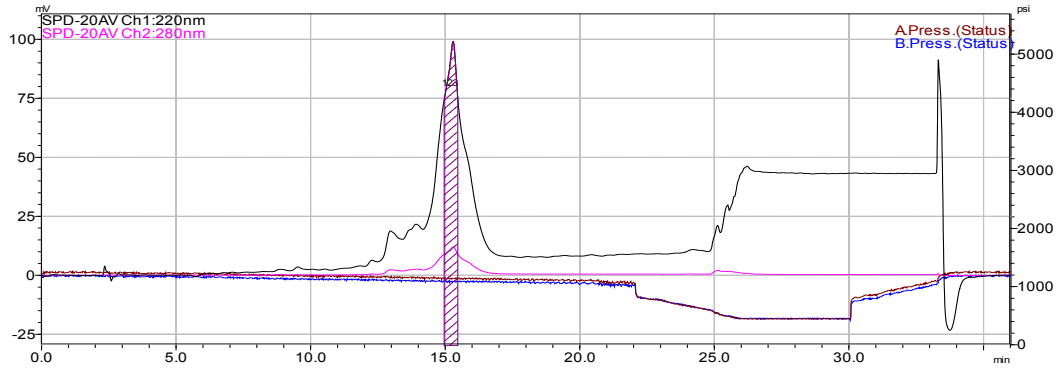


Figure 2-178.Semi-preparative reverse phase HPLC repurification data for **SH3 T20NPEG**. Protein solution was injected onto a C18 semi-preparative column and eluted using a linear gradient of 23-43% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 22 minutes, followed by a 7 minute rinse (95% B), and a 6 minute column re-equilibration (23% B) with a flow rate of 6 mL/min.

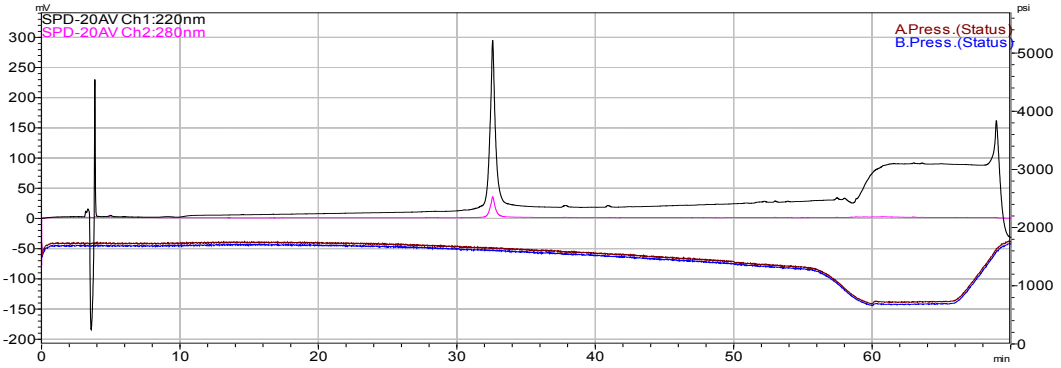


Figure 2-179.Analytical HPLC data for **SH3 T20N**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

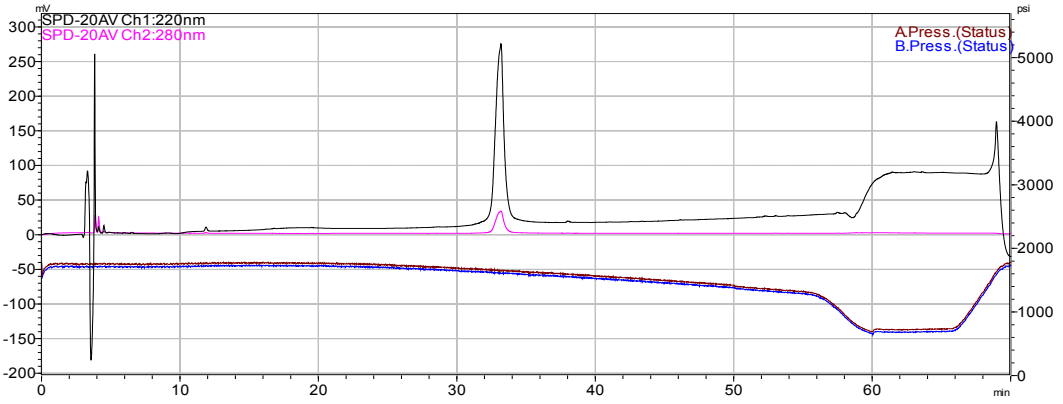


Figure 2-180.Analytical HPLC data for **SH3 T20NPEG**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

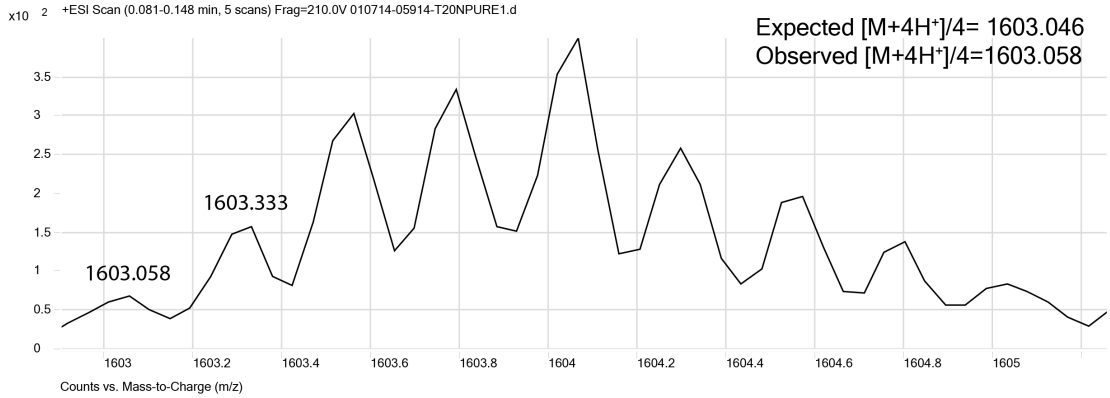


Figure 2-181. ESI TOF MS data for **SH3 T20N**. Expected [M+4H$^+$]/4 = 1603.046 Da. Observed [M+4H$^+$]/4 = 1603.058 Da.

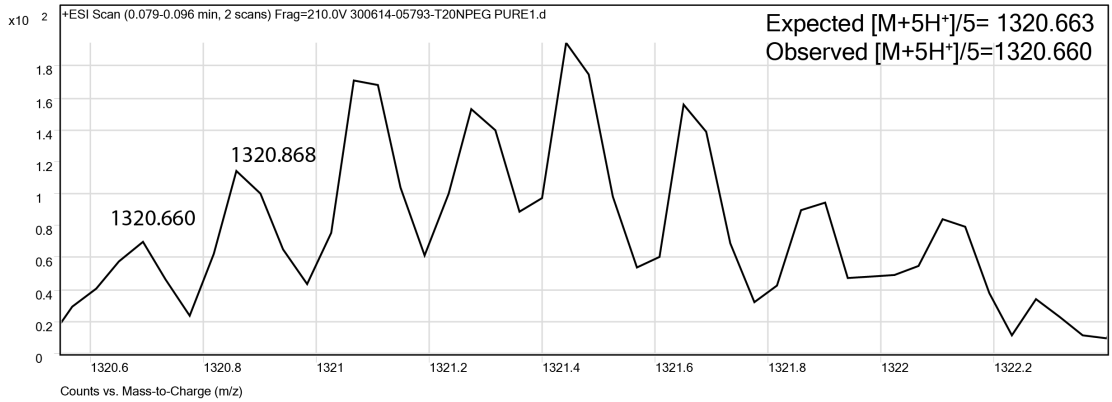


Figure 2-182.ESI TOF MS data for **SH3 T20NPEG**. Expected $[M+5H^+]/5$ = 1320.663 Da. Observed $[M+5H^+]/5$ = 1320.660 Da.

We made measurements on **SH3 T20N** and **SH3 T20NPEG** with an Aviv 420 Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of 0.1 cm. 50 μM 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 (ε_{Trp} = 5690 $M^{-1}cm^{-1}$, ε_{Tyr} = 1280 $M^{-1}cm^{-1}$).[60] CD spectra of were obtained from 340 to 200 nm at 25ºC (**Figure 2-183**). Variable temperature CD data were obtained in triplicate for 50 μM solutions of **SH3 T20N** and **SH3 T20NPEG**20 mM sodium phosphate (pH 7) by monitoring molar ellipticity at 222 nm from 1 to 95°C at 2 °C intervals, with 120 s equilibration time between data points and 30 s averaging times.

Data from the three replicate variable temperature CD experiments on each protein (**Figure 2-183**) were fit to the following model for two-state thermally induced unfolding transitions:

$$[\theta] = \frac{D_0 + K_f(N_0 + N_1 \cdot T)}{1 + K_f}$$

Equation 13

where T is temperature in Kelvin, D_0 is the *y*-intercept of the horizontal post-transition baseline; N_0 is the *y*-intercept and N_1 is the slope of the pre-transition baseline (which was held at $N_1 = 0$ for **SH3 T20N**); and K_f is the temperature-dependent folding equilibrium constant. K_f is related to the temperature-dependent free energy of folding $\Delta G_f(T)$ according to the following equation:

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

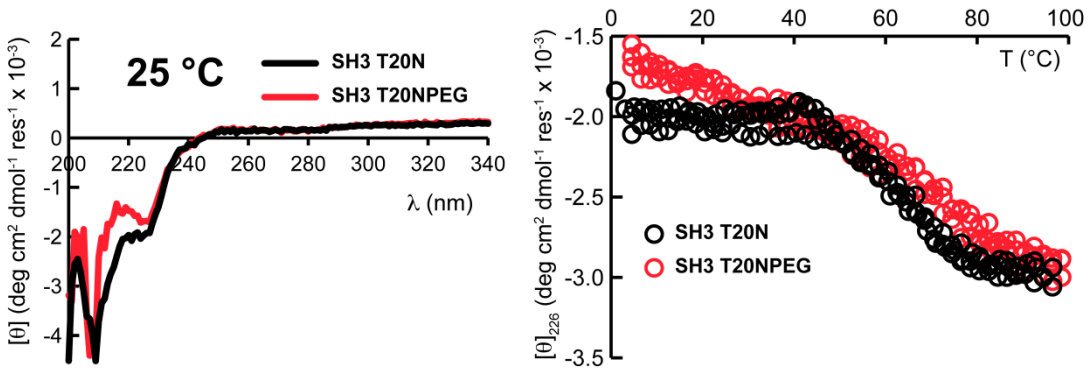
Equation 14

where R is the universal gas constant (0.0019872 kcal/mol/K). $\Delta G_f(T)$ was fit to the following equation:

$$\Delta G_f = \frac{\Delta H(T_m) \cdot (T_m - T)}{T_m}$$

Equation 15

where the fit parameters are T_m(the midpoint of the unfolding transition; the temperature at which $\Delta G_f = 0$); and $\Delta H(T_m)$, the change in enthalpy upon folding at T_m. This equation is derived from equation 5 above, with ΔC_p (the change in heat capacity upon folding) is held equal to 0 kcal mol^{-1} K^{-1}; in preliminary fitting efforts of the data for **SH3 T20N** and **SH3 T20NPEG** to equation 5, we found that the fit was over parameterized, with high-error near-zero values for ΔC_p. Therefore, we eliminated the heat capacity term, and used equation 15 to fit the data for **SH3 T20N** and **SH3 T20NPEG**.The parameters for equations 13-15 (Figure 2-183) were used to calculate the values of the folding free energy ΔG_f.



Protein	T_m / K	$\Delta H(T_m)$ / kcal mol^{-1}	R^2	rmsd error
SH3 T20N	334.2 ± 0.3	-32 ± 1	0.9932	0.033
SH3 T20NPEG	347.2 ± 0.8	-31 ± 3	0.9915	0.039

Figure 2-183.CD spectra (50 µM) and variable temperature CD data (50 µM) for **SH3 T20N** (black) and **SH3 T20NPEG** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S12–S14 appear in the table, along with standard errors.

2.5 References

(1) Koepf, E. K.; Petrassi, H. M.; Ratnaswamy, G.; Huff, M. E.; Sudol, M.; Kelly, J. W. *Biochemistry.* **1999**, *38*, 14338.

(2) Koepf, E. K.; Petrassi, H. M.; Sudol, M.; Kelly, J. W. *Protein Sci.* **1999**, *8*, 841.

(3) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373.

(4) Kaul, R.; Angeles, A. R.; Jäger, M.; Powers, E. T.; Kelly, J. W. *J. Am. Chem. Soc.* **2001**, *123*, 5206.

(5) Deechongkit, S.; Kelly, J. W. *J. Am. Chem. Soc.* **2002**, *124*, 4980.

(6) Kaul, R.; Deechongkit, S.; Kelly, J. W. *J. Am. Chem. Soc.* **2002**, *124*, 11900.

(7) Kowalski, J. A.; Kiu, K.; Kelly, J. W. *Biopolymers* **2002**, *63*, 111.

(8) Nguyen, H.; Jäger, M.; Moretto, A.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci USA.* **2003**, *100*, 3948.

(9) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101.

(10) Nguyen, H.; Jäger, M.; Kelly, J. W.; Gruebele, M. *J. Phys. Chem. B.* **2005**, *109*, 15182.

(11) Sekijima, Y.; Wiseman, R. L.; Matteson, J.; Hammarström, P.; Miller, S. R.; Sawkar, A. R.; Balch, W. E.; Kelly, J. W. *Cell* **2005**, *121*, 73.

(12) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. USA.* **2006**, *103*, 10648.

(13) Jäger, M.; Dendle, M.; Fuller, A. A.; Kelly, J. W. *Protein Sci.* **2007**, *16*, 2306.

(14) Jäger, M.; Nguyen, H.; Dendle, M.; Gruebele, M.; Kelly, J. W. *Protein Sci.* **2007**, *16*, 1495.

(15) Jäger, M.; Deechongkit, S.; Koepf, E. K.; Nguyen, H.; Gao, J.; Powers, E. T.; Gruebele, M.; Kelly, J. W. *Biopolymers* **2008**, *90*, 751.

(16) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proc. Natl. Acad. Sci. USA.* **2008**, *105*, 2369.

(17) 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. *Proc. Natl. Acad. Sci. USA.* **2009**, *106*, 11067.

(18) Gao, J.; Bosco, D. A.; Powers, E. T.; Kelly, J. W. *Nat. Struct. Mol. Biol.* **2009**, *16*, 684.

(19) Jäger, M.; Dendle, M.; Kelly, J. W. *Protein Sci.* **2009**, *18*, 1806.

(20) Price, J. L.; Shental-Bechor, D.; Dhar, A.; Turner, M. J.; Powers, E. T.; Gruebele, M.; Levy, Y.; Kelly, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 15359.

(21) Culyba, E. K.; Price, J. L.; Hanson, S. R.; Dhar, A.; Wong, C. H.; Gruebele, M.; Powers, E. T.; Kelly, J. W. *Science* **2011**, *331*, 571.

(22) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14127.

(23) Ranganathan, R.; Lu, K. P.; Hunter, T.; Noel, J. P. *Cell* **1997**, *89*, 875.

(24) Price, J. L.; Powers, E. T.; Kelly, J. W. *ACS Chem. Biol.* **2011**, *6*, 1188.

(25) 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. *Bioconjugate Chem.* **2013**, *24*, 796.

(26) Shental-Bechor, D.; Levy, Y. *Proc. Natl. Acad. Sci. USA.* **2008**, *105*, 8256.

(27) Shental-Bechor, D.; Levy, Y. *Curr. Opin. Struct. Biol.* **2009**, *19*, 524.

(28) Hagai, T.; Levy, Y. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 2001.

(29) Shental-Bechor, D.; Smith, M. T. J.; MacKenzie, D.; Broom, A.; Marcovitz, A.; Ghashut, F.; Go, C.; Bralha, F.; Meiering, E. M.; Levy, Y. *Proc. Nat. Acad. Sci. USA* **2012**, *109*, 17839.

(30) Chao, S.-H.; Matthews, S. S.; Paxman, R.; Aksimentiev, A.; Gruebele, M.; Price, J. L. *J. Phys. Chem. B.* **2014**, *118*, 8388.

(31) Cioni, P.; Strambini, G. B. *Biophys. J.* **2002**, *82*, 3246.

(32) Prabhu, N. V.; Sharp, K. A. *Annu. Rev. Phys. Chem.* **2004**, *56*, 521.

(33) Yu, H.; Rosen, M. K.; Schreiber, S. L. *FEBS Lett.* **1993**, *324*, 87.

(34) Meng, W.; Guo, X.; Qin, M.; Pan, H.; Cao, Y.; Wang, W. *Langmuir* **2012**, *28*, 16133.

(35) Horne, W. S.; Boersma, M. D.; Windsor, M. A.; Gellman, S. H. *Angew. Chem. Int. Ed.* **2008**, *47*, 2853.

(36) Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A. M.; Kraynov, V. S.; Sim, B. C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; Becky, G. C.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9060.

(37) Zhang, C.; Yang, X.-l.; Yuan, Y.-h.; Pu, J.; Liao, F. *Biodrugs* **2012**, *26*, 209.

(38) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. D* **2010**, *66*, 486.

(39) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F. ß.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, ñ. n.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, Inc.: Wallingford, CT, USA, 2009.

(40) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(41) Harihara, P. C.; Pople, J. A. *Theor. Chim. Acta.* **1973**, *28*, 213.

(42) Harihara, P. C.; Pople, J. A. *Mol Phys.* **1974**, *27*, 209.

(43) Gordon, M. S. *Chem. Phys. Lett.* **1980**, *76*, 163.

(44) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654.

(45) Binning, R. C.; Curtiss, L. A. *J. Comput. Chem.* **1990**, *11*, 1206.

(46) Blaudeau, J. P.; Curtiss, L. A. *Int. J. Quantum Chem.* **1997**, *61*, 943.

(47) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. *J. Chem. Phys.* **1998**, *109*, 1223.

(48) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. *J. Comput. Chem.* **2001**, *22*, 976.

(49) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(50) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; III, T. E. C.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A.; University of California: San Francisco, 2014.

(51) da Silva, A. W. S.; Vranken, W. F. *BMC Res. Notes* **2012**, *5*, 367.

(52) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(53) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950.

(54) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(55) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463.

(56) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(57) Herzner, H.; Kunz, H. *Carbohydr. Res.* **2007**, *342*, 541.

(58) 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. *Bioconjug. Chem.* **2013**, *24*, 796.

(59) Price, J. L.; Powers, E. T.; Kelly, J. W. *ACS Chem. Biol.* **2011**, *6*, 1188.

(60) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.

(61) Ballew, R. M.; Sabelko, J.; Gruebele, M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 5759.

(62) Ballew, R. M.; Sabelko, J.; Reiner, C.; Gruebele, M. *Rev. Sci. Instrum.* **1996**, *67*, 3694.

(63) Ervin, J.; Sabelko, J.; Gruebele, M. *J. Photochem. Photobiol., B.* **2000**, *54*, 1.

(64) Kramers, H. A. *Physica* **1940**, *7*, 284.

(65) Hänggi, P.; Talkner, P.; Borovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251.

(66) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7220.

(67) Bieri, O.; Wirz, J.; Hellrung, B.; Schutokowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9597.

(68) Ansari, A.; Jones, C. M.; Henry, E. R.; Hofrichter, J.; Eaton, W. A. *Science* **1992**, *256*, 1796.

(69) Weast, R. C. *CRC Handbook of Chemistry and Physics*; CRC Press: Boca Raton, 1982.

3 CONJUGATION STRATEGY STRONGLY IMPACTS THE CONFORMATIONAL STABILITY OF A PEG-PROTEIN CONJUGATE

3.1 Introduction

Increasing protein conformational stability is a useful method[1] for enhancing the pharmacokinetic properties of proteins because unfolded or misfolded proteins can aggregate,[2] experience proteolytic degradation,[3] and be recognized by antibodies more readily[4] than are highly stable folded proteins. We recently found that protein PEGylation imparts enhanced resistance to proteolysis when PEG increases protein conformational stability.[1,5] We have also developed a predictive structure-based method for identifying such locations within two model proteins: the WW domain of the human protein Pin 1 and the Src SH3 domain.[5] We generated the required PEG-protein conjugates for these preliminary studies by attaching a short PEG oligomer to the side-chain amide nitrogen of an Fmoc-protected Asn residue, which was incorporated at desired positions within WW or SH3 via solid-phase peptide synthesis.

Our long-term goal is to identify and evaluate stabilizing PEGylation sites within larger and more therapeutically relevant proteins using this method. However, its reliance on Asn-PEGylation presents a major logistical challenge: Asn-PEG is not genetically encodable. Asn-PEGylated peptides and proteins can be prepared via solid-phase peptide synthesis[5] and/or native chemical ligation,[6] but not via biological expression. An attractive potential alternative to Asn-

PEGylation is to incorporate PEG site-specifically[7] via chemoselective reactions[8] with genetically encodable amino acids. For example, Cys thiols can be selectively modified by conjugate addition with maleimide-based PEG reagents, among other approaches.[9] Similarly, unnatural azide- or alkyne-bearing amino acids like azidohomoalanine[10] (Aha) or propargyloxyphenylalanine[11] (PrF) can be incorporated into expressed proteins,[12,13] which can then be modified selectively with the appropriate alkyne- or azide-functionalized PEG via the copper-catalyzed azide-alkyne cycloaddition (CuAAC).[14-16] In addition, transglutaminase enzymes can modify the amide nitrogen of a Gln side-chain with PEG, albeit without complete site-specificity.[17] However, the Cys-PEG thioether and the PrF- and Aha-PEG triazole linkers (Figure 3-1) differ substantially from the Asn-PEG amide linker used in our previous studies; and though Gln-PEG more closely resembles Asn-PEG, it places the pendant PEG further from the peptide backbone than does Asn-PEG. Our structure-based method for selecting PEGylation[5] sites relies on assumptions about the orientation of the PEGylated side chain relative to nearby OH groups; we wondered whether substantial changes to linker structure would alter the ability of PEG to enhance protein conformational stability.

3.2 Results and Discussion

We investigated the impact of linker structure on PEG-based protein stabilization by comparing the impact of Cys-, Aha-, PrF-, and Gln-PEGylation relative to Asn-PEGylation at a previously characterized stabilizing position within WW. To this end, we used solid-phase synthesis to prepare proteins **2**, **3**, **4**, and **5**, in which Cys, PrF, Aha, and Gln occupy position 19 of WW, respectively (Figure 1). We then prepared four Cys-PEGylated counterparts of protein **2**: (1) protein **2p-1**, in which a four-unit PEG is directly attached to the Cys thiol; (2) protein **2p-2**, in which PEG is attached to the imide nitrogen of the α-alkylthiosuccinimide linker obtained by

conjugate addition of a PEG-maleimide to the Cys thiol; (3) protein **2p-3**, which differs from **2p-2** in that the PEG is connected to the amide nitrogen of a more flexible α-alkylthioacetamide linker (we note that **2p-3** is a mixture of epimers at the α-carbon of position 19 as a result of the conjugation reaction; see the supporting information for details); and (4) protein **2p-4**, which differs from **2p-2** in that the PEG is connected to the imide nitrogen of the α-alkylthiosuccinimide via a three-carbon linker. In addition, we modified the azide of the Aha side chain in **3** with a PEG alkyne via CuAAC to give Aha-PEGylated protein **3p**. Similarly, we modified the alkyne of the PrF side chain in **4** with a PEG azide to give PrF-PEGylated protein **4p**. We prepared Gln-PEGylated protein **5p** by incorporating the appropriate PEG-linked Gln monomer during solid-phase synthesis.

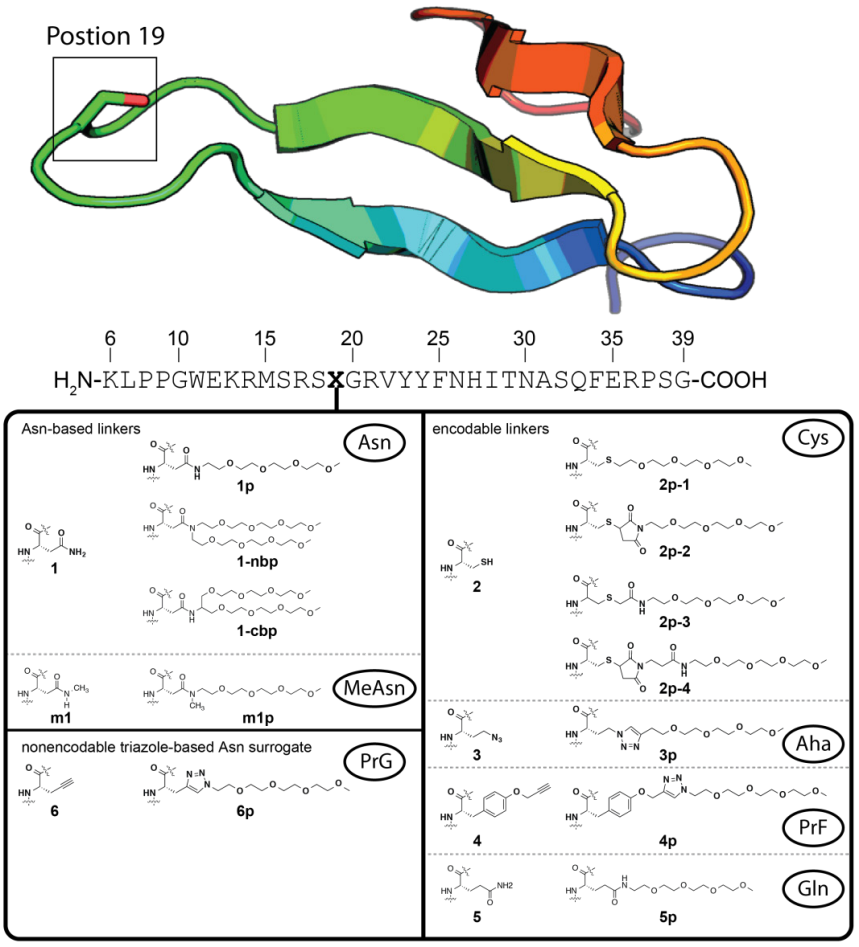


Figure 3-1. Amino acid sequences of PEGylated and non-PEGylated WW variants with side-chain structures at position 19 as indicated.

We used variable temperature circular dichroism (CD) to assess the stability of PEGylated variants **2p-1**, **2p-2**, **2p-3**, **2p-4**, **3p**, **4p**, and **5p** relative to their non-PEGylated counterparts. Importantly, non-PEGylated **2**, **3**, **4**, and **5** are relatively similar in stability to Asn-containing **1** (Table 1), indicating that changing Asn to Cys, Aha, PrF, or Gln at position 19 does not dramatically perturb WW stability. As described previously[5] Asn-PEGylation at position 19 increases WW stability by -0.74 ± 0.02 kcal mol^{-1} (compare **1p** vs. **1** in Table 3-1). In contrast, **2p-1** and **2p-2** are only slightly more stable than **2**, whereas **2p-3** and **2p-4** are each slightly less stable than **2**. The Aha-PEG triazole and Gln-PEG amide linkers in **3p** and **5p** provide nearly

identical PEG-based stabilization (-0.36 ± 0.01 kcal/mol for **3p** vs. -0.37 ± 0.01 kcal/mol for **5p**; Table 1), consistent with their similar lengths and with previous observations that peptide bonds can be replaced by triazoles without dramatic disruptions to protein structure.[14,18] In contrast, PrF-PEGylation does not substantially change WW stability (compare **4p** vs. **4** in Table 3-1). These observations suggest that the stabilizing impact of Asn-PEGylation is highly sensitive to the structure of the PEG-protein linker. Cys-PEGylated **2p-1**, **2p-2**, **2p-3**, and **2p-4**, and PrF-PEGylated **4p** apparently fail to recapitulate some essential structural feature of the Asn-PEGylated **1p**, whereas Aha-PEGylation (as in **3p**) and Gln-PEGylation (as in **5p**) more closely mimic the stabilizing effect of Asn-PEGylation.

Gln-PEG differs from Asn-PEG by the addition of a single methylene group between the amide linker and the peptide backbone; this subtle structural difference decreases the amount of PEG-based stabilization in **5p** by 0.37 ± 0.02 kcal/mol relative to **1p**. Previous work suggests that Asn-PEGylation stabilizes the folded conformation of **1p** by entropically favorable local desolvation of nearby amino acids on the protein surface. The longer Gln side chain may not place PEG close enough to the protein surface for optimal desolvation, leading ultimately to diminished PEG-based stabilization. Distance from the backbone might also explain why triazole-containing Aha-PEG and its isostere Gln-PEG provide similar amounts of PEG-based stabilization relative to each other, but inferior amounts of stabilization relative to the shorter Asn-PEG. To test this hypothesis, we replaced Aha in **3** with propargylglycine (PrG) to give protein **6**; we then prepared PEGylated **6p** by modifying the acetylene in **6** with a PEG-azide via CuAAC (Figure 3-1). A PrG-linked PEG should be one methylene unit closer to the protein backbone than an Aha-linked PEG and should therefore more closely resemble the Asn-linked PEG of **1p**.

Table 3-1. Impact of Linker Structure on PEG-based Stabilization of WW

Protein	T_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)
1	55.9 ± 0.2	0.38 ± 0.02	---
1p	63.8 ± 0.1	-0.36 ± 0.01	-0.74 ± 0.02
1-nbp	68.1 ± 0.1	-0.81 ± 0.02	-1.19 ± 0.02
1-cbp	67.3 ± 0.3	-0.75 ± 0.04	-1.13 ± 0.04
m1	60.7 ± 0.1	-0.06 ± 0.01	---
m1p	63.7 ± 0.2	-0.36 ± 0.01	-0.30 ± 0.01
2	55.6 ± 0.4	0.42 ± 0.02	---
2p-1	56.8 ± 0.2	0.26 ± 0.01	-0.16 ± 0.02
2p-2	56.3 ± 0.2	0.34 ± 0.01	-0.08 ± 0.02
2p-3	54.9 ± 0.2	0.43 ± 0.01	0.01 ± 0.02
2p-4	53.1 ± 0.4	0.67 ± 0.02	0.24 ± 0.02
3	54.0 ± 0.2	0.55 ± 0.01	---
3p	57.7 ± 0.2	0.18 ± 0.01	-0.36 ± 0.01
4	52.3 ± 0.3	0.68 ± 0.02	---
4p	53.7 ± 0.2	0.64 ± 0.02	-0.04 ± 0.02
5	54.2 ± 0.2	0.48 ± 0.01	---
5p	58.7 ± 0.2	0.11 ± 0.01	-0.37 ± 0.01
6	52.6 ± 0.3	0.73 ± 0.03	---
6p	58.6 ± 0.1	0.14 ± 0.01	-0.59 ± 0.03

Melting temperatures and folding free energies, are given ± standard error at 50 μM protein concentration in 20 mM aqueous sodium phosphate, pH 7 at 333.15K.

Consistent with this hypothesis, we observed that PrG-PEGylation increases the stability of **6p** by -0.59 ± 0.03 kcal/mol relative to **6**, a larger increment than observed previously for Aha-PEGylation. However, PrG-PEGylation is unlikely to be a useful alternative to Asn-PEGylation: it is not as stabilizing as Asn-PEGylation and is also not genetically encodable.

Having demonstrated the importance of linker distance from the peptide backbone, we wanted to understand the additional structural features that make Asn-PEG more stabilizing than the other linkers investigated here. Unlike its inferior isostere PrG-PEG, Asn-PEG contains a secondary amide group; we wondered whether hydrogen bonding by the amide proton plays an important role in PEG-based stabilization. Attempts to address this question via amide-to-ester

mutation[19,20] were unsuccessful due to the instability of the resulting ester protein-PEG linkage. Accordingly, we used an alternative approach, preparing protein **m1**, in which *N*-methyl Asn occupies position 19, along with its PEGylated counterpart **m1p**, in which the methylated Asn nitrogen is also attached to a four-unit PEG (Figure 1). NMR analysis indicates that the Fmoc-protected *N*-methyl *N*-PEGylated Asn monomer used to prepare protein **m1p** exists as a mixture of E and Z amide stereoisomers; and we expect the same to be true within protein **m1p**. Nevertheless, variable temperature CD data for **m1p** can be fit to a simple two-state folding model without evidence of heterogeneity, suggesting that the E and Z amide stereoisomers of **m1p** must have closely similar folding free energies.

Interestingly, *N*-methylation of Asn itself increases WW stability by -0.45 ± 0.02 kcal/mol (Figure 3-2; compare **m1** vs. **1** in Table 3-1). This result is in agreement with our previous observations that appending even a single ethylene oxide unit to Asn19 in WW can provide substantial stabilization,[1] and highlights the impact of changes to side-chain structure very close to the peptide backbone on protein conformational stability. Further elongation from the methyl group in **m1** to the four-unit PEG in **1p** provides an additional -0.30 ± 0.01 kcal/mol of stability. Subsequent N-methylation of the Asn-PEG amide nitrogen in **1p** gives protein **m1p** (T_m = 63.7 ± 0.2 °C), which is indistinguishable from **1p** (T_m = 63.8 ± 0.1 °C) in terms of stability. This result indicates that the Asn-PEG amide proton in **1p** is not an essential feature of stabilizing PEG-protein linkers. Elongating the N-methyl group of **m1p** to a second four-unit PEG results in N-branched bis-PEGylated variant **1-nbp** (Figure 3-1, Table 3-1), which is -1.19 ± 0.02 kcal/mol more stable than non-PEGylated **1** and -0.45 ± 0.02 kcal/mol more stable than mono-PEGylated **1p**. The difference in stability between **1-nbp** and **1p** cannot be attributed solely to the fact that **1-nbp** contains four more ethylene oxide units than **1p**: branched **1-nbp** is also more stable than

a previously characterized derivative of **1p** that has eight ethylene oxide units instead of four.[21] This observation highlights the possibility that branched PEGs may be more effective at increasing protein stability than linear PEGs of similar molecular weight.

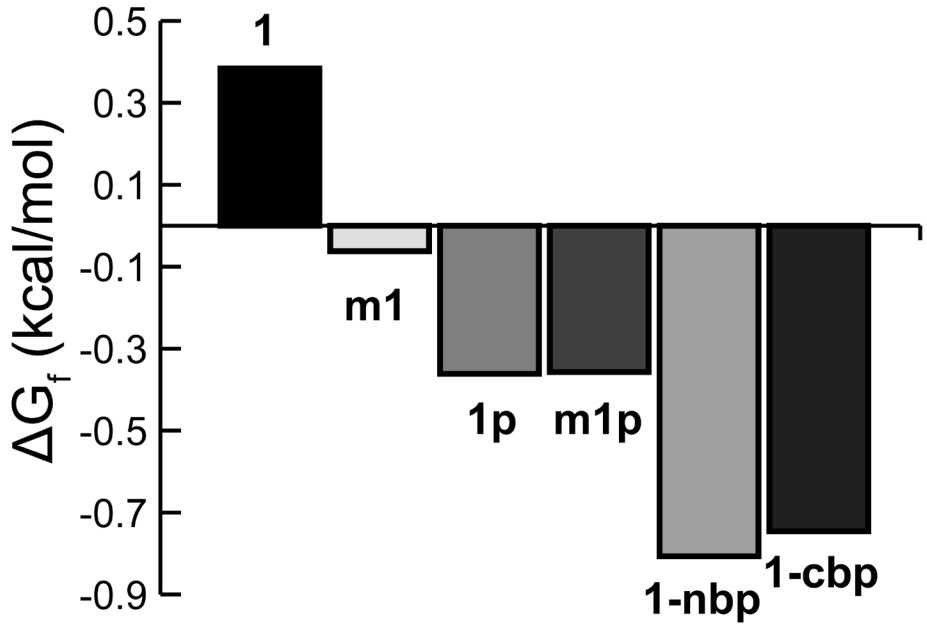


Figure 3-17. Folding free energies of WW variants in which Asn, N-Methyl-Asn, or their PEGylated derivatives occupy position 19.

We wondered whether the favorable impact of branching on PEG-based stabilization depends on the location of the branch point at the amide nitrogen of Asn or whether a different branch point might provide similar levels of stabilization. Accordingly, we prepared protein **1-cbp (**Figure 3-1) in which the branch point is located at the first carbon of the appended PEG chain; the resulting linkage is a secondary amide (as in **1p**) and not a tertiary amide (as in **1-nbp** and **m1p**). C-branched bis-PEGylated protein **1-cbp** is only marginally less stable than **1-nbp** (Figure 3-2, Table 3-1), and is -0.38 ± 0.04 kcal/mol more stable than mono-PEGylated **1p**, suggesting that PEG-branching can provide substantial stabilization independent of the identity (N vs. C) or geometry (trigonal planar vs. tetrahedral) of the branch point.

3.3 Conclusion

The picture that emerges from these data is that the identity of the PEG-protein linker is a critical determinant of PEG-based stabilization within WW. In particular, linkers that provide the largest PEG-based increases in stability have rigid planar functional groups (amides as in **1p**, **1-nbp**, **1-cbp**, and **5p**; triazoles as in **3p** and **6p**) relatively close to the peptide backbone, but do not absolutely require the presence of an amide proton. The substantially larger PEG-based stabilization observed for PrG-PEGylation relative to Aha-PEGylation and for Asn-PEGylation relative to Gln-PEGylation demonstrates that moving this rigid planar functional group even one atom further from the peptide backbone can prevent the pendant PEG from providing optimal stabilization. Despite their inferiority in this regard, Gln-PEG and Aha-PEG can be installed chemoselectively on encoded side-chains unlike their more stabilizing truncated counterparts Asn-PEG and PrG-PEG; the use of Gln-PEG or Aha-PEG as encodable alternatives to Asn-PEGylation therefore merits further investigation. Additionally, the superiority of branched vs. linear Asn-PEGs suggest important additional applications for branched PEGs in protein stabilization. In the short term, applying our predictive guidelines for Asn-PEGylation to larger therapeutic proteins may require a fragment condensation/ligation approach and/or the development of new chemoselective reactions that target Asn side chains. Alternatively, it will be interesting to see whether the rules developed previously for identifying stabilizing Asn-PEGylation sites are also useful for identifying stabilizing Aha- or Gln-PEGylation sites. More broadly, our results suggest that efforts to generate optimally stable PEG-protein conjugates must consider conjugation chemistry as an important variable.

3.4 Supporting Information

3.4.1 WW Variant Synthesis

Proteins **1-nbp**, **1-cbp**, **m1**, **m1p**, **2**, **2p-1**, **3**, **4**, **5**, **6**, and **6p** were synthesized as C-terminal acids, by microwave-assisted solid-phase peptide synthesis, using a standard Fmoc Nα protection strategy as described previously.[22,23] 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 or Advanced ChemTech except for previously synthesized N-[(9H-Fluoren-9-ylmethoxy)-O-2-propyn-1yl-L-Tyrosine[24] (used in the synthesis of **4**), along with new compounds Fmoc-L-Cys(PEG4)-OH [16-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-2,5,8,11-tetraoxa-14-thiaheptadecan-17-oic acid, (**S1),** used in the synthesis of **2p-1**]; Fmoc-Methyl-L-Asn-OH [N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^4-methylasparagine, **(S3)**, used in the synthesis of **m1**]; Fmoc-L-GlnPEG$_4$-OH [18-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-15-oxo-2,5,8,11-tetraoxa-14-azanonadecan-19-oic acid, (**S5),** used in the synthesis of **6p**]; Fmoc-Methyl-L-AsnPEG$_4$-OH [*(S)-17-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-14-methyl-15-oxo-2,5,8,11-tetraoxa-14-azaoctadecan-18-oic acid*, (**S7),** used in the synthesis of **m1p**]; Fmoc-L-Asn(PEG$_4$)$_2$-OH [(*S*)-17-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-13-(2,5,8,11-tetraoxadodecyl)-15-oxo-2,5,8,11-tetraoxa-14-azaoctadecan-18-oic acid, (**S9),** used in the synthesis of **1-cbp**]; Fmoc-L-Asn(PEG)$_2$-OH [**(***S*)-17-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-15-oxo-14-(2,5,8,11-tetraoxatridecan-13-yl)-2,5,8,11-tetraoxa-14-

azaoctadecan-18-oic acid, (**S14),** used in the synthesis of **1-nbp**], each of which were each synthesized as described below.

WW variants 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). 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 1.25 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 stock coupling solution of 100 mL NMP, 3.17 g HBTU (0.01 mol, 0.1 M) and 1.53 g HOBt (0.01 mol, 0.1 M) for a final concentration of 0.1 M HBTU and 0.1 M 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-L-Cys(PEG4)-OH (S1), Fmoc-methyl-L-Asn-OH (S3), Fmoc-L-GlnPEG$_4$-OH (S5), Fmoc-methyl-L-AsnPEG$_4$-OH (S7), Fmoc-L-

(AsnPEG$_4$)$_2$-OH (S9), and Fmoc-L-Asn(PEG$_4$)$_2$-OH (S14) monomers, 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 vacuum 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.

3.4.2 Protein Side-chain Functionalization

3.4.3 Preparation of proteins 3p, 4p, and 5p

PEGylated protein **3p** was prepared from proteins **3 and** known PEG-alkyne 2,5,8,11-tetraoxatetradec-13-yne[25] via the copper(I) catalyzed azide-alkyne cycloaddition reaction. Protein **3p** was prepared by dissolving 2,5,8,11-tetraoxatetradec-13-yne (12.6 mg, 62.5μmol),

copper (I) iodide (12.0 mg, 62.5 μmol) and sodium ascorbate (12.3 mg, 62.5 μmol) in 2 mL 20% piperdine in DMF (v/v), and stirring the resulting solution with protected resin bound protein **3** (12.5 μmol scale) overnight. The resin was then washed with DMF, followed by dichloromethane. Protein **3p** was globally deprotected and cleaved from resin as described above, and purified by reverse-phase HPLC as described below.

PEGylated proteins **4p and 5p** were prepared from proteins **4 and 5, respectively, along with** commercially available PEG-azide 13-azido-2,5,8,11-tetraoxatridecane via the copper(I) catalyzed azide-alkyne cycloaddition reaction. Proteins **4p and 5p** were prepared by dissolving 13-azido-2,5,8,11-tetraoxatridecane (14.5 mg, 62.5μmol), copper (I) iodide (12.0 mg, 62.5 μmol) and sodium ascorbate (12.3 mg, 62.5 μmol) in 2mL 20% piperdine in DMF (v/v), and stirring the resulting solution with protected resin bound protein **4 and 5** (12.5 μmol), respectively, overnight. The resin was then washed with DMF, followed by dichloromethane. Proteins **4p and 5p** were globally deprotected and cleaved from resin as described above, and purified by reverse-phase HPLC as described below.

3.4.4 Preparation of proteins 2p-2, 2p-3, and 2p-4

PEGylated protein **2p-2** was prepared from fully deprotected protein **2** and PEG4-maleimide **S12** [i.e., 1-(2,5,8,11-tetraoxatridecan-13-yl)-1*H*-pyrrole-2,5-dione; synthesized as described below] via conjugate addition. Protein **2** (0.3 μmol) and sodium acetate (0.246 mg, 3 μmol) were dissolved in 2 mL of methanol. PEG4-maleimide **S12** (1.07 mg, 3 μmol) was added to the methanol solution and stirred at room temperature for 1.5 hours open to air. Protein **2p-2** was purified by reverse-phase HPLC as described below.

PEGylated protein **2p-3** was prepared from fully deprotected WW variant **dehydroWW** and 2-mercaptoacetamidePEG$_4$ **S11** [i.e., 2-mercapto-*N*-(2,5,8,11-tetraoxatridecan-13-yl)acetamide; synthesized as described below] via conjugate addition. WW variant **dehydroWW** was prepared by adding a solution of known 2,5-dibromohexanediamide[26] (16.8 mg in 128 μL DMF) to a solution of fully deprotected protein **2** (2 mL of a 564.5 μM solution in 12.5 mM sodium phosphate buffer, pH 8). The resulting mixture was then allowed to stir at room temperature for 30 minutes, at 37 °C for 1 hour, and again at room temperature overnight. This process converts the Cys at position 19 of **2** into a dehydroalanine residue. WW variant **dehydroWW** was purified by reverse-phase HPLC as described below. The dehydroalanine residue of **dehydroWW** was then functionalized with 2-mercaptoacetamidePEG$_4$ **S11** via conjugate addition by dissolving 2 mg of **dehydroWW** and 80 mg of **S11** in 1 mL of 25 mM sodium phosphate buffer (pH 8). The resulting mixture was stirred at room temperature for 24 hours and purified by reverse-phase HPLC as described below to give protein **2p-3**. It is important to note that functionalizing **dehydroWW** with 2-mercaptoacetamidePEG$_4$ results in an diastereomeric mixture of peptides, each containing one of two possible Cys-functionalized epimers at position 19. This explains why there are two peaks in the analytical HPLC trace for protein **2p-3** in Figure S25. Nevertheless, variable temperature CD data for **2p-3** (see below) can be fit to a simple two-state folding model without evidence of heterogeneity, suggesting that the epimers of **2p-3** must have closely similar folding free energies.

PEGylated protein **2p-4** was prepared from fully deprotected protein **2** and commercially available m-dPEG®4-MAL (Quanta Biodesign) via conjugate addition. Protein **2** (~2 mg) and sodium acetate (0.4 mg, 5 μmol) were dissolved in 1 mL of methanol. m-dPEG®4-MAL (1.8

mg, 5μmol) was added to the methanol solution and stirred at room temperature for 2 hours open to air. Protein **2p-4** was purified by reverse-phase HPLC as described below.

3.4.5 WW Variant 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 spectrometry (ESI-TOF, spectra appear below in Figures 3-3 through 3-20), and purity was analyzed by Analytical HPLC (Figures 3-21 through 3-38).

3.4.6 ESI-TOF data

ESI-TOF for proteins **1-nbp**, **1-cbp, m1, m1p, 2, 2p-1, 2p-2, 2p-3, 2p-4, 3, 3p, 4, 4p, 5, 5p, 6, 6p, and dehydroWW** are shown in Figures 3-3 through 3-20.

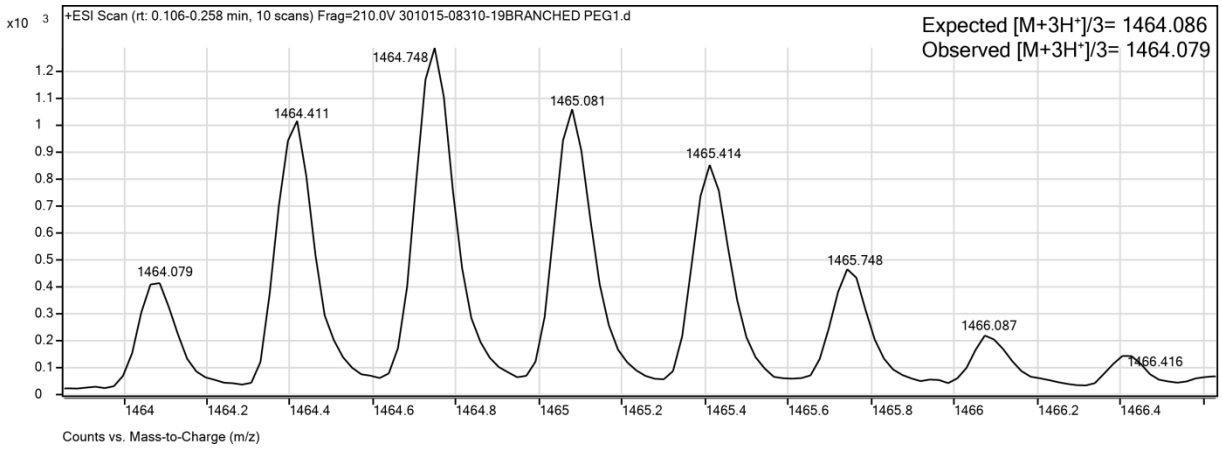


Figure 3-3. ESI TOF spectrum for WW variant **1-nbp**

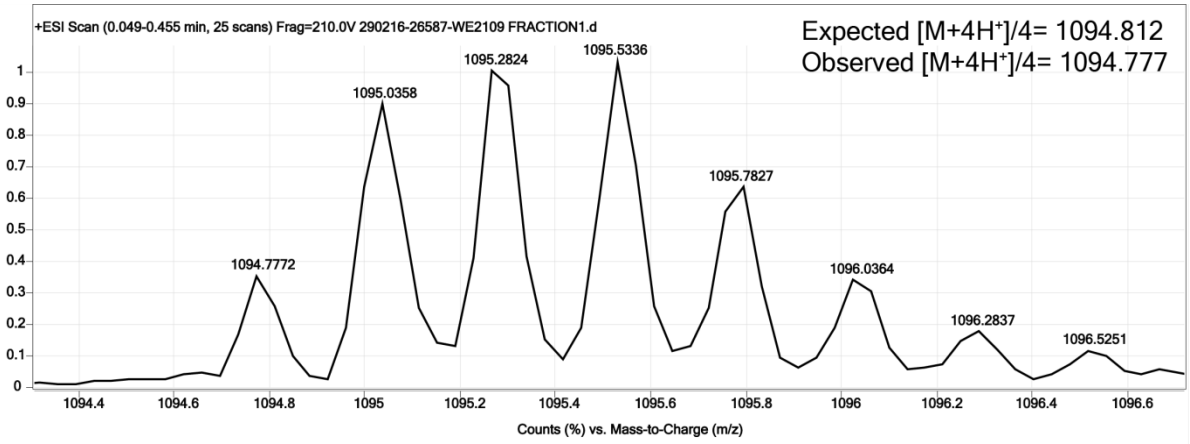


Figure 3-4. ESI TOF spectrum for WW variant **1-cbp**

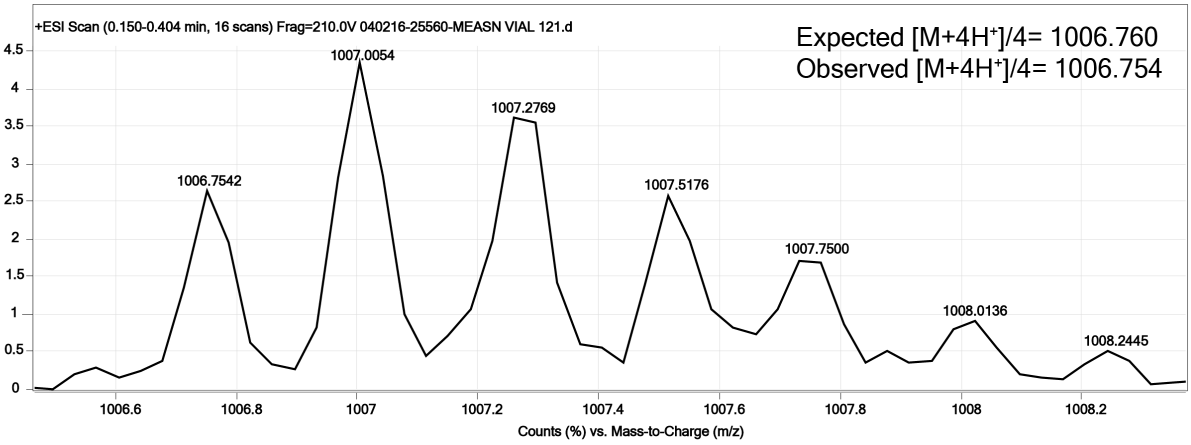


Figure 3-5. ESI TOF spectrum for WW variant **m1**.

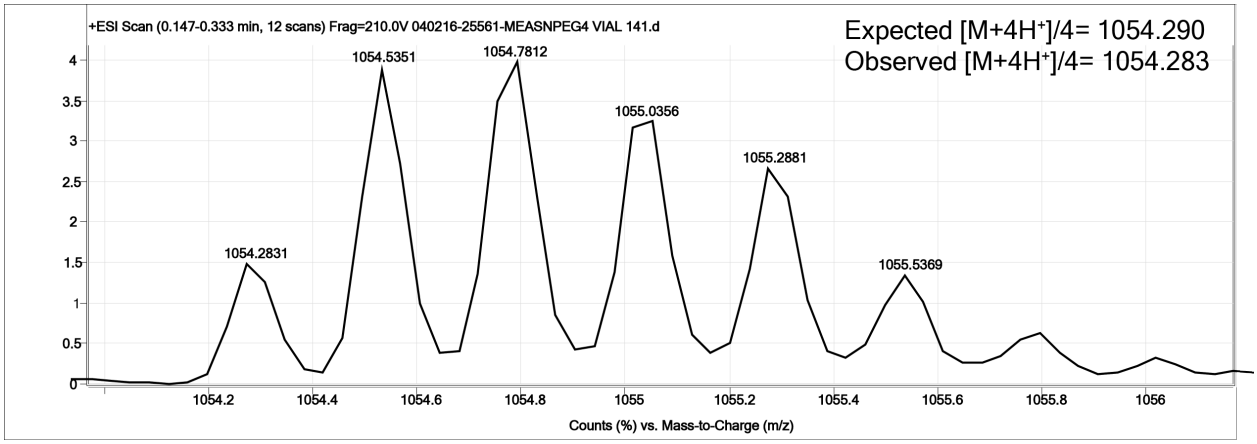


Figure 3-6. ESI TOF spectrum for WW variant **m1p**

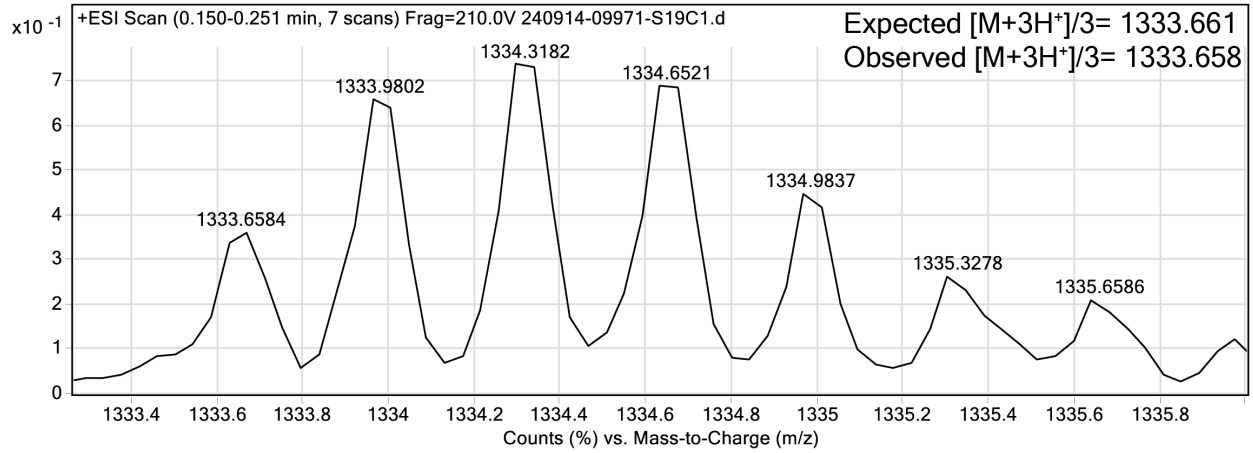


Figure 3-7. ESI TOF spectrum for WW variant **2**

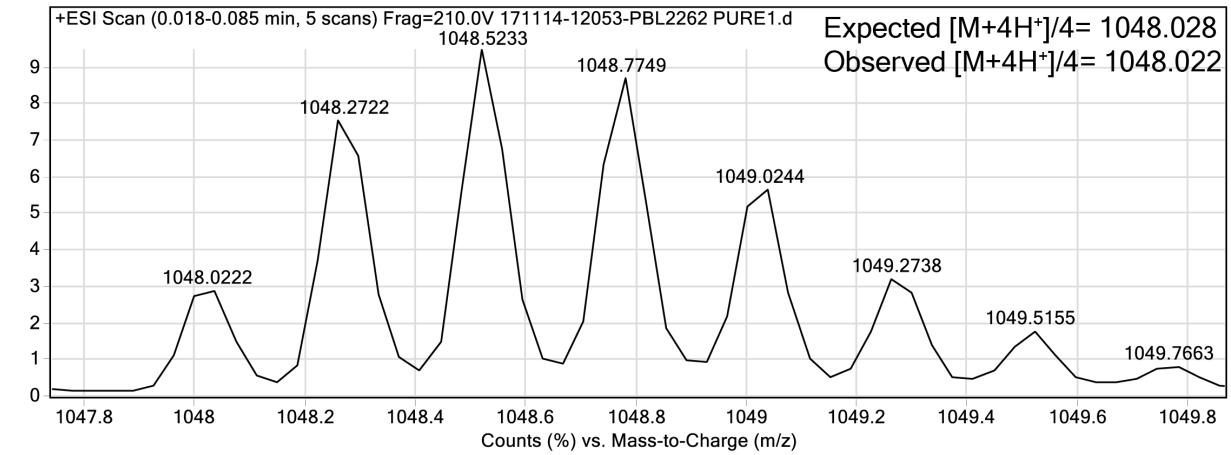


Figure 3-8. ESI TOF spectrum for WW variant **2p-1**

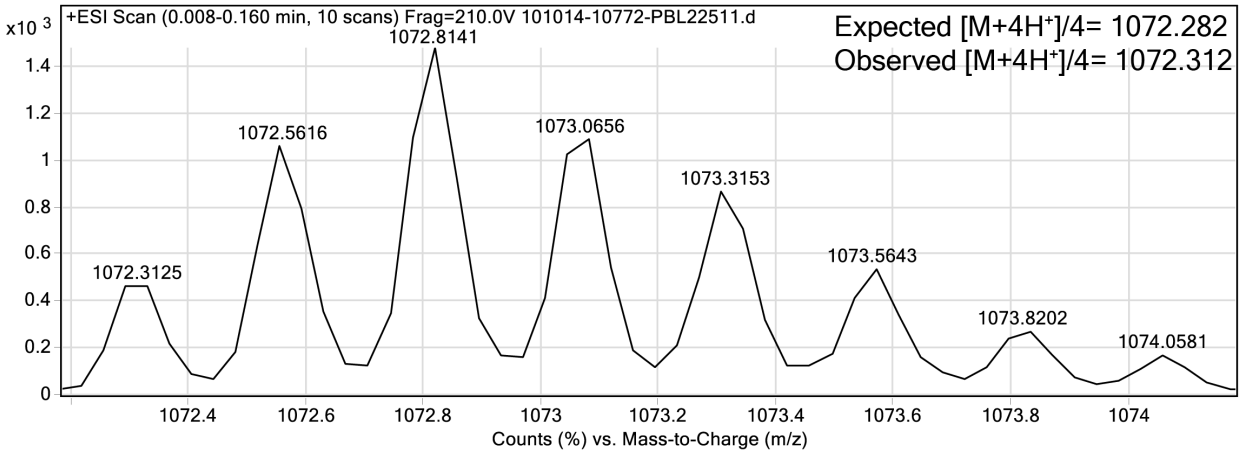


Figure 3-9. ESI TOF spectrum for WW variant **2p-2**

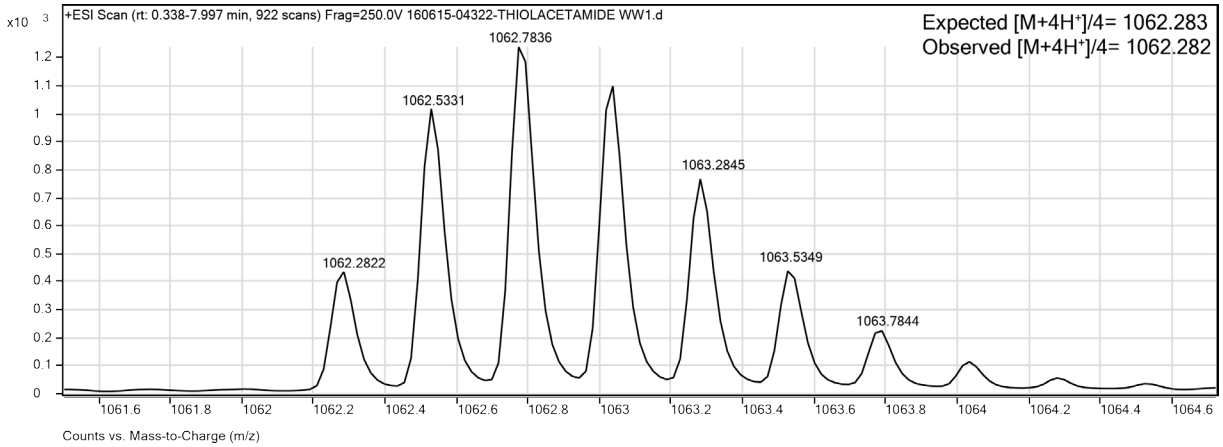


Figure 3-10. ESI TOF spectrum for WW variant **2p-3**

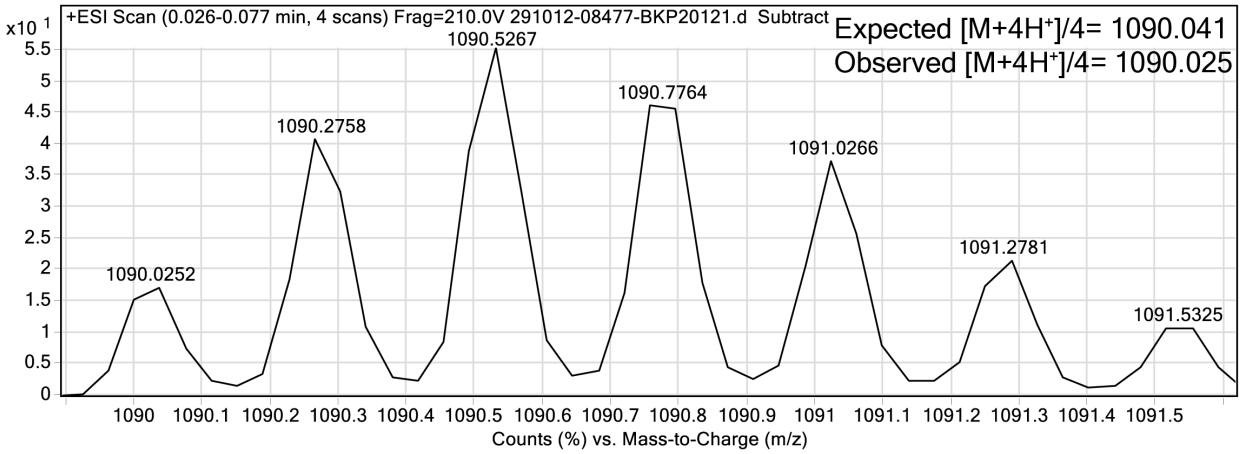


Figure 3-11. ESI TOF spectrum for WW variant **2p-4**

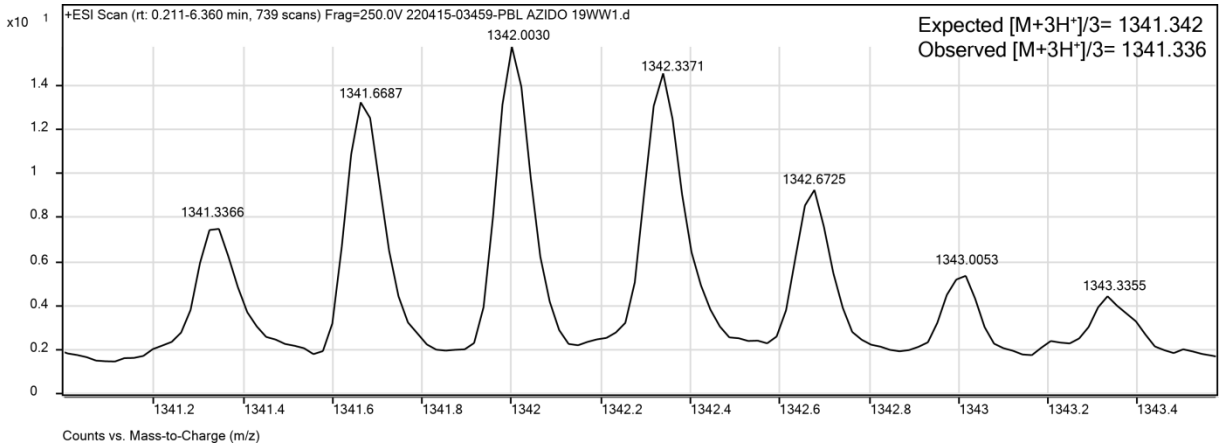


Figure 3-12. ESI TOF spectrum for WW variant **3**

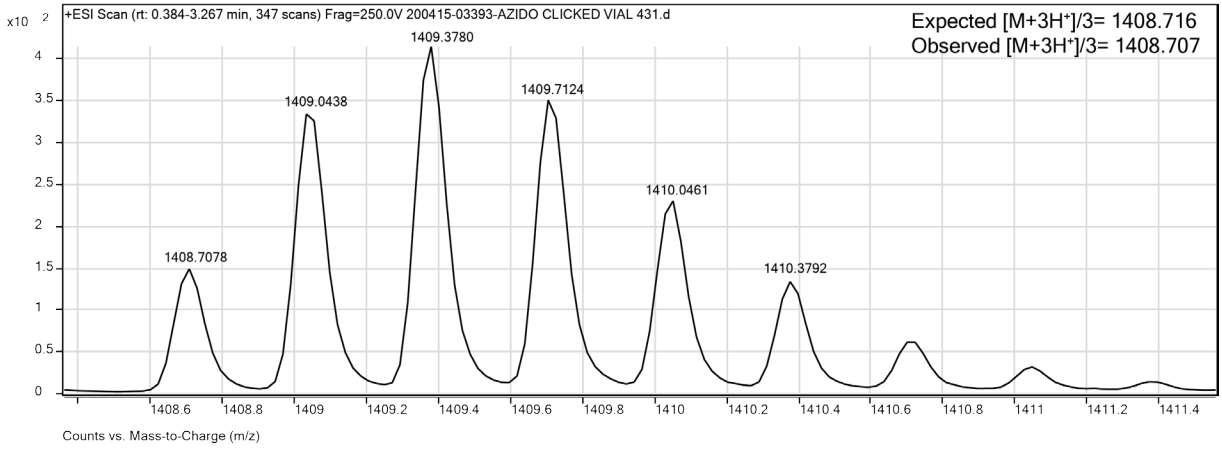


Figure 3-13. ESI TOF spectrum for WW variant **3p**

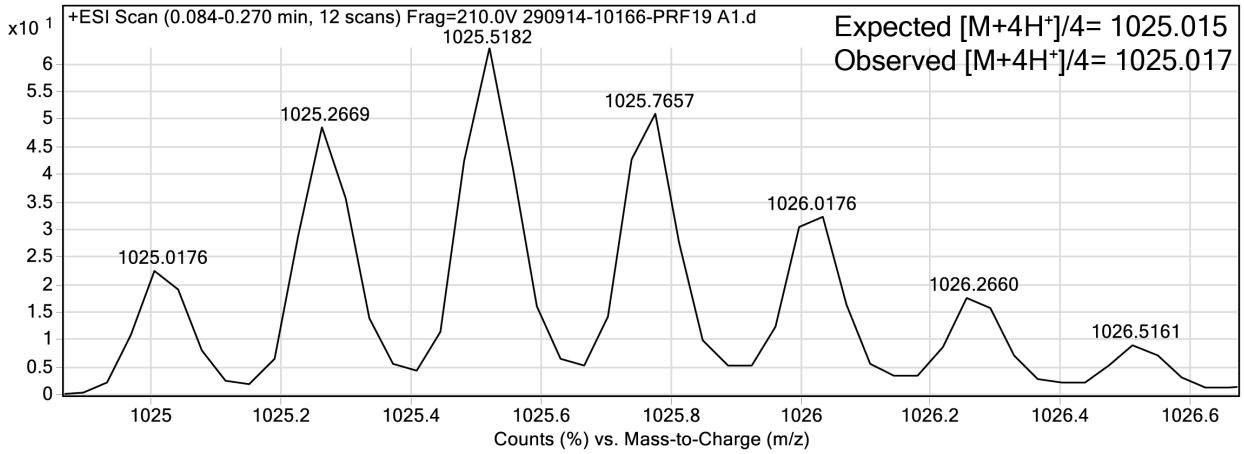


Figure 3-14. ESI TOF spectrum for WW variant **4**

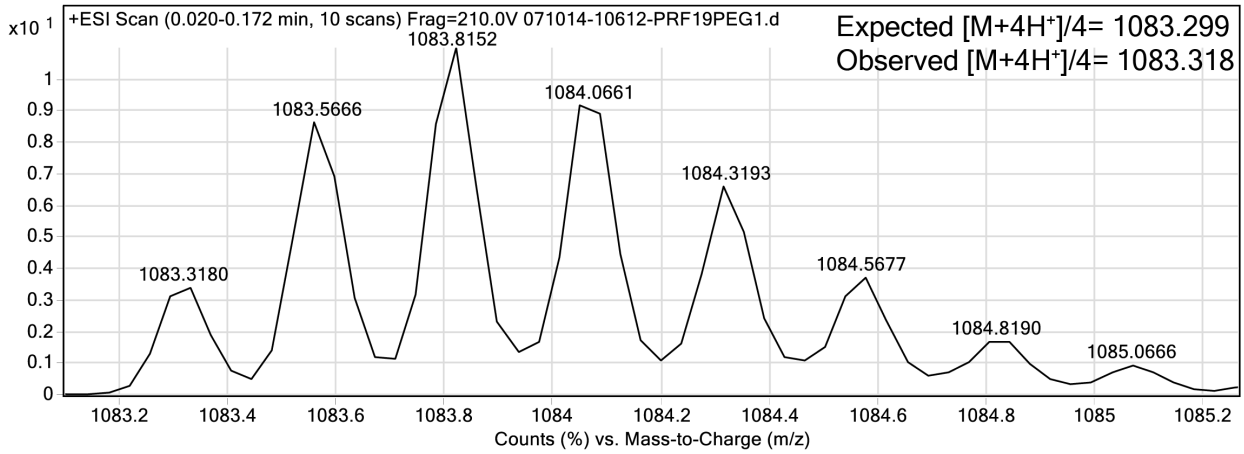


Figure 3-15. ESI TOF spectrum for WW variant **4p**

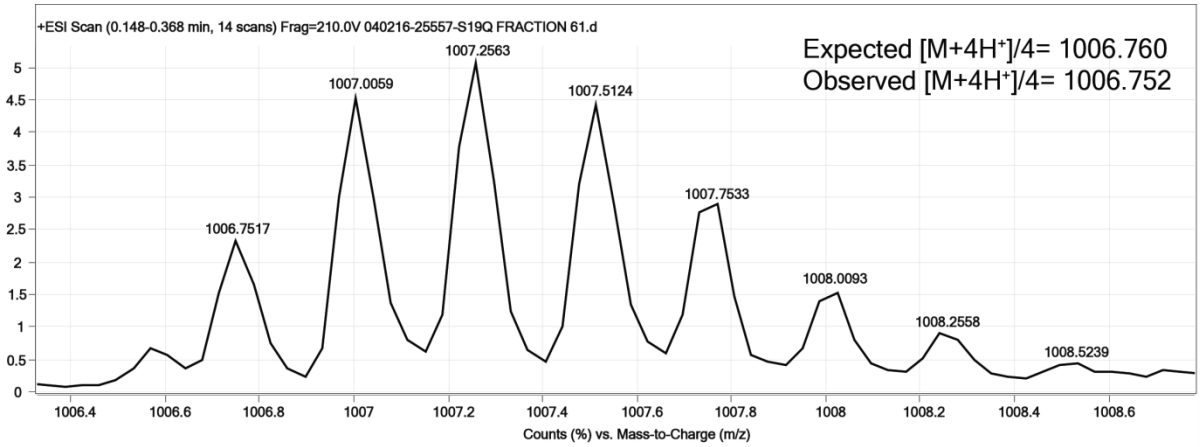


Figure 3-16. ESI TOF spectrum for WW variant **5**

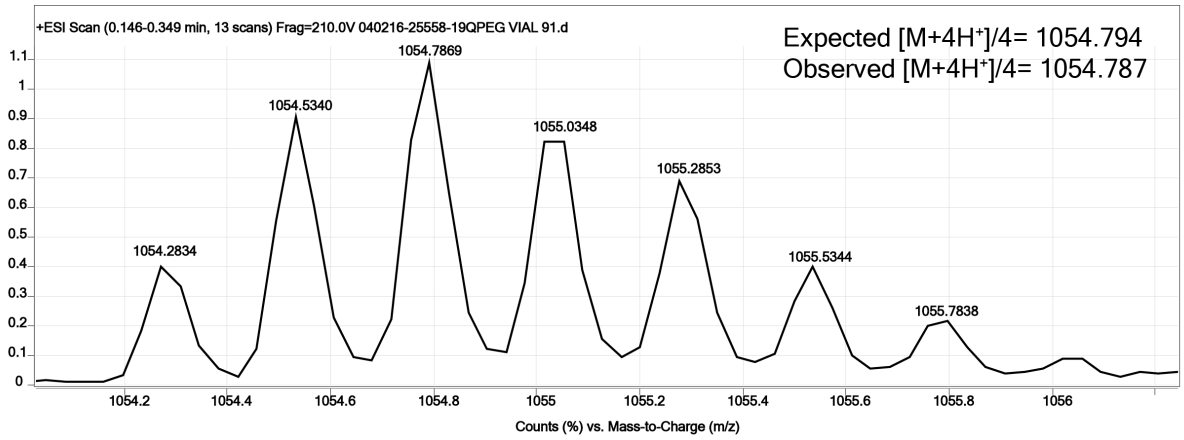


Figure 3-17. ESI TOF spectrum for WW variant **5p**

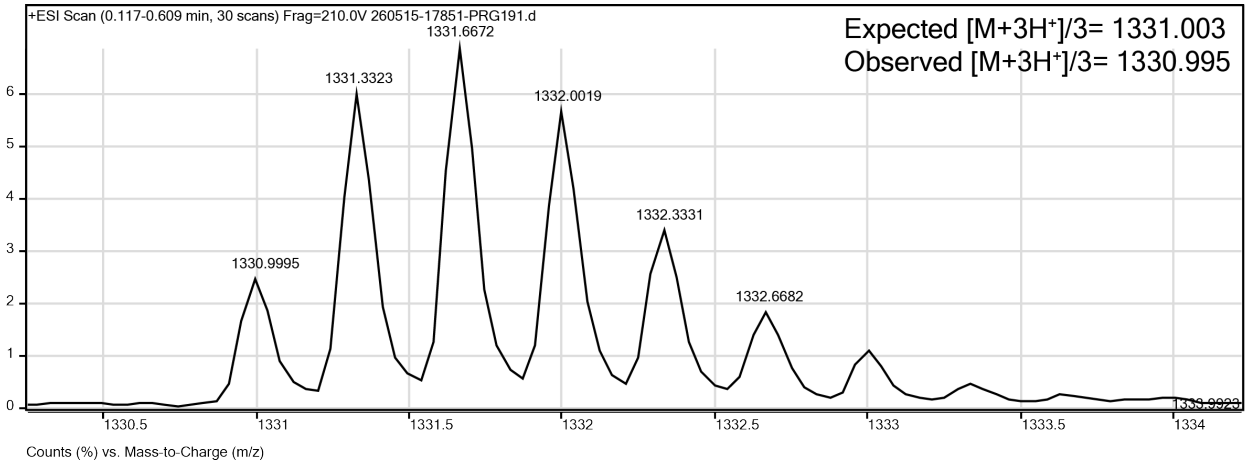


Figure 3-18. ESI TOF spectrum for WW variant **6**

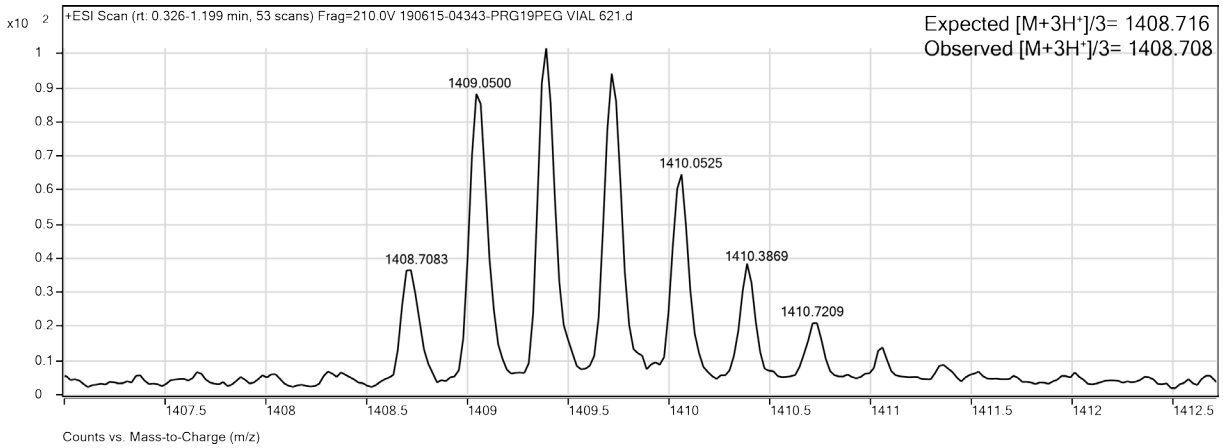


Figure 3-19. ESI TOF spectrum for WW variant **6p**

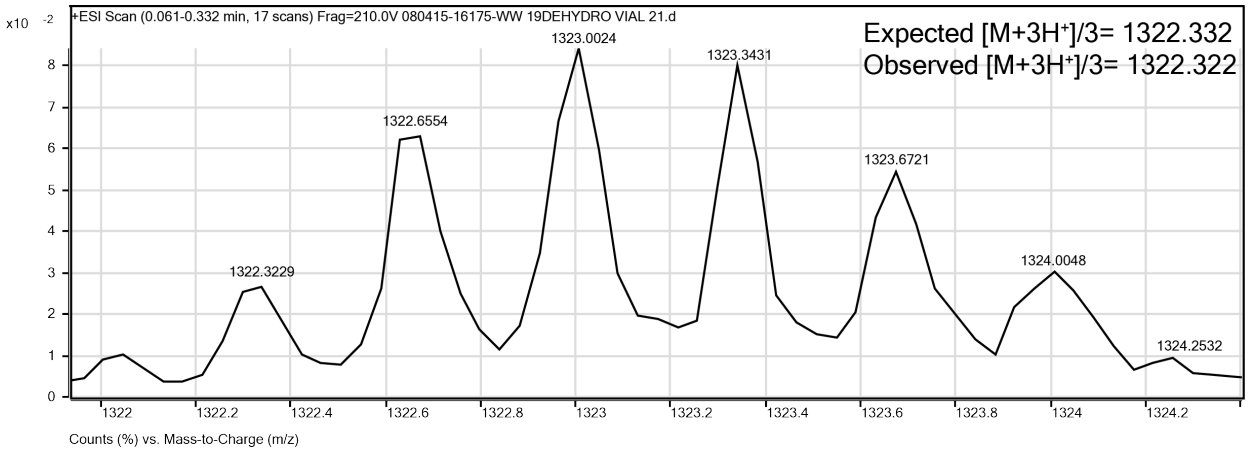


Figure 3-20. ESI TOF spectrum for WW variant **dehydroWW**

3.4.7 Analytical HPLC data

HPLC traces for proteins **1-nbp**, **1-cbp, m1, m1p, 2, 2p-1, 2p-2, 2p-3, 2p-4, 3, 3p, 4, 4p, 5, 5p, 6, 6p,** and **dehydroWW** are shown in Figures 3-21 through 3-38.

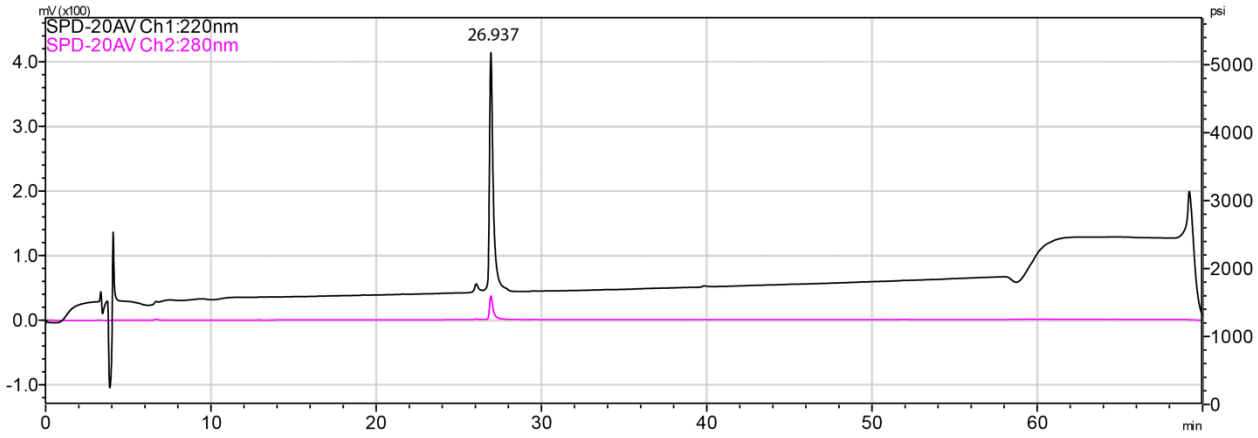


Figure 3-21. Analytical HPLC Data for WW variant **1-nbp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

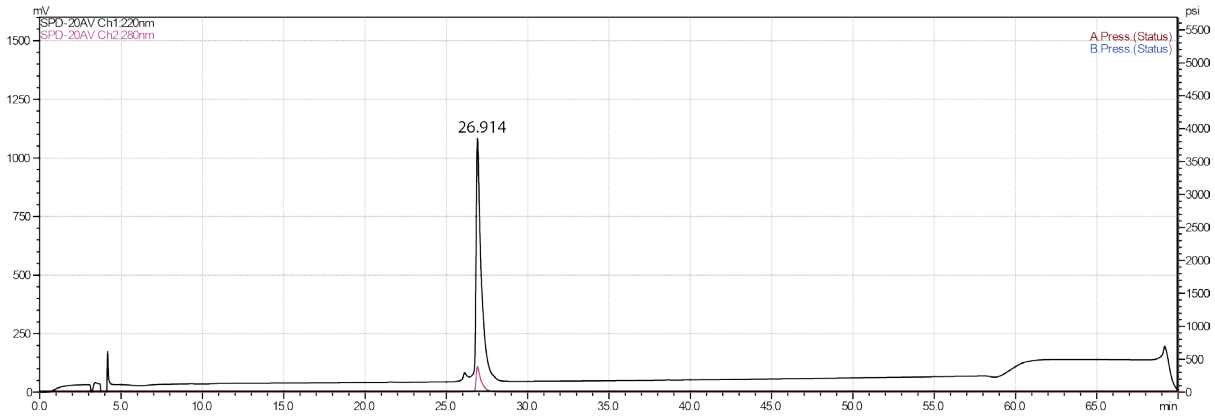


Figure 3-22. Analytical HPLC Data for WW variant **1-cbp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

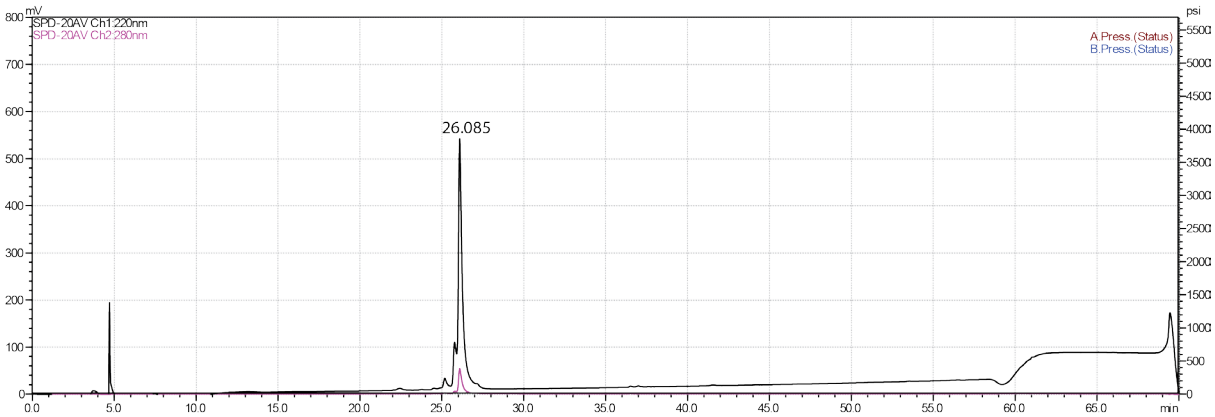


Figure 3-23. Analytical HPLC Data for WW variant **m1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

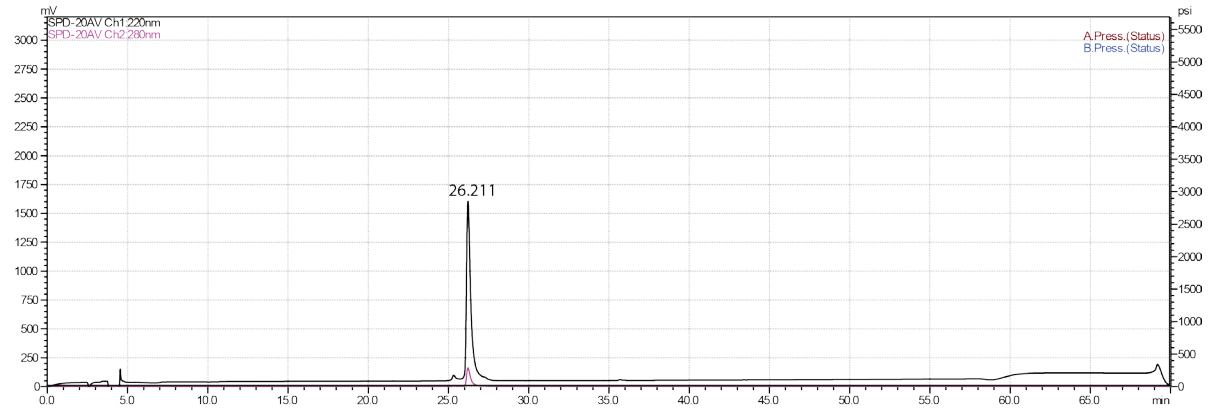


Figure 3-24. Analytical HPLC Data for WW variant **m1p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

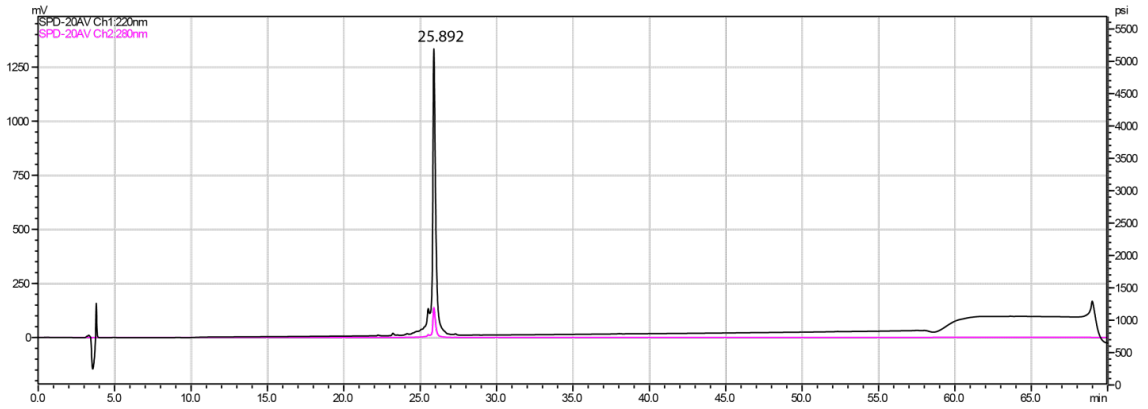


Figure 3-25. Analytical HPLC Data for WW variant **2**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

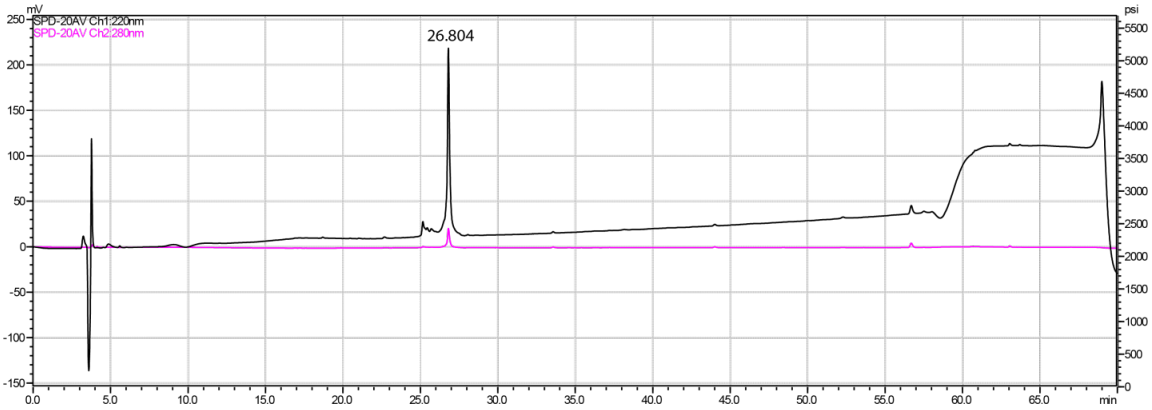


Figure 3-26. Analytical HPLC Data for WW variant **2p-1**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

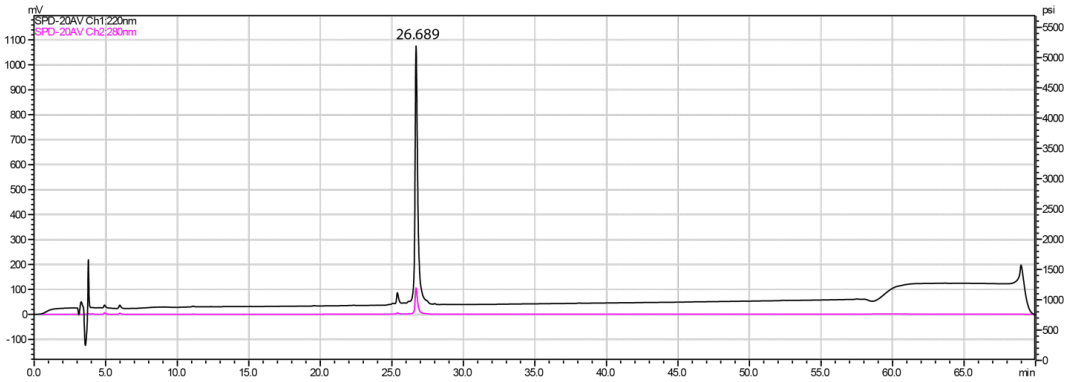


Figure 3-27. Analytical HPLC Data for WW variant **2p-2**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

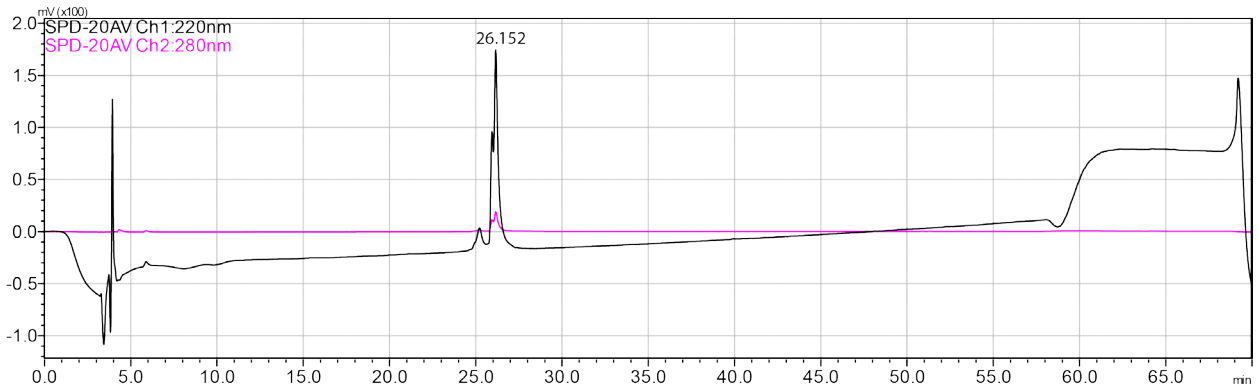


Figure 3-28. Analytical HPLC Data for WW variant **2p-3**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min. It is important to note that the functionalizing dehydroWW with 2-mercaptoacetamidePEG_4 results in a diastereomeric mixture, each peptide containing one of two possible Cys-functionalized epimers at position 19. This explains why there are two peaks in the analytical HPLC trace.

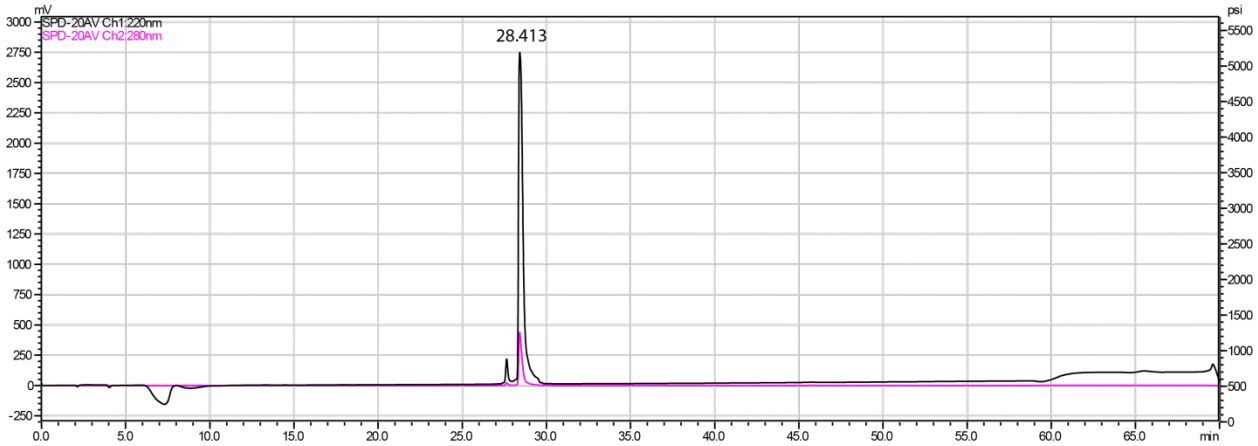


Figure 3-29. Analytical HPLC Data for WW variant **2p-4**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

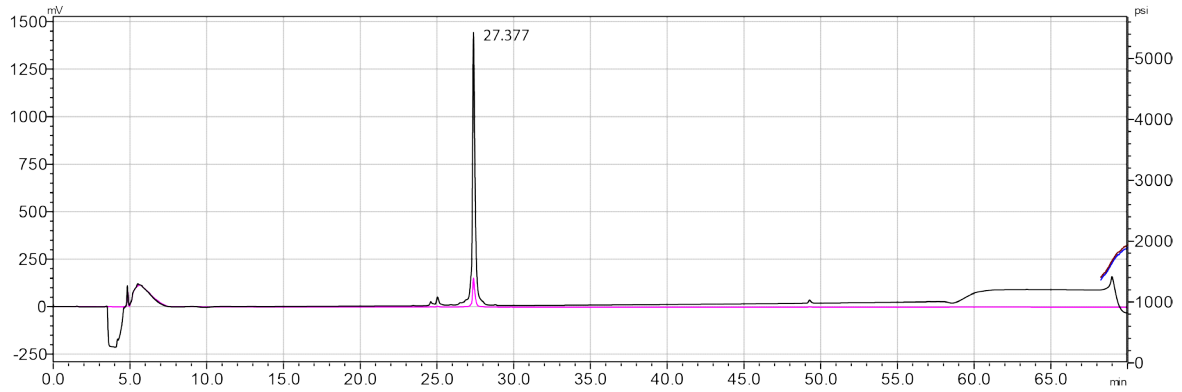


Figure 3-30. Analytical HPLC Data for WW variant **3**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

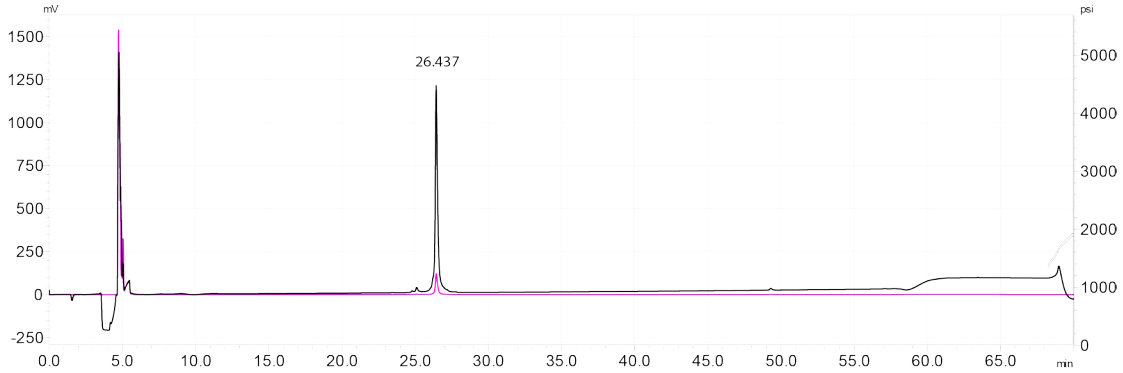


Figure 3-31. Analytical HPLC Data for WW variant **3p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

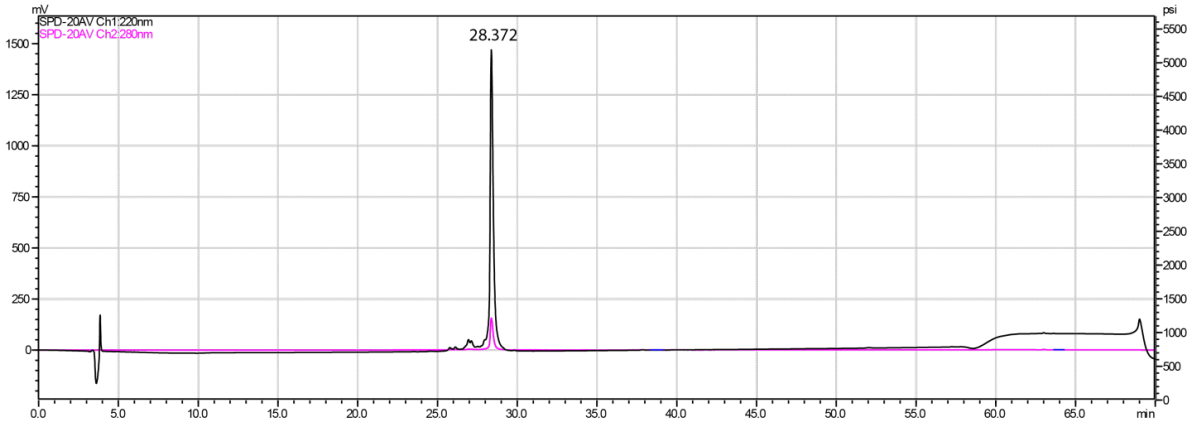


Figure 3-32. Analytical HPLC Data for WW variant **4.** Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

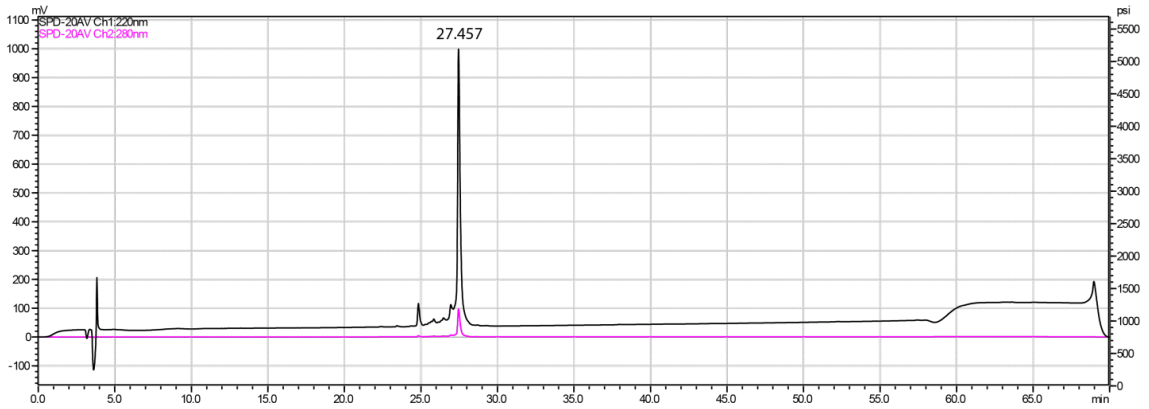


Figure 3-33. Analytical HPLC Data for WW variant **4p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

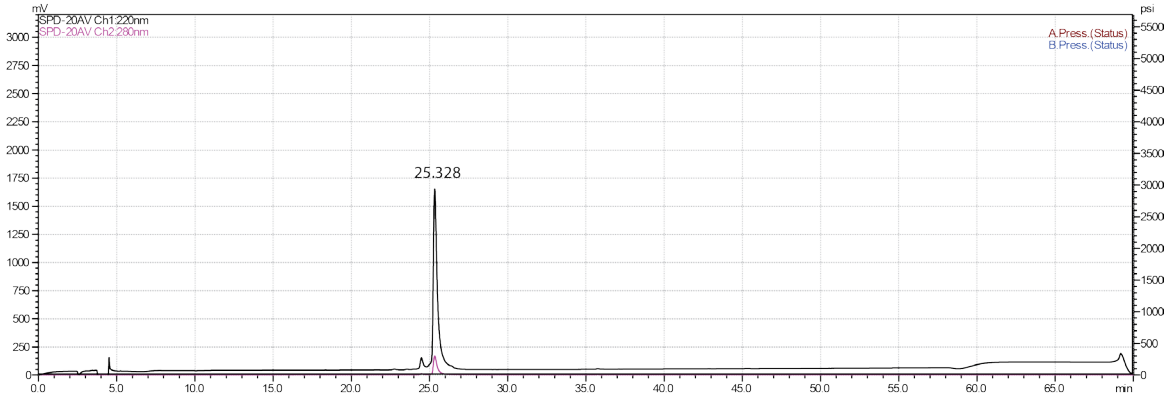


Figure 3-34. Analytical HPLC Data for WW variant **5**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

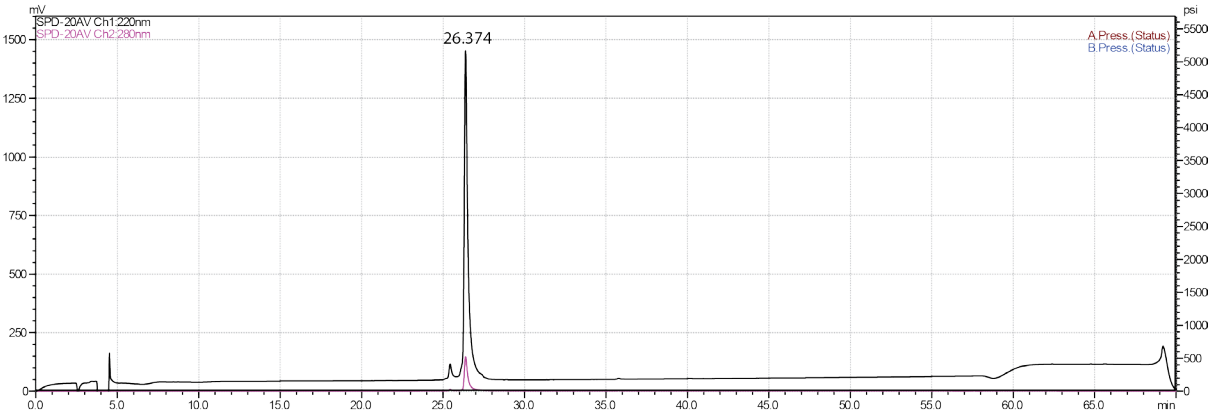


Figure 3-35. Analytical HPLC Data for WW variant **5p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

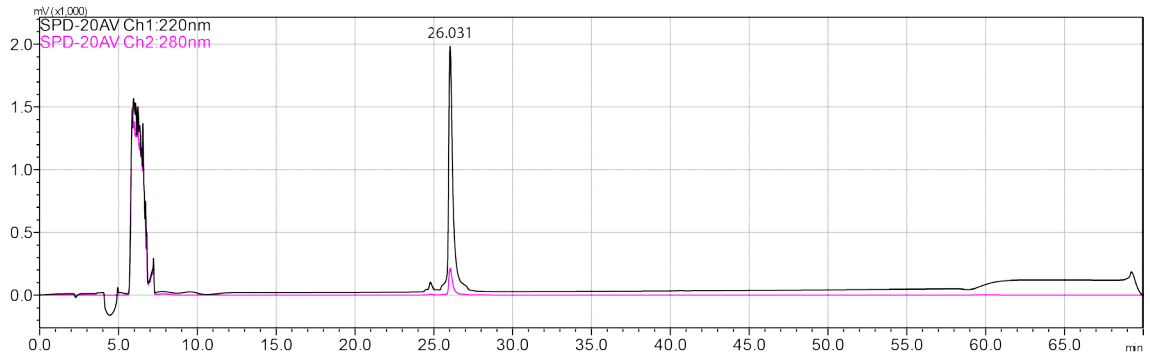


Figure 3-36. Analytical HPLC Data for WW variant **6**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

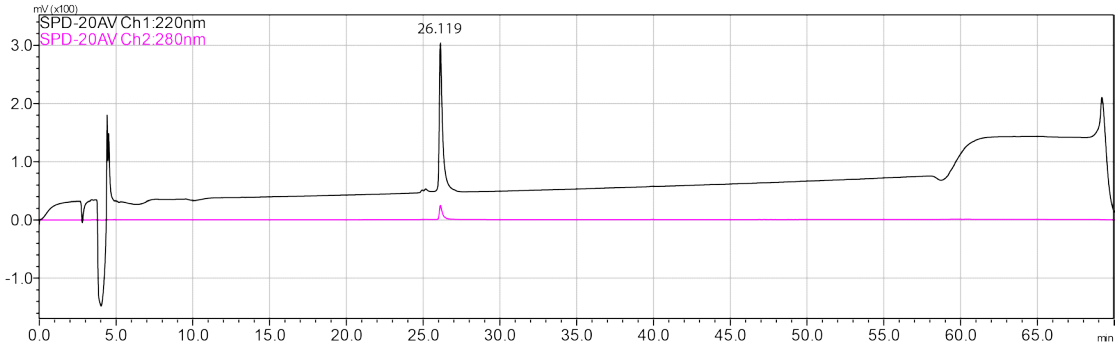


Figure 3-37. Analytical HPLC Data for WW variant **6p**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

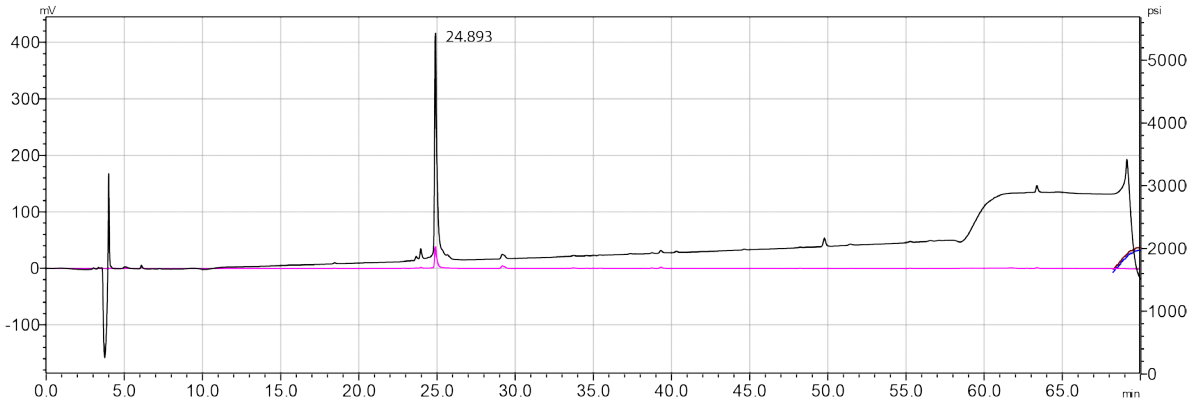
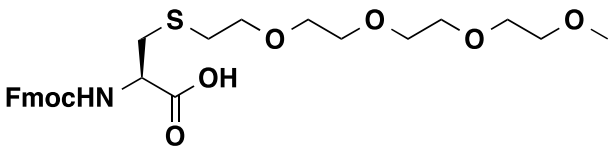


Figure 3-38. Analytical HPLC Data for WW variant **dehydroWW**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

3.4.8 Organic Synthesis

3.4.9 Fmoc-L-Cys(PEG4)-OH (S1)



To a suspension of Fmoc-l-Cys-OH (200.0 mg, 0.582 mmol) in CH_2Cl_2 (50 mL) was added commercially available 13-bromo-2,5,8,11-tetraoxatridecane (700mg, 2.59mmol). The reaction mixture was stirred at 25 °C for 24 hours, and the solvent was removed under reduced pressure. The crude product was then dissolved in acetonitrile and purified via semi-preparative reverse-phase HPLC to yield 20 mg (6.4% yield) of **S1**. A representative chromatogram from semi-preparative HPLC purification of Fmoc-L-Cys(PEG4)-OH is shown in Figure 3-39.

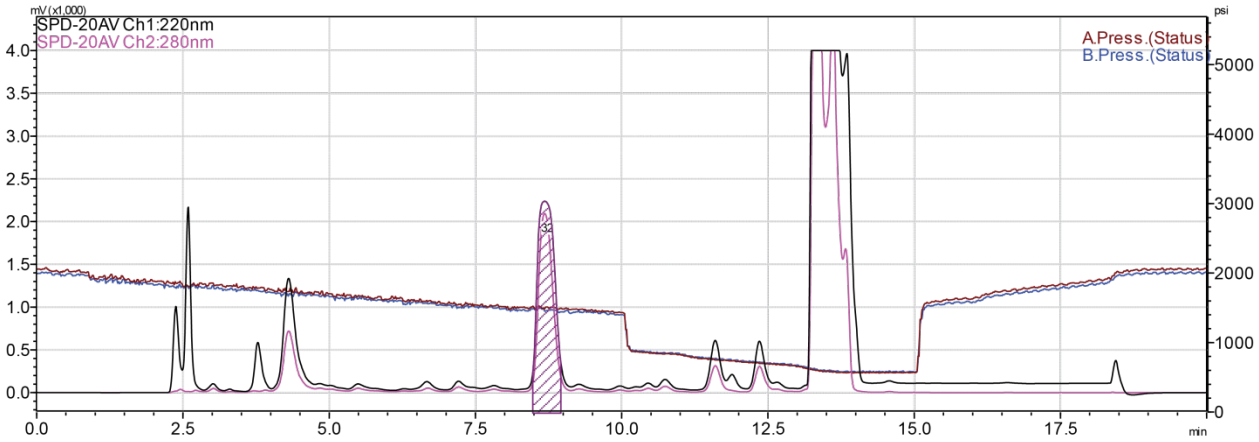


Figure 3-39. Semi-prep reverse-phase HPLC purification of Fmoc-L-Cys(PEG4)-OH. Fmoc-L-Cys(PEG)-OH was injected onto a C18 semi-prep column and eluted using a linear gradient of 40-62% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 10 minutes, followed by a 5 minute rinse (95% B), and a 50 minute column re-equilibration (40% B) with a flow rate of 1 mL/min.

High-resolution electrospray ionization time-of-flight mass spectrometry (ESI-TOF MS) data for Fmoc-L-Cys(PEG4)-OH are shown in Figure 3-40.

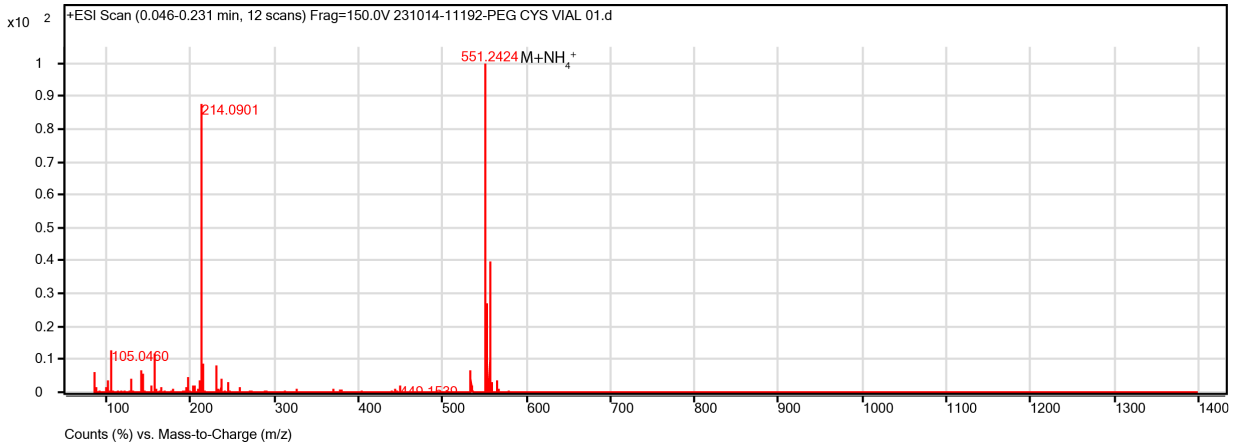


Figure 3-40. ESI-TOF MS data for Fmoc-L-Cys(PEG4)-OH. Calculated m/z for $C_{27}H_{39}N_2O_8S$ ($M+NH_4^+$) is 551.24, found 551.2424.

^{1}H NMR (500 MHz, DMSO): δ 12.88 (1H, broad s, (COO**H**) δ 7.89 (2H, d, J = 7.78 Hz, Fmoc aryl C-**H**); 7.73 (2H, d, J = 7.16 Hz, Fmoc aryl C-**H**); 7.41 (2H, t, J = 7.30 Hz, Fmoc aryl C-**H**); 7.32 (2H, t, J = 7.41Hz, Fmoc aryl C-**H**); 4.25-4.21 (1H, m, -CONHCα**H**(COOH)-CβH_2-); 4.32-4.28 (2H, m, Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.15-4.11 (1H, m, Fmoc Ar2C**H**-CH_2-O-); 3.55-3.39 (14H, m, -S-CH_2-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-); 3.21 (3H, s, -O-C**H$_3$**); 2.98 (1H, dd, J = 13.72 Hz, 4.39 Hz, -CαH(COOH)-Cβ(**Ha**)Hb-S-); 2.79-2.74 (1H, m, -CαH(COOH)-Cβ(Ha)(**Hb**)-S-); 2.68 (2H, t, J = 6.88 -S-C**H$_2$**-CH_2-O-). The full ^{1}H NMR spectrum for Fmoc-L-Cys(PEG4)-OH is shown in Figure 3-41.

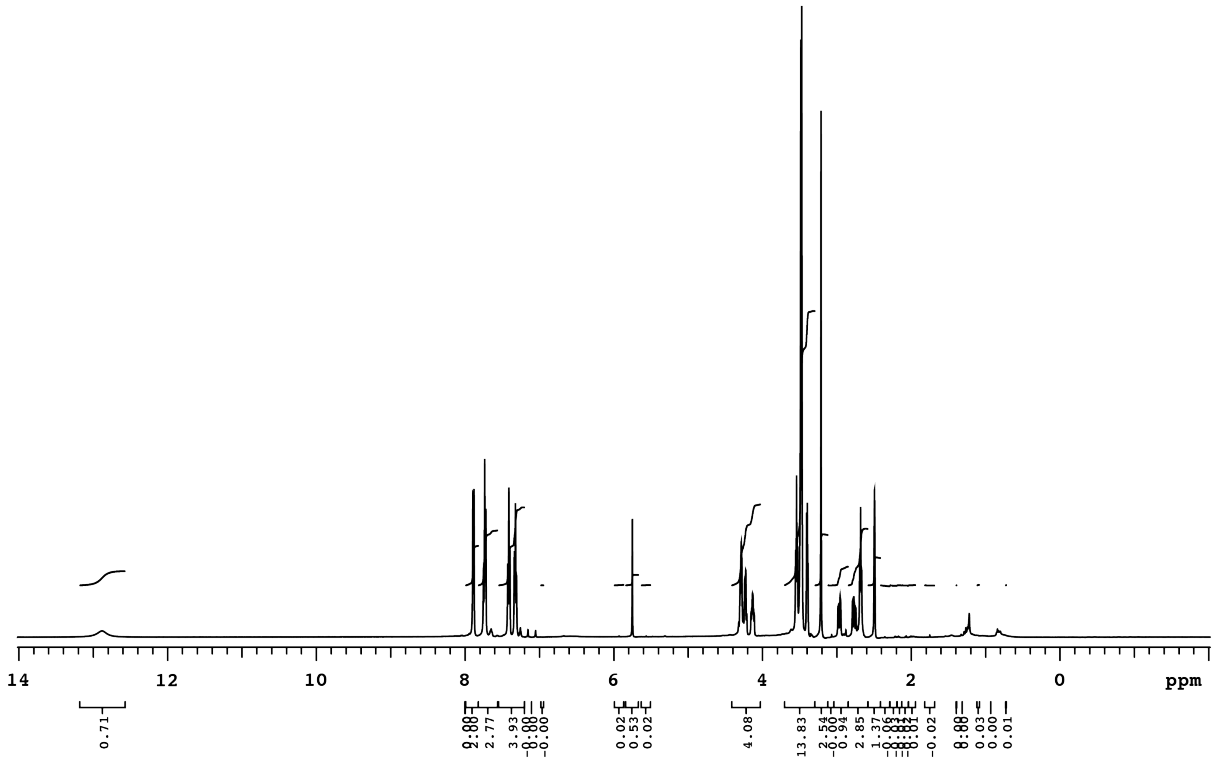


Figure 3-41. ^{1}H NMR spectrum of Fmoc-L-Cys(PEG4)-OH.

^{13}C NMR (126 MHz, $CDCl_3$): δ 172.78 (-NH-CαH(**C**OOH)-CβH_2-); 156.48 (Fmoc-O-**C**ONH-); 144.25, 141.17 (Fmoc aryl **ipso** C's); 128.10, 127.54, 125.77, 125.74, 120.58 (Fmoc Ar **C**-H); 71.71, 70.65, 70.25, 70.20, 70.18, 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-); 66.197 (Fmoc Ar_2CH-**C**H_2-O-); 58.49 (-O-**C**H_3); 47.06 (-NH-**C**αH(COOH); 33.61 (-CαH(COOH)-**Cβ**H_2-S-); 54.69 (Fmoc Ar_2**C**H-CH_2-O-); 31.53 (S-**C**H_2-CH_2-O-); The full ^{13}C NMR spectrum for Fmoc-L-Cys(PEG4)-OH is shown in Figure 3-42.

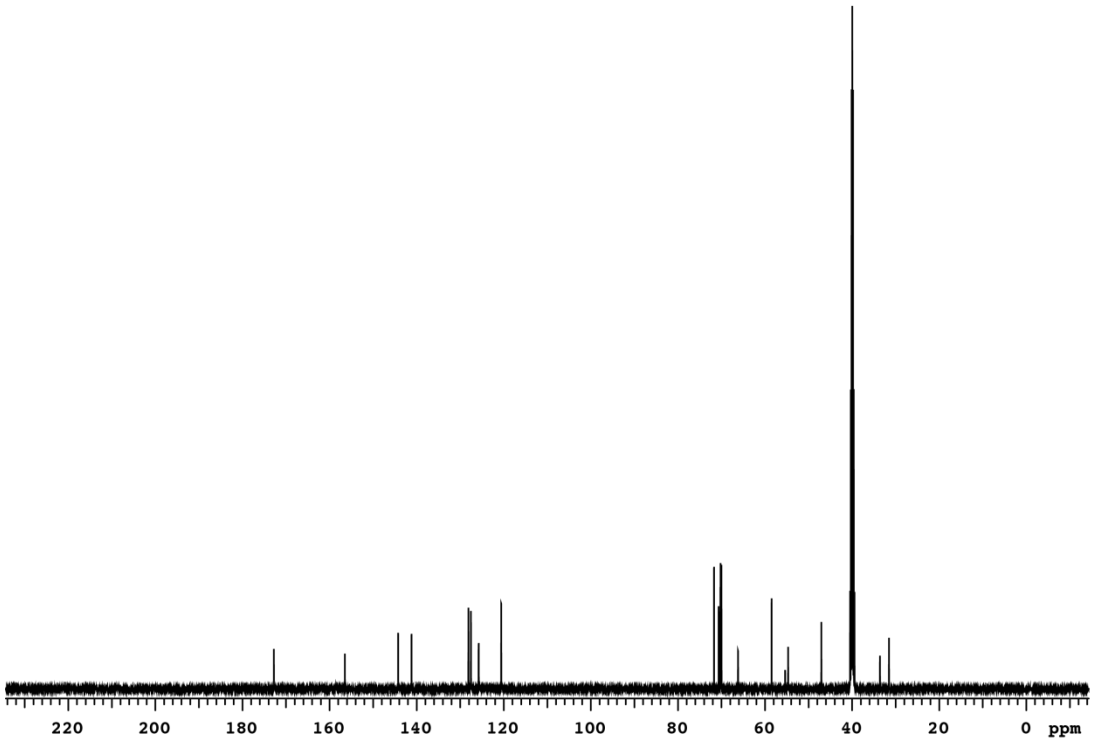


Figure 3-42. ^{13}C NMR spectrum for Fmoc-L-Cys(PEG4)-OH.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-L-Cys(PEG4)-OH were with the assistance of a 2D HSQC experiment (Figure 3-43), using the one-bond C-H correlations shown in Table 3-2.

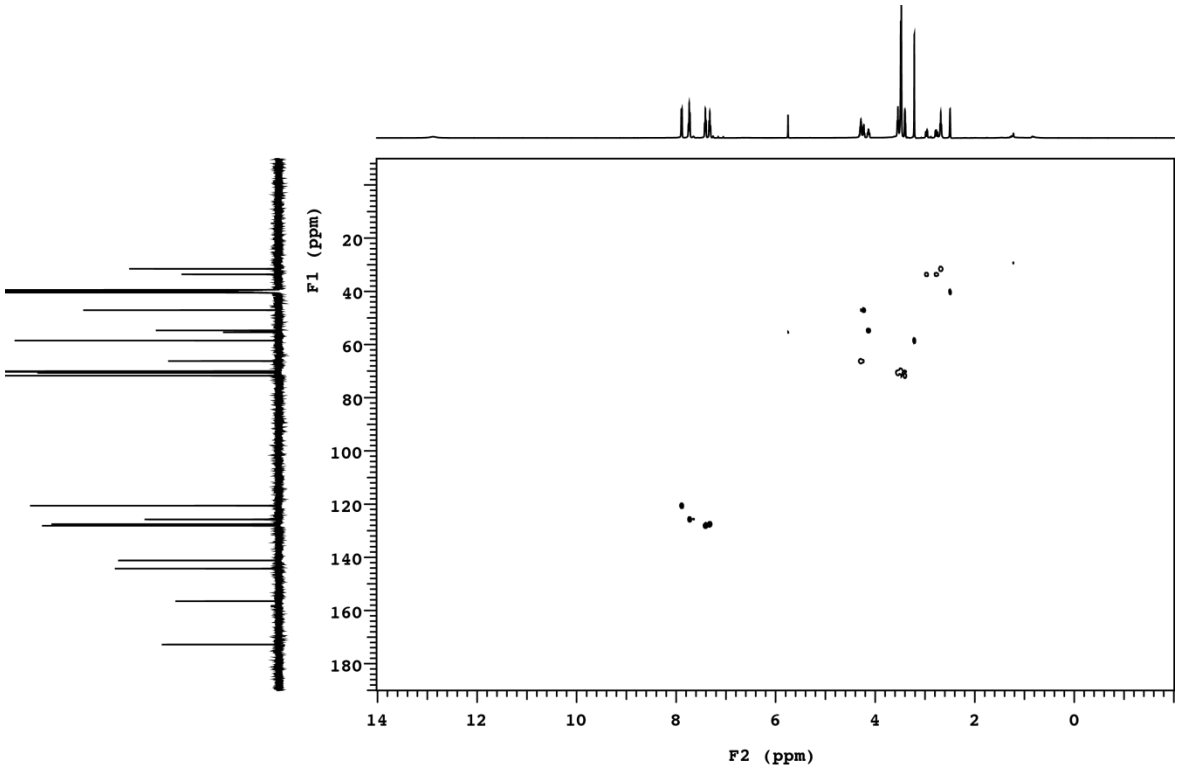
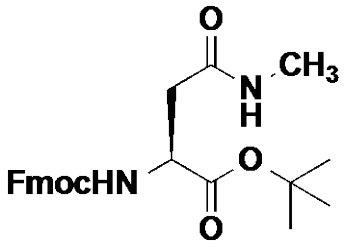


Figure 3-43. 2D HSQC spectrum of Fmoc-L-Cys(PEG4)-OH.

Table 3-2. One-Bond C-H correlations identified from HSQC data for Fmoc-L-Cys(PEG4)-OH.

^{1}H δ	^{1}C δ	Assignment
7.896	128.105	Fmoc aryl C-H
7.737	127.544	Fmoc aryl C-H
7.413	125.771	Fmoc aryl C-H
7.323	120.58	Fmoc aryl C-H
4.259-4.213	47.067	-CONH**CαH**(COOH)-S-
4.324-4.280	66.197	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
4.157-4.111	54.691	Fmoc Ar2**CH**-CH2-O-
3.556-3.390	71.711-69.988	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-
2.682	31.532	S-**CH$_2$**-CH$_2$-O
3.214	58.494	-O-**CH$_3$**
2.986-2.746	33.611	-CαH(COOH)-**CβH$_2$**-

3.4.10 Fmoc-Methyl-L-Asn-OtBu (S2)



To a solution of (*S*)-3-((((*9H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert*-butoxy)-4-oxobutanoic acid (Fmoc-L-Asp-OtBu, 1.0g, 2.4 mmol) in NMP (15 mL) was added 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU, 1.5g, 3.6mmol), N-hydroxybenzotriazole hydrate (HOBt, 0.55g, 3.6mmol), and *N*,*N*-Diisopropylethylamine (DIEA, 1.27mL, 7.29mmol), and the resulting mixture was stirred for 15 min at room temperature. Then, Methylamine hydrochloride (0.18g, 2.7mmol) was added, and the resulting mixture was stirred at room temperature for 24 hours. To the reaction mixture was first added saturated sodium bicarbonate (100mL) and then saturated Brine (100 mL) was added to the reaction/bicarbonate mixture and extracted three times with ethyl acetate (100 mL). The combined ethyl acetate extracts were dried over $MgSO_4$, and concentrated by rotary evaporation to afford an oil. Fmoc-Methyl-L-Asn-OtBu [*tert*-butyl N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^4-methyl-*L*-asparaginat] was purified by flash chromatography over silica in ethyl acetate/hexanes. The product was concentrated via rotary evaporation (followed by rewashing with saturated bicarbonate three times, and saturated brine three times, and co-evaporation three times with benzene (200mL) to remove residual NMP) and dried in vacuo to give a thick oily solid (98.0% yield).

^{1}H NMR (500 MHz, $CDCl_3$): δ 7.78(2H, d, J = 7.6 Hz, Fmoc aryl C-**H**); 7.61 (2H, m, Fmoc aryl C-**H**); 7.41 (2H, t, J = 7.2 Hz, Fmoc aryl C-**H**); 7.32 (2H, t, J = 7.2Hz, Fmoc aryl C-**H**); 6.09 (1H, d, amide); 5.73 (1H, s, amide); 4.47 (1H, m, -CONHCα**H**(COOtBu)-CβH_2-); 4.40, 4.34 (2H, dd,

Fmoc Ar_2CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.23 (1 H, t, J = 7.0 Hz, Fmoc Ar2C**H**-CH_2-O-); 2.89 (1H, dd, J = 15.3 Hz, 3.4 Hz,-CαH(COOH)-Cβ(**Ha**)Hb-$CONHCH_3$); 2.72 (1H, dd, J = 15.1 Hz, 2.9 Hz, -CαH(COOH)-Cβ(Ha)**Hb**-$CONHCH_3$); 2.80 (3H, d, J = 3.6 Hz, -NH-C**H$_3$);** 1.49 (9H, s, -O-C-(C**H$_3$**)$_3$. The full ^{1}H NMR spectrum for Fmoc-Methyl-L-Asn-OtBu is shown in Figure 3-44

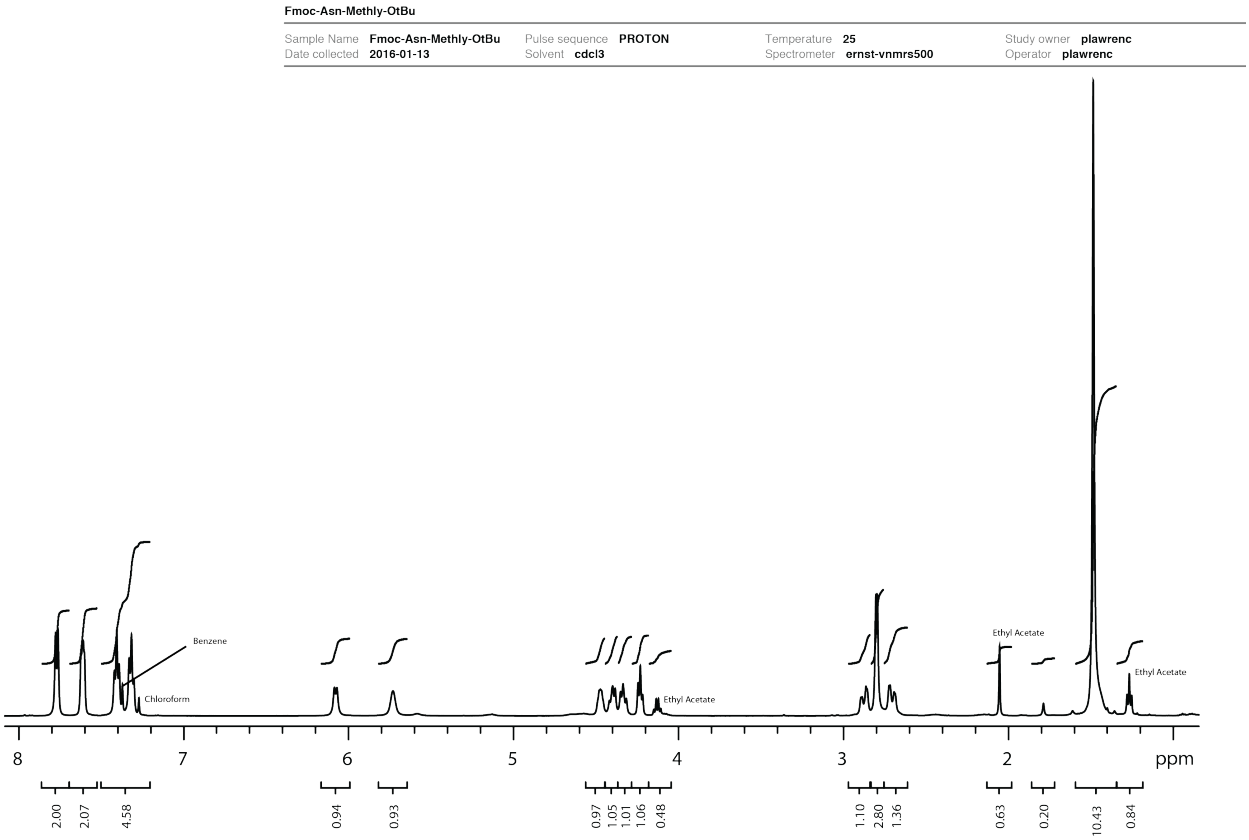


Figure 3-44. ^{1}H NMR spectrum for Fmoc-Methyl-L-Asn-OtBu.

^{13}C NMR (126 MHz, $CDCl_3$): δ 170.39, 170.04 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.23 (Fmoc-O-**C**ONH-); 143.92, 143.78, 141.28, (Fmoc aryl **ipso** C's); 127.71, 127.07, 125.22, 119.98, (Fmoc aryl **C**-H); 82.42 (-O-**C**$(CH_3)_3$); 67.12 (Fmoc Ar_2CH-**C**H_2-O-); 26.28 (-NH-**C**H_3); 51.47 (-NH-**C**αH(COOtBu)-CβH_2-); 47.12 (Fmoc Ar_2**C**H-CH_2-O-); 37.97 (-

CαH(COOH)-**C**βH_2-CONHCH_3-); 27.91 (-O-C(**C**H_3)$_3$). The full ^{13}C NMR spectrum for Fmoc-Methyl-L-Asn-OtBu is shown in Figure 3-45

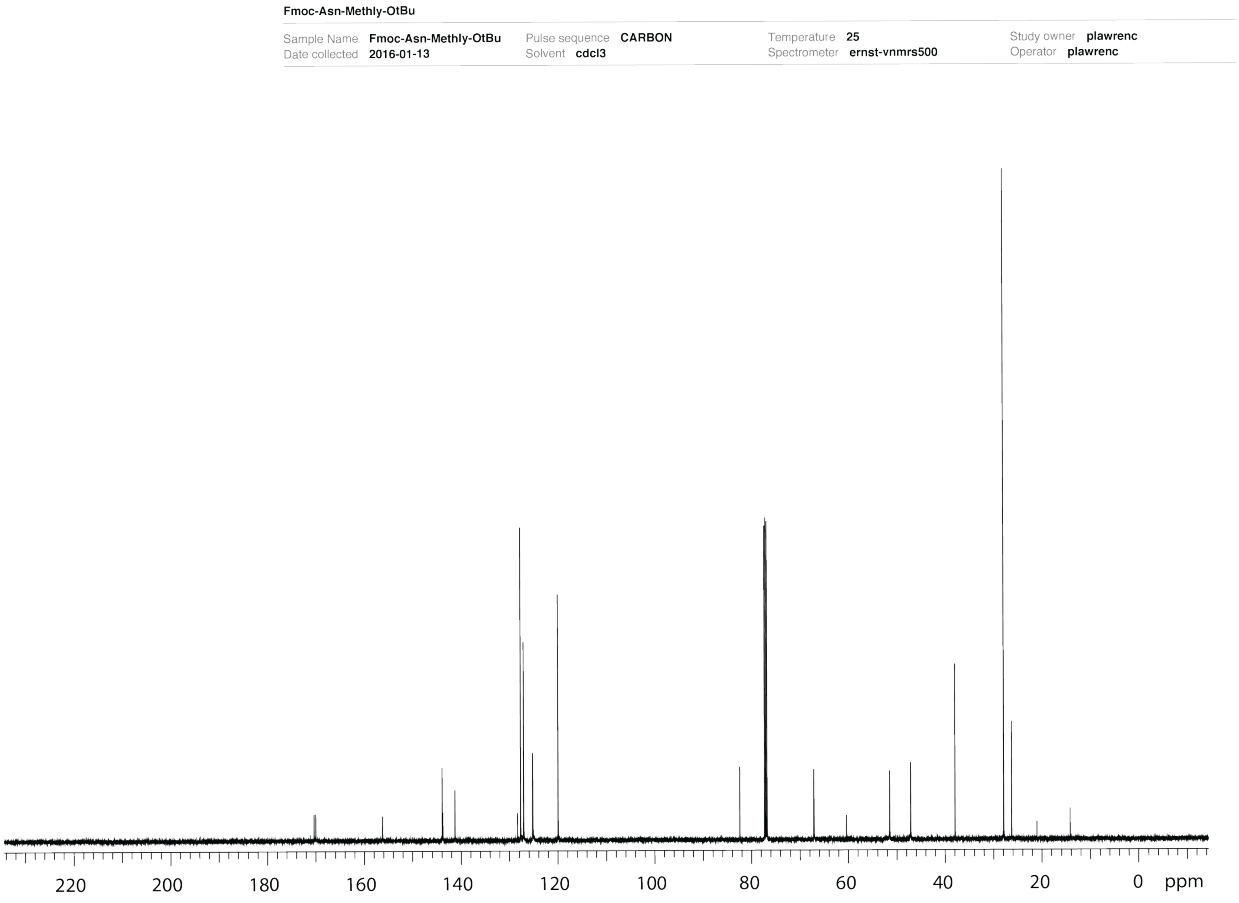


Figure 3-45. ^{13}C NMR spectrum for Fmoc-Methyl-L-Asn-OtBu.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-Methyl-L-Asp-OtBu were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-46, using the one-bond C-H correlations shown in Table 3-3.

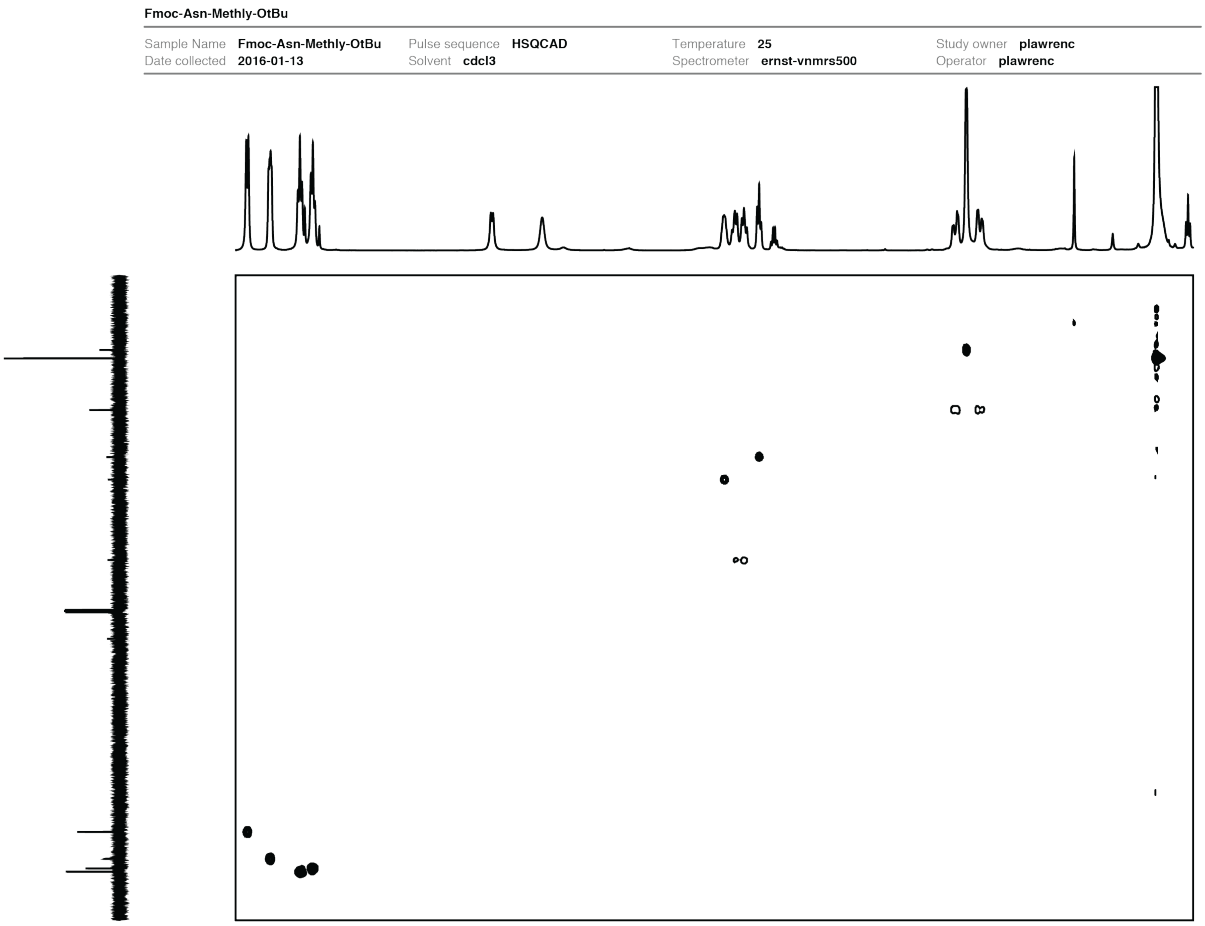


Figure 3-46. 2D HSQC spectrum of Fmoc-Methyl-L-Asn-OtBu.

Table 3-3. One-Bond C-H correlations identified from HSQC data for Fmoc-Methyl-L-Asn-OtBu.

^{1}H δ	^{1}C δ	Assignment
7.777	119.976	Fmoc aryl C-H
7.609	125.221	Fmoc aryl C-H
7.406	127.706	Fmoc aryl C-H
7.316	127.073	Fmoc aryl C-H
4.474	51.469	-CONH**CαH**(COOH)-CβH2-
4.232	47.123	Fmoc Ar**CH**-CH2-O-
4.400, 4.337	67.115	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
2.803, 2.796	26.279	NH-CH_3
2.894, 2.723	37.97	-CαH(COOH)-**CβH$_2$**-
1.486	27.914	-C(**CH$_3$**)$_3$

High-resolution ESI-TOF MS data for Fmoc-Methyl-L-Asn-OtBu are shown in Figure 3-47.

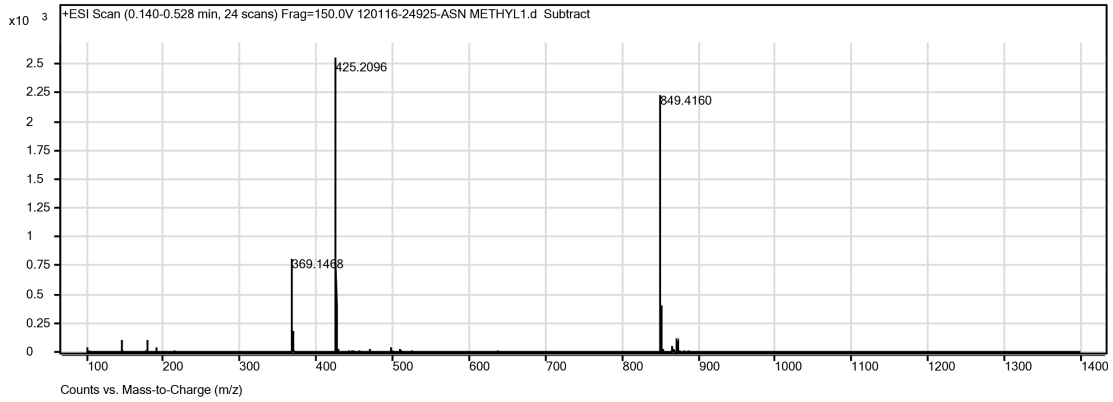
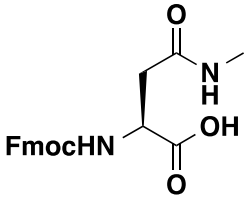


Figure 3-47. ESI-TOF MS data for Fmoc-Methyl-L-Asn-OtBu. Calculated m/z for $C_{24}H_{29}N_2O_5$ ($M+H^+$) is 425.21, found 425.2096

3.4.11 Fmoc-Methyl-L-Asn-OH (S3)



To a solution of TFA (95% in water, 50ml) was added 1.01 g Fmoc-Methyl-L-Asn-OtBu and the solution was stirred for 4 h under an argon atmosphere. The product Fmoc-Methyl-L-Asn-OH [N^2-(((9*H*-fluoren-9-yl)methoxy)carbonyl)-N^4-methylasparagine, **S3**] was concentrated by rotary evaporation, and used without further purification.

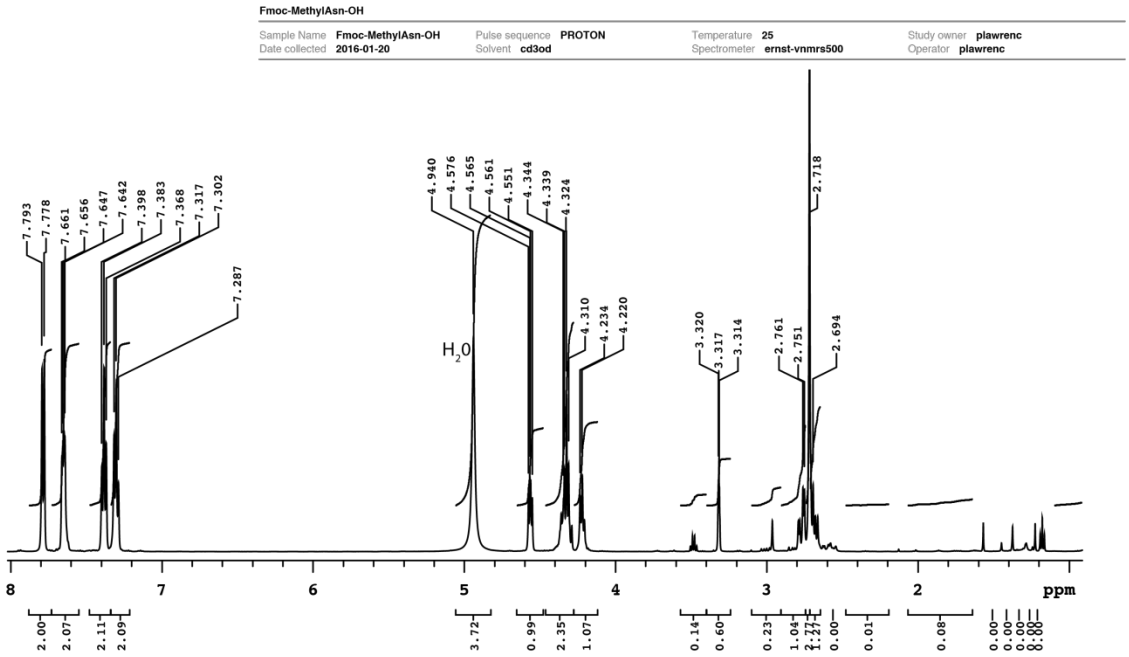


Figure 3-48. ^{1}H NMR spectrum for Fmoc-Methyl-L-Asn-OH.

^{1}H NMR (500 MHz, CD_3OD): δ 7.79(2H, d, J = 7.8 Hz, Fmoc aryl C-**H**); 7.66 (2H, m, Fmoc aryl C-**H**); 7.40 (2H, t, J = 7.5 Hz, Fmoc aryl C-**H**); 7.32 (2H, t, J = 7.5 Hz, Fmoc aryl C-**H**); 4.47 (1H, dd, J = 5.4 Hz, 12.7 Hz, -CONHCα**H**(COOtBu)-CβH_2-); 4.36 (2H, m, Fmoc Ar_2CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.23 (1 H, t, Fmoc Ar2C**H**-CH_2-O-); 2.79 (1H, dd, J = 15.1 Hz, 5.0 Hz,-CαH(COOH)-Cβ(**Ha**)Hb-$CONHCH_3$); 2.69 (1H, dd, J = 15.1 Hz, 7.6 Hz, -CαH(COOH)-Cβ(Ha)**Hb**-$CONHCH_3$); 2.80 (3H, s, -NH-C**H$_3$**). The full ^{1}H NMR spectrum for Fmoc-Methyl-L-Asn-OH is shown in Figure 3-48.

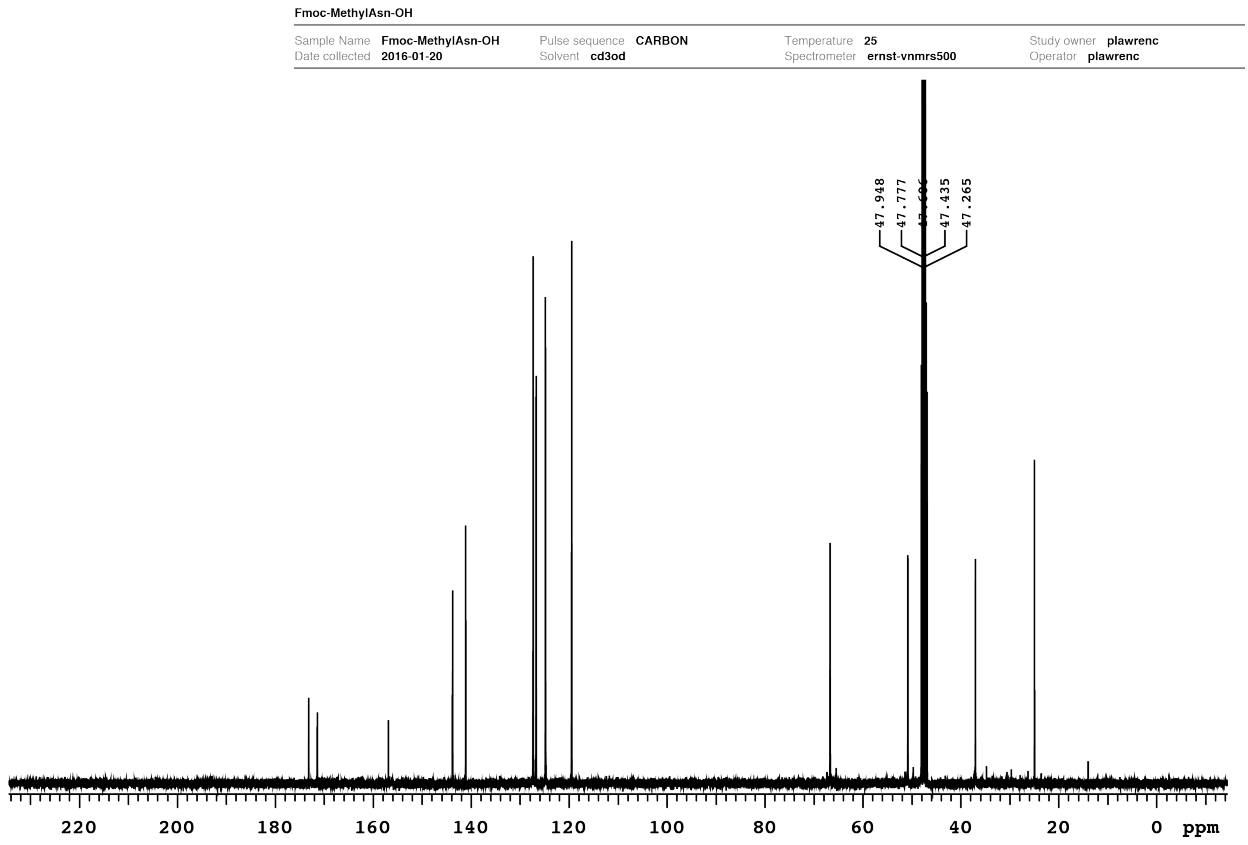


Figure 3-49. ^{13}C NMR spectrum for Fmoc-Methyl-L-Asn-OH.

^{13}C NMR (126 MHz, $CDCl_3$): δ 173.20, 171.44 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.93 (Fmoc-O-**C**ONH-); 143.85, 143.80, 141.15, (Fmoc aryl **ipso** C's); 127.41, 127.37, 126.75, 124.86, 119.51 (Fmoc aryl **C**-H); 66.72 (Fmoc Ar_2CH-**C**H_2-O-); 24.99 (-NH-**C**H_3); 50.86 (-NH-**C**αH(COOH)-CβH_2-); 46.90 (Fmoc Ar_2**C**H-CH_2-O-); 37.06 (-CαH(COOH)-**Cβ**H_2-CONHCH_3-). The full ^{13}C NMR spectrum for Fmoc-Methyl-L-Asn-OH is shown in Figure 3-49.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-Methyl-L-Asn-OH were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-50, using the one-bond C-H correlations shown in Table 3-4.

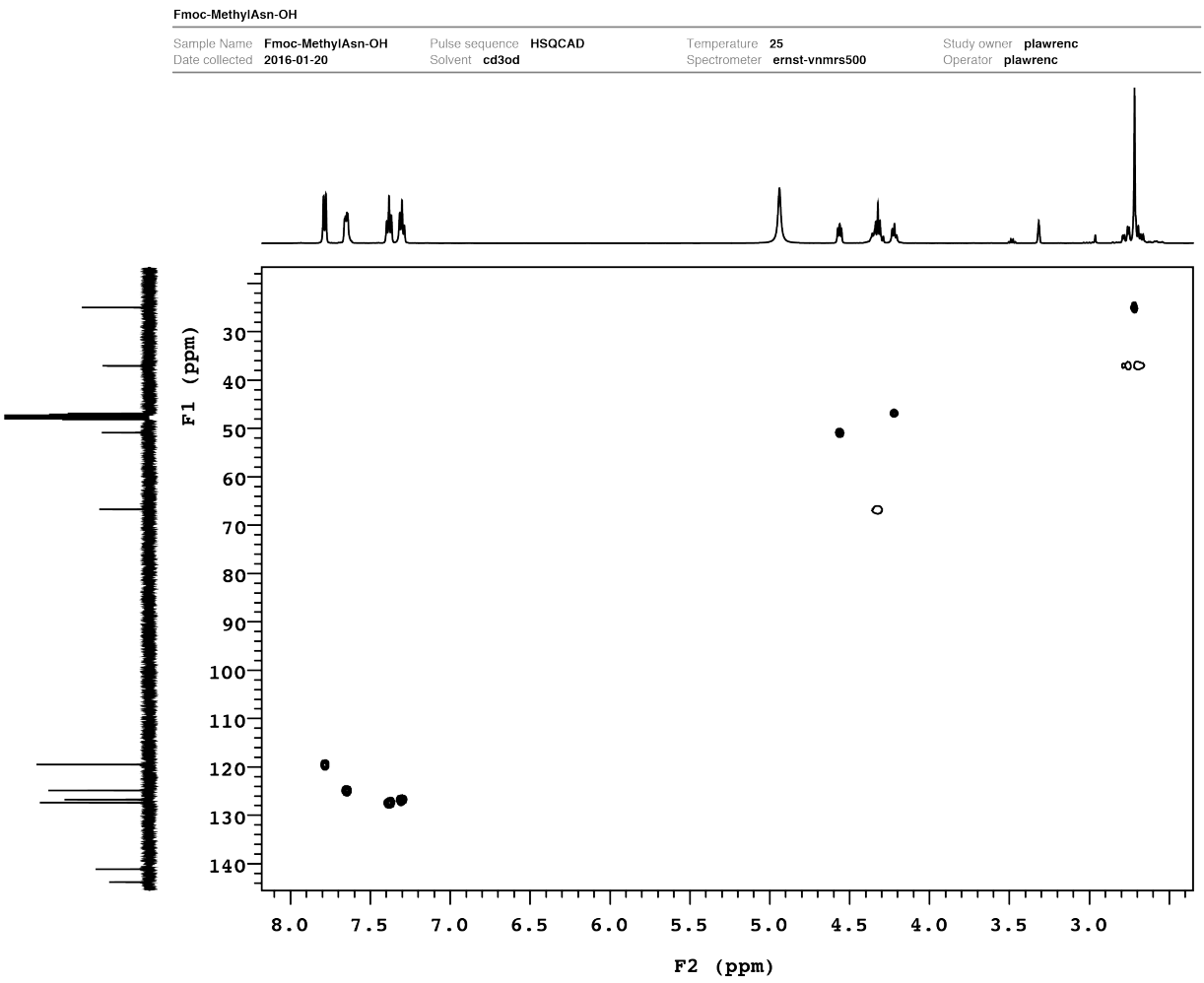


Figure 3-50. 2D HSQC spectrum of Fmoc-Methyl-L-Asn-OH.

Table 3-4. One-Bond C-H correlations identified from HSQC data for Fmoc-Methyl-L-Asn-OH.

^{1}H δ	^{1}C δ	Assignment
7.79	119.51	Fmoc aryl C-H
7.66	124.86, 126.75	Fmoc aryl C-H
7.4	127.41	Fmoc aryl C-H
7.32	127.37	Fmoc aryl C-H
4.58	50.86	-CONH**CαH**(COOH)-CβH2-
4.23	46.9	Fmoc Ar**CH**-CH2-O-
4.36	66.72	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
2.72	24.99	NH-CH_3
2.79, 2.69	37.06	-CαH(COOH)-**CβH$_2$**-

High-resolution ESI-TOF MS data for Fmoc-Methyl-L-Asn-OH are shown in Figure 3-51

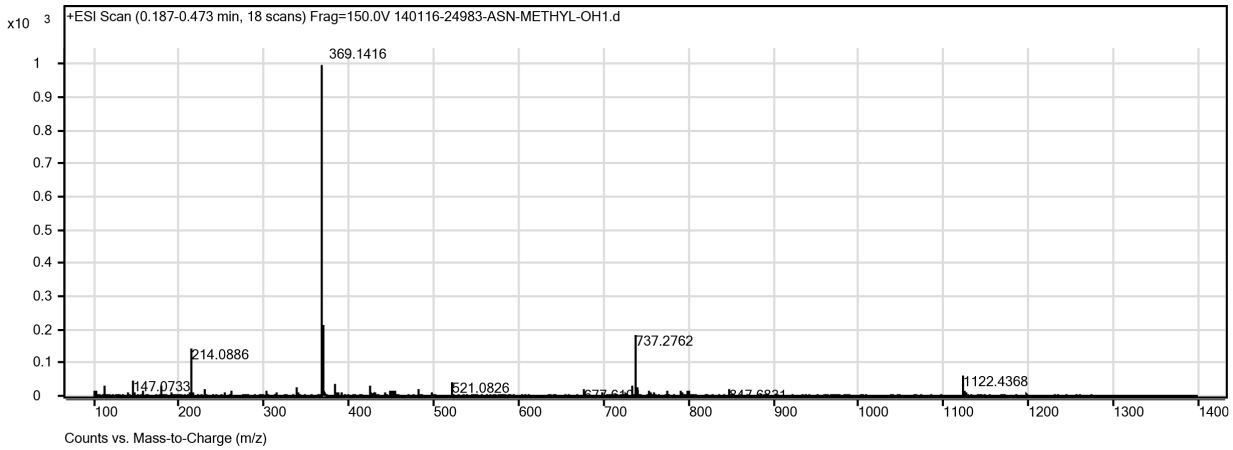
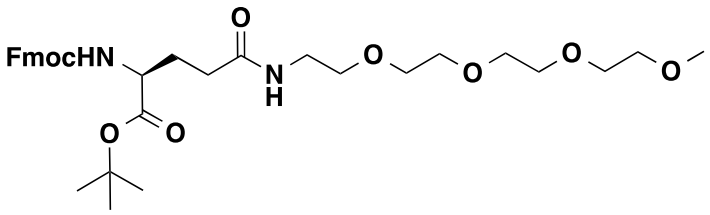


Figure 3-51. ESI-TOF MS data for Fmoc-Methyl-L-Asn-OH. Calculated m/z for $C_{20}H_{21}N_2O_5$ ($M+H^+$) is 369.15, found 369.1416

3.4.12 Fmoc-L-Glu-OtBu (S4)



To a solution of (*S*)-4-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-5-(*tert*-butoxy)-5-oxopentanoic acid (Fmoc-L-Glu-OtBu, 1.0g, 2.35 mmol) in NMP (15 mL) was added 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU, 1.45g, 3.52mmol), N-hydroxybenzotriazole hydrate (HOBt, 0.539g, 3.52mmol), and *N*,*N*-Diisopropylethylamine (DIEA, 1.22mL, 7.05mmol), and the resulting mixture was stirred for 15 min at room temperature. Then, 2,5,8,11-tetraoxatridecan-13-amine (0.54g, 2.59mmol) was added, and the resulting mixture was stirred at room temperature for 24 hours. To the reaction mixture was first added saturated sodium bicarbonate (100mL) and then saturated Brine (100 mL) was added to the reaction/bicarbonate mixture and extracted three times with ethyl acetate (100 mL). The combined ethyl acetate extracts were dried over $MgSO_4$, and concentrated by rotary evaporation to afford an oil. Fmoc-L-GlnPEG_4-OtBu *tert*-butyl (*S*)-18-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-15-oxo-2,5,8,11-tetraoxa-14-azanonadecan-19-oate was purified by flash chromatography over silica in ethyl acetate/hexanes. The product was concentrated via rotary evaporation (followed by rewashing with saturated bicarbonate three times, and saturated brine three times, and co-evaporation three times with benzene (200mL) to remove residual NMP) and dried in vacuo to give a thick oily solid (81% yield).

^{1}H NMR (500 MHz, $CDCl_3$): δ 7.75(2H, d, J = 7.6 Hz, Fmoc aryl C-**H**); 7.60 (2H, t, Fmoc aryl C-**H**); 7.38 (2H, t, J = 7.5 Hz, Fmoc aryl C-**H**); 7.30 (2H, t, J = 7.5 Hz, Fmoc aryl C-**H**); 6.46 (1H, s, amide); 5.87 (1H, d, J = 8.23 Hz, amide); 4.43 (2H, m, Fmoc Ar_2CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.23

(2H, m, -CONHCα**H**(COOtBu)-CβH_2-, and Fmoc Ar2C**H**-CH_2-O-); 3.59-3.45 (16H, m, -CONH-(C**H_2**-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-); 3.32 (3H, s, -O-C**H_3**) 2.28-1.97 (4H, m, -CαH(COOH)-Cβ**H_2**-Cγ**H_2**-CONHCH_2-); 1.46 (9H, s, -O-C-(C**H_3**)$_3$. The full ^{1}H NMR spectrum for Fmoc-L-GlnPEG$_4$-OtBu is shown in Figure 3-18.

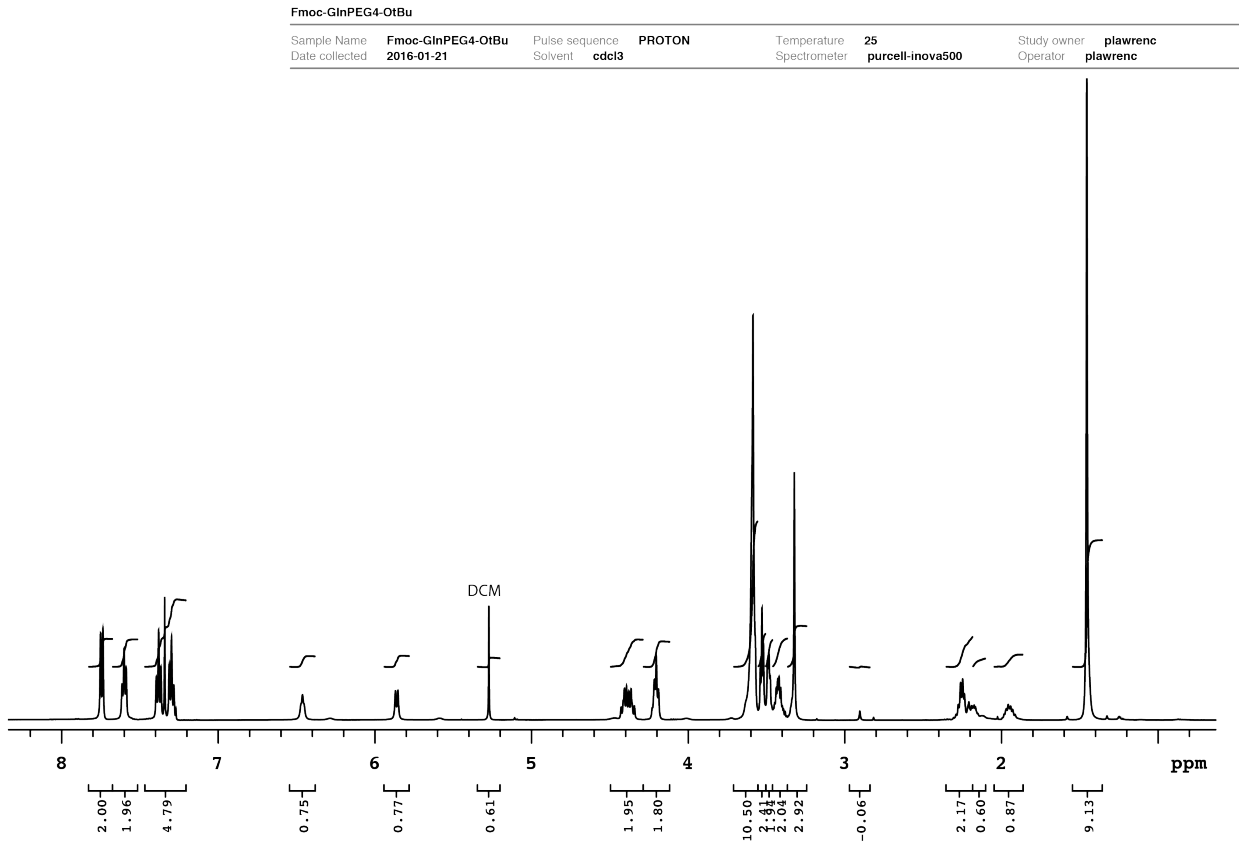


Figure 3-18. ^{1}H NMR spectrum for Fmoc-L-GlnPEG$_4$-OtBu.

^{13}C NMR (75.4 MHz, $CDCl_3$): δ 172.13, 171.24 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.30 (Fmoc-O-**C**ONH-); 143.97, 143.77, 141.30, 141.26 (Fmoc aryl **ipso** C's); 128.32, 127.79, 127.07, 125.15, 125.12, 119.96 (Fmoc aryl **C**-H); 82.18 (-O-**C**$(CH_3)_3$); 71.83-69.72 and 39.342 (-NH-**C**H_2-**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-); 66.79 (Fmoc Ar$_2$CH-**C**H_2-

O-); 58.93, (-O-**C**H_3); 54.18 (-NH-**C**αH(COOtBu)-CβH_2-); 47.18 (Fmoc Ar_2**C**H-CH_2-O-); 32.44 and 28.50 (-CαH(COOH)-**C**βH_2-**C**γH_2-CONHCH_2-); The full ^{13}C NMR spectrum for Fmoc-L-GlnPEG$_4$-OtBu is shown in Figure 3-19.

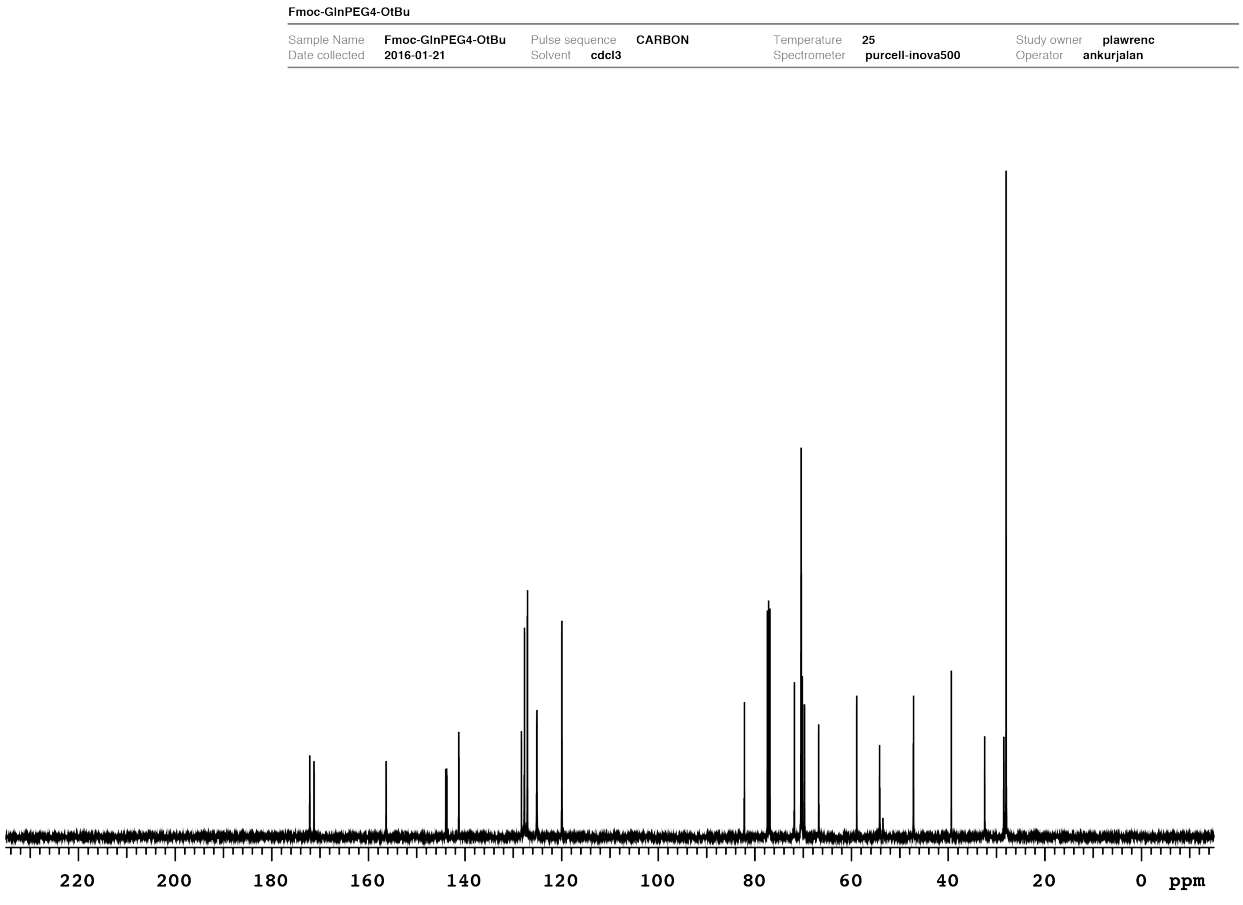


Figure 3-19. ^{13}C NMR spectrum for Fmoc-L-GlnPEG$_4$-OtBu.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-L-GlnPEG$_4$-OtBu were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-20, using the one-bond C-H correlations shown in Table 3-4

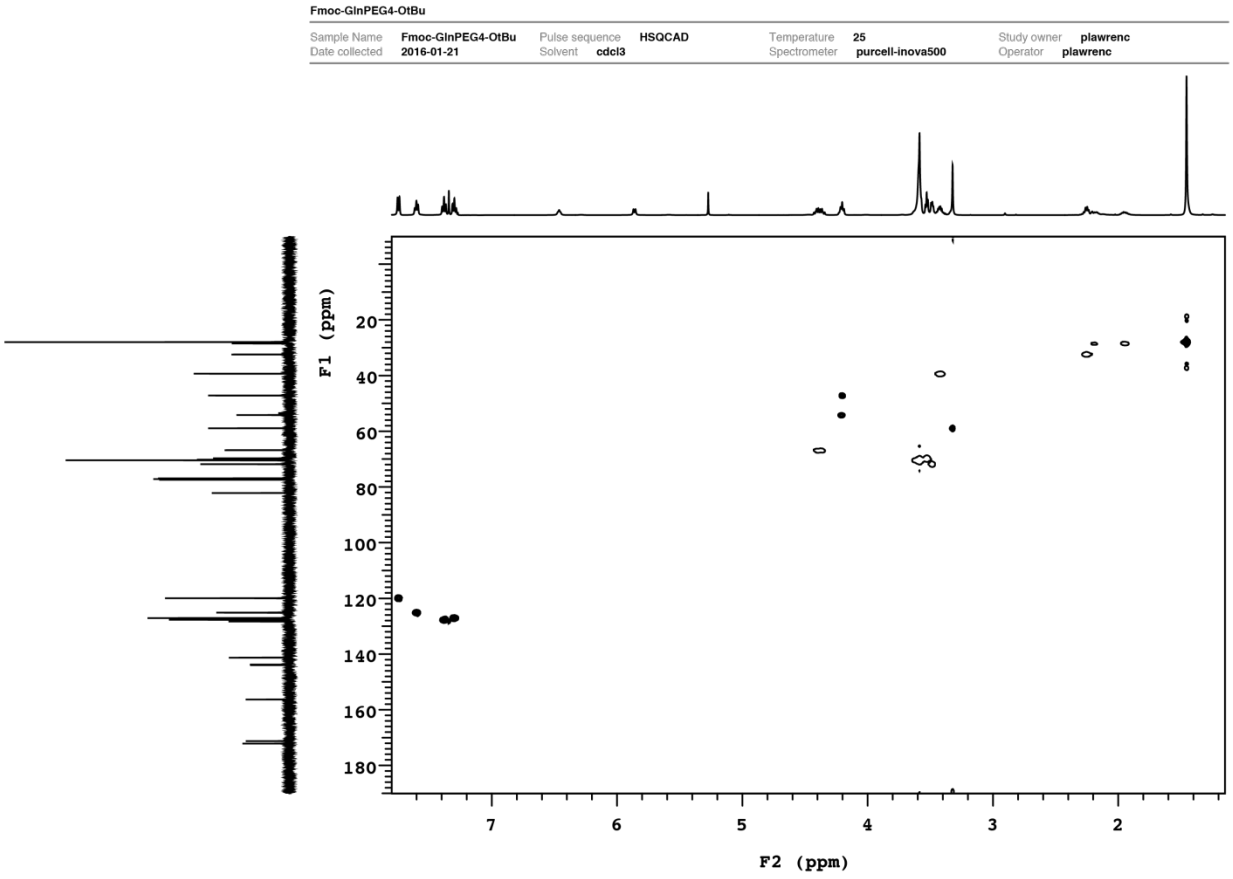


Figure 3-20. 2D HSQC spectrum of Fmoc-L-GlnPEG$_4$-OtBu.

Table 3-4. One-Bond C-H correlations identified from HSQC data for Fmoc-L-GlnPEG$_4$-OtBu.

^{1}H δ	^{1}C δ	Assignment
7.75	119.95	Fmoc aryl C-H
7.60	125.15	Fmoc aryl C-H
7.38	128.32, 127.69	Fmoc aryl C-H
7.30	127.07	Fmoc aryl C-H
4.23	54.18, 47.18	-CONH**CαH**(COOH)-CβH2-Fmoc Ar**CH**-CH2-O-
4.43	66.79	Fmoc Ar2CH-**CH$_2$**-O
3.587-3.448	71.827-69.722, and 39.342	-CONH-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-,
3.32	58.93	-O-**CH3**
2.276-1.969	32.441, 28.490	-CONHCαH(COOH)-**CβH2-CγH2**-CONHO-
1.46	27.99	-C(**CH$_3$**)$_3$

High-resolution ESI-TOF MS data for Fmoc-L-GlnPEG$_4$-OtBu are shown in Figure 3-21.

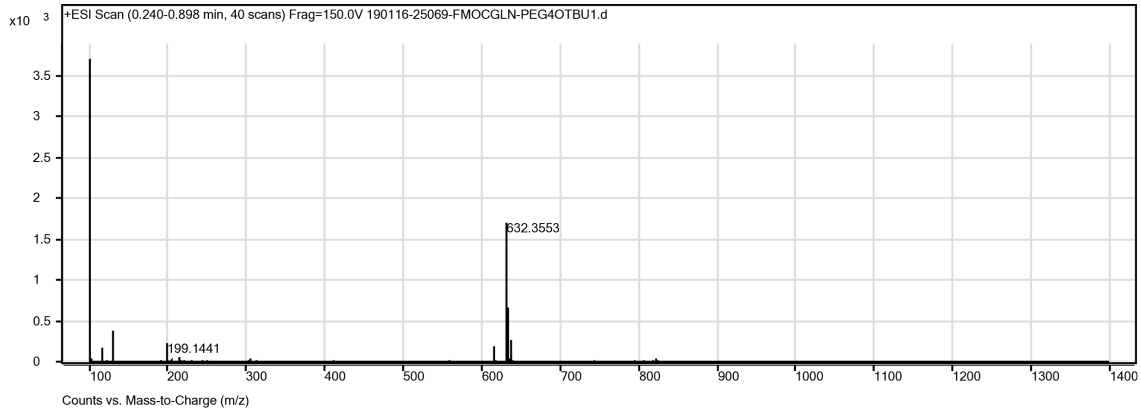
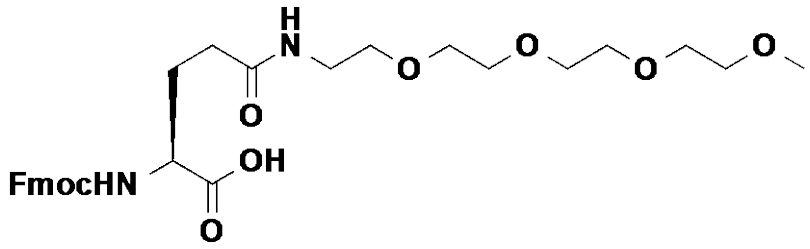


Figure 3-21. ESI-TOF MS data for Fmoc-L-GlnPEG$_4$-OtBu. Calculated m/z for $C_{33}H_{50}N_3O_9$ ($M+NH_4^+$) is 632.35, found 632.3553

3.4.13 Fmoc-L-GlnPEG$_4$-OH (S5)



To a solution of TFA (95% in water, 50ml) was added 1.17 g Fmoc-L-GlnPEG$_4$-OtBu (**S4**), and the solution was stirred for 4 h under an argon atmosphere. The product Fmoc-L-GlnPEG$_4$-OH [18-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-15-oxo-2,5,8,11-tetraoxa-14-azanonadecan-19-oic acid, **S5**] was concentrated by rotary evaporation, and used without further purification.

^{1}H NMR (500 MHz, $CDCl_3$):δ 11.01 (1H, broad s, (COO**H**); δ 7.77(2H, d, J = 7.4 Hz, Fmoc aryl C-**H**); 7.61 (2H, d, J = 7.2 Hz, Fmoc aryl C-**H**); 7.42 (2H, t, J = 7.4 Hz, Fmoc aryl C-**H**); 7.33 (2H, t, J = 7.4 Hz, Fmoc aryl C-**H**); 6.07 (1H, d, amide); 4.45 (2H, m, Fmoc Ar$_2$CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.40 (1H, m, -CONHCα**H**(COOH); and 4.23 (1H, dd, Fmoc Ar2C**H**-CH_2-O-); 3.63-3.42 (16H, m, -CONH-(C**H$_2$**-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-); 3.38 (3H, s, -O-C**H$_3$**) 2.46 (2H, t, J = 6.9 Hz, -CαH(COOH)-CβH_2-Cγ**H$_2$**-$CONHCH_2$-); 2.27-2.14 (2H, m, -CαH(COOH)-Cβ**H$_2$**-CγH_2-$CONHCH_2$-) . The full ^{1}H NMR spectrum for Fmoc-L-GlnPEG$_4$-OH is shown in Figure 3-22.

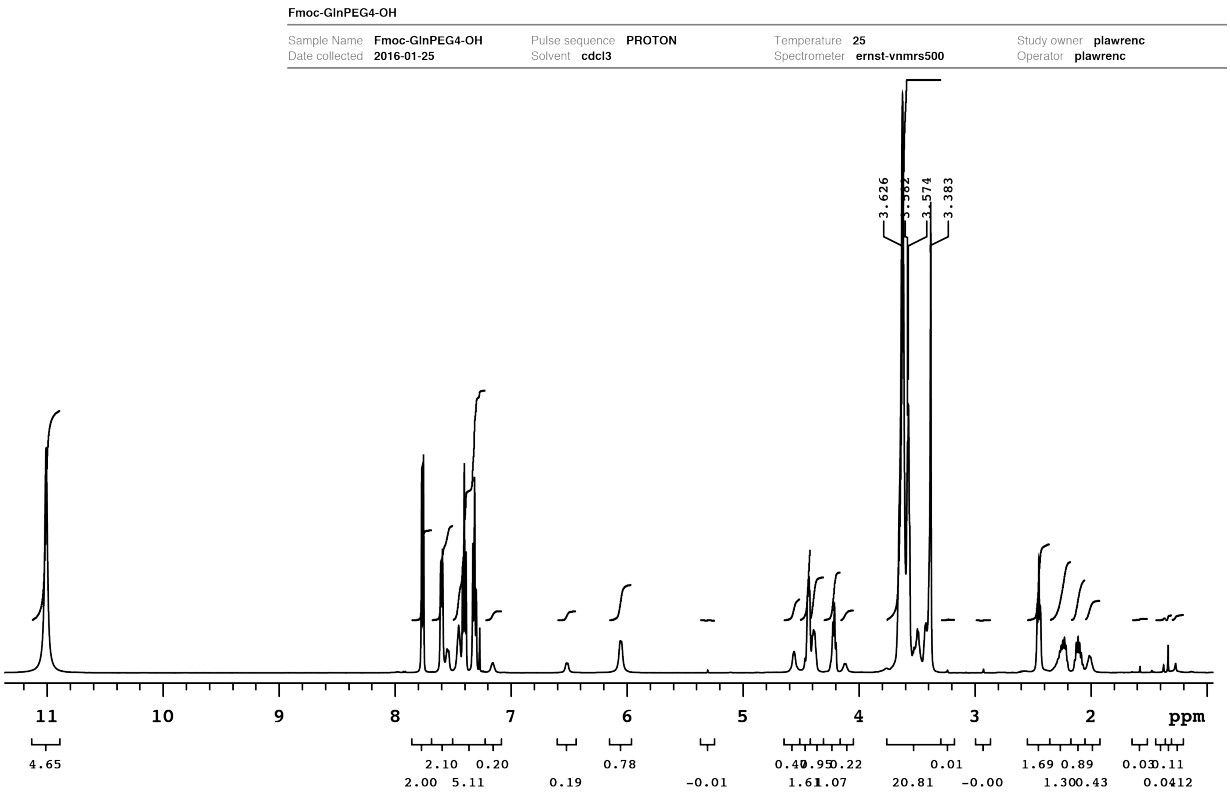


Figure 3-22. ^{1}H NMR spectrum for Fmoc-L-GlnPEG$_4$-OH.

^{13}C NMR (75.4 MHz, $CDCl_3$): δ 174.86, 174.27, 159.54, 158.88, 158.55 (-NH-CαH(**C**OOtBu)-CβH$_2$- and -CβH$_2$-**C**ON-CH$_2$-); 156.84 (Fmoc-O-**C**ONH-); 143.62, 143.49, 141.32, 141.28 (Fmoc aryl **ipso** C's); 127.81, 127.15, 125.04, 120.00, (Fmoc aryl **C**-H); 71.53-68.95 and 39.83 (-NH-**C**H$_2$-**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-); 67.41 (Fmoc Ar$_2$CH-**C**H$_2$-O-); 58.62, (-O-**C**H$_3$); 53.11 (-NH-**C**αH(COOtBu)-CβH$_2$-); 46.98 (Fmoc Ar$_2$**C**H-CH$_2$-O-); 31.91 (-CαH(COOH)-CβH$_2$-**C**γH$_2$-CONHCH$_2$-); 28.39 (-CαH(COOH)-**Cβ**H$_2$-CγH$_2$-CONHCH$_2$-) The full ^{13}C NMR spectrum for Fmoc-L-GlnPEG$_4$-OH is shown in Figure 3-23.

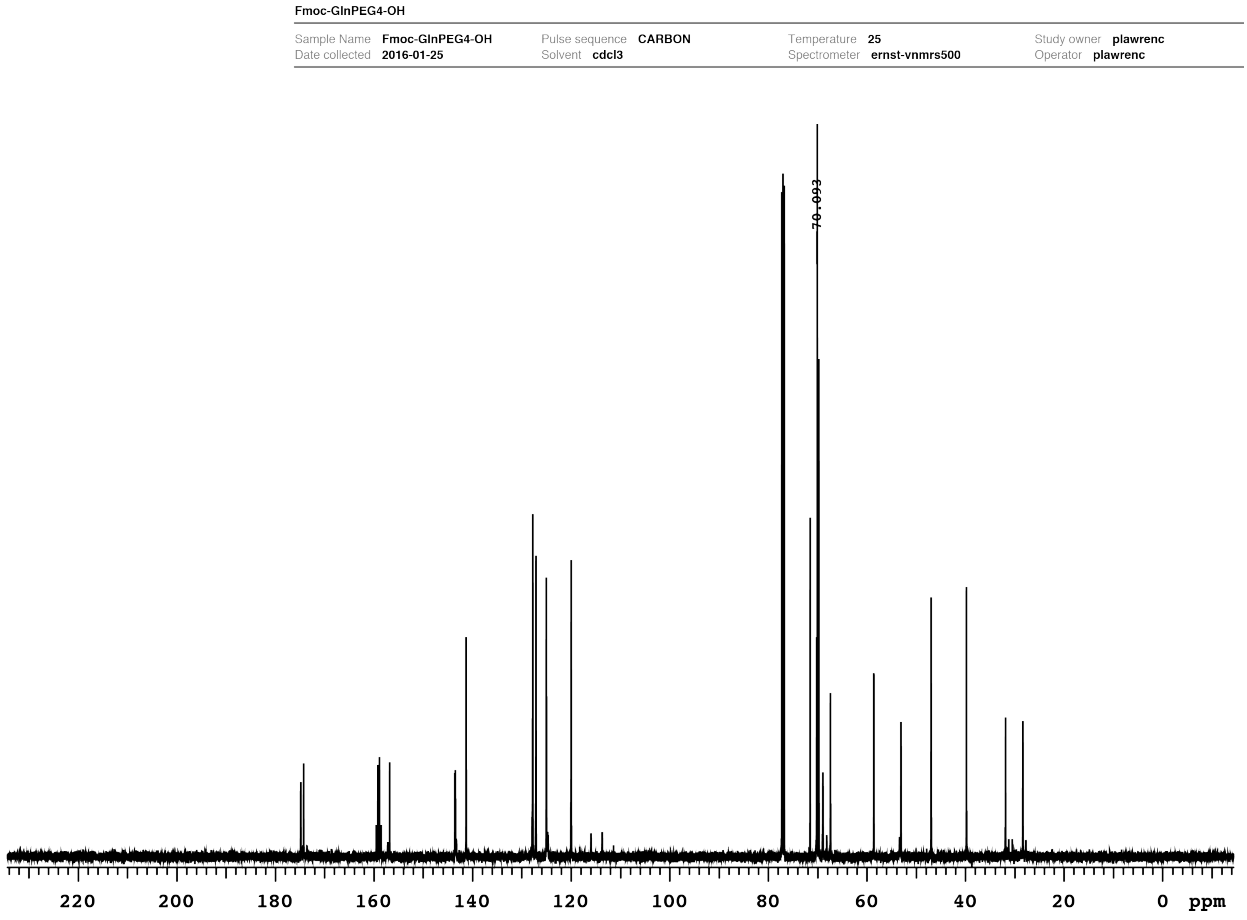


Figure 3-23. ^{13}C NMR spectrum for Fmoc-L-GlnPEG$_4$-OH.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-L-GlnPEG$_4$-OtBu were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-24, using the one-bond C-H correlations shown in Table 3-5.

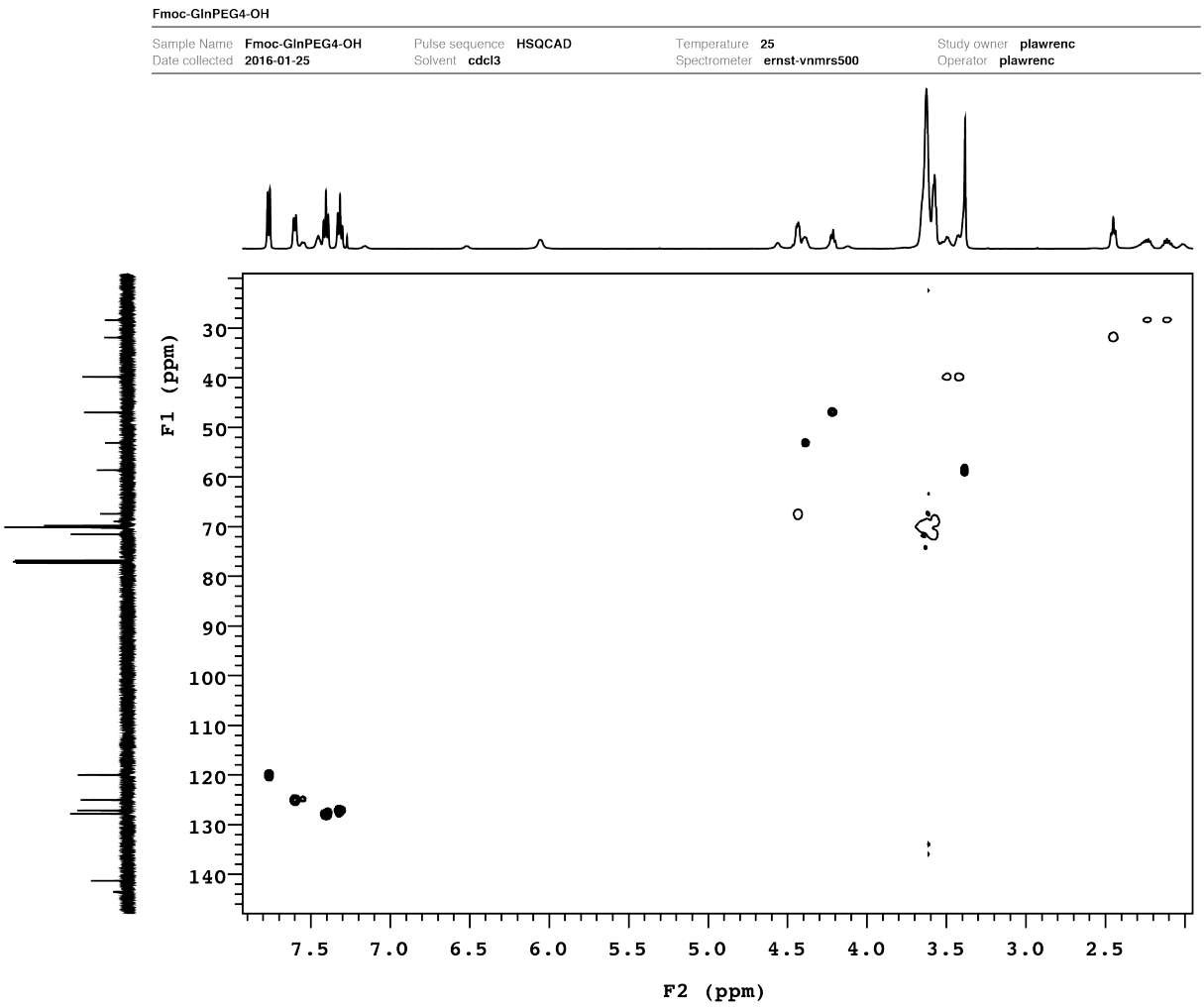


Figure 3-24. 2D HSQC spectrum of Fmoc-L-GlnPEG$_4$-OH.

Table 3-5. One-Bond C-H correlations identified from HSQC data for Fmoc-L-GlnPEG$_4$-OH.

^{1}H δ	^{1}C δ	Assignment
7.77	120.00	Fmoc aryl C-H
7.61	125.04	Fmoc aryl C-H
7.42	127.81	Fmoc aryl C-H
7.33	127.15	Fmoc aryl C-H
4.402	53.11	-CONH**CαH**(COOH)-CβH2-
4.23	46.98	Fmoc Ar**CH**-CH2-O-
4.45	67.41	Fmoc Ar2CH-**CH$_2$**-O
3.626-3.424	71.532-68.947, and 39.826	-CONH-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-,
3.38	58.62	-O-**CH3**
2.46	31.91	-CONHCαH(COOH)-CβH2-**CγH2**-CONHO-
2.270-2.141	28.39	-CONHCαH(COOH)-**CβH2**-CγH2-CONHO-

High-resolution ESI-TOF MS data for Fmoc-L-GlnPEG$_4$-OH are shown in Figure 3-25.

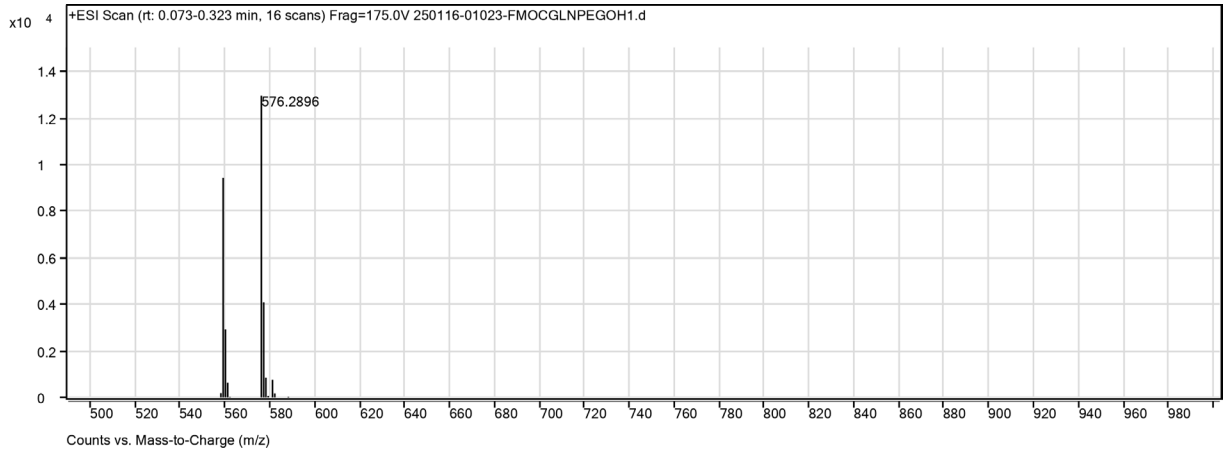
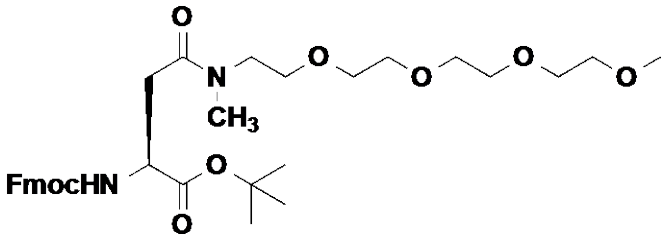


Figure 3-25. ESI-TOF MS data for Fmoc-L-GlnPEG$_4$-OH. Calculated m/z for $C_{29}H_{42}N_3O_9$ ($M+NH_4^+$) is 576.29, found 576.2896

3.4.14 Fmoc-Methyl-L-AsnPEG$_4$-OtBu (S6)



To a solution of (*S*)-3-((((*9H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert*-butoxy)-4-oxobutanoic acid (Fmoc-L-Asp-OtBu, 1.0g, 2.4 mmol) in NMP (15 mL) was added 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU, 1.5g, 3.6mmol), N-hydroxybenzotriazole hydrate (HOBt, 0.55g, 3.6mmol), and *N*,*N*-Diisopropylethylamine (DIEA, 1.27mL, 7.29mmol), and the resulting mixture was stirred for 15 min at room temperature. Then, previously synthesized *N*-methyl-2,5,8,11-tetraoxatridecan-13-amine[27] (0.59g, 2.7mmol) was added, and the resulting mixture was stirred at room temperature for 24 hours. To the reaction mixture was added saturated sodium bicarbonate (100mL) and then saturated Brine (100 mL) and extracted three times with ethyl acetate (100 mL each). The combined ethyl acetate extracts were dried over $MgSO_4$, and concentrated by rotary evaporation to afford an oil. Fmoc-Methyl-L-AsnPEG$_4$-OtBu [*tert-butyl (S)-17-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-14-methyl-15-oxo-2,5,8,11-tetraoxa-14-azaoctadecan-18-oate*] was purified by flash chromatography over silica in ethyl acetate/hexanes. The product was concentrated via rotary evaporation (followed by rewashing with saturated bicarbonate three times, and saturated brine three times, and co-evaporation three times with benzene (200mL) to remove residual NMP) and dried in vacuo to give a thick oily solid (49.5% yield).

^{1}H NMR (500 MHz, $CDCl_3$): δ 7.77(2H, d, J = 7.6 Hz, Fmoc aryl C-**H**); 7.66 (2H, m, Fmoc aryl C-**H**); 7.41 (2H, t, J = 7.4 Hz, Fmoc aryl C-**H**); 7.33 (2H, t, J = 7.4Hz, Fmoc aryl C-**H**); 6.16 (1H, d, amide); 4.56 (1H, m, -CONHCα**H**(COOtBu)-CβH_2-); 4.47 (1H, m, Fmoc Ar_2CH-C**H$_{(a)}$**H$_{(b)}$-

O-); 4.29 (1H, m, Fmoc Ar_2CH-$CH_{(a)}$**H$_{(b)}$**-O-); 4.25 (1 H, m, J = 7.0 Hz, Fmoc Ar2C**H**-CH_2-O-); 3.64-3.48 (16H, m, -CONH-(C**H$_2$**-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-); 3.38 (3H, s, -O-C**H$_3$**) 3.20-3.11 (1H, m, -CαH(COOH)-Cβ(**Ha**)Hb-$CONHCH_3$); 2.87-2.74 (1H, m, -CαH(COOH)-Cβ(Ha)**Hb**-$CONHCH_3$); 3.06 (1.5H, s, -NH-C**H$_3$** Cis or trans amide); 2.97 (1.5H, s, -NH-C**H$_3$** Cis or trans amide**)** 1.47 (9H, s, -O-C-(C**H$_3$**)$_3$. The full ^{1}H NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OtBu is shown in Figure 3-26.

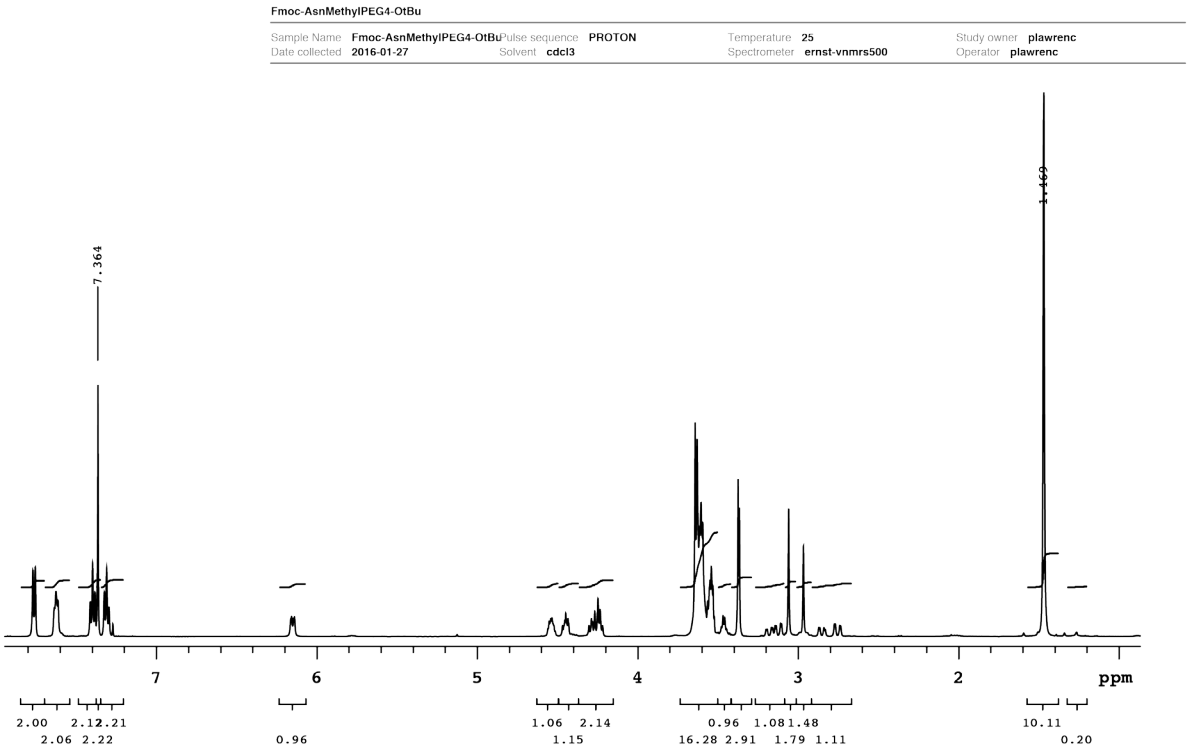


Figure 3-26. ^{1}H NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OtBu.

^{13}C NMR (126 MHz, $CDCl_3$): δ 170.61, 170.40, 170.35, 170.31, (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.37 (Fmoc-O-**C**ONH-); 144.06, 144.04, 143.87, 143.85, 141.28, 141.27, 141.25 (Fmoc aryl **ipso** C's); 128.34, 127.66, 127.07, 127.04, 125.33, 125.30, 125.21,

119.94 (Fmoc aryl **C**-H); 81.83, 81.763 (-O-**C**$(CH_3)_3$); 71.91-68.75 and 49.66, 47.79 (-NH-**C**H_2-**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-) 67.05 (Fmoc Ar_2CH-**C**H_2-O-); 36.91, 33.83 (-NH-**C**H_3); 59.03, (-O-**C**H_3) 51.30, 51.12 (-NH-**C**αH(COOtBu)-CβH_2-); 47.17 (Fmoc Ar_2**C**H-CH_2-O-); 35.94, 35.50 (-CαH(COOH)-**Cβ**H_2-CONHCH_3-); 27.93, 27.91 (-O-C(**C**$H_3)_3$). The full ^{13}C NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OtBu is shown in Figure 3-27.

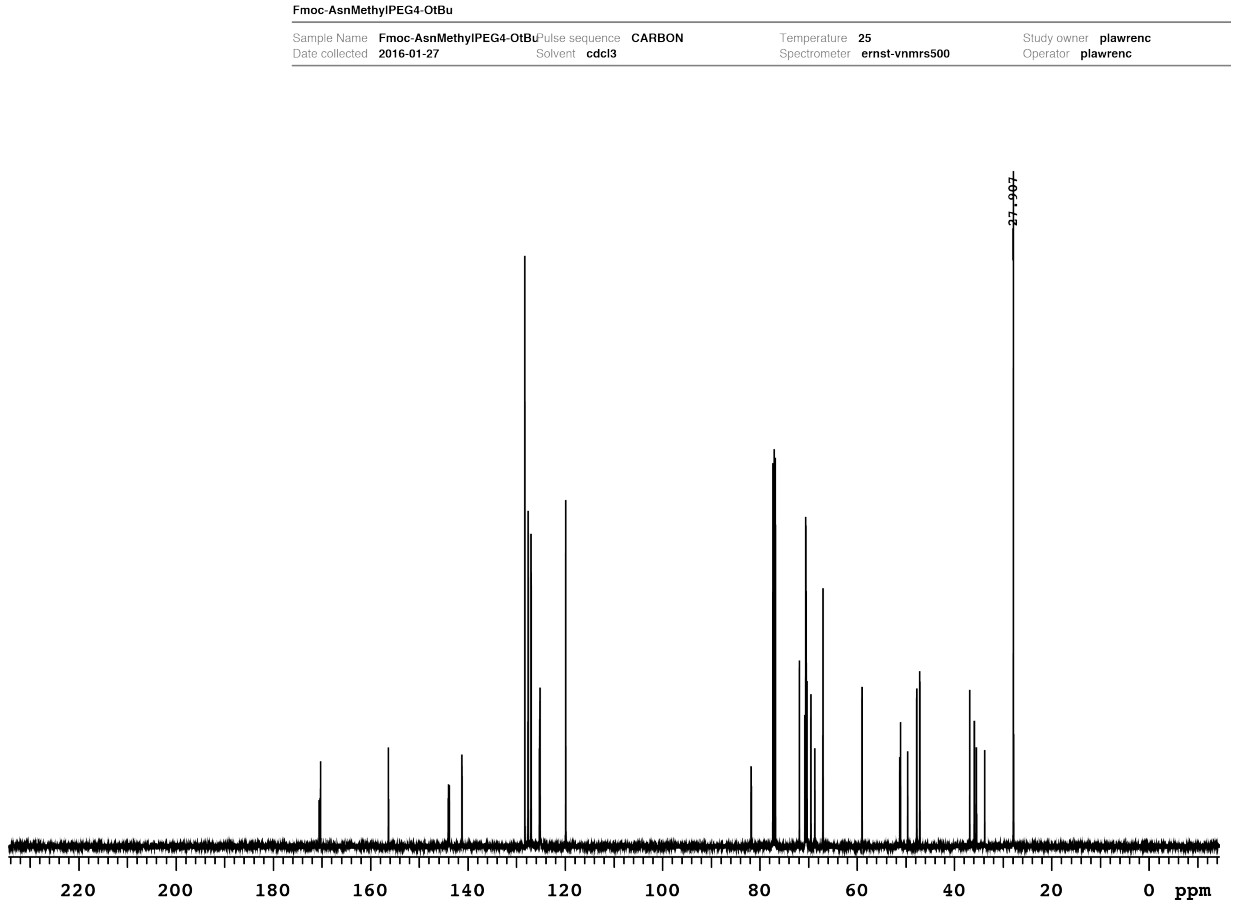


Figure 3-27. ^{13}C NMR spectrum for Fmoc-Methyl-L-AspPEG$_4$-OtBu.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-Methyl-L-AsnPEG$_4$-OtBu were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-28, using the one-bond C-H correlations shown in Table 3-6

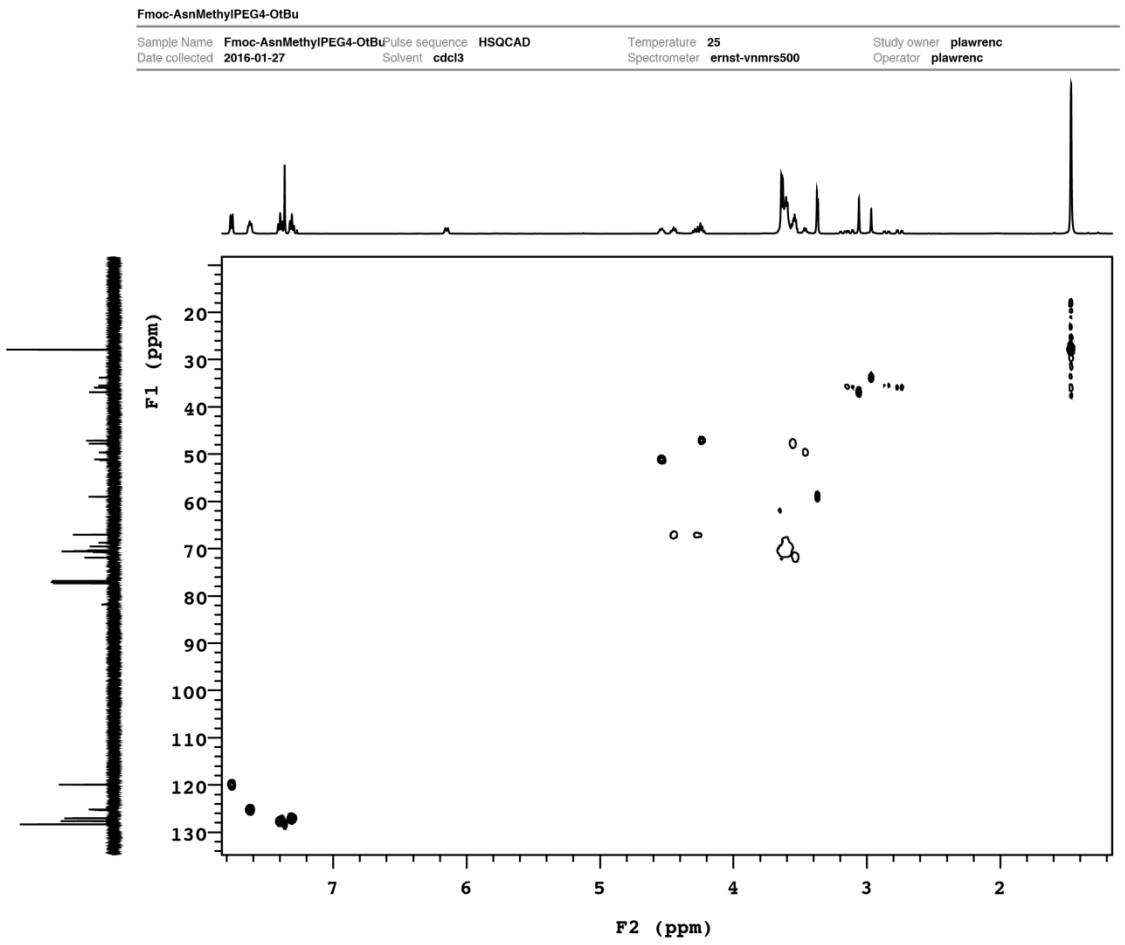


Figure 3-28. 2D HSQC spectrum of Fmoc-Methyl-L-AsnPEG$_4$-OtBu.

Table 3-6. One-Bond C-H correlations identified from HSQC data for Fmoc-Methyl-L-AsnPEG$_4$-OtBu.

^{1}H δ	^{1}C δ	Assignment
7.77	119.035	Fmoc aryl C-H
7.64	125.30, 125.206	Fmoc aryl C-H
7.41	128.34, 127.66, 127.07	Fmoc aryl C-H
7.33	127.04, 125.33	Fmoc aryl C-H
4.56	51.298, 51.131	-CONH**CαH**(COOH)-CβH2-
4.25	47.17	Fmoc Ar**CH**-CH2-O-
4.469, 4.29	67.05	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
3.06, 2.97	36.91, 33.83	-N-**CH3**
3.20-3.11, 2.87-2.74'	35.94, 35.50	-CαH(COOH)-**CβH$_2$**-
3.642-3.479	71.911-68.746 and 49.658 and 47.791	-CONH-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-,
3.38	59.03	-O-**CH3**
1.47	27.929, 27.907	-C(**CH$_3$**)$_3$

High-resolution ESI-TOF MS data for Fmoc-Methyl-L-AsnPEG$_4$-OtBu are shown in Figure 3-29.

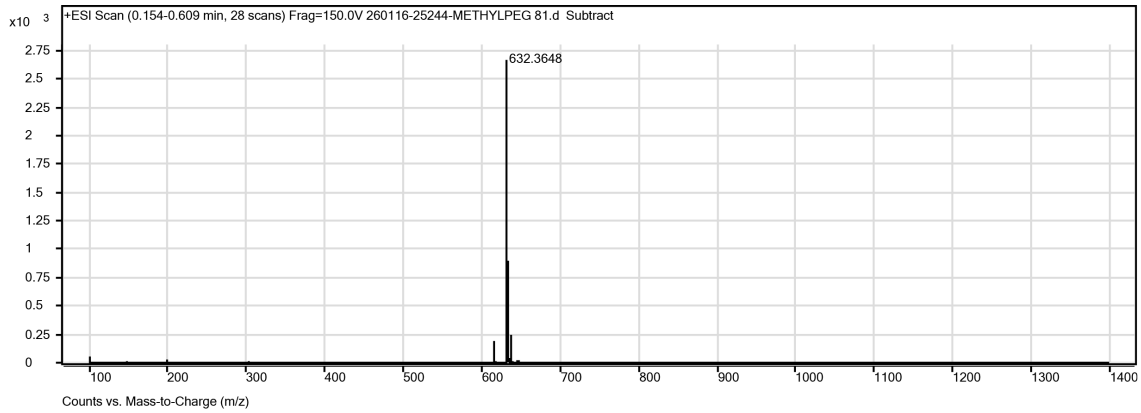
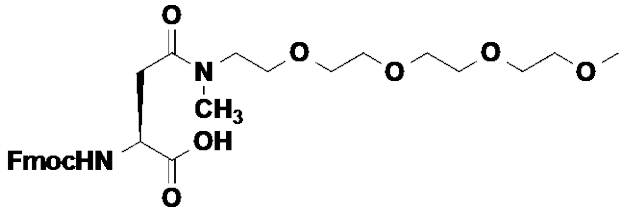


Figure 3-29. ESI-TOF MS data for Fmoc-Methyl-L-AspPEG$_4$-OtBu. Calculated m/z for $C_{33}H_{50}N_3O_9$ ($M+NH_4^+$) is 632.35, found 632.3648

3.4.15 Fmoc-Methyl-L-AsnPEG$_4$-OH (S7)



To a solution of TFA (95% in water, 50ml) was added 0.74 g Fmoc-Methyl-L-AsnPEG$_4$-OtBu (**S6**), and the solution was stirred for 4 h under an argon atmosphere. The product Fmoc-Methyl-L-AsnPEG$_4$-OH [*(S)-17-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-14-methyl-15-oxo-2,5,8,11-tetraoxa-14-azaoctadecan-18-oic acid*, **S7**] was concentrated by rotary evaporation, and used without further purification.

^{1}H NMR (500 MHz, $CDCl_3$): δ 10.39 (1H, broad s, (COO**H**); 7.77 (2H, d, J = 7.5 Hz, Fmoc aryl C-**H**); 7.64 (2H, m, Fmoc aryl C-**H**); 7.42 (2H, t, J = 7.2 Hz, Fmoc aryl C-**H**); 7.33 (2H, m, Fmoc aryl C-**H**); 6.57 (1H, d, amide); 4.71 (1H, broad s,-CONHCα**H**(COOtBu)-CβH_2-); 4.48 (1H, m, Fmoc Ar_2CH-C**H$_{(a)}$**$H_{(b)}$-O-); 4.38 (1H, m, Fmoc Ar_2CH-$CH_{(a)}$**H$_{(b)}$**-O-); 4.24 (1 H, dd, Fmoc Ar2C**H**-CH_2-O-); 3.65-3.63 (16H, m, -CONH-(C**H$_2$**-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-); 3.42 (3H, s, -O-C**H$_3$**) 3.25-2.873 (2H, m, -CαH(COOH)-Cβ(**Ha**)**Hb**-CONHCH$_3$); 3.09 (1.5H, s, -NH-C**H$_3$** Cis or trans amide); 2.99 (1.5H, s, -NH-C**H$_3$** Cis or trans amide**)**. The full ^{1}H NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OH is shown in Figure 3-30.

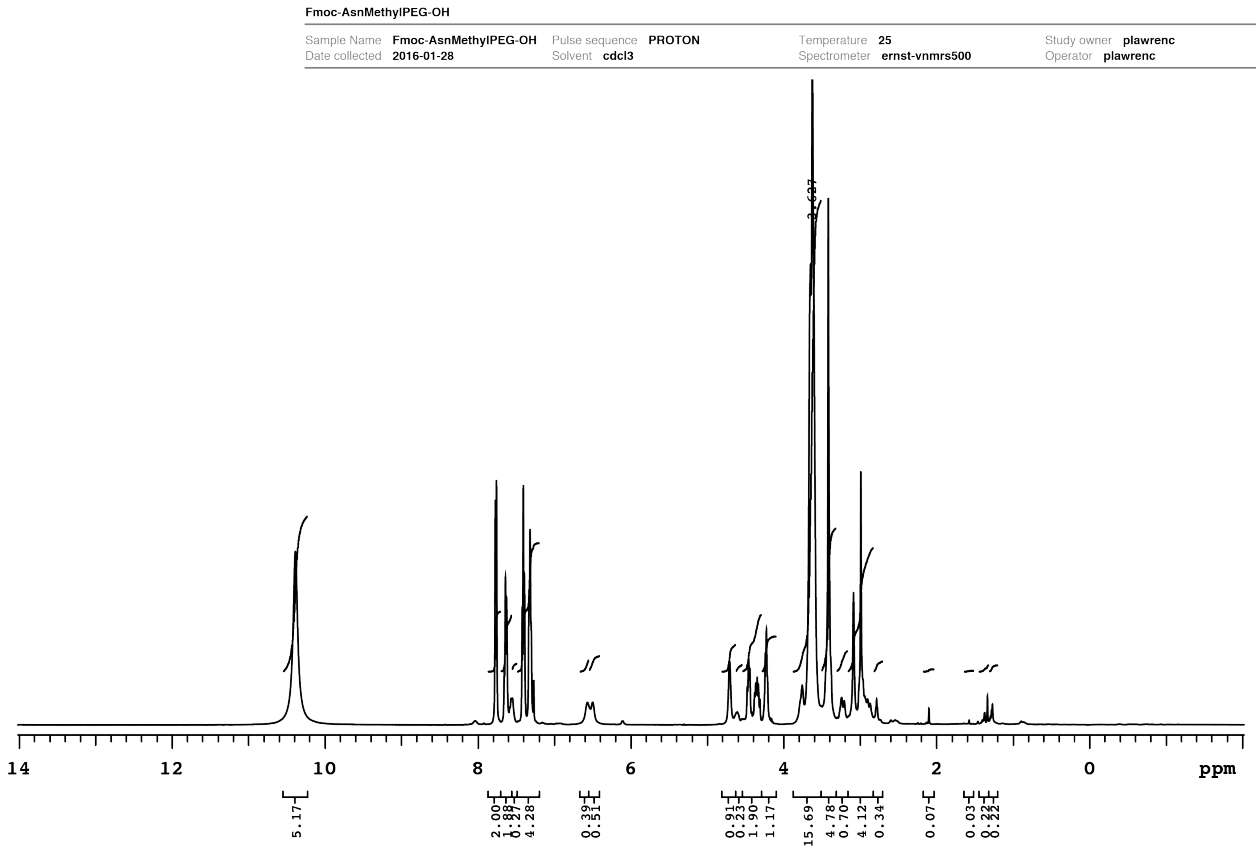


Figure 3-30. ^{1}H NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OH.

^{13}C NMR (126 MHz, $CDCl_3$): δ 174.55, 174.22, 172.38, 171.87 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.99, 156.84 (Fmoc-O-**C**ONH-); 143.68, 143.62, 143.47, 143.45, 141.27 (Fmoc aryl **ipso** C's); 127.82, 127.81, 127.15, 127.12, 125.270, 125.19, 125.17, 125.14, 120.00 (Fmoc aryl **C**-H); 71.50-69.60 and 67.76, 67.68 (-NH-**C**H_2-**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-) 67.80 (Fmoc Ar_2CH-**C**H_2-O-); 36.84, 34.17 (-NH-**C**H_3); 58.63, (-O-**C**H_3) 50.46, 50.27 (-NH-**C**αH(COOtBu)-CβH_2-); 48.05 (Fmoc Ar_2**C**H-CH_2-O-); 35.85, 35.97 (-CαH(COOH)-**Cβ**H_2-CONHCH_3-). The full ^{13}C NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OH is shown in Figure 3-31.

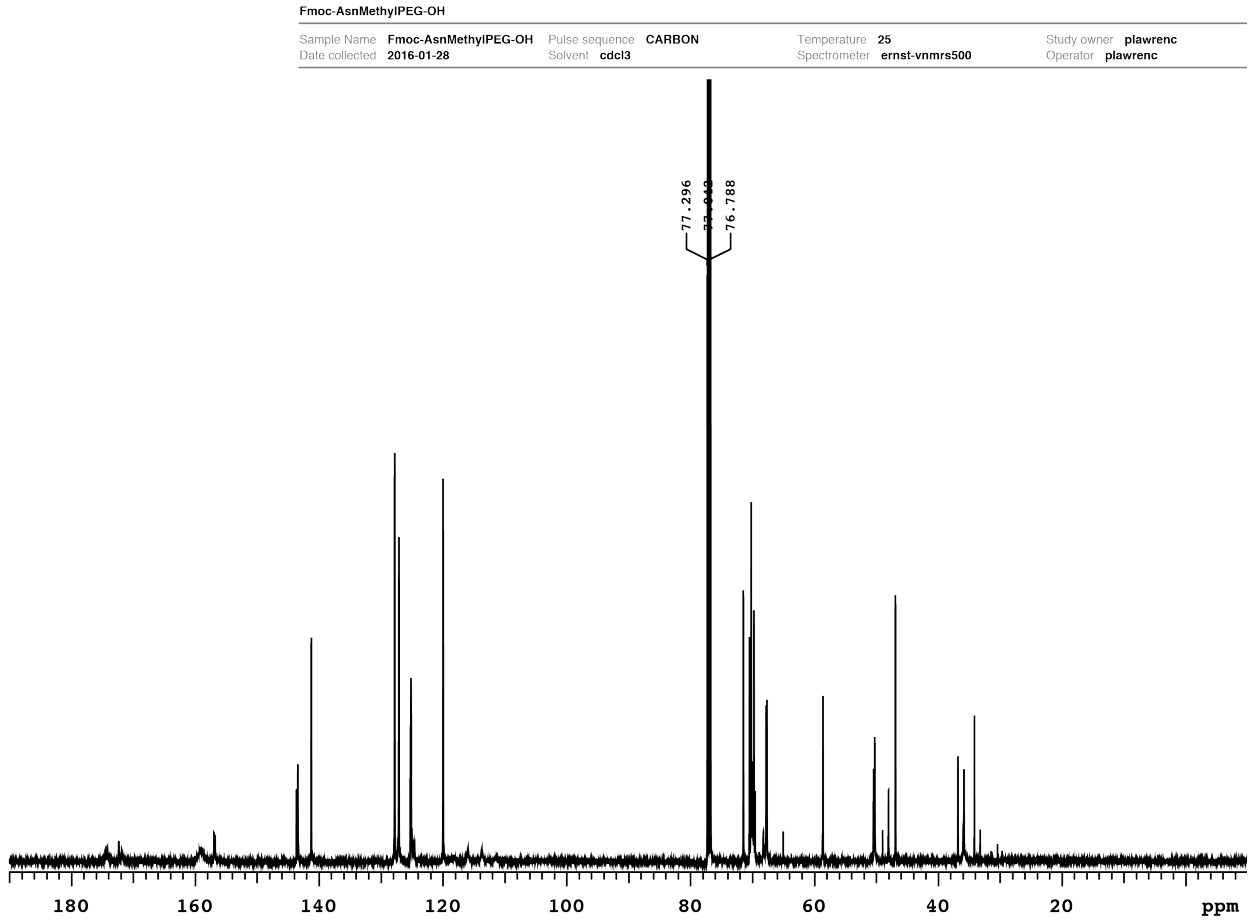


Figure 3-31. ^{13}C NMR spectrum for Fmoc-Methyl-L-AsnPEG$_4$-OH.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-Methyl-L-AsnPEG$_4$-OH were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-32, using the one-bond C-H correlations shown in Table 3-7.

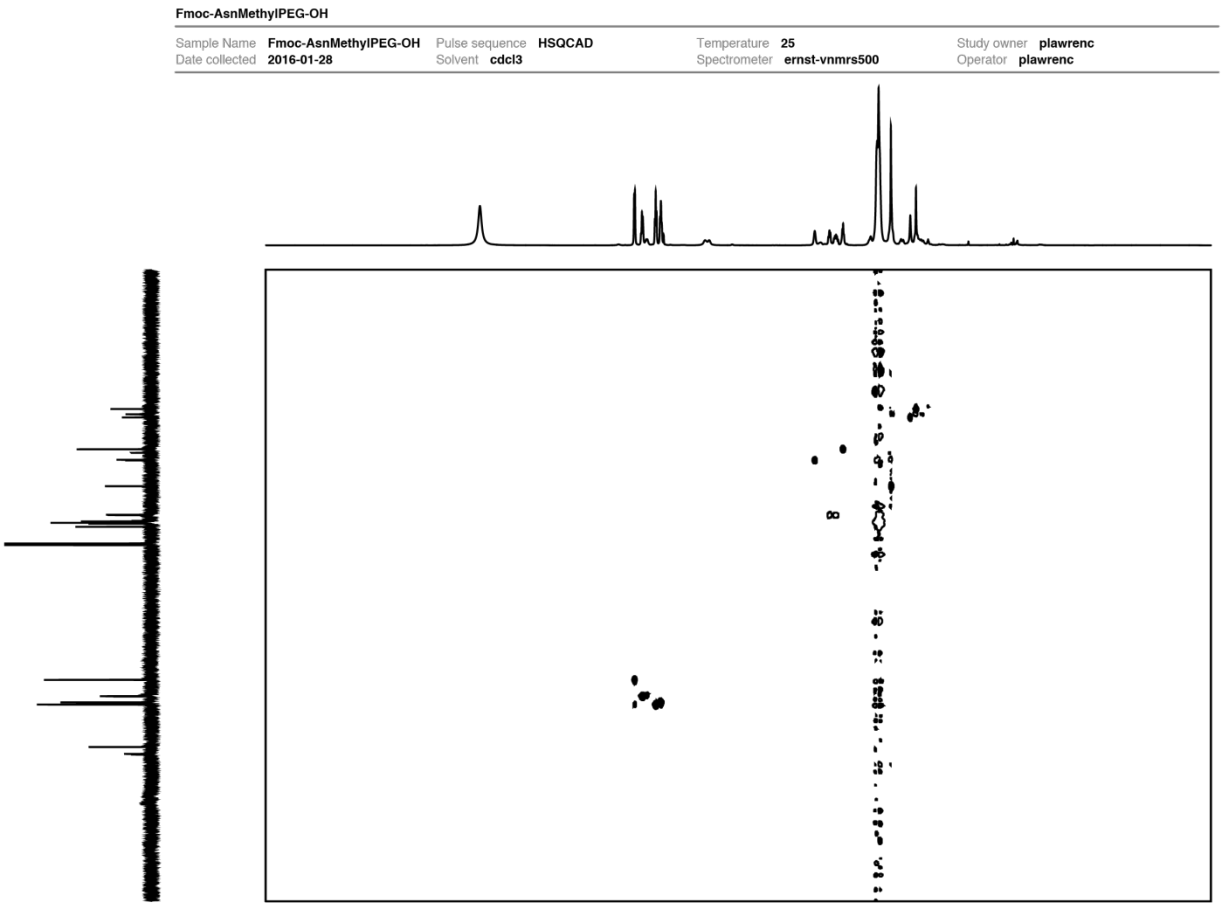


Figure 3-32. 2D HSQC spectrum of Fmoc-Methyl-L-AsnPEG$_4$-OH.

Table 3-7. One-Bond C-H correlations identified from HSQC data for Fmoc-Methyl-L-AsnPEG$_4$-OH.

^{1}H δ	^{1}C δ	Assignment
7.77	119.999	Fmoc aryl C-H
7.64	125.270, 125.190	Fmoc aryl C-H
7.42	127.82, 127.81	Fmoc aryl C-H
7.33	127.17, 127.12	Fmoc aryl C-H
4.71	50.463, 50.2.66	-CONH**CαH**(COOH)-CβH2-
4.24	48.053	Fmoc Ar**CH**-CH2-O-
4.479, 4.38	67.798	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
3.09, 2.99	36.84, 34.172	-N-**CH3**
3.245-2.873	35.853, 35.974	-CαH(COOH)-**CβH$_2$**-
3.652-3.627	71.490-69.596, 67.760, 67.680	-CONH-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-,
3.42	58.629	-O-**CH3**

High-resolution ESI-TOF MS data for Fmoc-Methyl-L-AsnPEG$_4$-OH are shown in Figure 3-33.

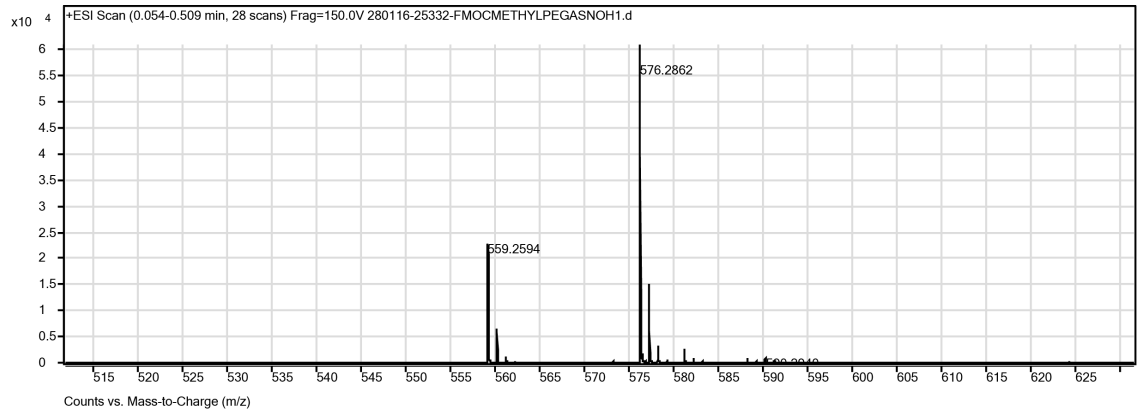
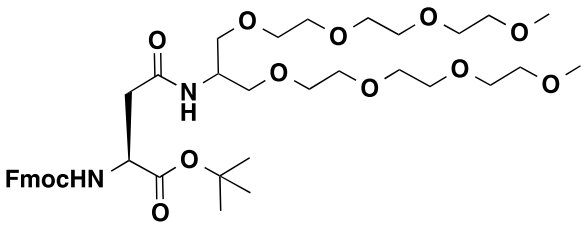


Figure 3-33. ESI-TOF MS data for Fmoc-Methyl-L-AsnPEG$_4$-OH. Calculated m/z for $C_{29}H_{42}N_3O_9$ ($M+NH_4^+$) is 576.29, found 576.2862

3.4.16 Fmoc-L-Asn(PEG$_4$)$_2$-OtBu (S8)



To a solution of (*S*)-3-((((*9H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert*-butoxy)-4-oxobutanoic acid (Fmoc-L-Asp-OtBu, 0.62g, 1.5 mmol) in NMP (15 mL) was added 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU, 0.93g, 2.26 mmol), N-hydroxybenzotriazole hydrate (HOBt, 0.35g, 2.26 mmol), and *N*,*N*-Diisopropylethylamine (DIEA, 0.78 mL, 4.5 mmol), and the resulting mixture was stirred for 15 min at room temperature. Then, *2,5,8,11,15,18,21,24-octaoxapentacosan-13-amine*[28] (0.63g, 1.65mmol) was added, and the resulting mixture was stirred at room temperature for 24 hours. To the reaction mixture was first added saturated sodium bicarbonate (100mL) and then saturated Brine (100 mL) was added to the reaction/bicarbonate mixture and extracted three times with ethyl acetate (100 mL). The combined ethyl acetate extracts were dried over $MgSO_4$, and concentrated by rotary evaporation to afford an oil. Fmoc-L-Asn(PEG$_4$)$_2$-OtBu [*tert-butyl (S)-17-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-13-(2,5,8,11-tetraoxadodecyl)-15-oxo-2,5,8,11-tetraoxa-14-azaoctadecan-18-oate* was purified by flash chromatography over silica in ethyl acetate/hexanes. The product was concentrated via rotary evaporation (followed by rewashing with saturated bicarbonate three times, and saturated brine three times, and co-evaporation three times with benzene (200mL) to remove residual NMP) and dried in vacuo to give a thick oily solid.

^{1}H NMR (300 MHz, $CDCl_3$): δ 7.73(2H, d, J = 7.4 Hz, Fmoc aryl C-**H**); 7.60 (2H, m, Fmoc aryl C-**H**); 7.38 (2H, t, Fmoc aryl C-**H**); 7.29 (2H, t, Fmoc aryl C-**H**); 6.68 (1H, d, amide); 6.26 (1H,

d, amide); 4.49-4.43 (1H, m, -CONHCα**H**(COOtBu)-CβH_2-); 4.40-4.36 (1H, m, Fmoc Ar_2CH-C**H**$_{(a)}$$H_{(b)}$-O-); 4.30-4.24 (1H, m, Fmoc Ar_2CH-$CH_{(a)}$**H**$_{(b)}$-O-); 4.21 (2 H, m, Fmoc Ar2C**H**-CH_2-O- and -CONH-C**H**-(CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-)$_2$; 3.59-3.48 (32H, m, -CONH-CH-(C**H**$_2$-C**H**$_2$-O-C**H**$_2$-C**H**$_2$-O-C**H**$_2$-C**H**$_2$-O-C**H**$_2$-C**H**$_2$-O-)$_2$; 3.32 (6H, s, -O-C**H**$_3$) 2.88-2.80 (1H, dd, -CαH(COOH)-Cβ(**Ha**)Hb-$CONHCH_3$); 2.70-2.64 (1H, dd, -CαH(COOH)-Cβ(Ha)**Hb**-$CONHCH_3$); 1.44 (9H, s, -O-C-(C**H**$_3$)$_3$. The full ^{1}H NMR spectrum for c is shown in Figure 3-34.

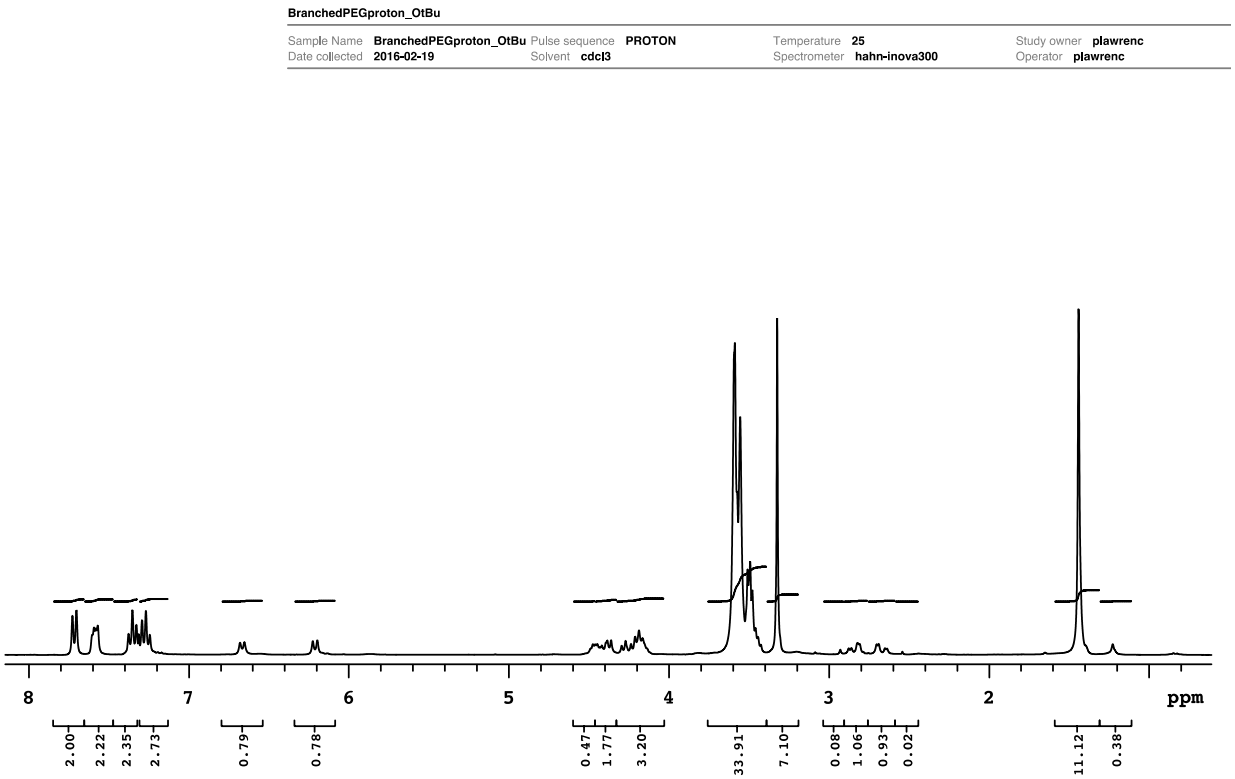


Figure 3-34. ^{1}H NMR spectrum for Fmoc-L-Asn(PEG_4)$_2$-OtBu.

^{13}C NMR (126 MHz, $CDCl_3$): δ 170.22, 169.77 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.99, 156.84 (Fmoc-O-**C**ONH-); 143.98, 143.83, 141.208 (Fmoc aryl **ipso** C's);

127.66, 127.07, 125.25, 125.17, 119.90 (Fmoc aryl **C**-H); 71.85-69.27 (-NH-CH-(**C**H_2-**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-)$_2$ 66.96 (Fmoc Ar_2CH-**C**H_2-O-); 58.93, (-O-**C**H_3); 51.45 (-NH-**C**αH(COOtBu)-CβH_2-); 47.11, 48.67 (Fmoc Ar_2**C**H-CH_2-O- and -CONH-**C**H-(CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-)$_2$); 37.65 (-CαH(COOtBu)-**Cβ**H_2-CONHCH-); 27.92 (-O-C(**C**H_3)$_3$). The full ^{13}C NMR spectrum for Fmoc-L-Asn(PEG_4)$_2$-OtBu is shown in Figure 3-35.

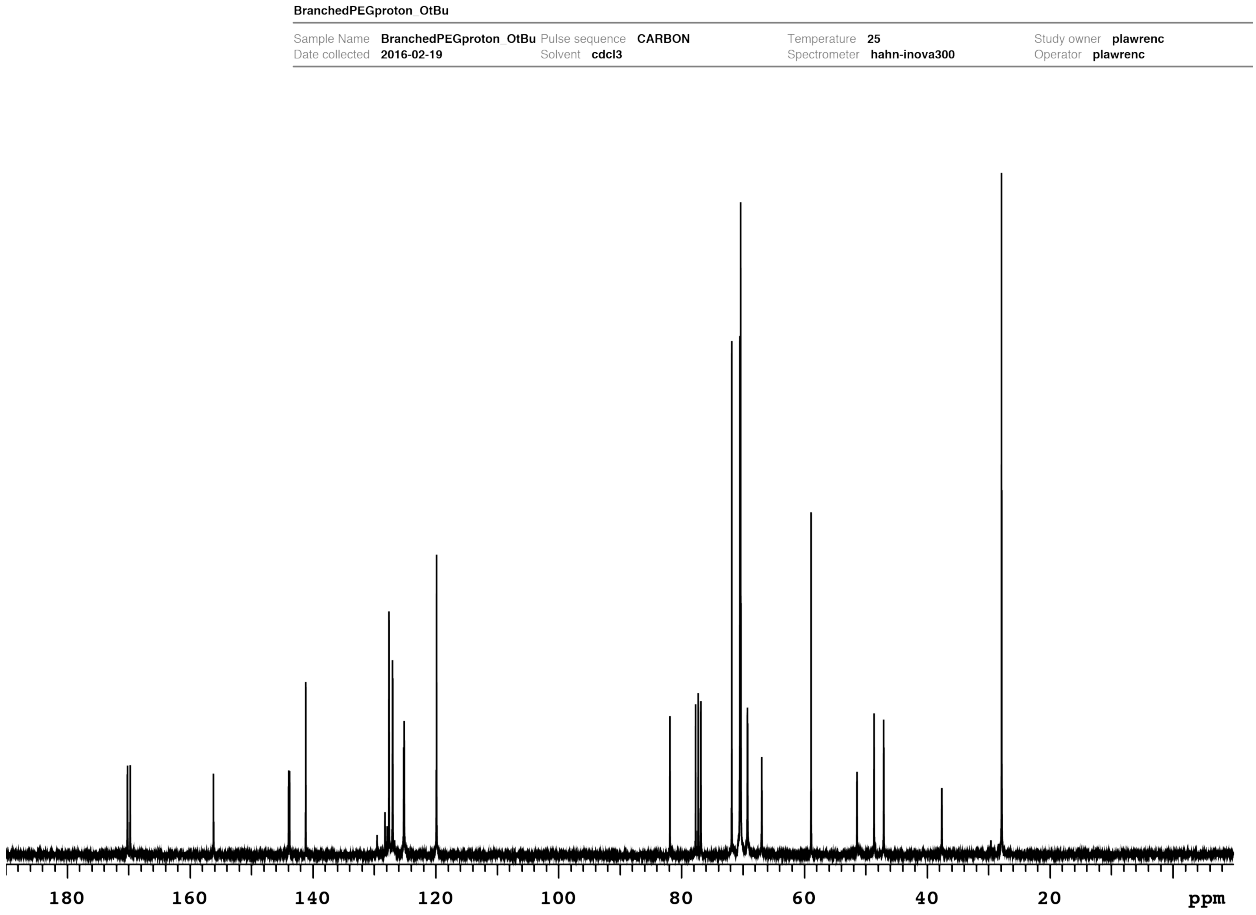


Figure 3-35. ^{13}C NMR spectrum for Fmoc-L-Asn(PEG_4)$_2$-OtBu

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-L-Asn(PEG_4)$_2$-OtBu were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-36, using the one-bond C-H correlations shown in Table 3-8

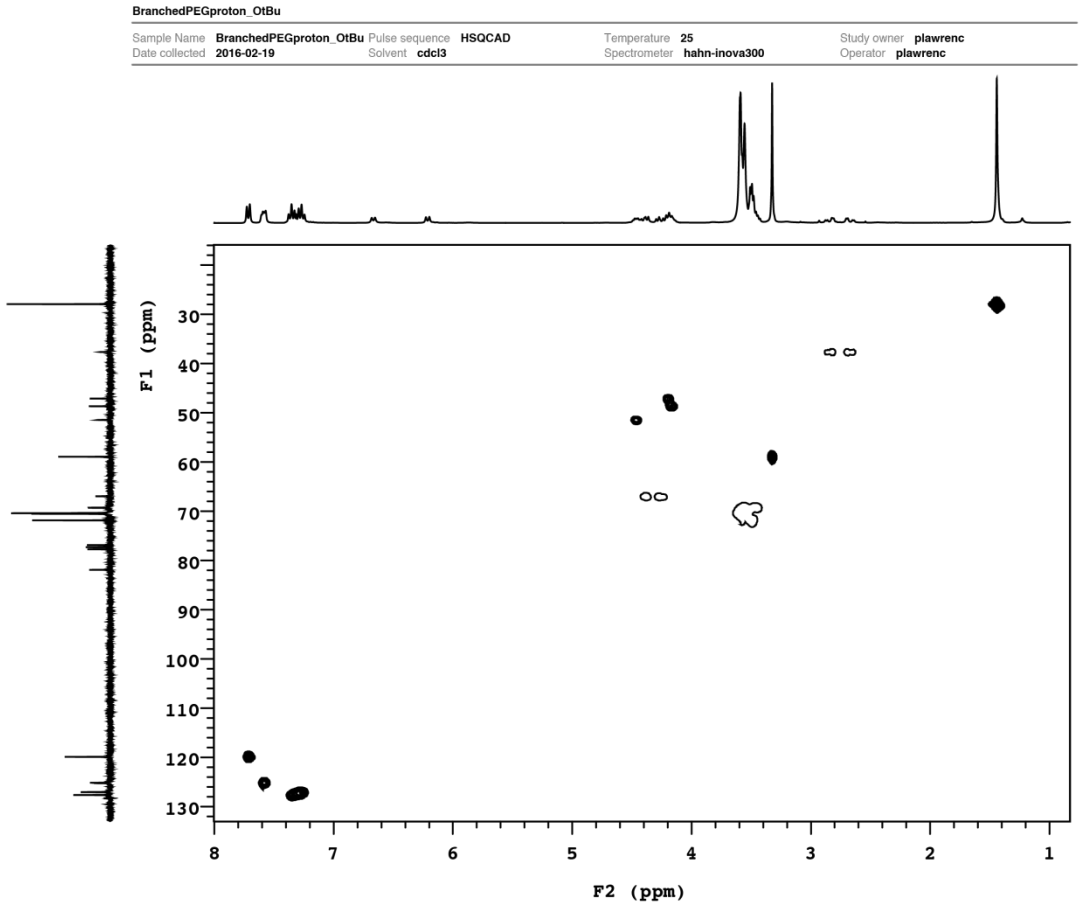


Figure 3-36. 2D HSQC spectrum **of** Fmoc-L-Asn(PEG_4)$_2$-OtBu.

Table 3-8. One-Bond C-H correlations identified from HSQC data for Fmoc-L-Asn(PEG_4)$_2$-OtBu.

^{1}H δ	^{1}C δ	Assignment
7.727	119.904	Fmoc aryl C-H
7.604	125.253, 125.17	Fmoc aryl C-H
7.377	127.657	Fmoc aryl C-H
7.292	127.069	Fmoc aryl C-H
4.490-4.432	51.453	-CONH**CαH**(COOH)-CβH2-
4.395-4.362, 4.296-4.238	66.958	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
4.212	47.108, 48.67	Fmoc Ar**CH**-CH2-O- and -CON-C**H**-PEG
3.589-3.480	71.849-69.277	-CONH-CH-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**,
3.325	58.931	-O-**CH$_3$**
2.878-2.809, 2.703-2.638	37.65	-CαH(COOH)-**CβH$_2$**-
1.442	27.921	-C(**CH$_3$**)$_3$

High-resolution ESI-TOF MS data for Fmoc-L-Asn(PEG_4)$_2$-OtBu are shown in Figure 3-37

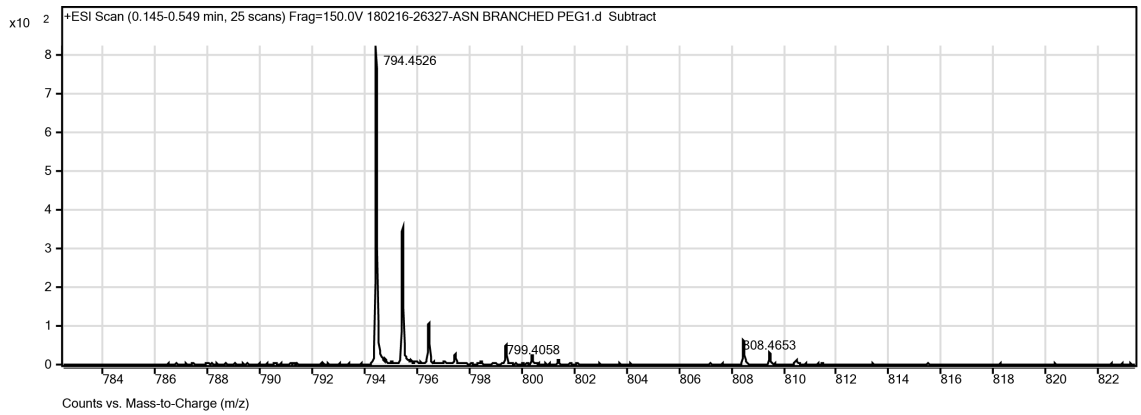
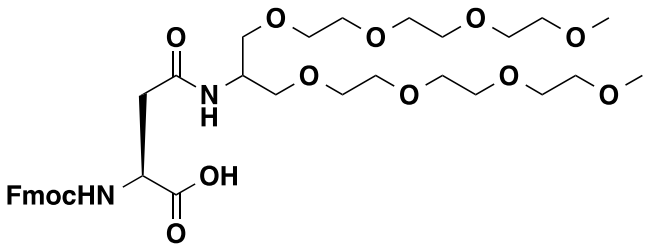


Figure 3-37. ESI-TOF MS data for Fmoc-L-Asn(PEG_4)$_2$-OtBu. Calculated m/z for $C_{40}H_{64}N_3O_{13}$ ($M+NH_4^+$) is 794.44, found 794.4526.

3.4.17 Fmoc-L-Asn(cPEG$_4$)$_2$-OH (S9)



To a solution of TFA (95% in water, 50ml) was added Fmoc-L-Asn(cPEG$_4$)$_2$-OtBu (**S8**), and the solution was stirred for 4 h under an argon atmosphere. The product Fmoc-L-Asn(cPEG$_4$)$_2$-OH [(*S*)-17-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-13-(2,5,8,11-tetraoxadodecyl)-15-oxo-2,5,8,11-tetraoxa-14-azaoctadecan-18-oic acid, **S9**] was concentrated by rotary evaporation, and used without further purification.

^{1}H NMR (500 MHz, $CDCl_3$): δ 10.91 (1H, s, (COO**H**); 7.76(2H, d, J = 7.5 Hz, Fmoc aryl C-**H**); 7.62 (2H, t, J = 6.76 Fmoc aryl C-**H**); 7.40 (2H, t, J = 7.4 Hz, Fmoc aryl C-**H**); 7.32 (2H, t, J = 7.3Hz, Fmoc aryl C-**H**); 6.34 (1H, d, amide); 6.99 (1H, d, amide); 4.63 (1H, m, -CONHCα**H**(COOtBu)-CβH_2-); 4.47 (2H, m, Fmoc Ar$_2$CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.25-4.20 (2H, m, Fmoc Ar$_2$C**H**-CH$_{(a)}$H$_{(b)}$-O- and -CONH-C**H**-(CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-)$_2$); 3.63-3.53 (32H, m, -CONH-CH-(C**H$_2$**-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-)$_2$); 3.37 and 3.36 (3H, s, -O-C**H$_3$**); 2.95-2.76 (2H, m, -CαH(COOH)-Cβ(**Ha**)**Hb**-CON-); . The full ^{1}H NMR spectrum for Fmoc-L-Asn(PEG$_4$)$_2$-OH is shown in Figure 3-38.

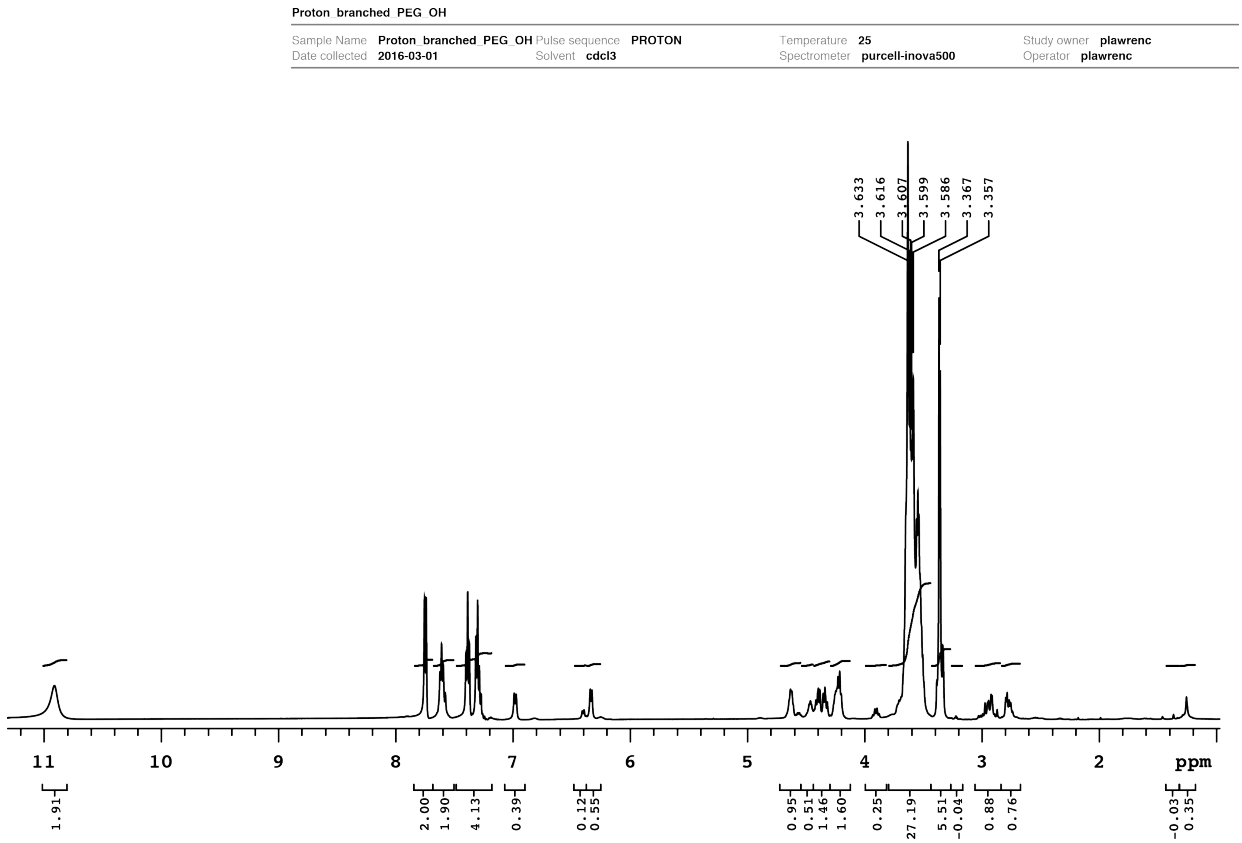


Figure 3-38. ^{1}H NMR spectrum for Fmoc-L-Asn$(PEG_4)_2$-OH

^{13}C NMR (126 MHz, $CDCl_3$): δ 172.35, 171.32 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 159.14, 158.81, 156.22 (Fmoc-O-**C**ONH-); 143.83, 143.71, 141.32, 141.25 (Fmoc aryl **ipso** C's); 127.72, 127.118, 125.22, 125.17, 119.94 (Fmoc aryl **C**-H); 71.75-69.55 (-NH-CH-(**C**H_2-**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-)$_2$ 67.29 (Fmoc Ar_2CH-**C**H_2-O-); 58.84, 58.82 (-O-**C**H_3); 50.99 (-NH-**C**αH(COOtBu)-CβH_2-); 47.11, 49.17 (Fmoc Ar_2**C**H-CH_2-O- and -CONH-**C**H-(CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-CH_2-CH_2-O-)$_2$); 38.04 (-CαH(COOtBu)-**Cβ**H_2-CONHCH-). The full ^{13}C NMR spectrum for Fmoc-L-Asn$(PEG_4)_2$-OH is shown in Figure 3-39.

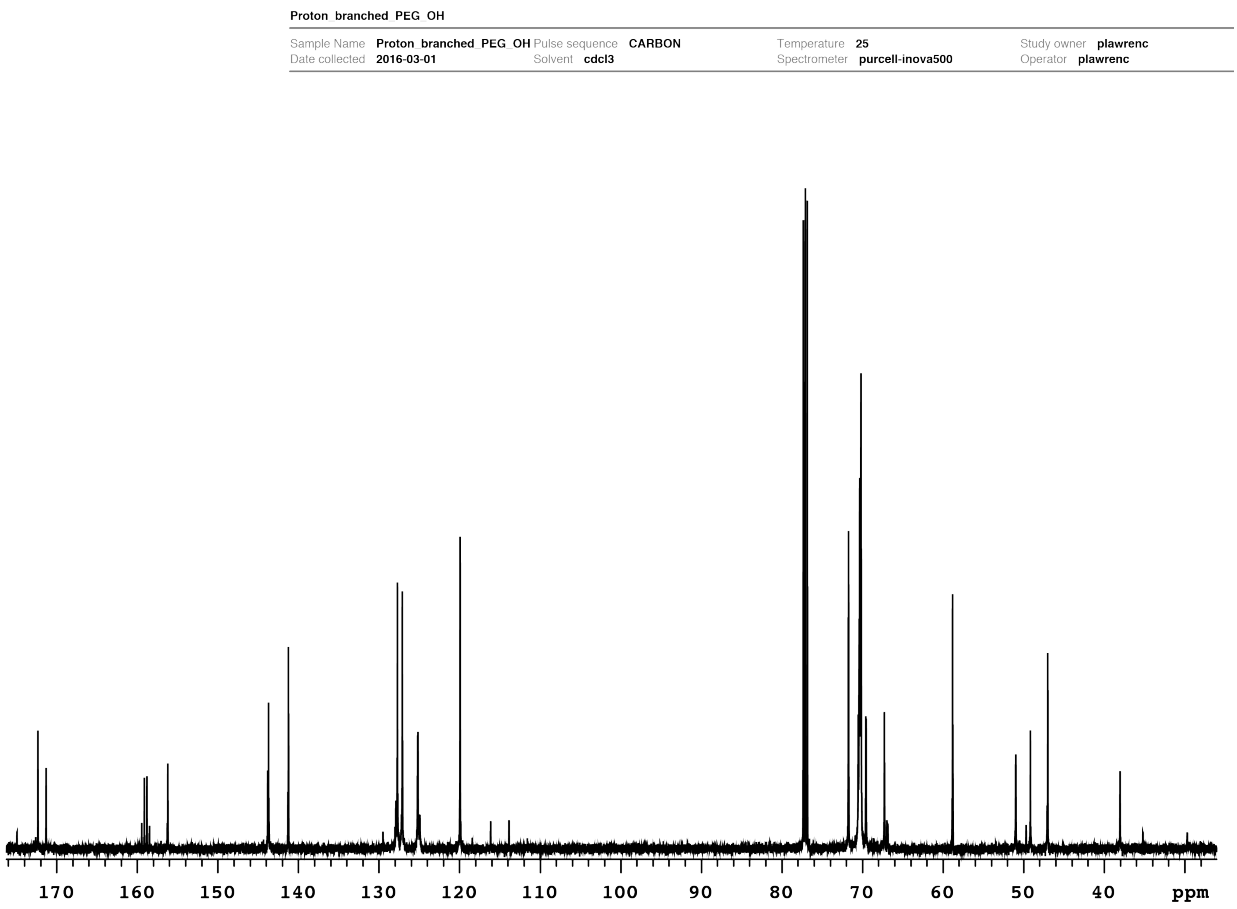


Figure 3-39. ^{13}C NMR spectrum for Fmoc-L-Asn$(PEG_4)_2$-OH.

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-L-Asn$(PEG_4)_2$-OH were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-40, using the one-bond C-H correlations shown in Table 3-9.

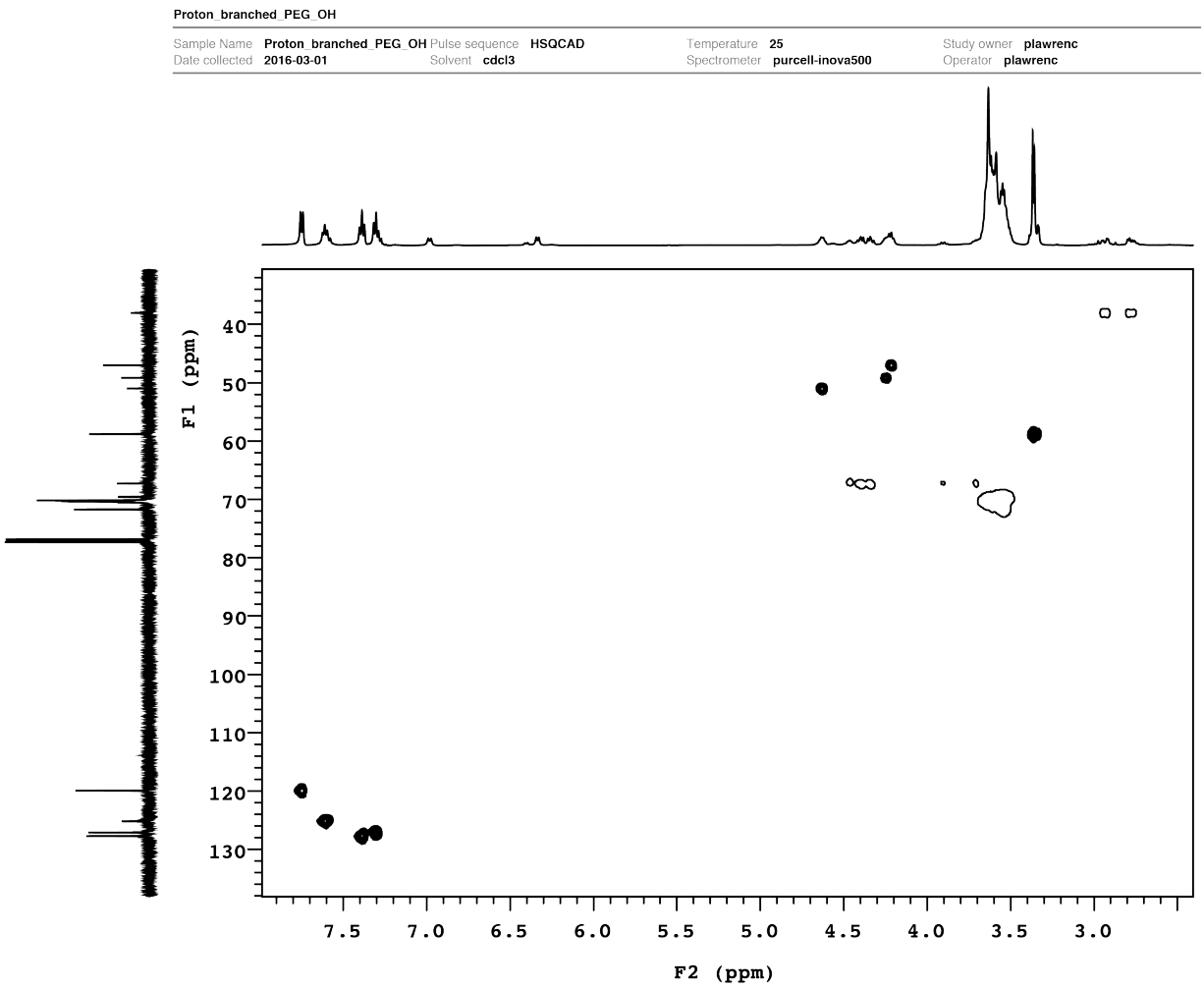


Figure 3-40. 2D HSQC spectrum of Fmoc-L-Asn(PEG_4)$_2$-OH.

Table 3-9. One-Bond C-H correlations identified from HSQC data for Fmoc-L-Asn(PEG_4)$_2$-OH.

^{1}H δ	^{1}C δ	Assignment
7.76	119.94	Fmoc aryl C-H
7.62	125.22, 125.17	Fmoc aryl C-H
7.40	127.72	Fmoc aryl C-H
7.32	127.12	Fmoc aryl C-H
4.63	50.99	CONH**CαH**(CO OH)-CβH2-
4.47-4.32	67.29	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
4.25-4.20	47.01, 49.17	Fmoc Ar**CH**-CH2-O- and -CONH-C**H**-PEG
3.63-3.53	71.75-69.55	-CONH-CH-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**,
3.37, 3.36	58.84, 58.82	-O-**CH$_3$**
4.47-4.32	67.29	-CαH(COOH)-**CβH$_2$**-

High-resolution ESI-TOF MS data for Fmoc-L-Asn(PEG_4)$_2$-OH are shown in Figure 3-41.

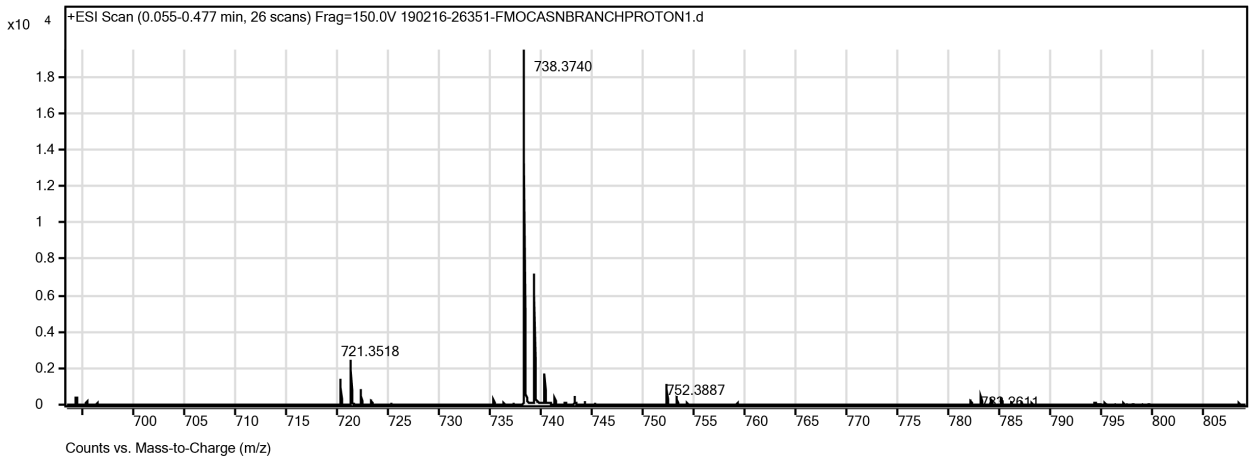
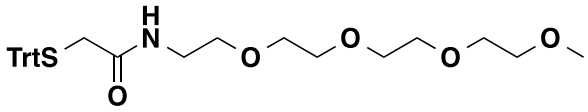


Figure 3-41. ESI-TOF MS data for Fmoc-L-Asn(PEG_4)$_2$-OH. Calculated m/z for $C_{36}H_{56}N_3O_{13}$ ($M+NH_4^+$) is 738.38, found 738.3740

3.4.18 2-(tritylthio)acetamidePEG4 (S10)



To a solution of known 2-(Tritylmercapto)acetic acid[29] (1.70g, 5.09 mmol) in NMP (30mL) was added 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU, 3.16g, 7.64mmol), N-hydroxybenzotriazole hydrate (HOBt, 1.16g, 7.64mmol), and *N*,*N*-diisopropylethylamine (DIEA, 2.21mL, 12.72mmol), and the resulting mixture was stirred for 15 min at room temperature. Then, 2,5,8,11-tetraoxatridecan-13-amine (1.58g, 7.64mmol) was added and the reaction was stirred at room temperature for 24 hours. The reaction mixture was then washed with saturated sodium bicarbonate (100 mL) and extracted three times with ethyl acetate (100 mL). The combined ethyl acetate extracts were washed three times with brine (100mL), dried over $MgSO_4$, and concentrated by rotary evaporation to afford an oil. The desired product was purified by flash chromatography over silica gel using ethyl acetate/hexanes. The desired product **(S10)** was concentrated via rotary evaporation and dried in vacuo.

^{1}H NMR (500 MHz, CDCl3): δ 7.37 (6H, d, J = 7.7 Hz, Trt aryl C-**H**), 7.23 (6H, t, J = 7.5 Hz, Trt aryl C-**H**); 7.16 (3H, t, J = 7.1 Hz, Trt aryl C-**H**); 6.48 (1H, broad s, -S-CH_2CON**H**-CH_2-); 3.54-3.45 (12H, m, -CONH-CH_2-CH_2O-C**H₂**-C**H₂**-O-C**H₂**-C**H₂**-O--C**H₂**-C**H₂-O-**); 3.35 (2H, t, J = 5.10, -CONH-CH_2-C**H₂**O-); 3.28 (3H, s, -O-C**H₃**); 3.14 (2H, m, -CONH-C**H₂**-CH_2O-); 3.01 (2H, s, -S-C**H₂**CONH-CH_2-). The full 1H NMR spectrum for 2-(tritylthio)acetamidePEG4 appears in Figure 3-42.

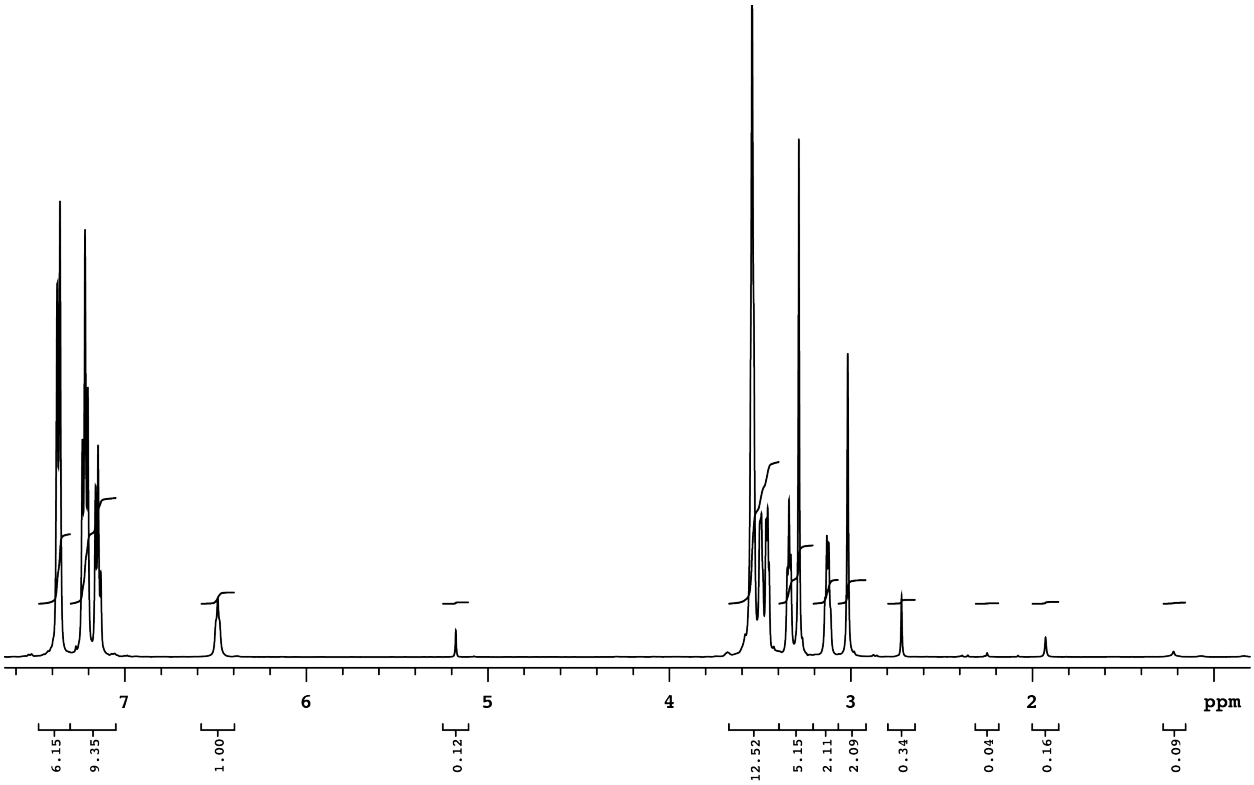


Figure 3-42. ^{1}H NMR spectrum of 2-(tritylthio)acetamidePEG4

^{13}C NMR (126 MHz, $CDCl_3$): δ 168.06 (-S-CH_2**C**ONH-CH_2-); 144.03 (Trt aryl **ipso C**); 128.10, 128.02, 126.97 (Trt aryl **C**); 71.83, 70.53, 70.48, 70.46, 70.41 70.19 (-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-); 69.35 (-CONH-CH_2-**C**H_2O-); 58.95 (-O-**C**H_3); 39.40 (-CONH-**C**H_2-CH_2O-); 36.06 (-S-**C**H_2CONH-CH_2-). The full 13C NMR spectrum for 2-(tritylthio)acetamidePEG4 appears in Figure 3-43.

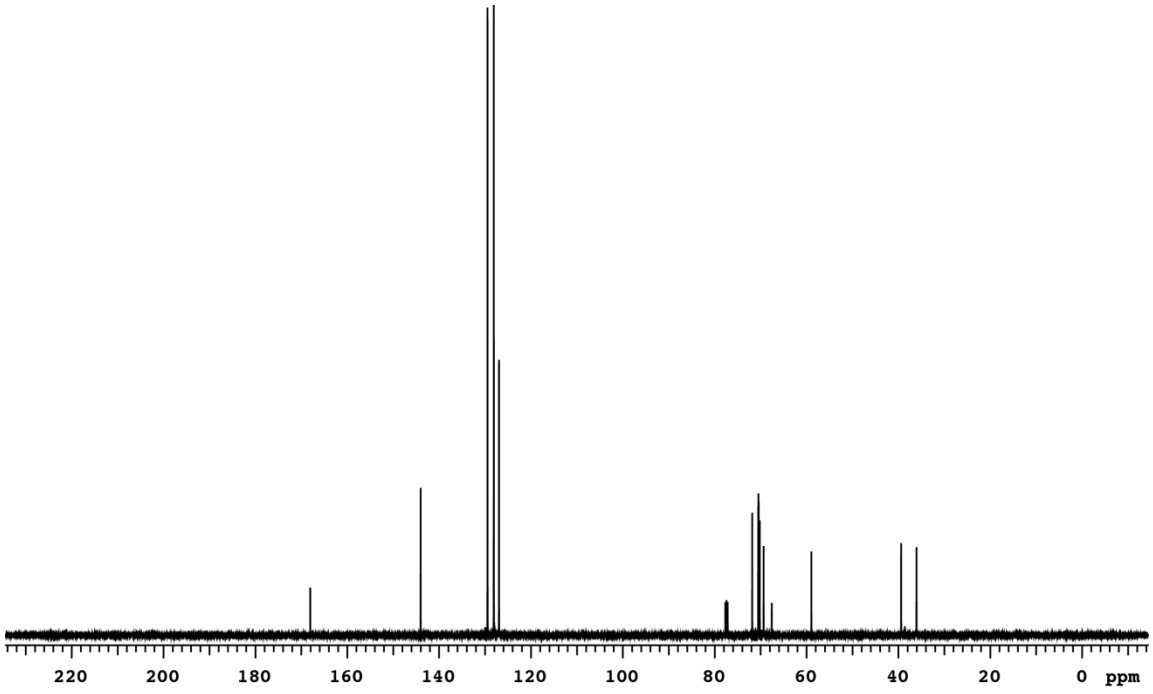


Figure 3-43. ^{13}C NMR spectrum of 2-(tritylthio)acetamidePEG4.

Assignments of the 1H and 13C NMR spectra for the 2-(tritylthio)acetamidePEG4 were made with the assistance of a 2D HSQC experiment Figure 3-44, using the one-bond C-H correlations shown in Table 3-10.

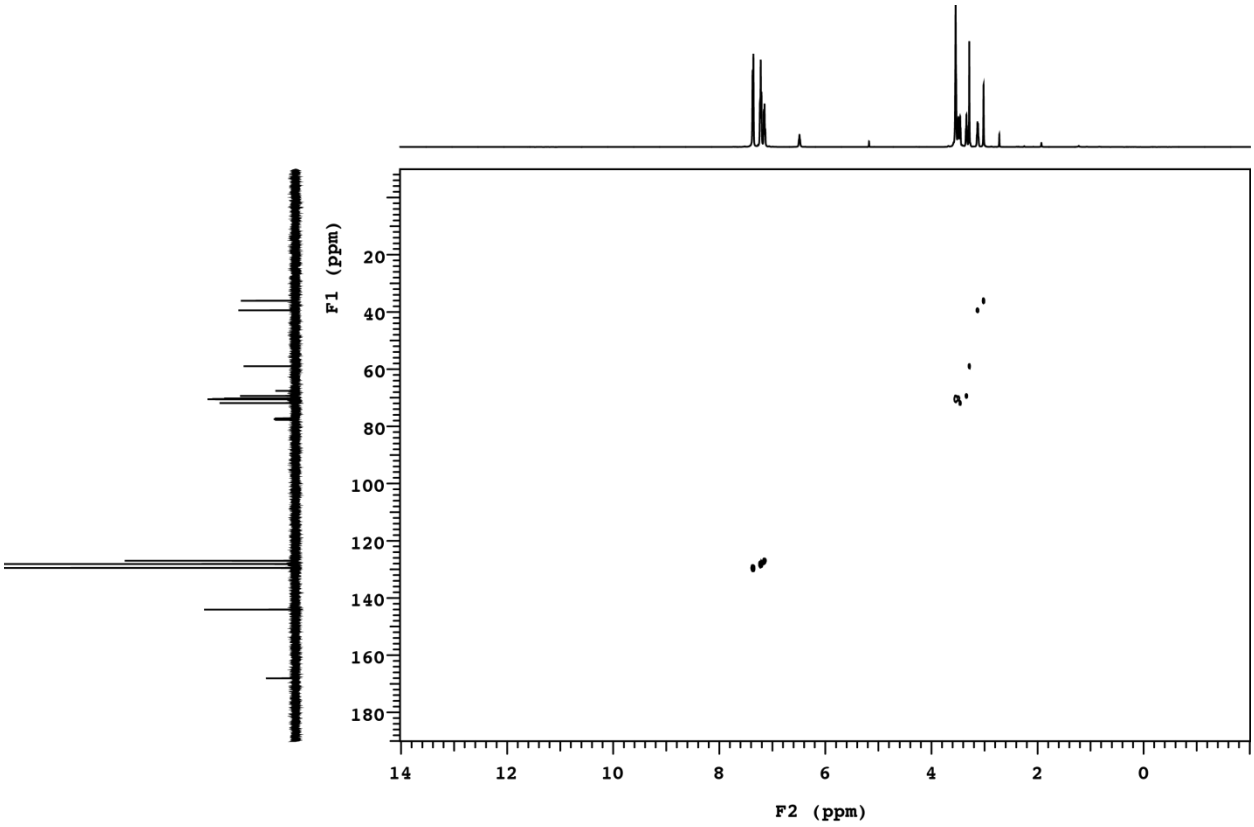


Figure 3-44. 2D HSQC spectrum of 2-(tritylthio)acetamidePEG4.

Table 3-10. One-Bond C-H correlations identified from HSQC experiment of 2-(tritylthio)acetamidePEG4.

^{1}H δ	^{1}C δ	Assignment
7.373	128.105	Trt aryl C-H
7.234	128.029	Trt aryl C-H
7.162	126.97	Trt aryl C-H
3.544-3.450	50.377	**CH_2-CH_2**-O-**CH_2-CH_2**-O--**CH_2-CH_2-**O-
4.589-3.613	71.83, 70.53, 70.48, 70.46, 70.41 70.19	-CONH-CH_2-CH_2O-
3.287	58.952	O-**CH_3**
3.142	39.409	-CONH-**CH_2-**
3.017	36.062	-S-**CH_2**CONH-

High-resolution ESI-TOF MS data for 2-(tritylthio)acetamidePEG4 (**S10)** are shown in Figure 3-45.

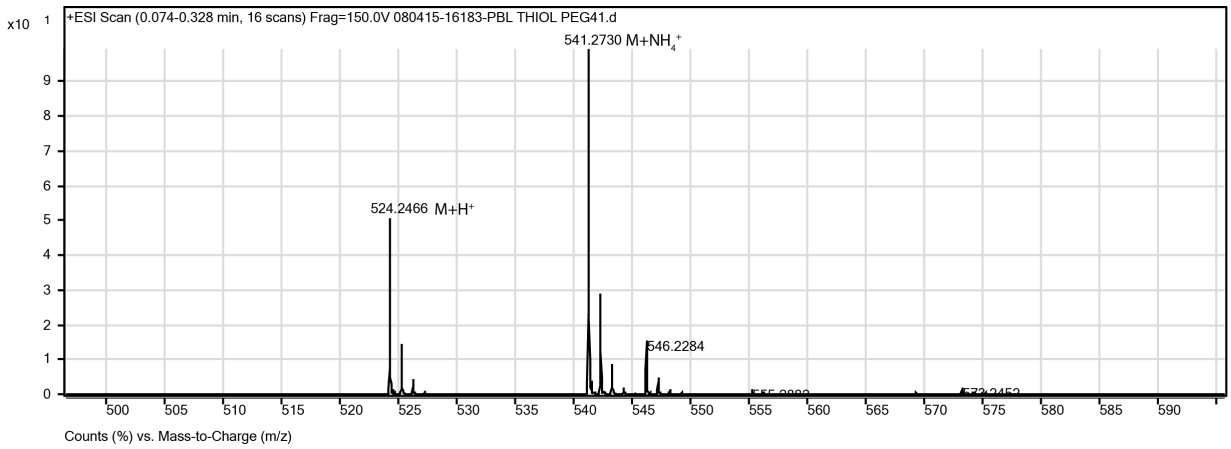
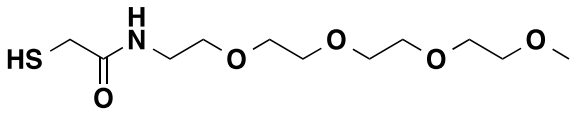


Figure 3-45. ESI-TOF MS data for 2-(tritylthio) acetamidePEG4. Calculated m/z for $C_{30}H_{38}NO_5S$ ($M+H^+$) 524.25, found 524.24

3.4.19 2-mercaptoPEG$_4$acetamide (S11)



To a solution of TFA (25mL) was added triisopropylsilane (100µL) and 2-(tritylthio)acetamidePEG4 (**S10**) (80mg, 0.28mmol) and the solution was stirred for 2 h under an argon atmosphere. The product was concentrated by rotary evaporation, dissolved in H_2O and purified by RP-HPLC (Figure 3-46). Pooled fractions containing the desired product were flash-frozen over acetone/dry ice; solvent was removed via lyophilization.

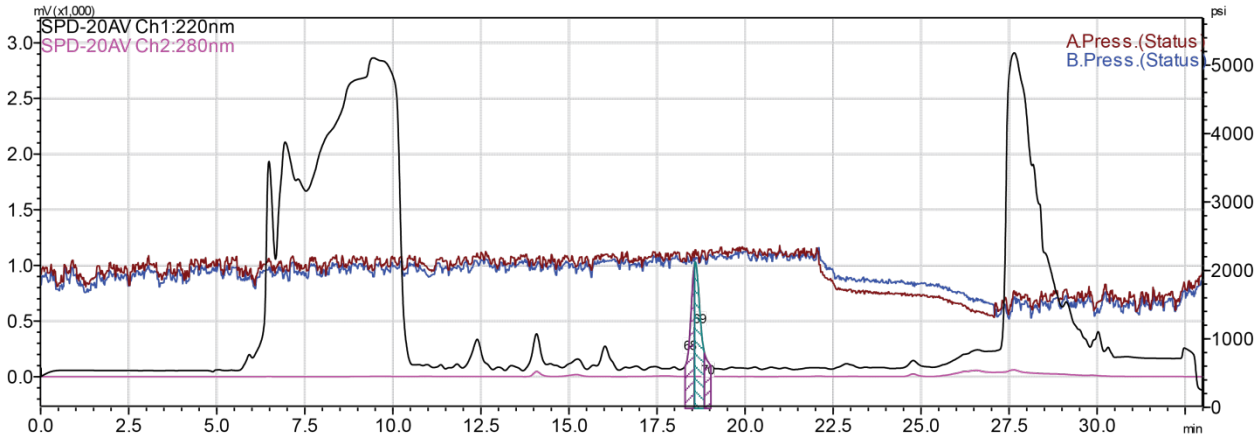


Figure 3-46. Preparative reverse-phase HPLC purification of 2-mercaptoPEG$_4$acetamide. 2-mercaptoPEG$_4$acetamide was injected onto a C18 preparative column and eluted using a linear gradient of 0-30% B (A=H_2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 22 minutes, followed by a 5 minute rinse (95% B), and a 6 minute column re-equilibration (0% B) with a flow rate of 15 mL/min.

^{1}H NMR (500 MHz, CDCl3): δ 7.35 (1H, s, -S-CH_2CON**H**-CH_2-); 3.66-3.54 (14H, m, -CONH-CH_2-CH_2O-C**H$_2$**-C**H$_2$**-O-C**H$_2$**-C**H$_2$**-O--C**H$_2$**-C**H$_2$-O-**- and -CONH-CH_2-C**H$_2$**O-); 3.49 (2H, m, -CONH-C**H$_2$**-CH_2O-); 3.37 (3H, s, -O-C**H$_3$**); 3.23 (2H, d, J= 8.46, -S-C**H$_2$**CONH-CH_2-); 2.01 (1H, t, J= 8.76, **H**S-CH_2CONH-). The full 1H NMR spectrum for 2-mercaptoPEG$_4$acetamide appears in Figure 3-47.

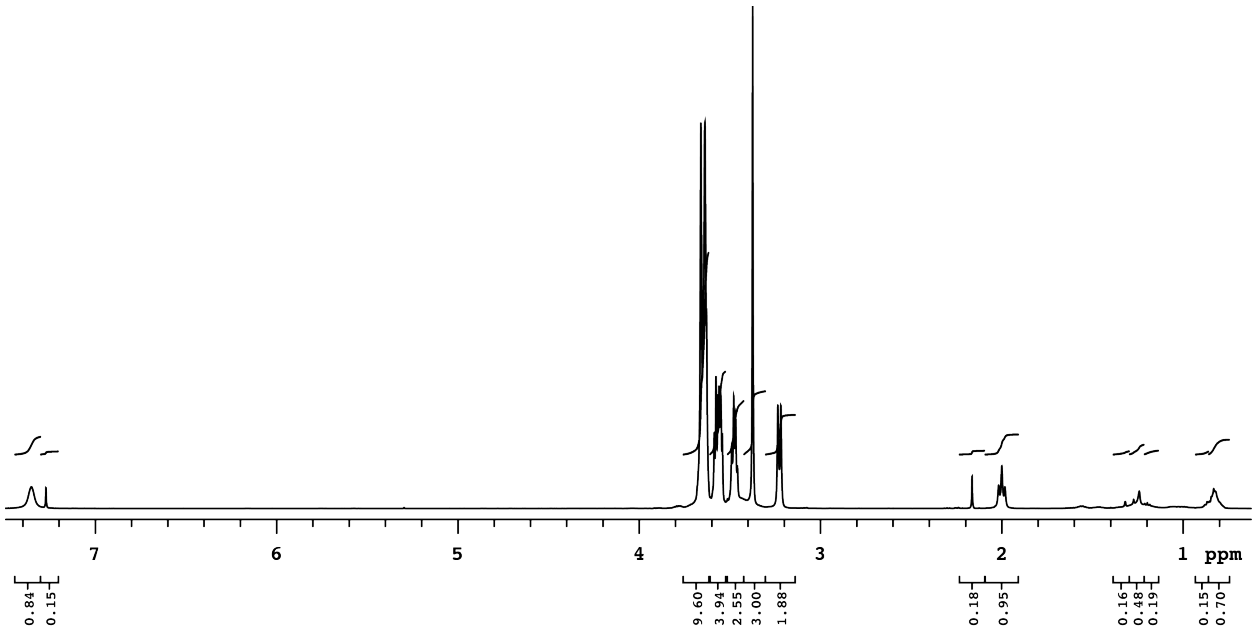


Figure 3-47. ^{1}H NMR spectrum of 2-mercaptoPEG$_4$acetamide S11

^{13}C NMR (126 MHz, $CDCl_3$): δ 169.92 (-S-CH_2**C**ONH-CH_2-); 71.80, 70.44, 70.37, 70.31, 70.26 70.18, 70.05, 69.68 (-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-); 69.68 (-CONH-CH_2-**C**H_2O-); 59.00 (-O-**C**H_3); 39.65 (-CONH-**C**H_2-CH_2O-); 28.19 (-S-**C**H_2CONH-CH_2-). The full 13C NMR spectrum for 2-mercaptoPEG$_4$acetamide appears in Figure 3-48.

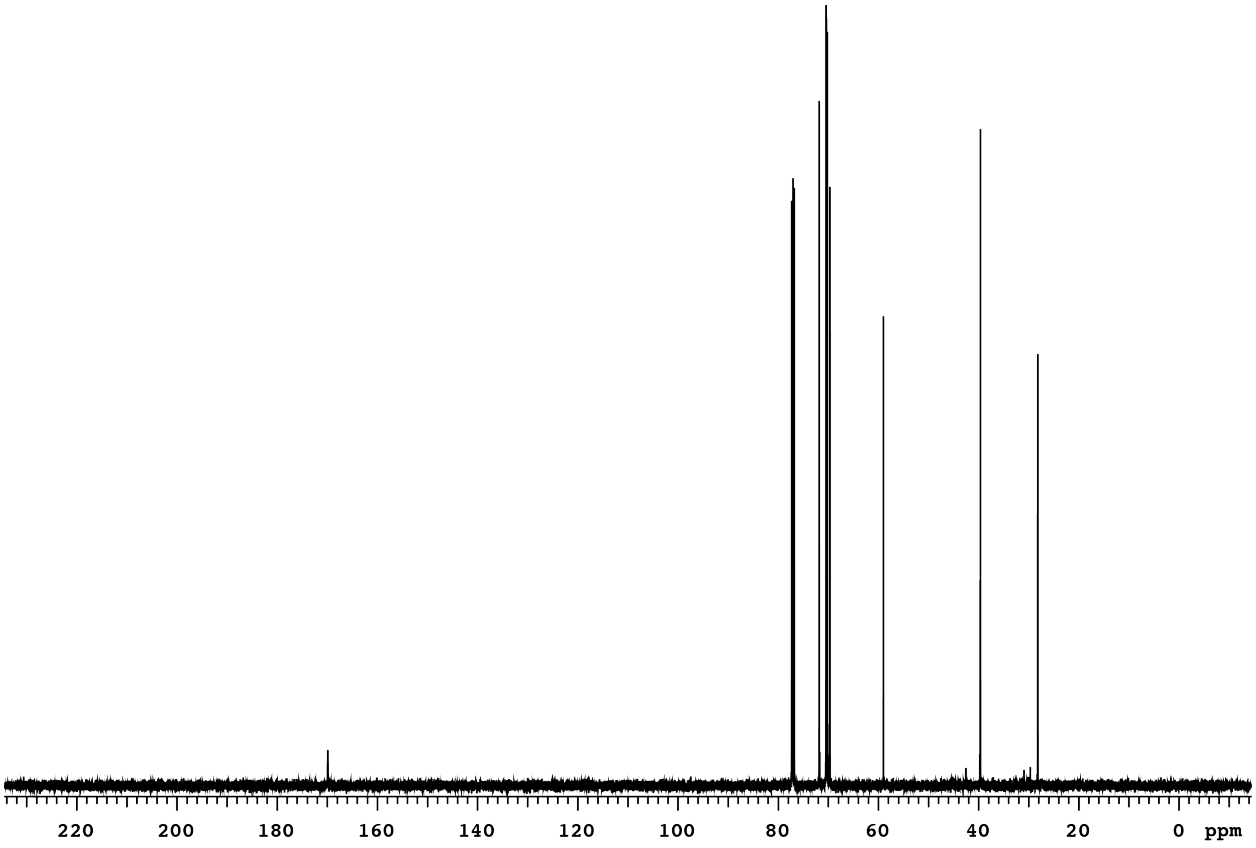


Figure 3-48. ^{13}C NMR spectrum of 2-mercaptoPEG$_4$acetamide S11

Assignments of the ^{1}H and ^{13}C NMR spectra for the 2-mercaptoPEG$_4$acetamide were made with the assistance of a 2D HSQC experiment Figure 3-49, using the one-bond C-H correlations shown in Table 3-11.

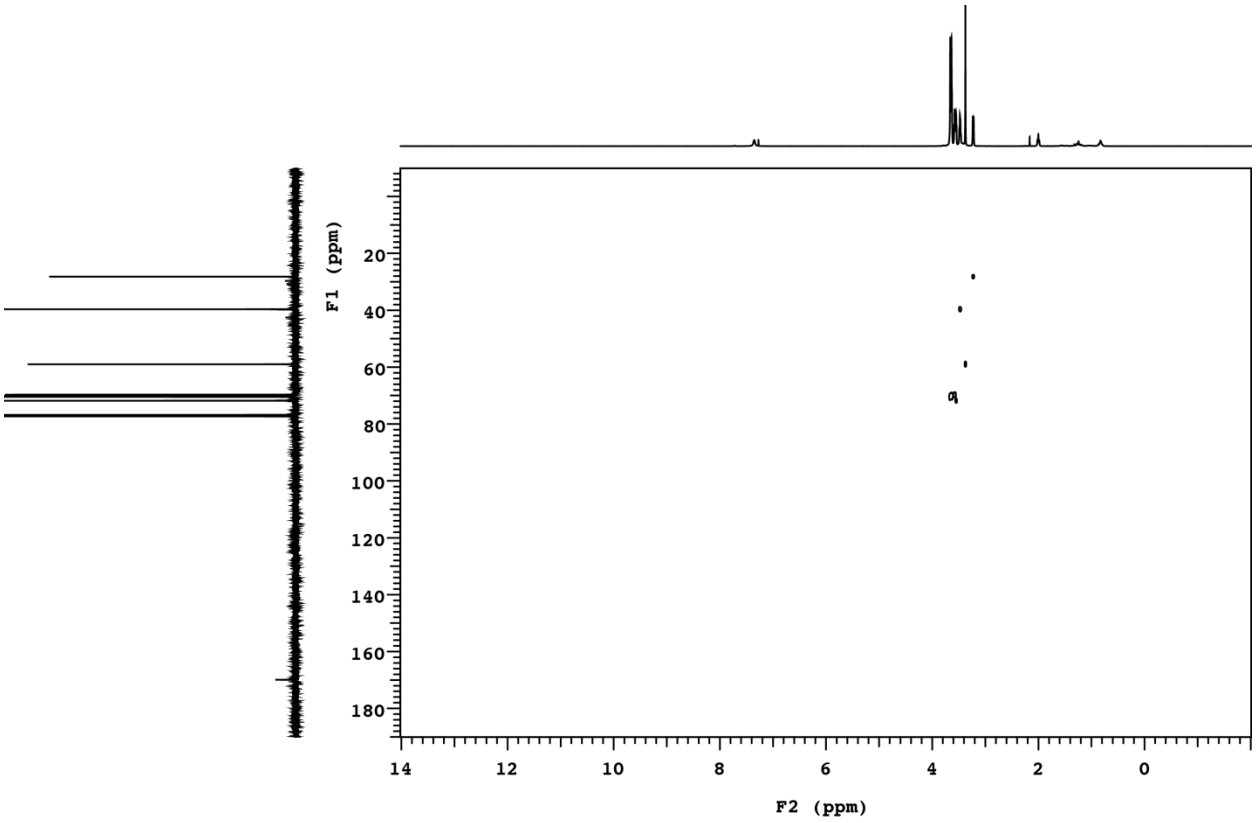


Figure 3-49. 2D HSQC spectrum of 2-mercaptoPEG$_4$acetamide S11

Table 3-11. One-Bond C-H correlations identified from HSQC experiment for 2-mercaptoPEG$_4$acetamide.

^{1}H δ	^{1}C δ	Assignment
3.660-3.542	71.80, 70.44, 70.37, 70.31, 70.26 70.18, 70.05, 69.68, 69.68	**CH_2-CH_2**-O-**CH_2-** **CH_2**-O--**CH_2-** **CH_2**-O-, -CONH-CH$_2$-**CH_2**O-
3.375	59.001	O-**CH_3**
3.49	39.651	-CONH-**CH_2-**
3.236	28.195	-S-**CH_2**CONH-

High-resolution ESI-TOF MS data for 2-mercaptoPEG$_4$acetamide are shown in Figure 3-50.

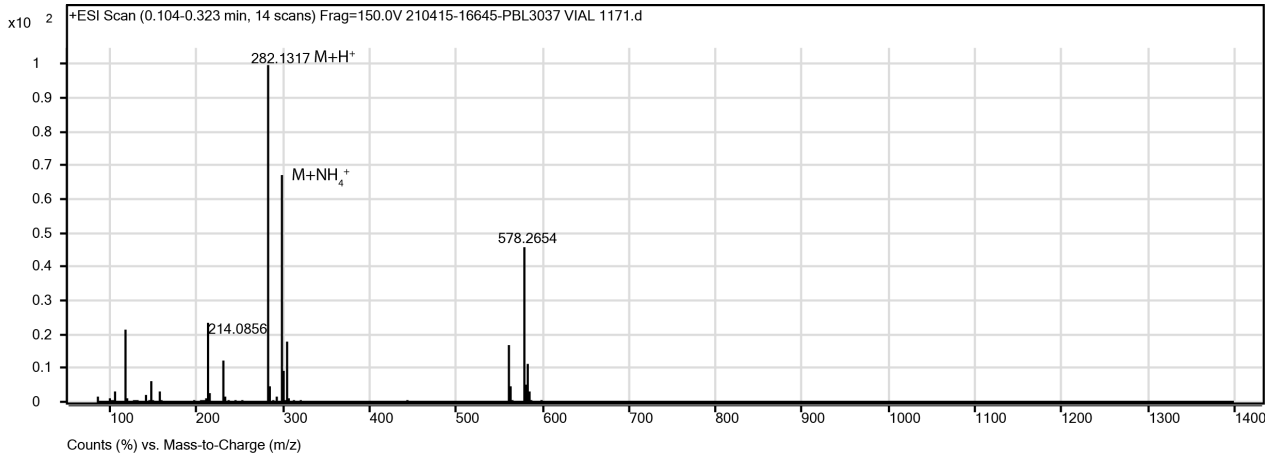
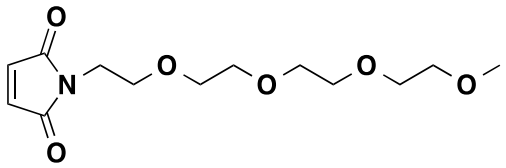


Figure 3-50. ESI-TOF MS data for **Error! Reference source not found.** Calculated m/z for $C_{11}H_{24}NO_5S$ ($M+H^+$) 282.14, found 282.13

3.4.20 4PEG-maleimide (S12)



Commercially available 4PEG-amine (100mg, 0.48mmol) was dissolved in a saturated solution of $NaHCO_3$ at 0 °C. N-methoxy-carbonyl maleimide (74.8mg, 0.48mmol) was added slowly. The mixture reacted for 45 minutes at 0 °C and then for an additional 45 minutes at room temperature. The product was extracted from the aqueous mixture with CH_2Cl_2 (2 X 15 mL), and the organic layer was dried over Na_2SO_4 and concentrated in vacuo. The desired product **(S12)** was purified by flash chromatography over silica gel (10% MeOH/DCM).

^{1}H NMR (300 MHz, CDCl3): δ 6.71 (2H, s, alkene **H's**); 3.74 (2H, m, -N-C**H$_2$**-CH_2O-); 3.67-3.53 (14H, m, -N-CH_2-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**-); 3.39 (3H, s, -O-C**H$_3$**). The full 1H NMR spectrum for 4PEG-maleimide appears in Figure 3-51.

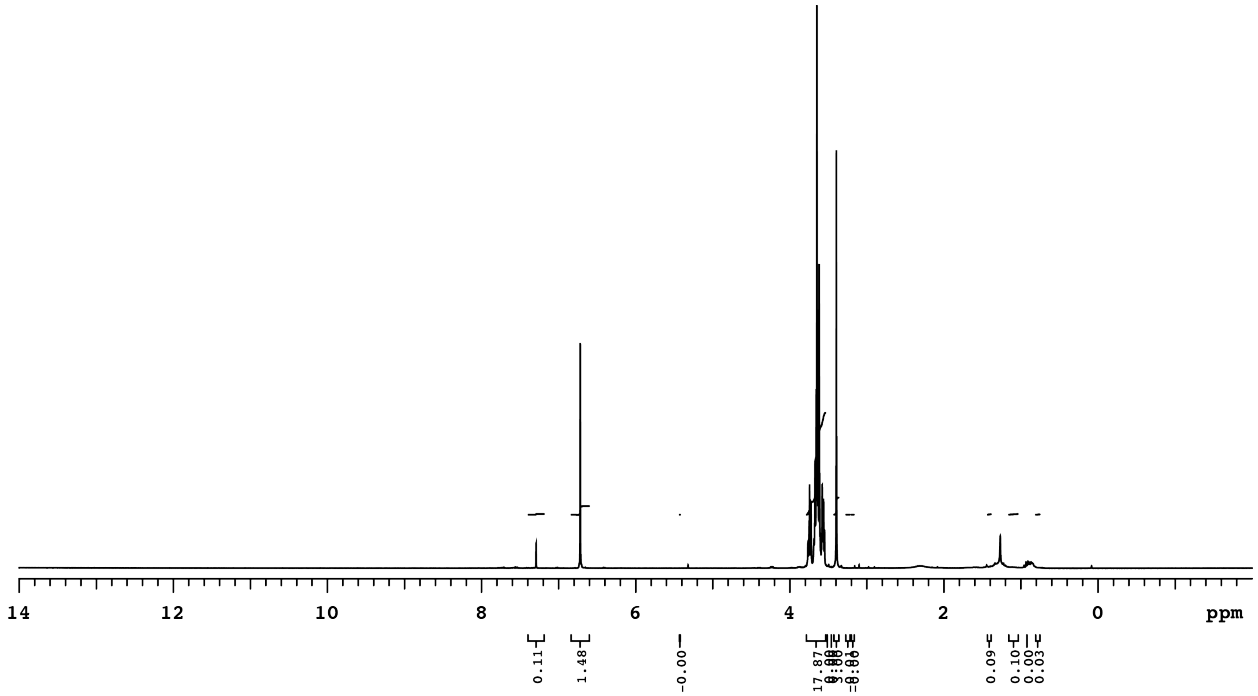


Figure 3-51. ^{1}H NMR spectrum of 4PEG-maleimide.

^{13}C NMR (75.4 MHz, $CDCl_3$): δ 170.68 (carbonyl **C's**); 134.16 (alkene **C's**); 71.93, 70.60, 70.58, 70.54, 70.51, 70.02, 67.84 (-N-CH_2-**C**H_2O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-O-**C**H_2-**C**H_2-**O**-); 59.04 (3H, s, -O-**C**H_3); 37.12 (-N-**C**H_2-CH_2O-). The full 13C NMR spectrum for 4PEG-maleimide appears in Figure 3-52.

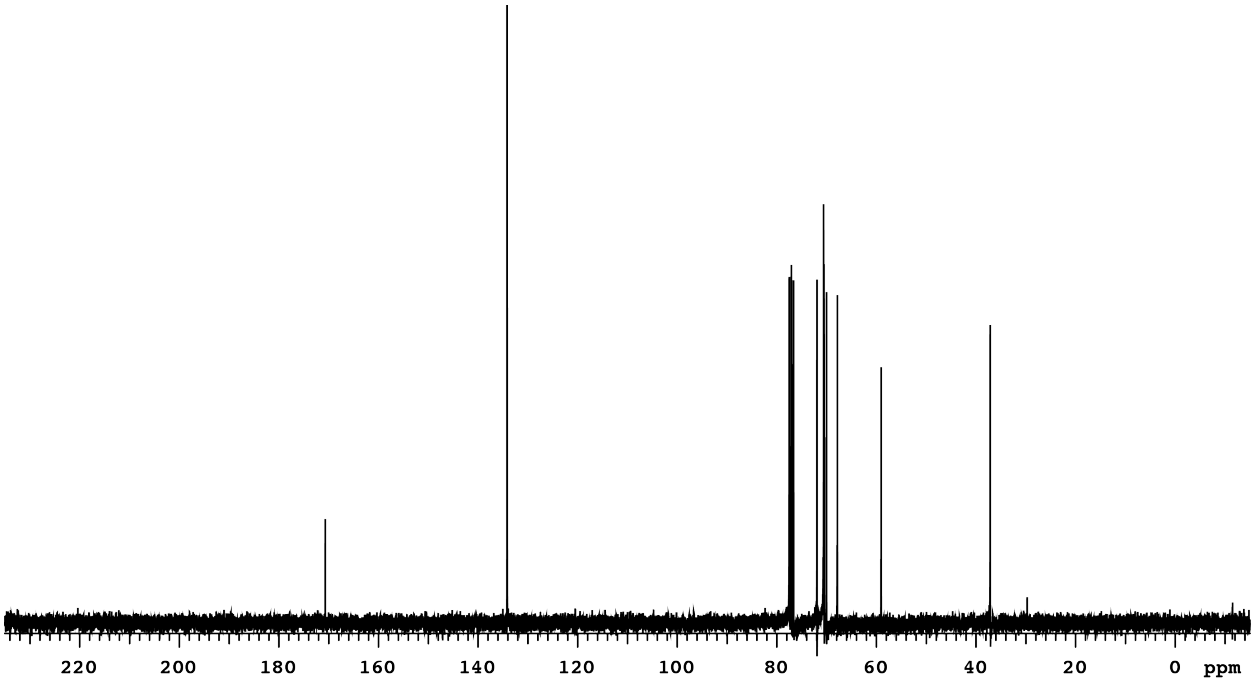


Figure 3-52**.** ^{13}C NMR spectrum **of** 4PEG-maleimide

Assignments of the ^{1}H and ^{13}C NMR spectra for the 4PEG-maleimide were made with the assistance of a 2D HSQC experiment Figure 3-53, using the one-bond C-H correlations shown in Table 3-12. One-Bond C-H correlations identified from HSQC experiment for 4PEG-maleimide.

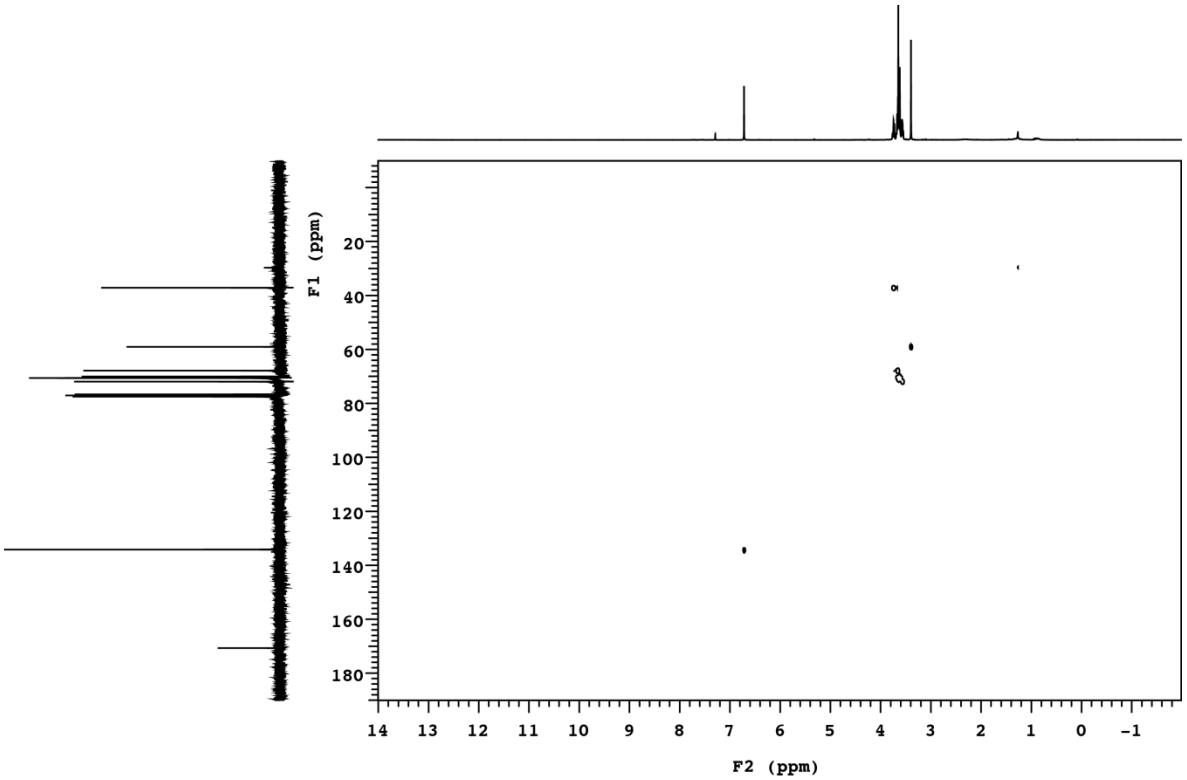


Figure 3-53**.** 2D HSQC spectrum **of** 4PEG-maleimide.

Table 3-12. One-Bond C-H correlations identified from HSQC experiment for 4PEG-maleimide.

^{1}H δ	^{1}C δ	Assignment
6.719	134.165	alkene
3.74	37.12	-N-C**H$_2$**-CH$_2$O-
3.67-3.53	71.93, 70.60, 70.58, 70.54, 70.51, 70.02, 67.84	-N-CH$_2$-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**-
3.395	59.049	-O-C**H$_3$**

High-resolution ESI-TOF MS data for 4PEG-maleimide are shown in Figure 3-54.

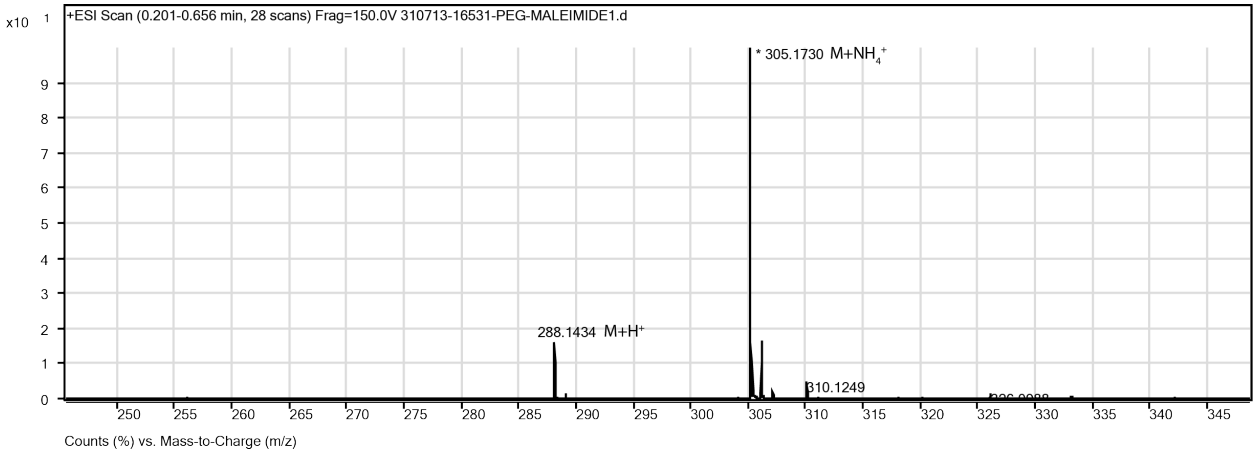
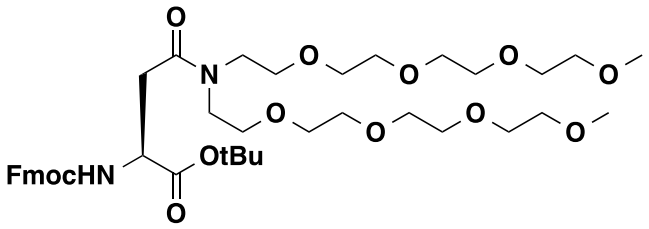


Figure 3-54. ESI-TOF MS data for 4PEG-maleimide. Calculated m/z for $C_{13}H_{22}NO_6$ ($M+H^+$) 288.14, found 288.14

3.4.21 Fmoc-L-Asn(PEG4)$_2$-OtBu (S13)



To a solution of commercially available (*S*)-3-((((*9H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-(*tert*-butoxy)-4-oxobutanoic acid (Fmoc-L-Asp-OtBu, 0.5139g, 1.25 mmol) in NMP (50 mL) was added 2-(6-Chloro-1H-benzotriazole-1-yl)-1,1,3,3-tetramethylaminium hexafluorophosphate (HCTU, 0.775g, 1.875mmol), N-hydroxybenzotriazole hydrate (HOBt, 0.287g, 1.875mmol), and *N*,*N*-Diisopropylethylamine (DIEA, 0.652mL, 3.75mmol), and the resulting mixture was stirred for 15 min at room temperature. Then, commercially available di-(2,5,8,11-tetraoxatridecan-13-yl)-amine (0.5g, 1.26mmol) was added, and the resulting mixture was stirred at room temperature for 24 hours. The reaction mixture was

then washed with saturated sodium bicarbonate (100 mL) and extracted three times with ethyl acetate (100 mL). The combined ethyl acetate extracts were washed three times with brine (100mL), dried over $MgSO_4$, and concentrated by rotary evaporation to afford an oil. Fmoc-L-Asp(PEG4)$_2$-OtBu (*tert*-butyl [(*S*)-17-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-15-oxo-14-(2,5,8,11-tetraoxatridecan-13-yl)-2,5,8,11-tetraoxa-14-azaoctadecan-18-oate], **S13**) was purified by flash chromatography over silica in ethyl acetate/hexanes. The product was concentrated via rotary evaporation and dried in vacuo to give a thick oily solid.

^{1}H NMR (500 MHz, $CDCl_3$): δ 7.72 (2H, d, J = 7.5 Hz, Fmoc aryl C-**H**); 7.57 (2H, m, Fmoc aryl C-**H**); 7.34 (2H, t, J = 7.4 Hz, Fmoc aryl C-**H**); 7.26 (2H, t, J = 7.5Hz, Fmoc aryl C-**H**); 6.169 (1H, d, -CON**H**-CαH(COOtBu)-); 4.51 (1H, m, -CONHCα**H**(COOtBu)-CβH_2-); 4.45 (2H, m, Fmoc Ar$_2$CH-C**H$_{(a)}$H$_{(b)}$**-O-); 4.18 (1H, t, J = 6.945, Fmoc Ar$_2$C**H**-CH$_{(a)}$H$_{(b)}$-O-); 3.58-3.33 (32H, m, -CON-(C**H$_2$**-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-)$_2$); 3.31 (3H, s, -O-C**H$_{3a}$**); 3.30 (3H, s, -O-C**H$_{3b}$**); 3.14 (1H, dd, -CαH(COOH)-Cβ(**Ha**)Hb-CON-); 2.82 (1H, dd, -CαH(COOH)-Cβ(Ha)**Hb**-CON-); 1.40 (9H, s, -O-C-(C**H$_3$**)$_3$. Peaks at 1.96, 2.31, 2.78 correspond to residual NMP. The full ^{1}H NMR spectrum for Fmoc-L-Asn(PEG4)$_2$-OtBu is shown in Figure 3-55.

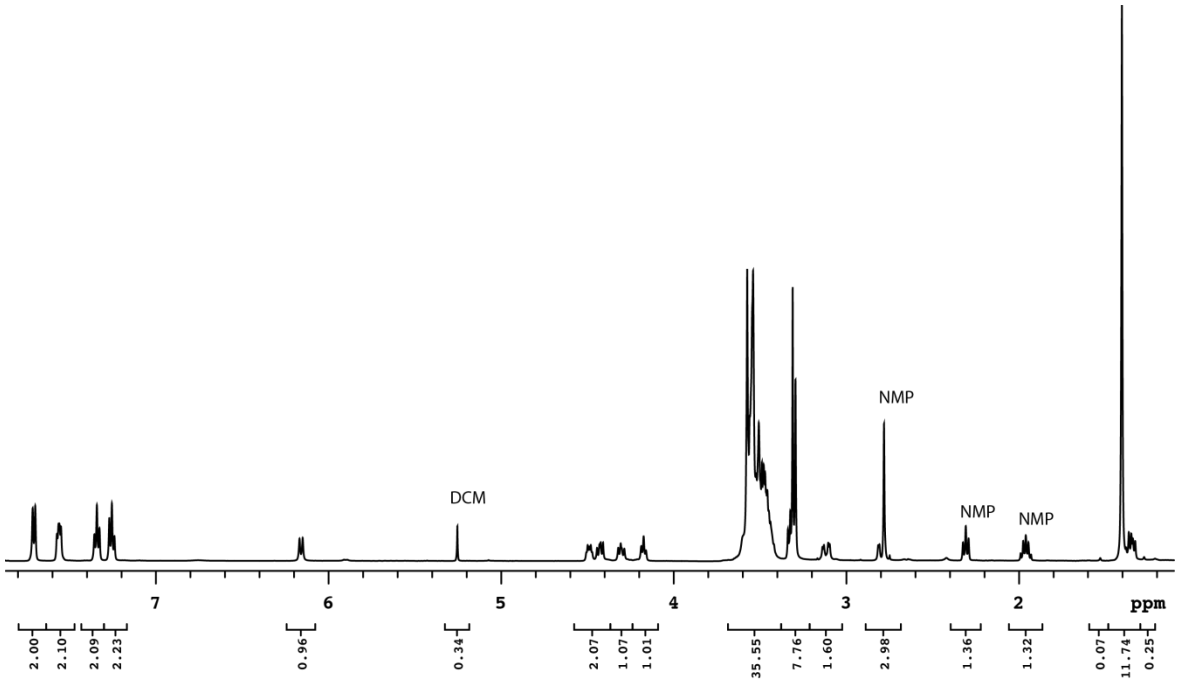


Figure 3-55. ^{1}H NMR spectrum for Fmoc-L-Asn(PEG4)$_2$-OtBu.

^{13}C NMR (126 MHz, $CDCl_3$): δ 171.20, 170.36 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.34 (Fmoc-O-**C**ONH-); 143.94, 143.78, 141.23, 141.22 (Fmoc aryl **ipso** C's); 127.65, 127.10, 127.03, 125.11, 125.02, 119.92, 119.91 (Fmoc aryl **C**-H); 81.68 (-O-**C**$(CH_3)_3$); 71.84-46.19 (-**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-); 66.68 (Fmoc Ar_2CH-**C**H_2-O-); 58.94, 58.91 (-O-**C**H_3); 51.21 (-NH-**C**αH(COOtBu)-CβH_2-); 47.12 (Fmoc Ar_2**C**H-CH_2-O-); 35.65 (-CαH(COOH)-**Cβ**H_2-COO-); 27.90 (-O-C(**C**$H_3)_3$). The full ^{13}C NMR spectrum for Fmoc-L-Asn(PEG4)$_2$-OtBu is shown in Figure 3-56

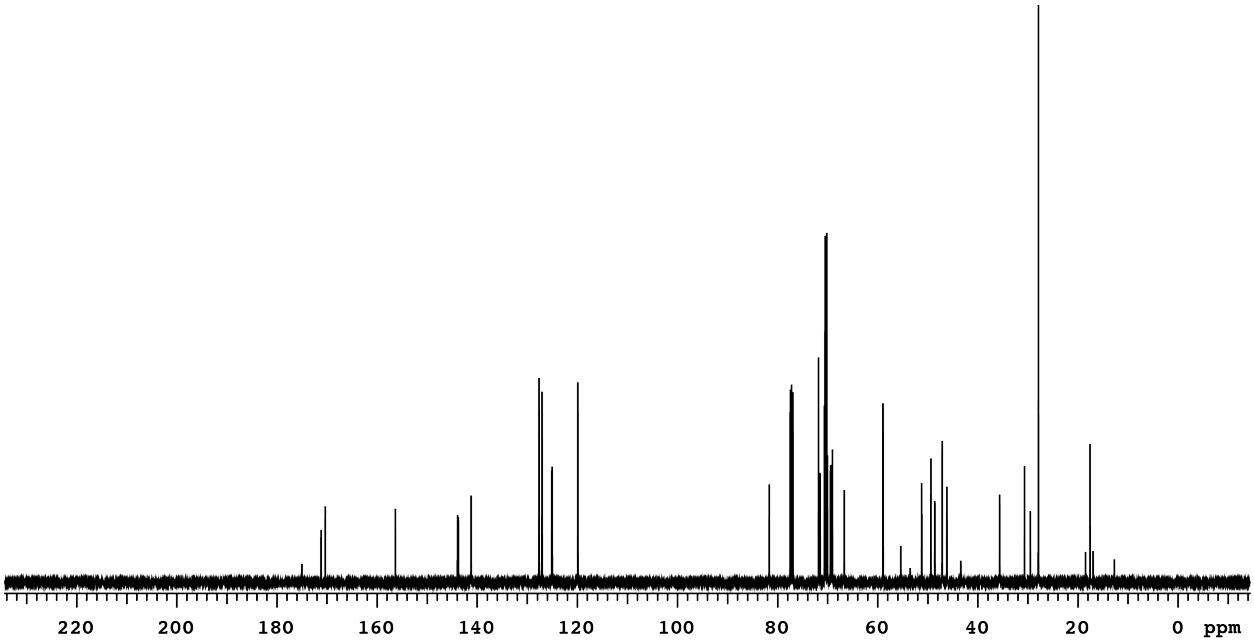


Figure 3-56. ^{13}C NMR spectrum for Fmoc-L-Asn(PEG4)$_2$-OtBu.

Assignments of the ^{1}H and ^{1}C NMR spectra for the Fmoc-L-Asn(PEG4)$_2$-OtBu were with the assistance of a 2D HSQC experiment Figure 3-57, using the one-bond C-H correlations shown in Table 3-13.

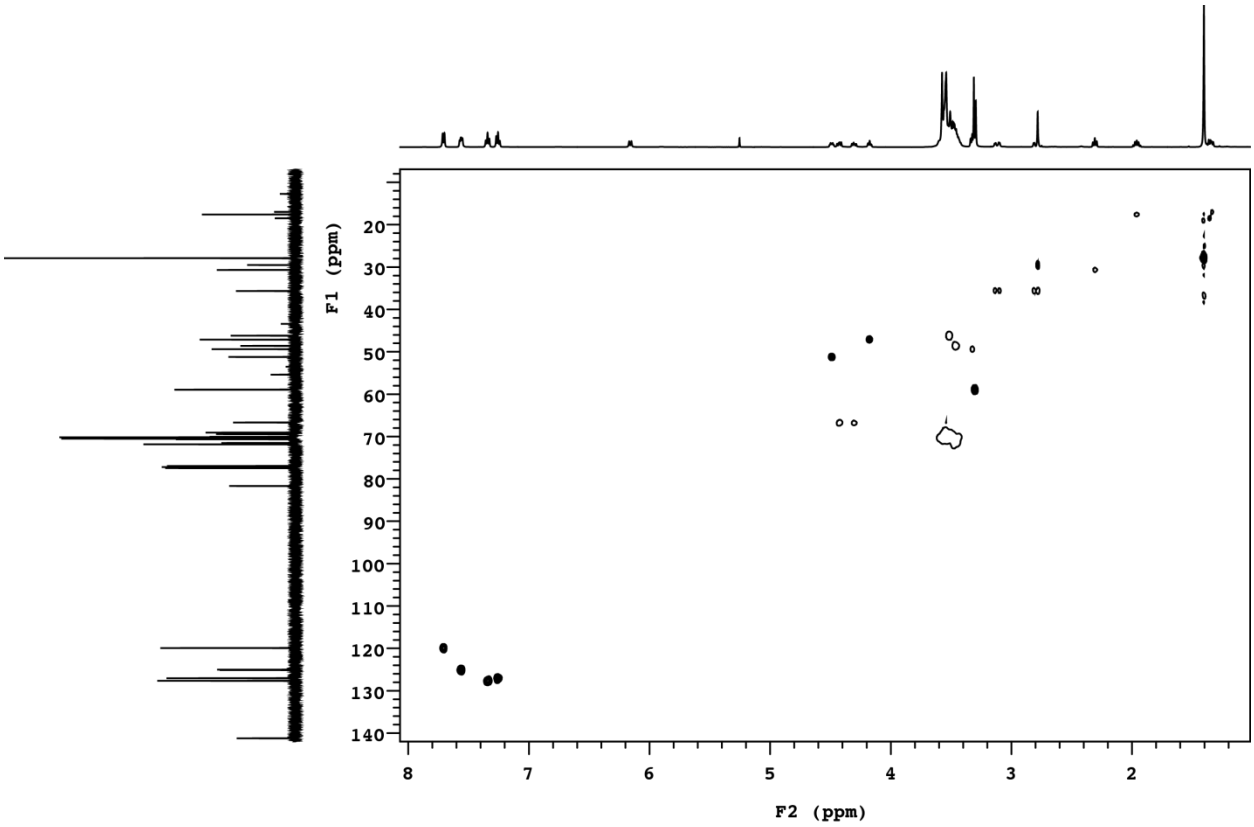


Figure 3-57. 2D HSQC spectrum of Fmoc-L-Asn(PEG4)$_2$-OtBu.

Table 3-13. One-Bond C-H correlations identified from HSQC data for Fmoc-L-Asn(PEG4)$_2$-OtBu.

^{1}H δ	^{1}C δ	Assignment
7.715	119.92, 119.91	Fmoc aryl C-H
7.574-7.551	125.11, 125.02	Fmoc aryl C-H
7.342	127.65	Fmoc aryl C-H
7.256	127.05, 127.03	Fmoc aryl C-H
4.51	51.21	-CONH**CαH**(COOH)-CβH2-
4.45	66.68	Fmoc Ar2CH-**CH$_{(a)}$H$_{(b)}$**-O
4.18	47.12	Fmoc Ar**CH**-CH2-O-
3.58-3.33	71.84-46.19	-CON-**CH$_2$**-**CH$_2$**-O-**CH$_2$**-**CH$_2$**-O-**CH$_2$**-**CH$_2$**-O-**CH$_2$**-**CH$_2$**,
3.31 and 3.30	58.94, 58.91	-O-**CH$_3$**
3.14, 2.82	35.65	-CαH(COOH)-**CβH$_2$**-
1.40	27.9	-C(**CH$_3$**)$_3$

High-resolution ESI-TOF MS data for Fmoc-L-Asn(PEG4)$_2$-OtBu are shown in Figure 3-58.

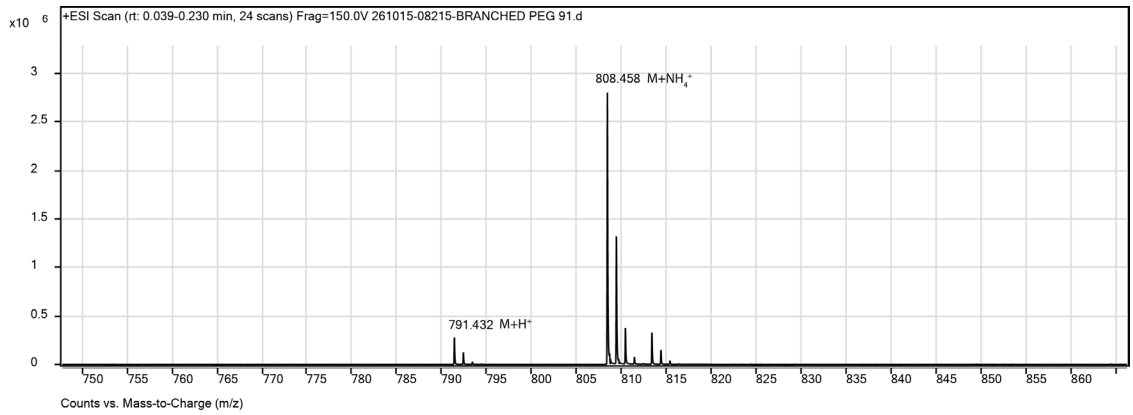
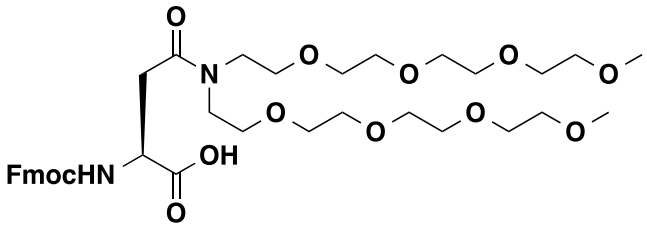


Figure 3-58. ESI-TOF MS data for Fmoc-L-Asn(PEG4)$_2$-OtBu. Calculated m/z for $C_{41}H_{66}N_3O_{13}$ ($M+NH_4^+$) is 808.46, found 808.458

3.4.22 Fmoc-L-Asn(PEG)$_2$-OH (S14)



To a solution of TFA (95% in water, 50ml) was added Fmoc-L-Asn(PEG)$_2$-OtBu (**S13**), and the solution was stirred for 4 h under an argon atmosphere. The product Fmoc-L-Asn(PEG)$_2$-OH [**(***S*)-17-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-15-oxo-14-(2,5,8,11-tetraoxatridecan-13-yl)-2,5,8,11-tetraoxa-14-azaoctadecan-18-oic acid, **S14**] was concentrated by rotary evaporation, and used without further purification.

^{1}H NMR (500 MHz, $CDCl_3$): δ 11.43 (1H, s, (COO**H**); 7.77(2H, d, J = 7.6 Hz, Fmoc aryl C-**H**); 7.63 (2H, t, J = 7.44 Fmoc aryl C-**H**); 7.40 (2H, t, J = 7.4 Hz, Fmoc aryl C-**H**); 7.31

(2H, t, J = 7.4Hz, Fmoc aryl C-**H**); 6.37 (1H, broad s, -CON**H**-CαH(COOtBu)-); 4.64 (1H, broad s, -CONHCα**H**(COOtBu)-CβH_2-); 4.43 (1H, m, Fmoc Ar_2CH-C**H**$_{(a)}$$H_{(b)}$-O-); 4.32 (1H, m, Fmoc Ar_2CH-$CH_{(a)}$**H**$_{(b)}$-O-); 4.23 (1H, t, J = 7.2, Fmoc Ar_2C**H**-$CH_{(a)}H_{(b)}$-O-); 3.66-3.50 (32H, m, -CON-(C**H**$_2$-C**H**$_2$-O-C**H**$_2$-C**H**$_2$-O-C**H**$_2$-C**H**$_2$-O-C**H**$_2$-C**H**$_2$-O-)$_2$); 3.40 (3H, s, -O-C**H**$_3$); 3.36 (1H, dd, -CαH(COOH)-Cβ(**Ha**)Hb-CON-); 2.96 (1H, dd, -CαH(COOH)-Cβ(Ha)**Hb**-CON-); Peaks at 2.10, 2.60, and 2.92 correspond to residual NMP. The full ^{1}H NMR spectrum for Fmoc-L-Asn$(PEG)_2$-OH is shown in Figure 3-59.

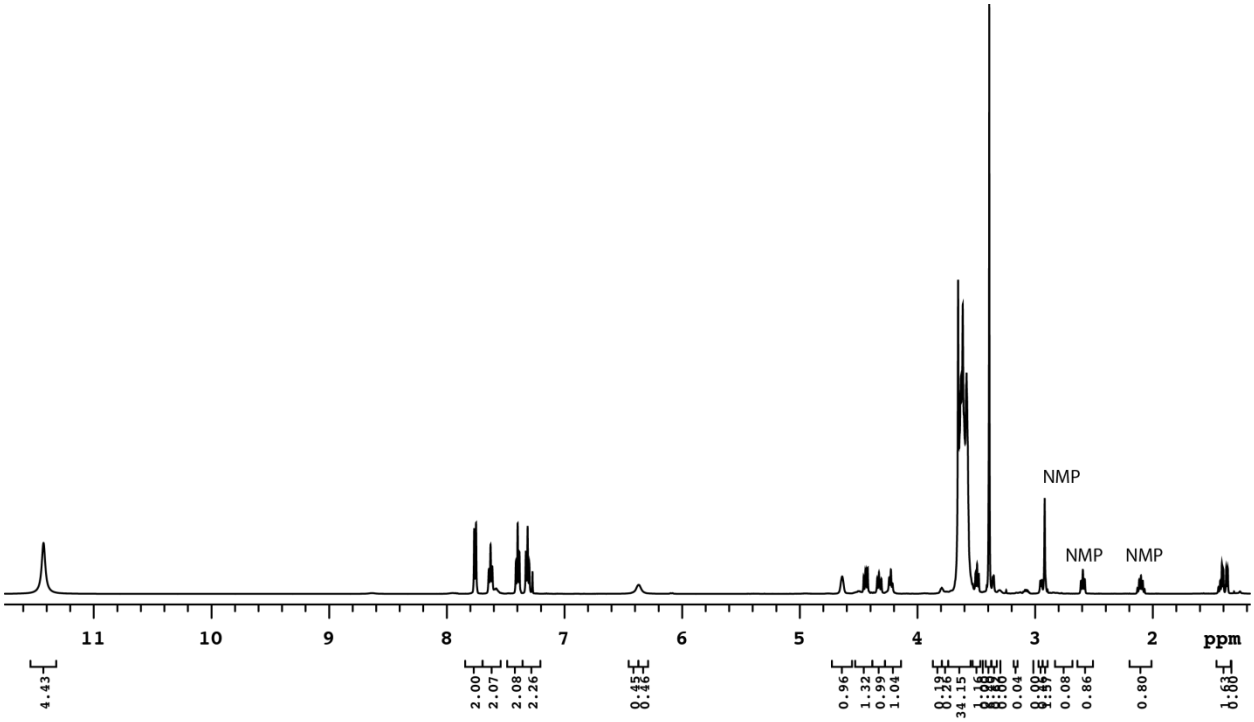


Figure 3-59. ^{1}H NMR spectrum for Fmoc-L-Asn$(PEG)_2$-OH.

^{13}C NMR (75.4 MHz, $CDCl_3$): δ 173.00, 172.61 (-NH-CαH(**C**OOtBu)-CβH_2- and -CβH_2-**C**ON-CH_2-); 156.38 (Fmoc-O-**C**ONH-); 143.86, 143.66, 141.26, 141.25 (Fmoc aryl **ipso** C's); 127.73, 127.12, 127.01, 125.26, 125.17, 119.94 (Fmoc aryl **C**-H); 71.73-46.32 (-**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-); 67.34 (Fmoc Ar_2CH-**C**H_2-O-); 58.83, 58.77, (-O-**C**H_3);

54.47 (-NH-**C**αH(COOtBu)-CβH_2-); 47.03 (Fmoc Ar_2**C**H-CH_2-O-); 35.73 (-CαH(COOH)-**Cβ**H_2-COO-); The full ^{13}C NMR spectrum for Fmoc-L-Asn$(PEG)_2$-OH is shown in Figure 3-60.

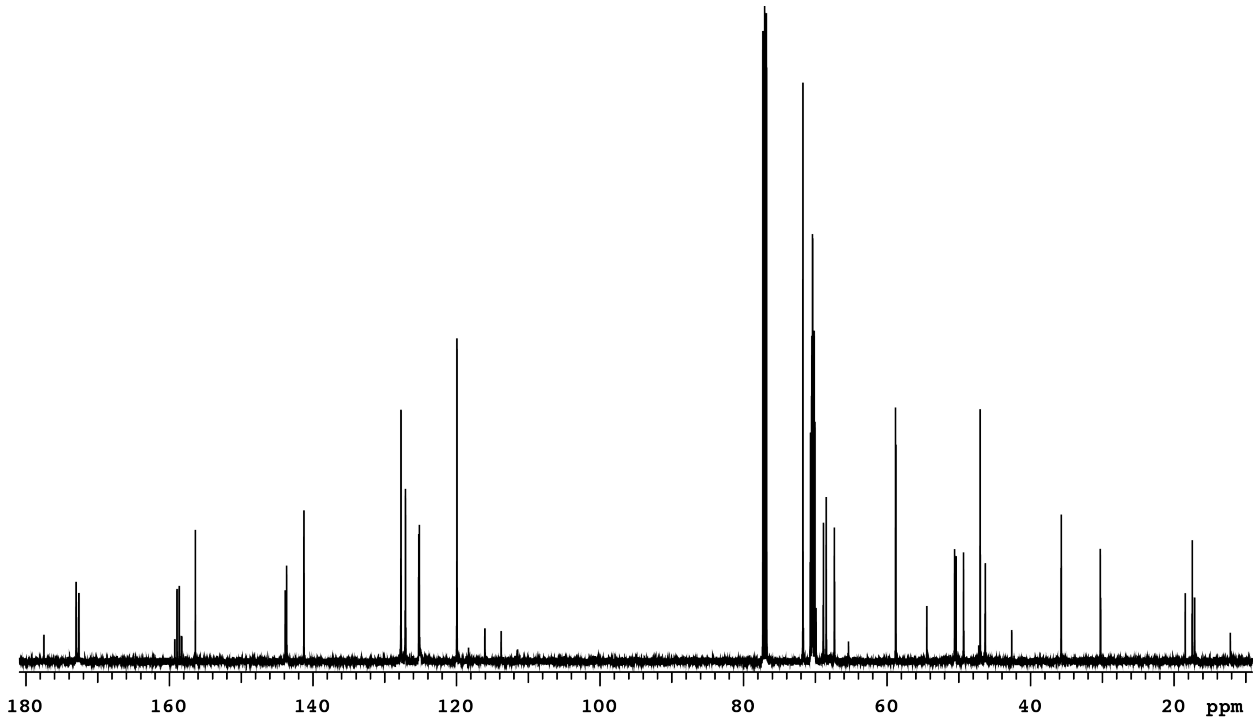


Figure 3-60. ^{13}C NMR spectrum for Fmoc-L-Asn$(PEG)_2$-OH

Assignments of the ^{1}H and ^{1}C NMRs for the Fmoc-L-Asn$(PEG)_2$-OH were made by analogy with published spectral data for related compounds, and with the assistance of a 2D HSQC experiment Figure 3-61, using the one-bond C-H correlations shown in Table 3-14.

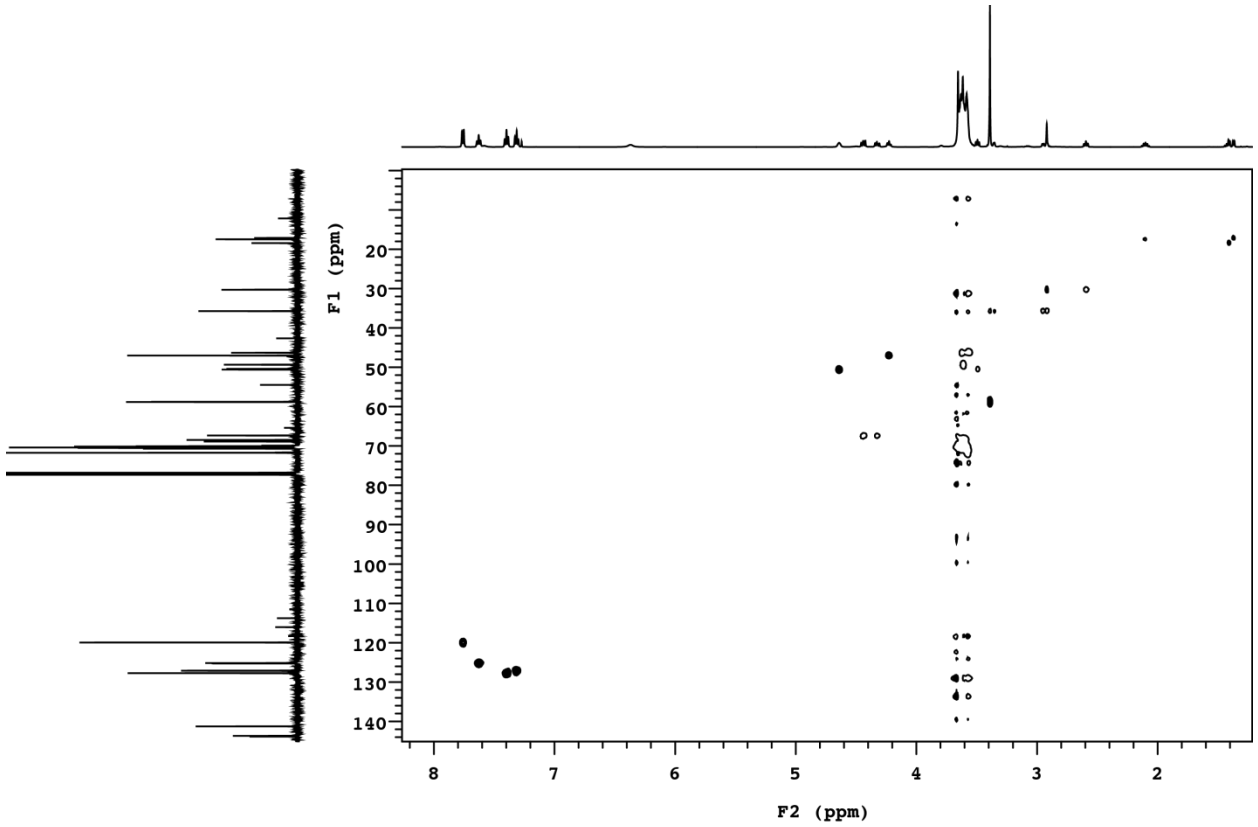


Figure 3-61. 2D HSQC spectrum of Fmoc-L-Asn(PEG)$_2$-OH.

Table 3-14. One-Bond C-H correlations identified from HSQC data for Fmoc-L-Asn(PEG)$_2$-OH.

^{1}H δ	^{1}C δ	Assignment
7.77	119.94	Fmoc aryl C-H
7.63	125.259, 125.168	Fmoc aryl C-H
7.40	127.73	Fmoc aryl C-H
7.31	127.118, 127.095	Fmoc aryl C-H
4.64	54.47	CONH**CαH**(CO OH)-CβH2-
4.46	67.34	Fmoc Ar2CH-C**H$_{(a)}$H$_{(b)}$**-O
4.23	47.03	Fmoc Ar**CH**-CH2-O-
3.655-3.479	71.729-46.315	-CON-C**H$_2$**-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-C**H$_2$**-C**H$_2$**,
3.39	58.834, 58.774	-O-**CH$_3$**
3.357, 2.952	35.73	-CαH(COOH)-**CβH$_2$**-

High-resolution ESI-TOF MS data for Fmoc-L-Asn(PEG)$_2$-OH are shown in Figure 3-62.

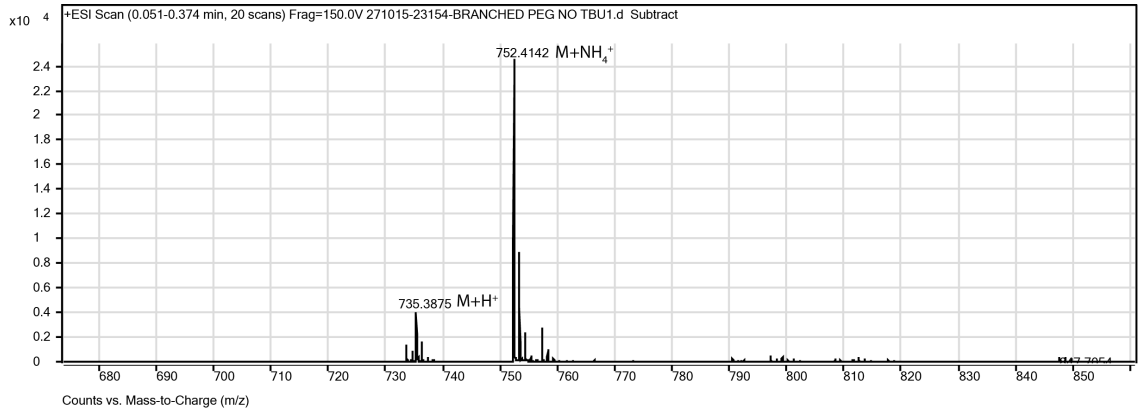


Figure 3-62. ESI-TOF MS data for Fmoc-L-Asn(PEG)$_2$-OH. Calculated m/z for $C_{37}H_{58}N_3O_{13}$ ($M+NH_4^+$) is 752.40, found 752.41

3.4.23 Analysis of WW Variant Folding Thermodynamics

3.4.24 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 (ε_{Trp} = 5690 $M^{-1}cm^{-1}$, ε_{Tyr} = 1280 $M^{-1}cm^{-1}$).[30] CD spectra of 50 μM solutions were obtained from 340 to 200 nm at 25ºC and 1°C. Variable temperature CD data were obtained in triplicate for 50 μM solutions of: **1-nbp, 1-cbp**, **2**, **2p-1**, **2p-2**, **2p-3**, **2p-4**, **3**, **3p**, **4**, **4p**, **6**, **6p**, **m1**, **m1p, 5,** and **5p,** 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.

3.4.25 Global Fitting of Variable Temperature CD Data

For **2**, **2p-4**, **4p**, and **5p** data from the three replicate variable temperature CD experiments on each protein were fit to the following model for two-state thermally induced unfolding transitions:

$$[\theta] = \frac{(D_0 + D_1 \cdot T) + K_f(N_0 + N_1 \cdot T)}{1 + K_f}, \text{ (S1)}$$

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. K_f is related to the temperature-dependent free energy of folding ΔG_f (T) according to the following equation:

$$K_f = \exp\left[\frac{-\Delta G_f(T)}{RT}\right], \text{ (S2)}$$

where R is the universal gas constant (0.0019872 kcal/mol/K). ΔG_f(T) was fit to the following equation:

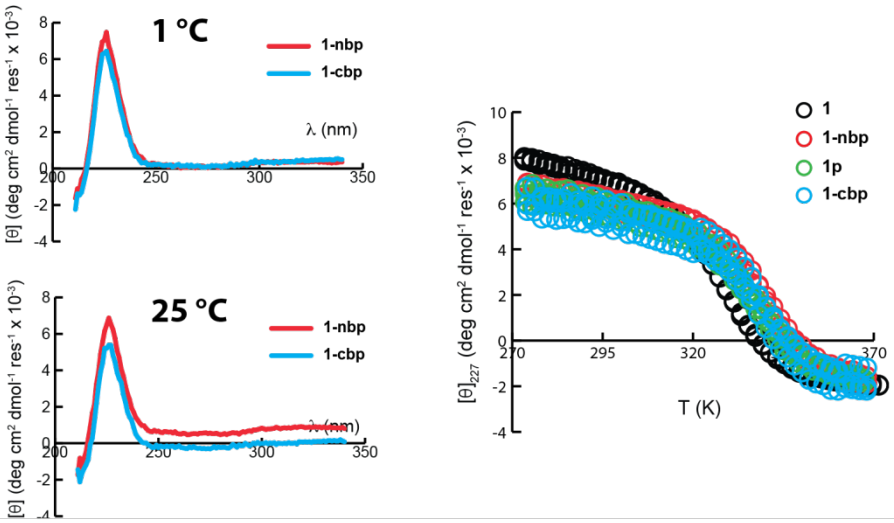
$$\Delta G_f = \frac{\Delta H(T_m)\cdot(T_m - T)}{T_m} + \Delta C_p \cdot (T - T_m - T \cdot \ln\left[\frac{T}{T_m}\right]) \text{ (S3)}$$

where the fit parameters are T_m (the midpoint of the unfolding transition; the temperature at which $\Delta G_f = 0$); $\Delta H(T_m)$, the change in enthalpy upon folding at T_m; and ΔC_p, the change in heat capacity upon folding. The parameters for equations S1-S3 were used to calculate the values of the folding free energy ΔG_f for WW variants in the main text and in tables below.

In preliminary fitting efforts of the data for **3**, **4**, **6**, **2p-1**, **2p-2**, **2p-3**, **3p**, **m1**, **m1p**, **5**, **5p**, **1-cbp**, and **1-nbp** to equation S3, we found that the fits were over parameterized, with high p-values for the ΔCp (heat capacity) and D_1 (slope of the post-transition baseline) terms. Therefore,

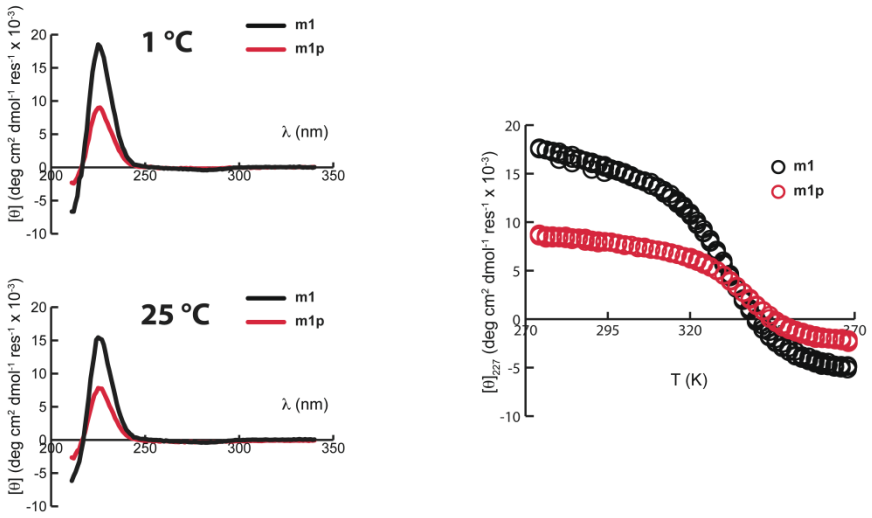
we eliminated the heat capacity term for variants **3**, **4**, and **6** while eliminating both the heat capacity and slope of the post-transition baseline terms for variants **2p-1**, **2p-2**, **2p-3**, **3p**, **m1**, **m1p**, **5**, **5p**, **1-cbp**, and **1-nbp**.

3.4.26 Plots of CD Spectra, Variable Temperature CD Data



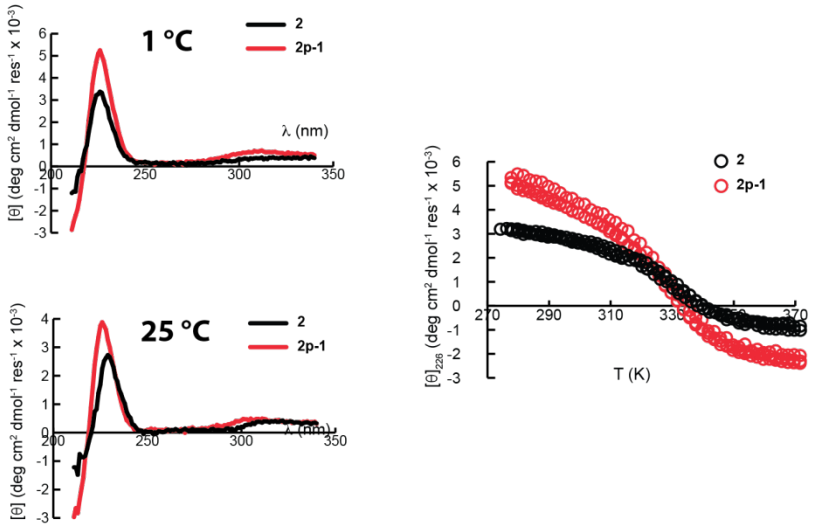
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
1	329.0 ± 0.2	-29.5 ± 0.3	-0.51 ± 0.05	0.9986	0.030
1p	337.0 0.1	-33.0 0.3	-0.72 0.01	0.9988	0.024
1-cbp	340.4 ± 0.3	-34.9 ± 1.3	±	0.9968	0.169
1-nbp	341.2 ± 0.1	-34.0 ± 0.5	±	0.9995	0.071

Figure 3-63. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **1** (black), **1-nbp** (red), **1p** (green), and **1-cbp** (blue) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



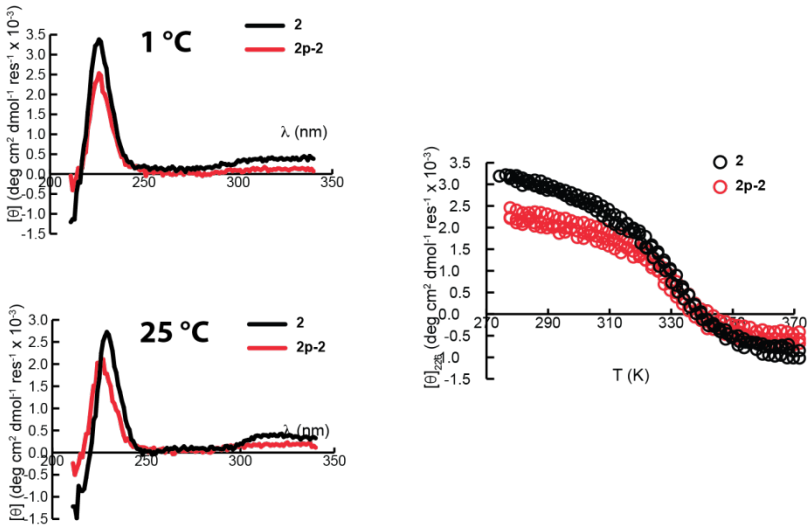
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
m1	333.9 ± 0.1	-28.7 ± 0.3	±	0.9998	0.123
m1p	336.8 ± 0.2	-34.1 ± 0.7	-0.73 - 0.04	0.9998	0.068

Figure 3-64. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **m1** (black) and **m1p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



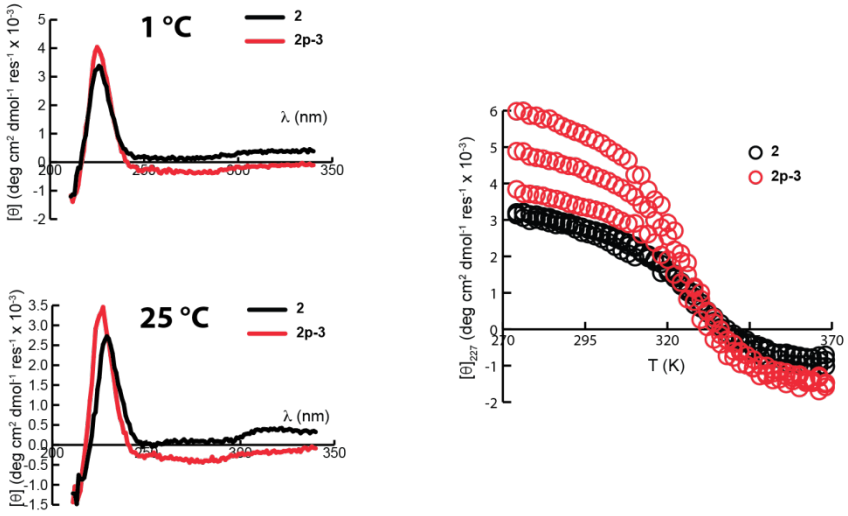
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
2	328.7 ± 0.4	-29.5 ± 1.3	-0.74 ± 0.06	0.9994	0.041
2p-1	329.9 ± 0.2	-26.2 ± 0.4	-	0.9996	0.058

Figure 3-65. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **2** (black) and **2p-1** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



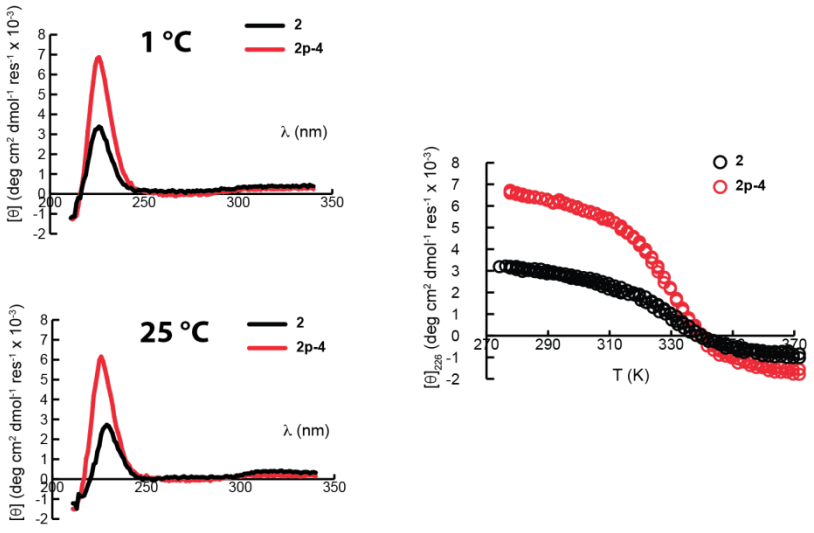
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
2	328.7 ± 0.4	-29.5 ± 1.3	-0.74 ± 0.06	0.9994	0.041
2p-2	329.5 ± 0.2	-30.3 ± 0.7	-	0.9989	0.038

Figure 3-66. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **2** (black) and **2p-2** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



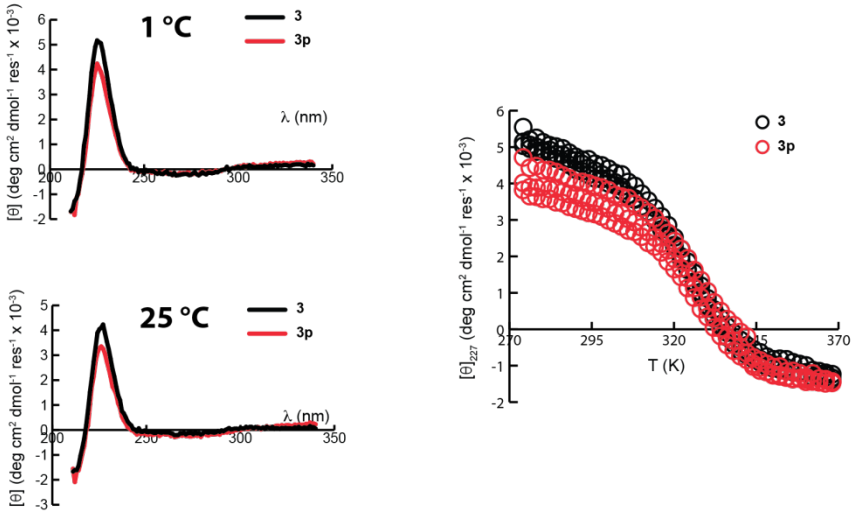
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
2	328.7 ± 0.4	-29.5 ± 1.3	-0.74 ± 0.06	0.9994	0.041
2p-3	328.0 ± 0.2	-27.5 ± 0.5	-	0.9992	0.072

Figure 3-67. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **2** (black) and **2p-3** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



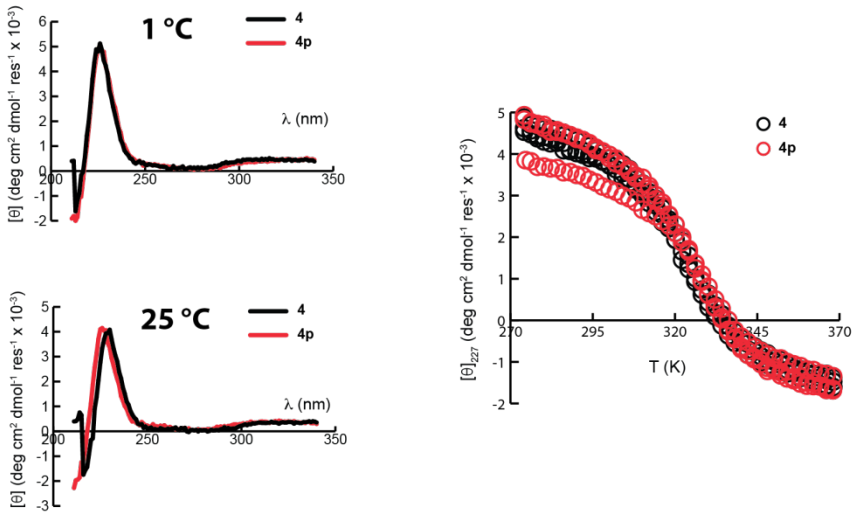
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
2	328.7 ± 0.4	-29.5 ± 1.3	-0.74 ± 0.06	0.9994	0.041
2p-4	326.2 ± 0.4	-30.0 ± 0.5	-0.41 - 0.13	0.9998	0.044

Figure 3-68. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **2** (black) and **2p-4** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



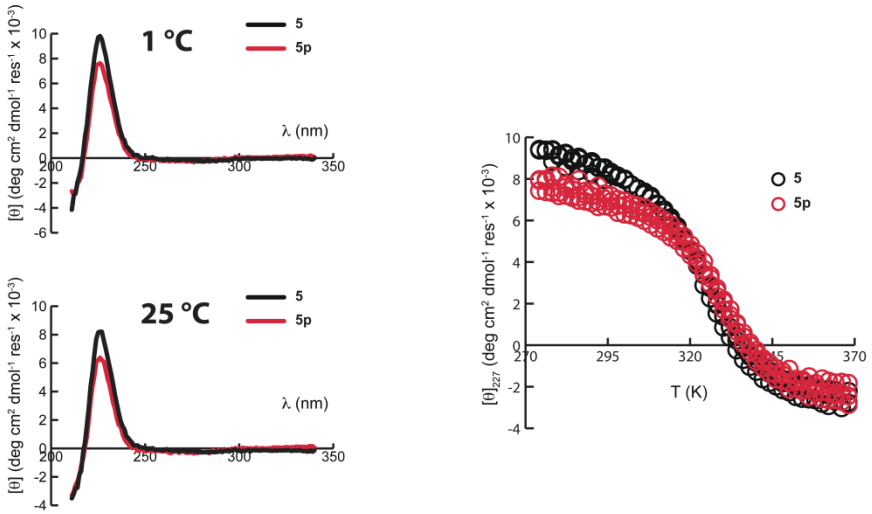
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
3	327.1 ± 0.2	-29.6 ± 0.6	±	0.9997	0.041
3p	330.8 ± 0.2	-26.1 ± 0.5	-	0.9992	0.063

Figure 3-69. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **3** (black) and **3p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors



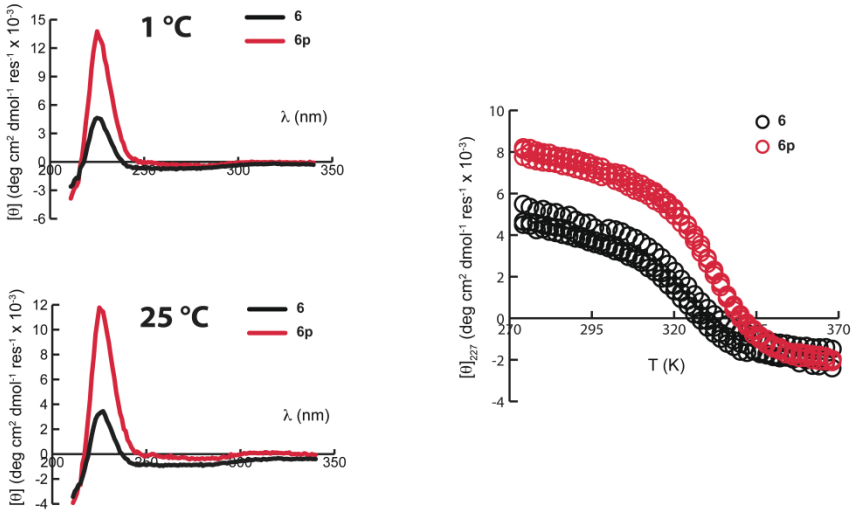
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
4	325.4 ± 0.3	-28.8 ± 0.8	±	0.9996	0.049
4p	326.9 ± 0.2	-31.0 ± 0.9	-0.79 - 0.05	0.9997	0.043

Figure 3-70. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **4** (black) and **4p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
5	327.4 ± 0.2	-27.4 ± 0.4	0.00 ± 0.00	0.9994	0.115
5p	331.9 ± 0.2	-29.1 ± 0.7	0.00 ± 0.00	0.9988	0.139

Figure 3-71. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **5** (black) and **5p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.



Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
6	325.8 ± 0.3	-32.1 ± 1.2	±	0.9991	0.082
6p	331.7 ± 0.1	-31.8 ± 0.5	-0.67 ± 0.04	0.9999	0.049

Figure 3-72. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **6** (black) and **6p** (red) in 20 mM sodium phosphate, pH 7. Fit parameters from equations S1–S3 appear in the table, along with standard errors.

3.5 References

(1) 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. *Bioconjugate Chem.* **2013**, *24*, 796.

(2) Dobson, C. M. *Nature* **2003**, *426*, 884.

(3) Parsell, D. A.; Sauer, R. T. *J. Biol. Chem.* **1989**, *264*, 7590.

(4) Ohkuri, T.; Nagatomo, S.; Oda, K.; So, T.; Imoto, T.; Ueda, T. *J. Immunol.* **2010**, *185*, 4199.

(5) Lawrence, P. B.; Gavrilov, Y.; Matthews, S. S.; Langlois, M. I.; Shental-Bechor, D.; Greenblatt, H. M.; Pandey, B. K.; Smith, M. S.; Paxman, R.; Torgerson, C. D.; Merrell, J. P.; Ritz, C. C.; Prigozhin, M. B.; Levy, Y.; Price, J. L. *J. Am. Chem. Soc.* **2014**, *136*, 17547.

(6) Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776.

(7) Nischan, N.; Hackenberger, C. P. R. *J. Org. Chem.* **2014**, *79*, 10727.

(8) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem. Int. Ed.* **2009**, *48*, 6974.

(9) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. Asian J.* **2009**, *4*, 630.

(10) Ngo, J. T.; Tirrell, D. A. *Acc. Chem. Res.* **2011**, *44*, 677.

(11) Bundy, B. C.; Swartz, J. R. *Bioconjugate Chem.* **2010**, *21*, 255.

(12) Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Curr. Opin. Chem. Biol.* **2010**, *14*, 774.

(13) Kim, C. H.; Axup, J. Y.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2013**, *17*, 412.

(14) Tornoe, C. W.; Christensen, C.; Meldal, M. *J. Org. Chem.* **2002**, *67*, 3057.

(15) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596.

(16) Deiters, A.; Cropp, T. A.; Summerer, D.; Mukherji, M.; Schultz, P. G. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5743.

(17) Fontana, A.; Spolaore, B.; Mero, A.; Veronese, F. M. *Adv. Drug Deliv. Rev.* **2008**, *60*, 13.

(18) Horne, W. S.; Yadav, M. K.; Stout, C. D.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2004**, *126*, 15366.

(19) Deechongkit, S.; Dawson, P. E.; Kelly, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 16762.

(20) Gao, J.; Kelly, J. W. *Protein Sci.* **2008**, *17*, 1096.

(21) 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. *Bioconjugate Chem.* **2013**, *24*, 796.

(22) Braun, A.; Kwee, L.; Labow, M. A.; Alsenz, J. *Pharmaceut. Res.* **1997**, *14*, 1472.

(23) Wang, W.; Roberts, C. J. *Aggregation of Therapeutic Proteins*; Wiley: Hoboken, NJ, USA, 2010.

(24) Moore, W. V.; Leppert, P. *J. Clin. Endocrinol. Metab.* **1980**, *51*, 691.

(25) Garofalo, A.; Parat, A.; Bordeianu, C.; Ghobril, C.; Kueny-Stotz, M.; Walter, A.; Jouhannaud, J.; Begin-Colin, S.; Felder-Flesch, D. *New J. Chem.* **2014**, *38*, 5226.

(26) Chooi, K. P.; Galan, S. R. G.; Raj, R.; McCullagh, J.; Mohammed, S.; Jones, L. H.; Davis, B. G. *J. Am. Chem. Soc.* **2014**, *136*, 1698.

(27) Riggs-Sauthier, J.; Zhang, W.; Viegas, T. X.; Bentley, M. D.; Nektar Therapeutics Al, Corporation, USA . 2008, p 74pp.

(28) Hansen, M.; Schnitzler, T.; Pisula, W.; Graf, R.; Müllen, K.; Spiess, H. *Angew. Chem. Int. Ed.* **2009**, *48*, 4621.

(29) Robbins, D. C.; Cooper, S. M.; Fineberg, S. E.; Mead, P. M. *Diabetes* **1987**, *36*, 838.

(30) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948.

4 LINKER SCAN

4.1 Introduction

In chapter 2 we developed a predictive structure-based method for identifying stabilizing PEGylation sites within two model proteins: the WW domain of the human protein Pin 1 and the Src SH3 domain. We generated the required PEG-protein conjugates for these studies by attaching a short PEG oligomer to the side-chain amide nitrogen of an Fmoc-protected Asn residue, which was incorporated at desired positions within WW or SH3 via solid-phase peptide synthesis.[1] However, Asn-PEG is not genetically encodable, which presents a major challenge for incorporating it site-specifically into larger and more therapeutically relevant proteins. Chapter 3 explores two genetically encodable alternatives to Asn-PEGylation in the context of the WW domain: PEGylation of a propargyloxyphenylalanine residue with a PEG-azide (ie., PrF-PEGylation); and PEGylation of an azidohomoalanine residue with a PEG-alkyne (ie., Aha-PEGylation). We found that neither PrF- nor Aha-PEGylation can stabilize WW as extensively as does Asn-PEGylation at position 19. We wondered whether PrF- or Aha-PEGylation would be similarly unable to recapitulate the impact of Asn-PEGylation at the other positions within WW that we studied in Chapter 2. Answering this question is key to understanding whether we can use our structure-based guidelines for identifying stabilizing Asn-PEGylation sites to identify stabilizing PrF- or Aha-PEGylation sites.

4.2 Results and Discussion

To address these questions, we generated proteins in which wild type residues at positions 16, 18, 19, 23, 27, 29, and 32 were replaced with PrF or Aha to give proteins **16PrF**, **18PrF**, **19PrF**, **23PrF**, **27PrF**, **29PrF**, **32PrF**, **16Aha**, **18Aha**, **19Aha**, **23Aha**, **27Aha**, **29Aha**, and **32Aha**. We then modified the PrF or Aha side-chains with a PEG-azide or PEG-alkyne, respectively, via copper-catalyzed azide-alkyne cycloaddition (CuAAC) to give PEGylated proteins **16PrFp**, **18PrFp**, **19PrFp**, **23PrFp**, **27PrFp**, **29PrFp**, **32PrFp**, **16AhaP**, **18AhaP**, **19AhaP**, **23AhaP**, **27AhaP**, **29AhaP**, and **32AhaP**. We used variable temperature circular dichroism (CD) to assess the stability of these PEGylated variants relative to their non-PEGylated counterparts

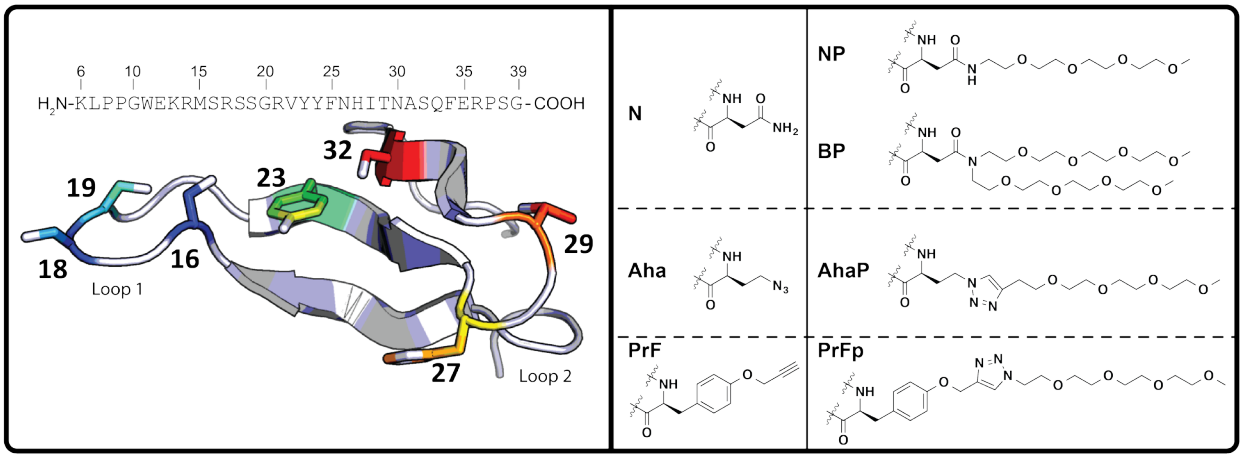


Figure 4-1. Sequence and ribbon diagram of WW (PBD ID: 1PIN). Positions where we incorporated N, NP, BP, Aha, AhaP, PrF, and PrFp are indicated.

The results of this analysis appear in figure 4-3 and table 4-1. Previous work indicates that Asn-PEGylation increases conformational stability at positions 16, 19, 29, and 32; has no effect at position 18, and is destabilizing at positions 23 and 27. Similarly, Aha-PEGylation moderately increases conformational stability at positions 16 and 19, and has minimal

stabilization at position 18, and is destabilizing at positions 23 and 27. However, unlike Asn-PEGylation, Aha-PEGylation has no effect at position 32. PrF-PEGylation has a substantially different impact on WW stability: it is moderately stabilizing at positions 23, 27, 29, and 32 while having no impact at positions 16, 18, and 19.

Table 4-1. Impact of the protein-PEG linkers on the WW thermodynamic stability at various sites.

Position	PrF vs. PrFp		Aha vs. AhaP		N vs. NP		N vs. BP	
	ΔT_m	$\Delta\Delta G_f$	ΔT_m	$\Delta\Delta G_f$	ΔT_m	$\Delta\Delta G_f$	ΔT_m	$\Delta\Delta G_f$
16	-0.7 ± 0.5	0.07 ± 0.05	4.1 ± 0.5	-0.39 ± 0.04	10.1 ± 0.3	-0.90 ± 0.03	19.7 ± 0.2	-1.70 ± 0.03
18	0.7 ± 0.4	-0.06 ± 0.04	1.7 ± 0.4	-0.16 ± 0.04	0.0 ± 0.7	0.00 ± 0.07	0.4 ± 0.3	-0.03 ± 0.02
19	1.4 ± 0.4	-0.04± 0.02	3.7 ± 0.3	-0.36 ± 0.01	7.9 ± 0.2	-0.74 ± 0.02	12.2 ± 0.2	-1.19 ± 0.02
23	3.5 ± 0.3	-0.29± 0.02	-2.9 ± 0.5	0.25 ± 0.05	-5.5 ± 1.3	0.40 ± 0.10	NA	NA
27	6.6 ± 0.2	-0.53± 0.02	-6.4 ± 0.3	0.51 ± 0.03	-4.0 ± 0.4	0.38 ± 0.04	-13.5 ± 0.4	1.27 ± 0.05
29	3.8 ± 0.5	-0.31± 0.04	-5.0 ± 0.6	0.39 ± 0.05	4.1 ± 0.4	-0.36 ± 0.04	8.0 ± 0.4	-0.67 ± 0.03
32	4.7 ± 0.4	-0.32± 0.03	-0.4 ± 0.06	0.03 ± 0.06	5.3 ± 0.3	-0.45 ± 0.02	9.2 ± 0.4	-0.81 ± 0.04

Observed data are given ± standard error at 50 μM protein concentration in 20 mM sodium phosphate buffer, pH 7 (except for proteins **16N, 16NP, 16BP, 18N, 18NP, 18BP, 19N, 19NP, 19BP, 23N, 23NP, 23BP, 32N, 32NP,** and **32BP**, which were characterized at 100 μM protein concentration). Observed values of $\Delta\Delta G_f$ were derived from variable-temperature CD experiments at the melting temperature of the corresponding non-PEGylated protein.

The data show that our structure-based method for selecting optimal Asn-PEGylation sites does not correctly predict all stabilizing and destabilizing sites for PrF- and Aha-PEGylation. For example: position 27 is destabilizing for both Asn- and Aha-PEGylation by 0.38 ± 0.04 kcal/mol and 0.51 ± 0.03 kcal/mol respectively while significantly stabilizing for PrF-PEGylation by -0.53 ± 0.02 kcal/mol. In contrast, position 16 is significantly stabilizing for Asn-PEGylation (-0.90 ± 0.043 kcal/mol) and moderately stabilizing for Aha-PEGylation (-0.39 ± 0.04 kcal/mol) while PrF-PEGylation had no effect (0.07 ± 0.05 kcal/mol).

We first wondered if PEG-based stabilization or destabilization is a result of how the unnatural amino acids PrF and Aha affects the melting temperature of the protein before the attachment of PEG. We wanted to know if stabilizing PEGylation sites are stabilizing because the non-PEGylated variant's melting temperature did not deviate significantly from the parent protein's melting temperature of 58.0 ± 0.07 °C, whereas destabilizing PEGylation sites are destabilizing because the non-PEGylated variant's melting temperature did deviate significantly from the parent protein's melting temperature. (see figure 4-2 and table 4-1). We find some non-PEGylated variants that have melting temperatures above 58 °C (**16PrF, 27Aha, 32PrF**, and **32Aha**) and others below 58 °C (**16Aha, 18PrF, 18Aha, 19PrF, 19Aha, 23Prf, 23Aha, 27PrF, 29PrF**, and **29Aha**). If the hypothesis is correct, then we should only see PEG-based stabilization for **16PrF, 27Aha, 32PrF**, and **32Aha**. Instead of finding PEG-based stabilization for all the variants with melting temperatures above 58 °C we find PEG-based stabilization for only **32PrFp** and PEG-based de-stabilization for proteins **16PrFp, 27AhaP**, and **32AhaP**. There is also PEG-based stabilization and de-stabilization for non-PEGylated variants with melting temperatures below 58 °C (see figure 4-2 and table 4-1). This analysis shows that there is no correlation between PEG-based stabilization and the stability of the non-PEGylated variant.

Interestingly, PrF- and Aha-PEGylation seems to have opposite effects at each position. In loop 1 (see figure 4-1) of the WW we notice that Aha-PEGylation increases the conformational stability while PrF-PEGylation has no effect. In loop two of the WW we see the opposite effect; Aha-PEGylation is now significantly destabilizing whereas PrF-PEGylation has a stabilizing effect. The data also show that Aha-PEGylation is destabilizing in beta strands (positions 23 and 32), whereas PrF-PEGylation is stabilizing.

Figure 4-2. Melting temperatures of the WW variants.

A possible explanation for these observations is that the PrF- and Aha-linkers are more sensitive to secondary structure than the Asn-linkage. Loop 1 is an unusual type II β-turn within a six-residue loop; it is possible that this structural context is ideal for Aha-PEG-based stabilization. In contrast, loop 2 is a type I β-turn that may have structural features that are optimal for PrF-PEGylation, but not for Aha-PEGylation. Asn-PEG is tolerated in both loops and seems to be less sensitive to secondary structure than PrF- and Aha-PEGylation; however, we do not know if branched Asn-PEG will behave similarly to mono Asn-PEG.

In chapter 3 we demonstrated that by adding an additional four-unit peg chain to the amide nitrogen of **19NP** increased WW stability by an additional -0.45 ± 0.02 kcal/mol. The resulting variant **19BP** (see figure 4-1) is -1.19 ± 0.02 kcal/mol more stable than non-PEGylated **19N**. We also wondered if branched Asn-PEG would stabilize WW at the same positions as non-branched PEG. To test this hypothesis, we used variable temperature circular dichroism (CD) to

assess the stability of the bis-PEGylated variants **16BP**, **18BP**, **19BP**, **23BP**, **27BP**, **29BP**, **32BP**, **16BP**, **18BP**, **19BP**, **23BP**, **27BP**, **29BP**, and **32BP** relative to their non-PEGylated counterparts (see chapter 2).

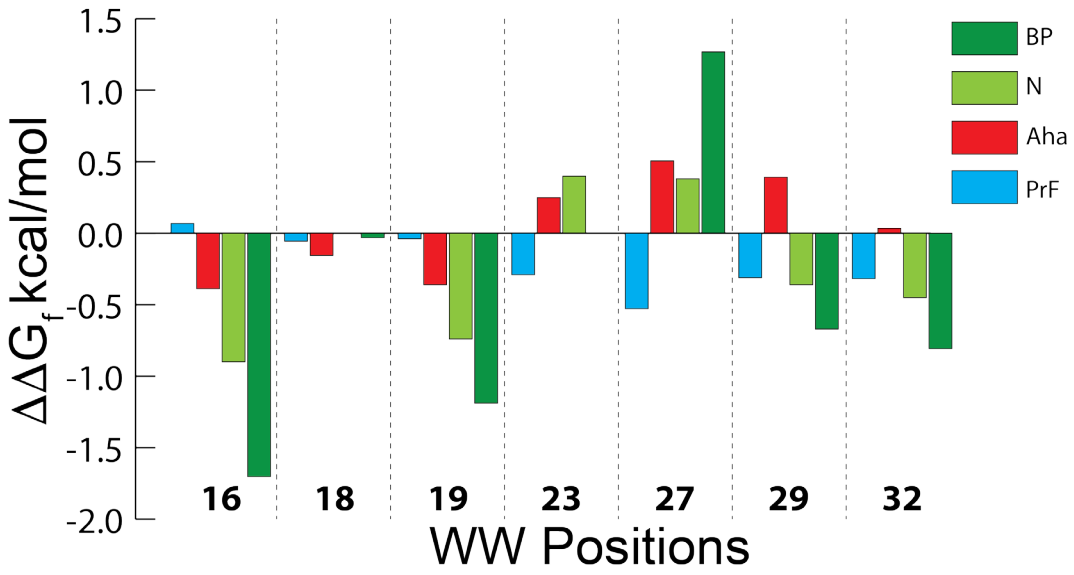


Figure 4-3. Graphical representation of the thermodynamic data found in table 1. $\Delta\Delta G_f$ data for protein **23BP** is not available due to the protein's instability.

The branched Asn-PEG is stabilizing at the same positions where non-branced Asn-PEG was stabilizing, but generally by larger amounts. The same is true for destabilizing positions. For example, protein **16BP** is -1.70 ± 0.03 kcal/mol more stable than non-PEGylated **16N** and -0.80 ± 0.04 kcal/mol more stable than mono-PEGylated **16NP**. Similarly, protein **27BP** is 1.27 ± 0.05 kcal/mol less stable than non-PEGylated **27N** and 0.89 ± 0.06 kcal/mol less stable than mono-PEGylated **27NP**.

4.3 Conclusion

The goal of this work was to find a genetically encodable linker that would follow the same rules that were developed previously for identifying stabilizing Asn-PEGylation sites. The two candidates were propargyloxyphenylalanine (PrF) and azidohomoalanine (Aha). We have shown that PrF- and Aha-PEGylation does not follow the same rules that govern Asn-PEGylation; however, Aha-PEGylation mimics the impact of Asn-PEGylation better than PrF-PEGylation at positions within the WW studied from chapter two. Aha-PEGylation may mimic the effects of Asn-PEGylation better than PrF-PEGylation because Aha-PEG and Asn-PEG are isosteres that share unique structural features that are different than PrF-PEG. Another explanation for why PrF- and Aha-PEGylation doesn't follow the same rules as Asn-PEGylation may be a result of PrF-PEG and Aha-PEG being more sensitive to local secondary structural contexts than Asn-PEG. It is interesting to note that most sites in the WW can become substantially stabilized via PEGylation sites when the correct linker is used. This work illustrates that the PEG linker is a critical determinant of PEG-based stabilization within WW and may explain why we see conflicting reports in the literature about the thermodynamic consequence of PEGylated proteins. The branched Asn-PEG scan did however follow the same rules as the mono Asn-PEG scan. Using the branched Asn-PEG linker does increase conformational stability significantly more than mono Asn-$PEG_{4.}$ In comparing **19BP** (eight ethylene oxide units) to a previously[2] synthesized variant **19NPEG$_{45}$** (45 eight ethylene oxide units) we find that the branched variant is more effective at increasing the conformational stability than a large linear PEG polymer. However, branched Asn-PEGylation still has limited application due to not being genetically encodable. Applying our predictive guidelines for branched Asn-PEGylation to larger therapeutic proteins may require a fragment condensation/ligation approach. Alternatively, using

our structure-based method for finding optimal PEGylation sites using PrF and Aha linkers only works when the structural context of the PEGylation site is accounted for.

4.4 Supporting Information

Protein synthesis, purification and characterization including HPLC, ESI-TOF MS, and CD spectropolarimetry were performed using identical condition to those reported in Chapter 3.

4.4.1 ESI-TOF

ESI-TOF spectra for proteins **16Aha, 16AhaP, 18Aha, 18AhaP, 23Aha, 23AhaP, 27Aha, 27AhaP, 29Aha, 29AhaP, 32Aha, 32AhaP, 16PrF, 16PrFp, 18PrF, 18PrFp, 23PrF, 23PrFp, 27PrF, 27PrFp, 29PrF, 29PrFp, 32PrF, 32PrFp, 16BP, 18BP, 23BP, 27BP, 29BP, and 32BP** are shown in Figure**s** 4-4 through 4-33.

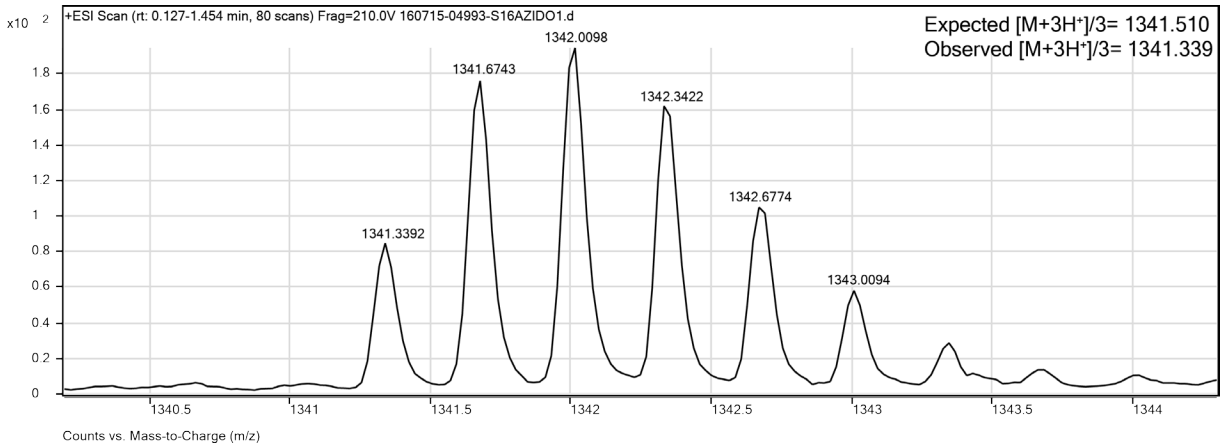


Figure 4-4. ESI-TOF spectrum for WW variant **16Aha**.

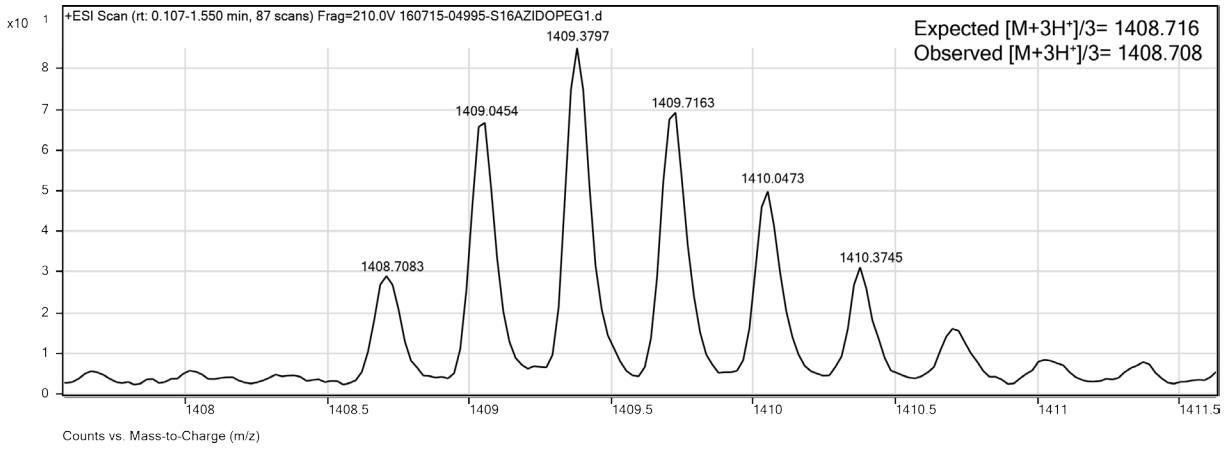


Figure 4-5. ESI-TOF spectrum for WW variant **16AhaP**

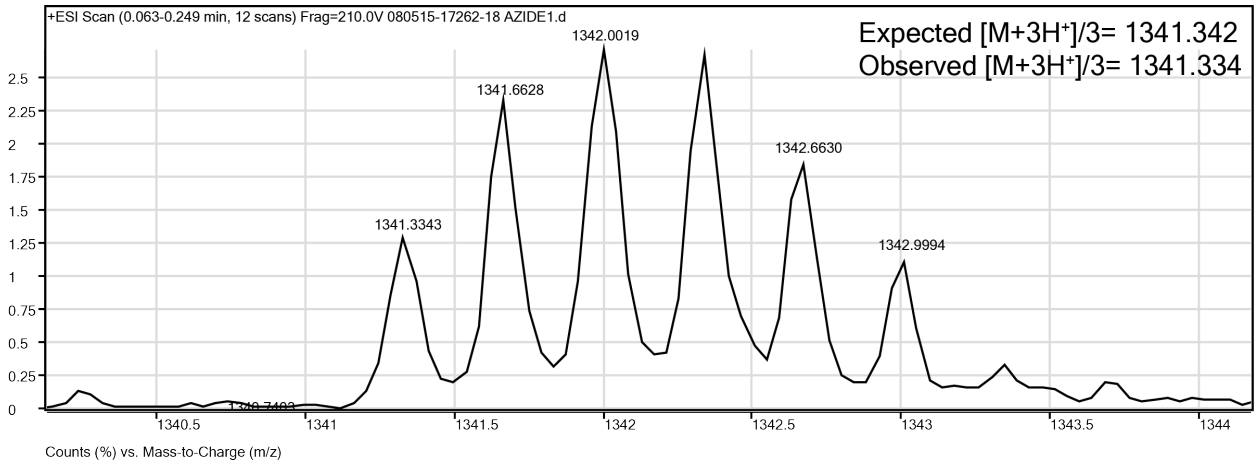


Figure 4-6. ESI-TOF spectrum for WW variant **18Aha**.

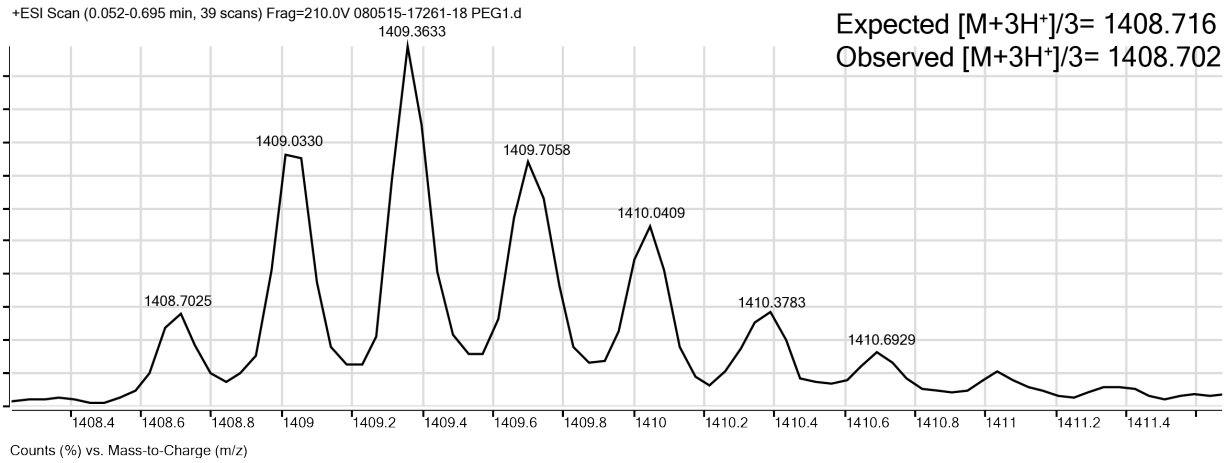


Figure 4-7. ESI-TOF spectrum for WW variant **18AhaP**

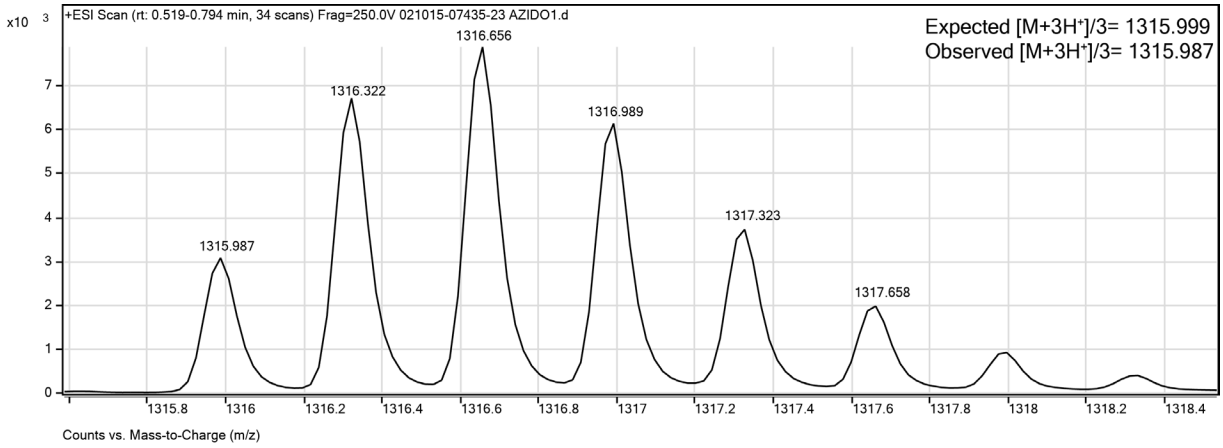


Figure 4-8. ESI-TOF spectrum for WW variant **23Aha**.

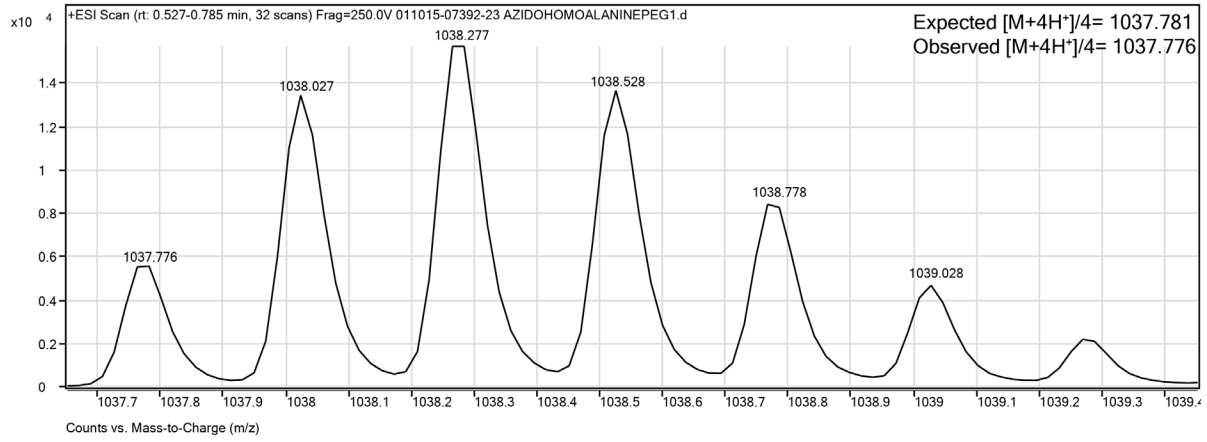


Figure 4-9. ESI-TOF spectrum for WW variant **23AhaP**.

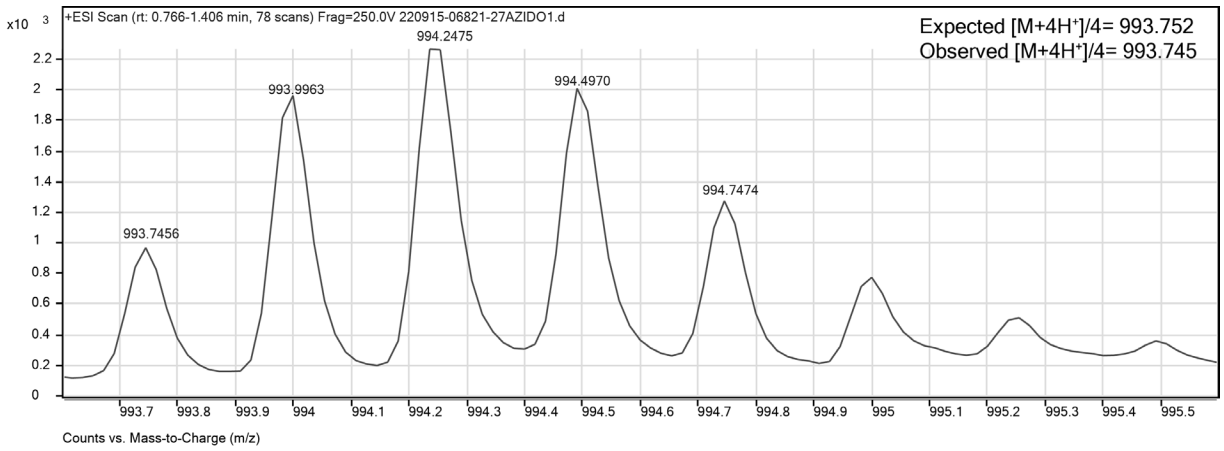


Figure 4-10. ESI-TOF spectrum for WW variant **27Aha**.

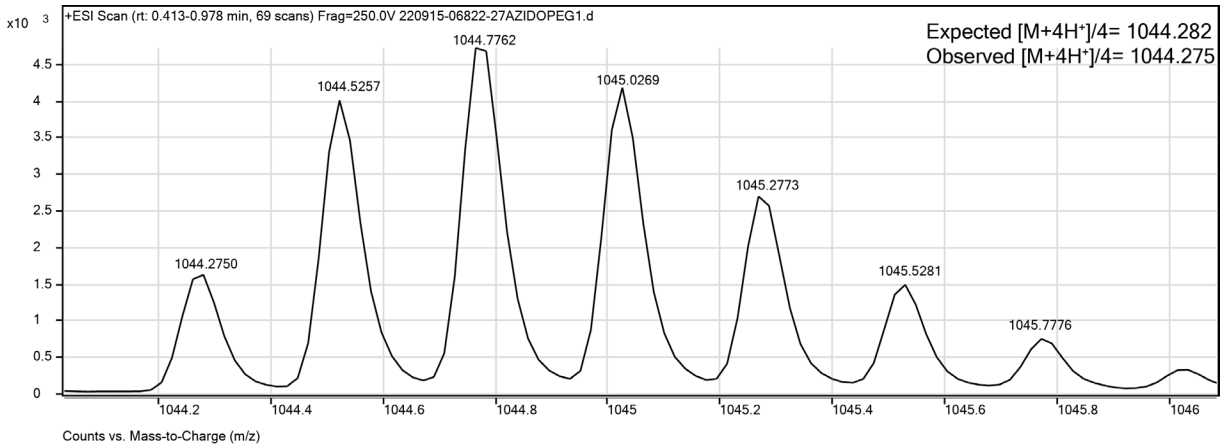


Figure 4-11. ESI-TOF spectrum for WW variant **27AhaP**.

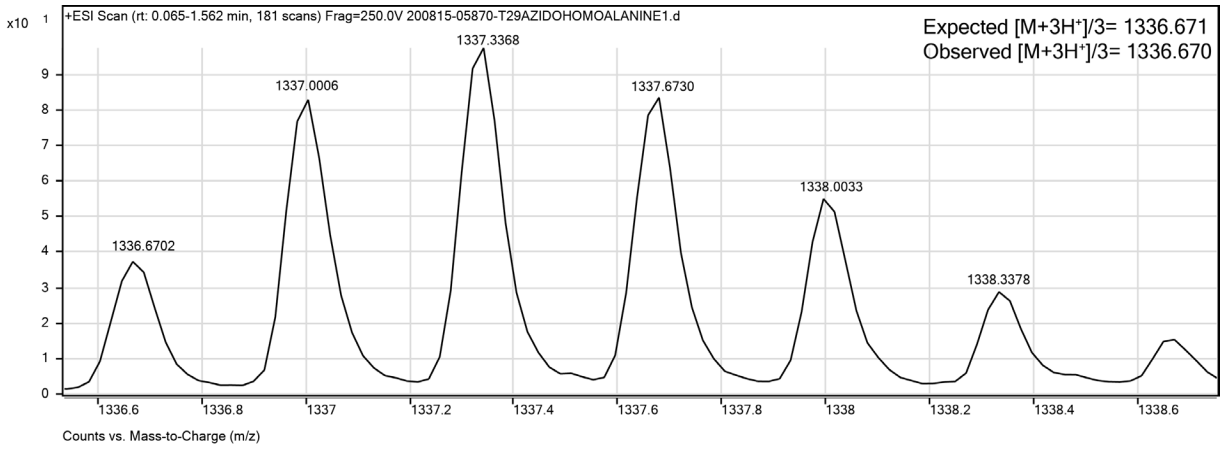


Figure 4-12. ESI-TOF spectrum for WW variant **29Aha**.

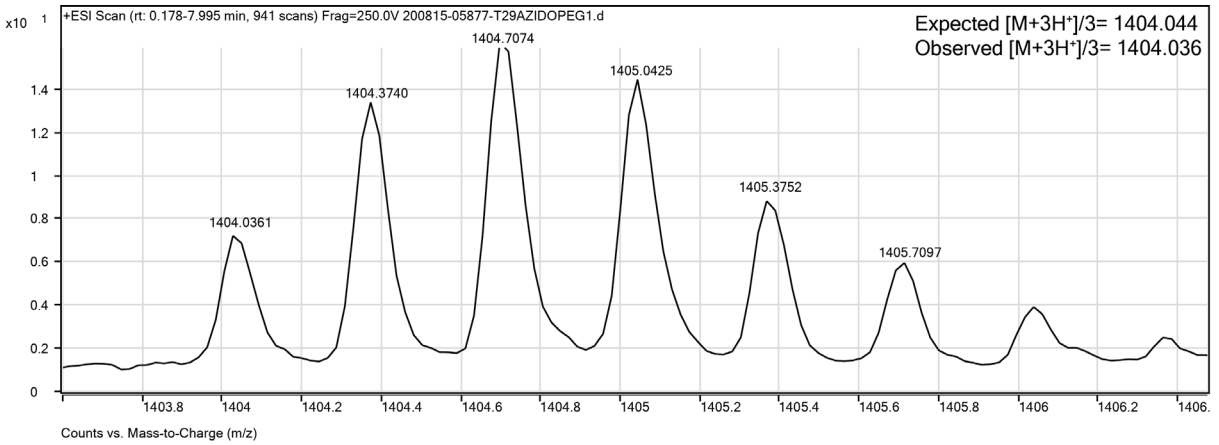


Figure 4-13. ESI-TOF spectrum for WW variant **29AhaP**.

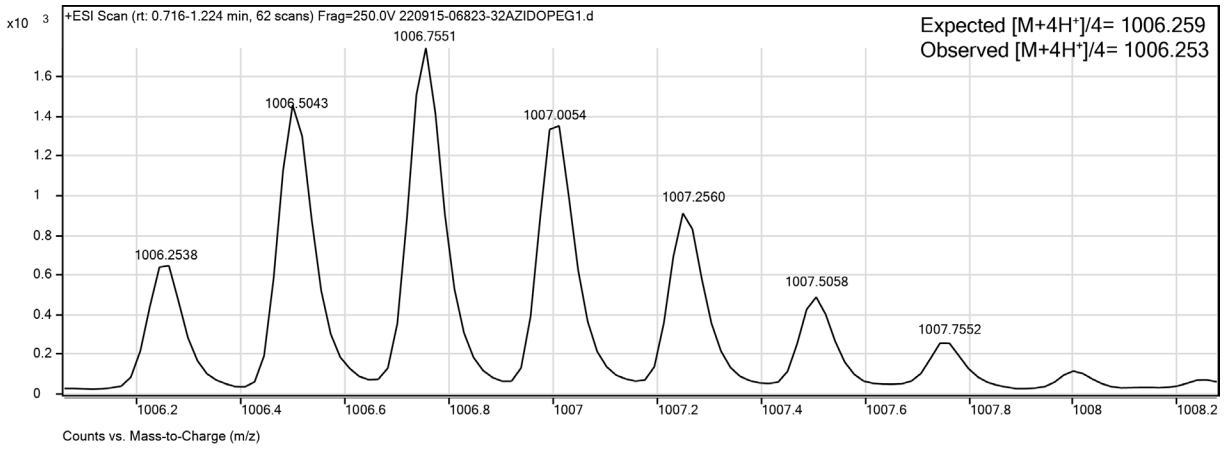


Figure 4-14. ESI-TOF spectrum for WW variant **32Aha**.

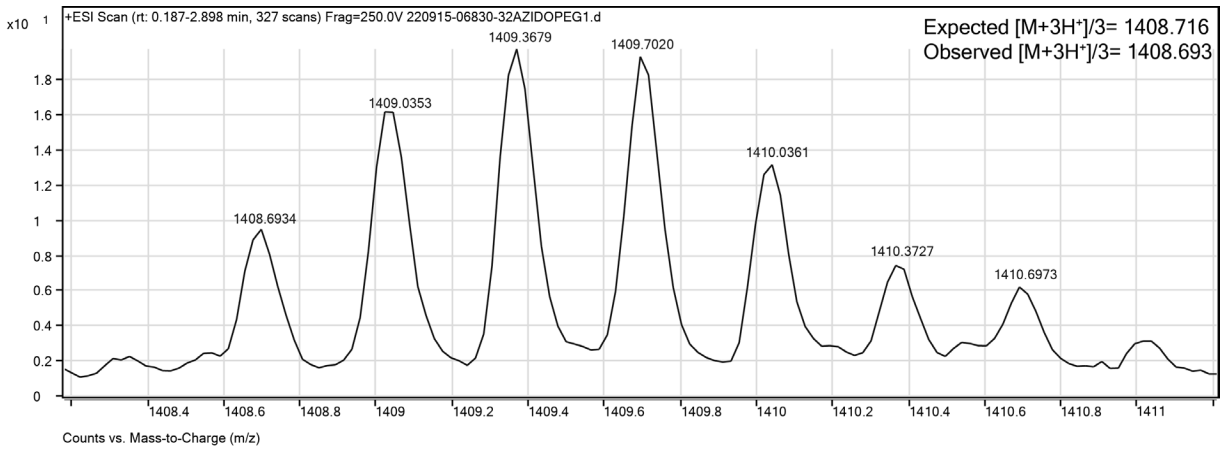


Figure 4-15. ESI-TOF spectrum for WW variant **32AhaP**.

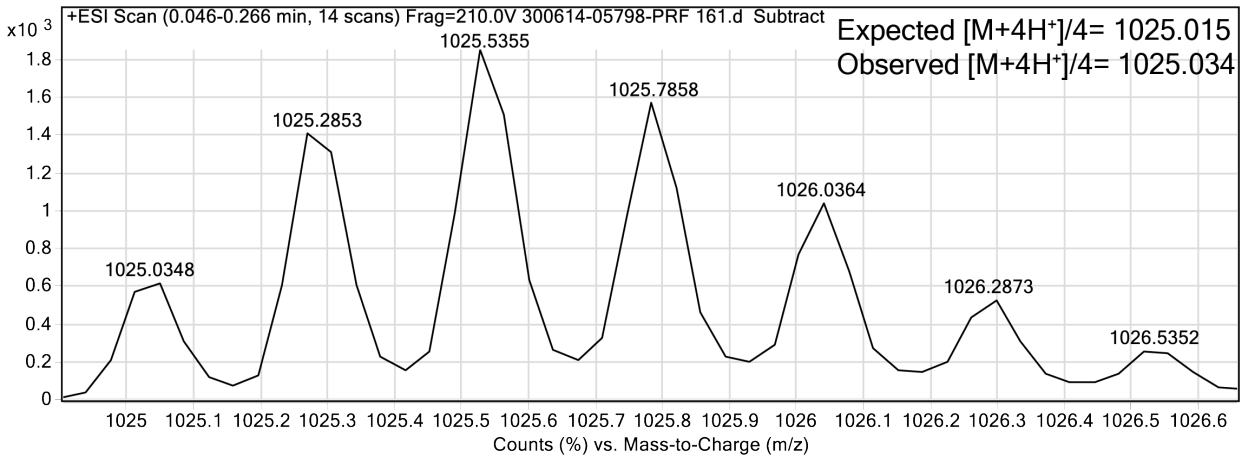


Figure 4-16. ESI-TOF spectrum for WW variant **16PrF**.

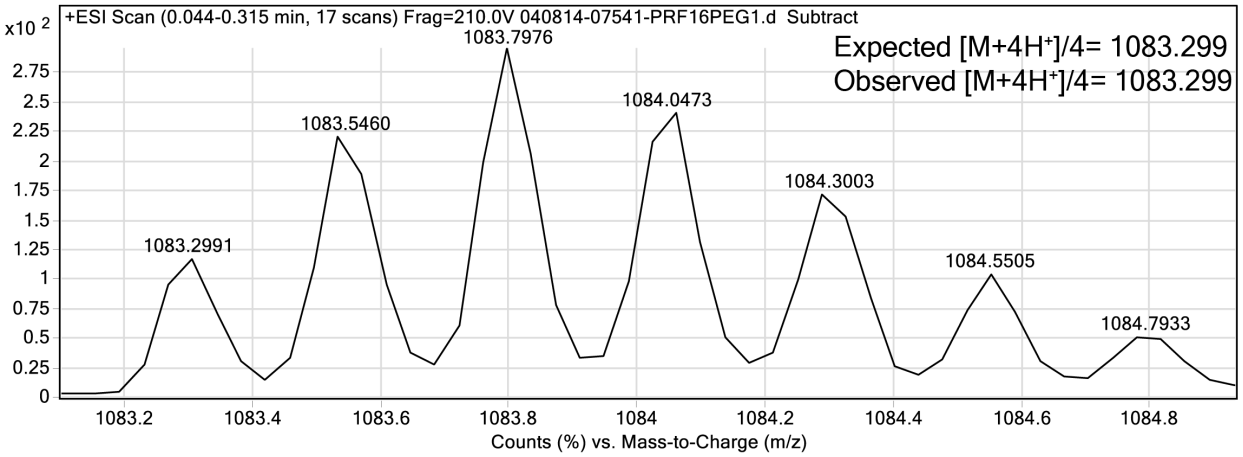


Figure 4-17. ESI-TOF spectrum for WW variant **16PrFp**.

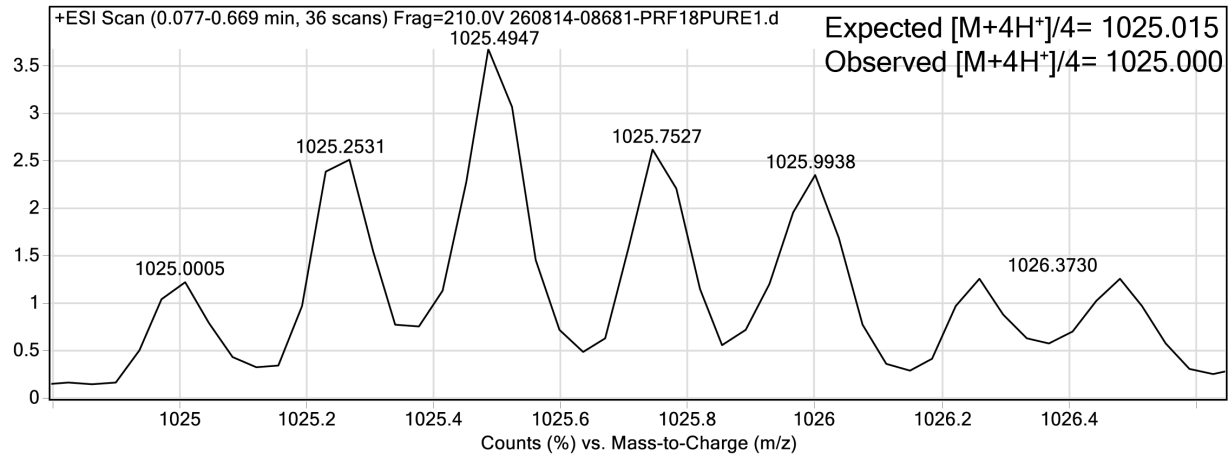


Figure 4-18. ESI-TOF spectrum for WW variant **18PrF**.

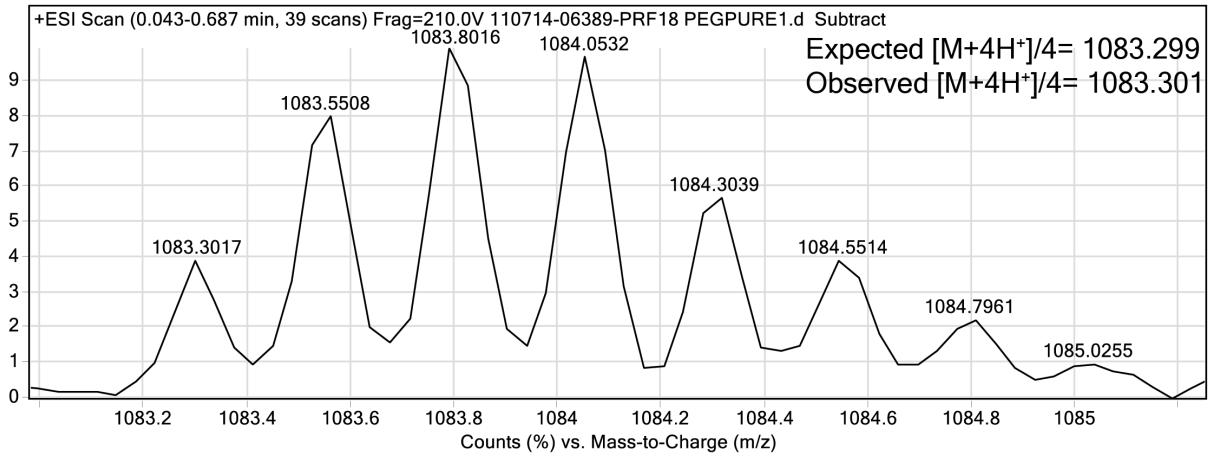


Figure 4-19. ESI-TOF spectrum for WW variant **18PrFp**.

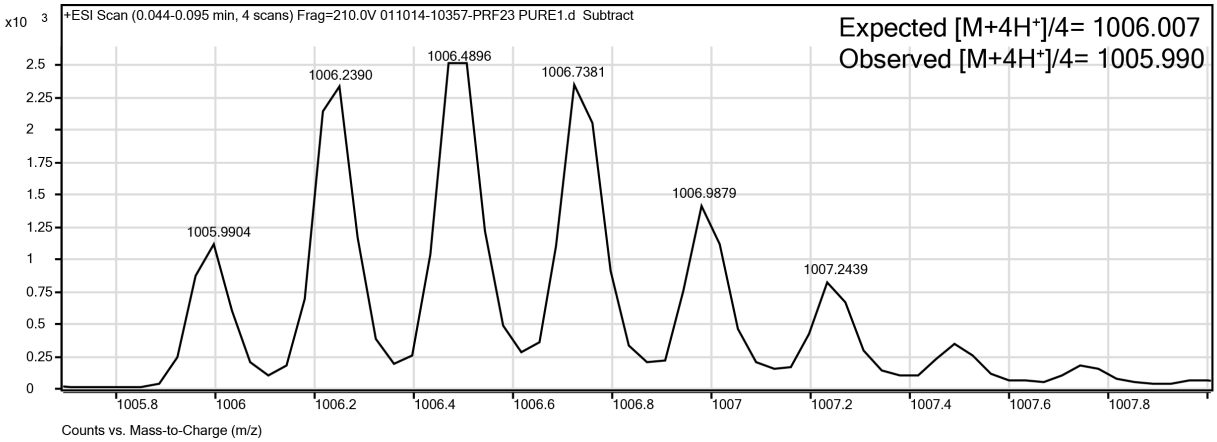


Figure 4-20. ESI-TOF spectrum for WW variant **23PrF**.

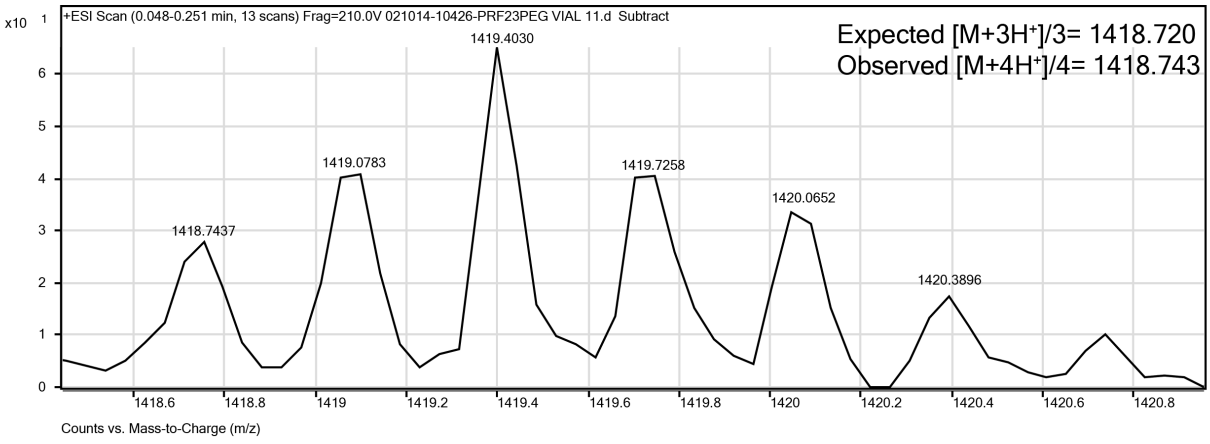


Figure 4-21. ESI-TOF spectrum for WW variant **23PrFp**.

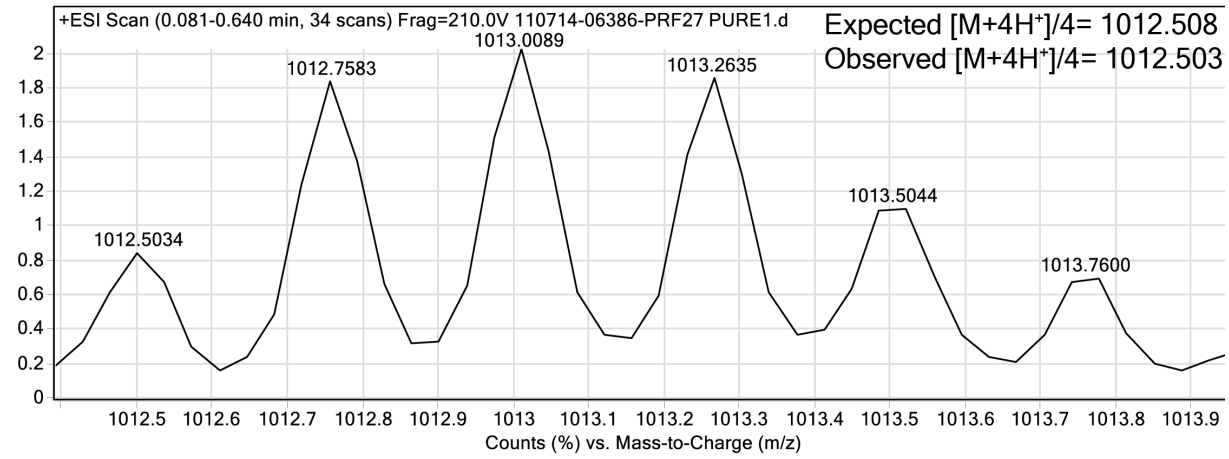


Figure 4-22. ESI-TOF spectrum for WW variant **27PrF**.

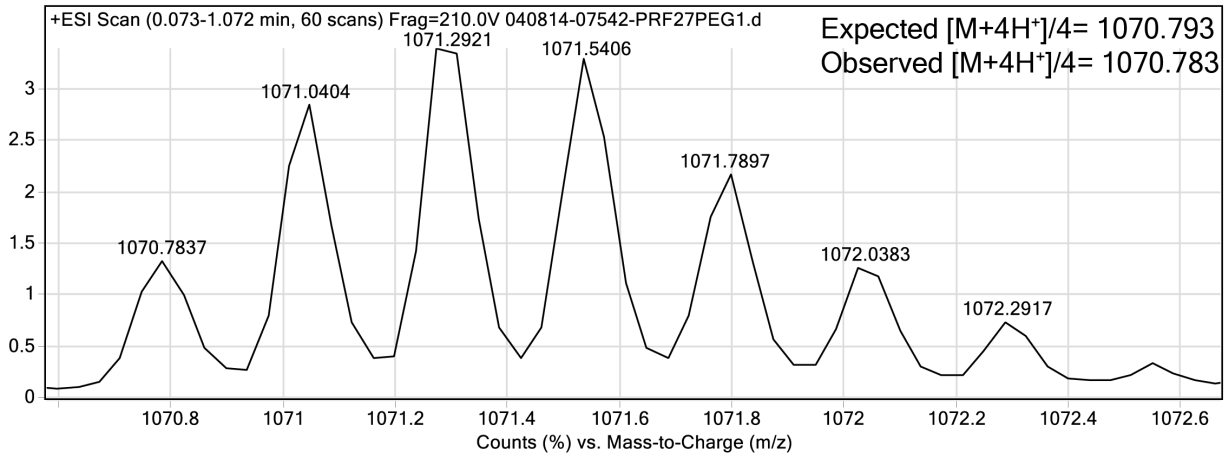


Figure 4-23. ESI-TOF spectrum for WW variant **27PrFp**.

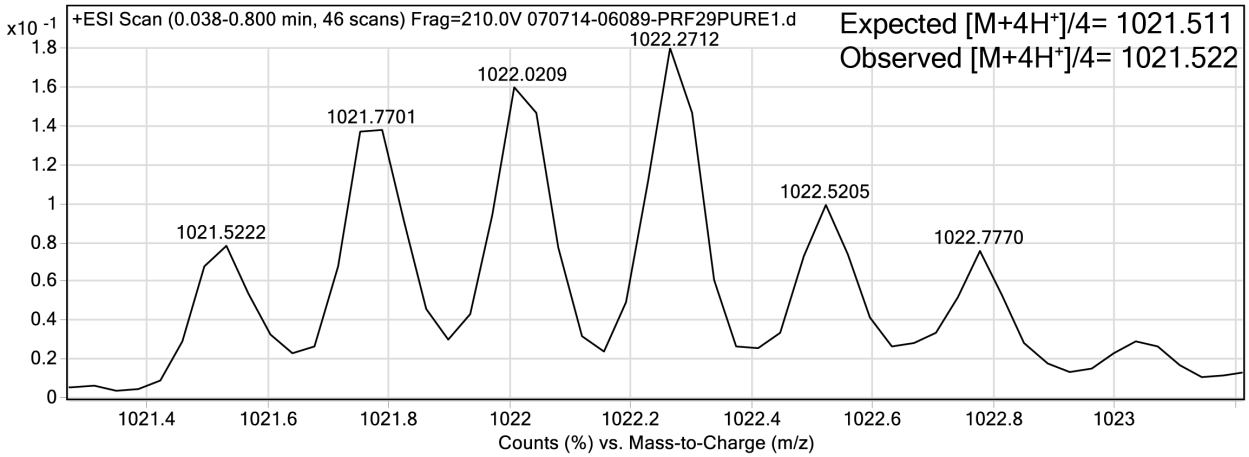


Figure 4-24. ESI-TOF spectrum for WW variant **29PrF**.

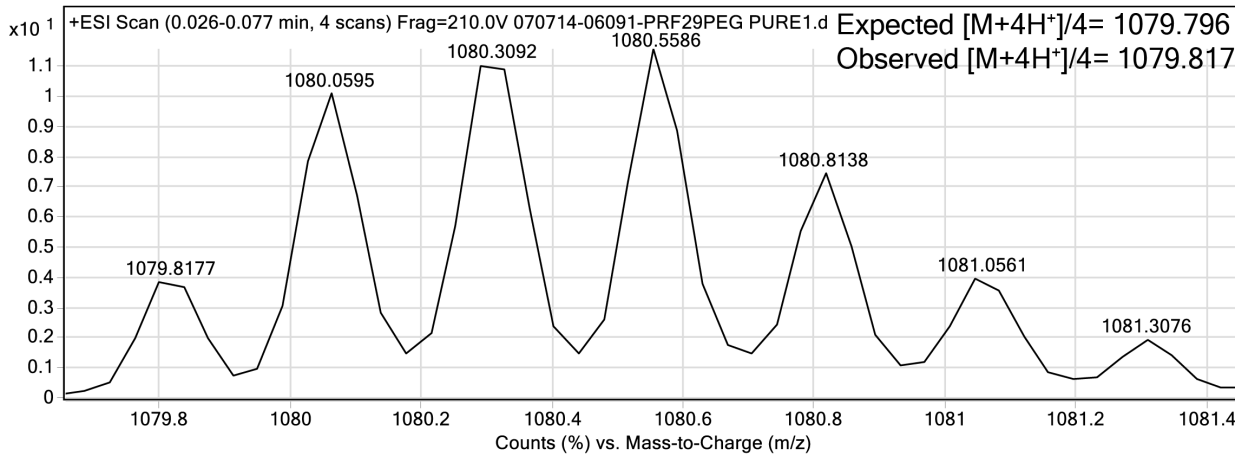


Figure 4-25. ESI-TOF spectrum for WW variant **29PrFp**.

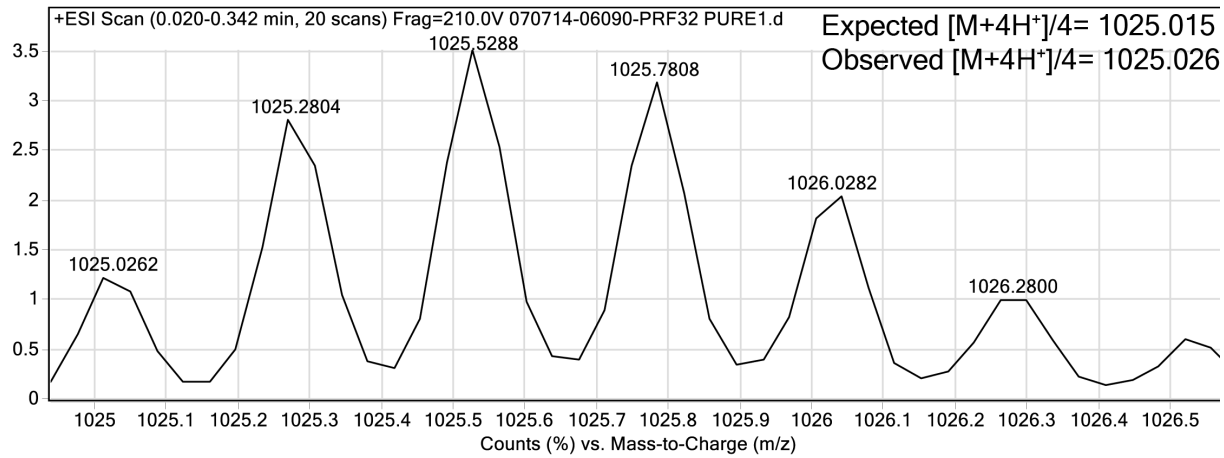


Figure 4-26. ESI-TOF spectrum for WW variant **32PrF**.

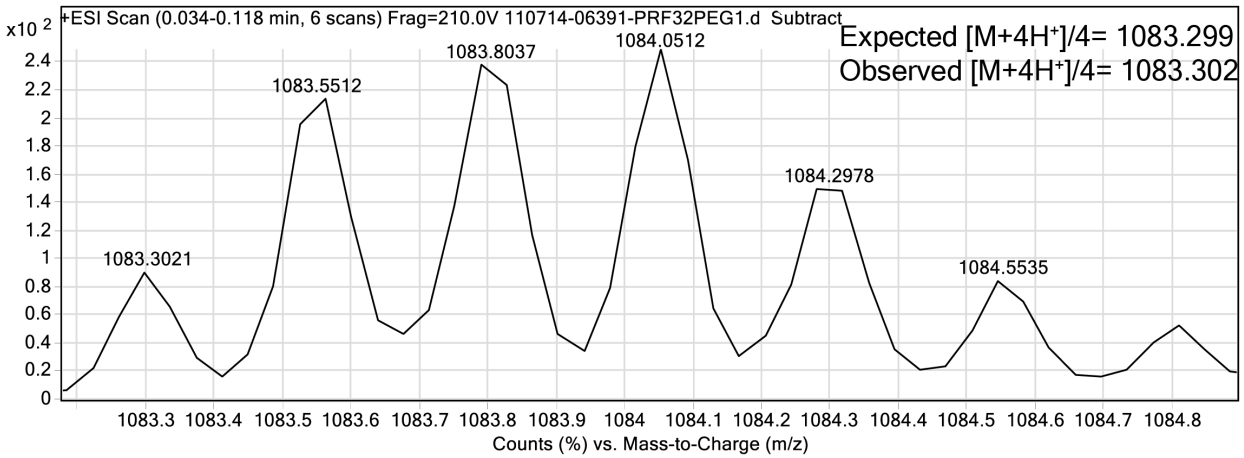


Figure 4-27. ESI-TOF spectrum for WW variant **32PrFp**.

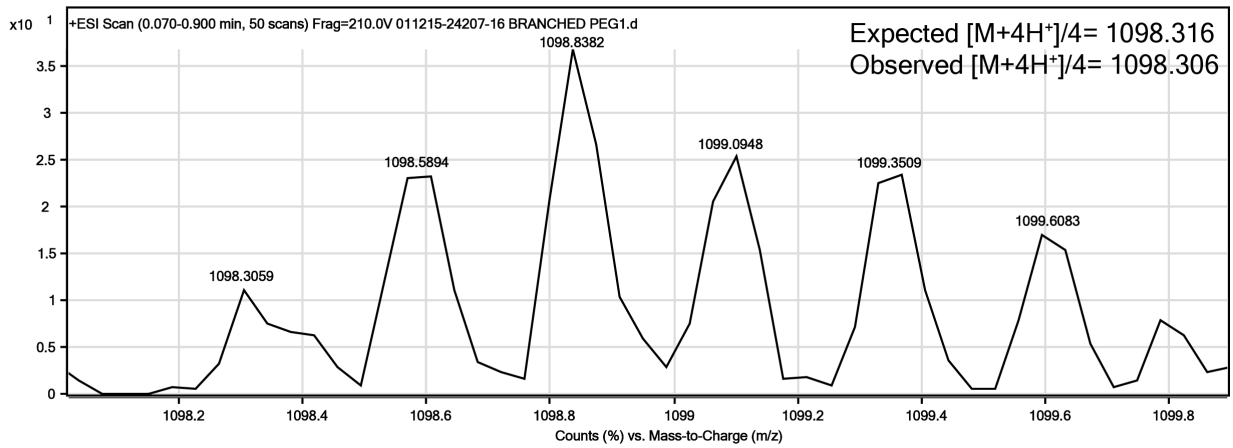


Figure 4-28. ESI-TOF spectrum for WW variant **16BP**.

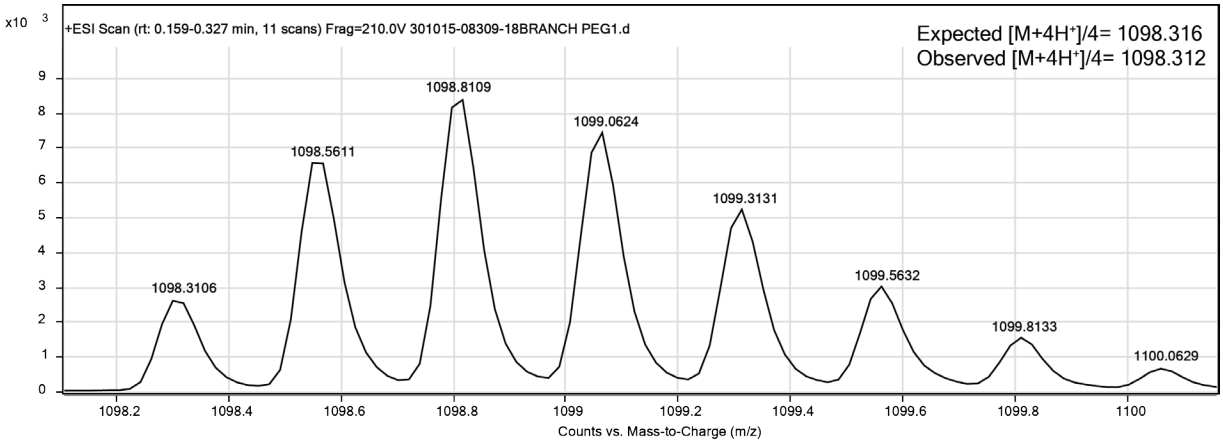


Figure 4-29. ESI-TOF spectrum for WW variant **18BP**.

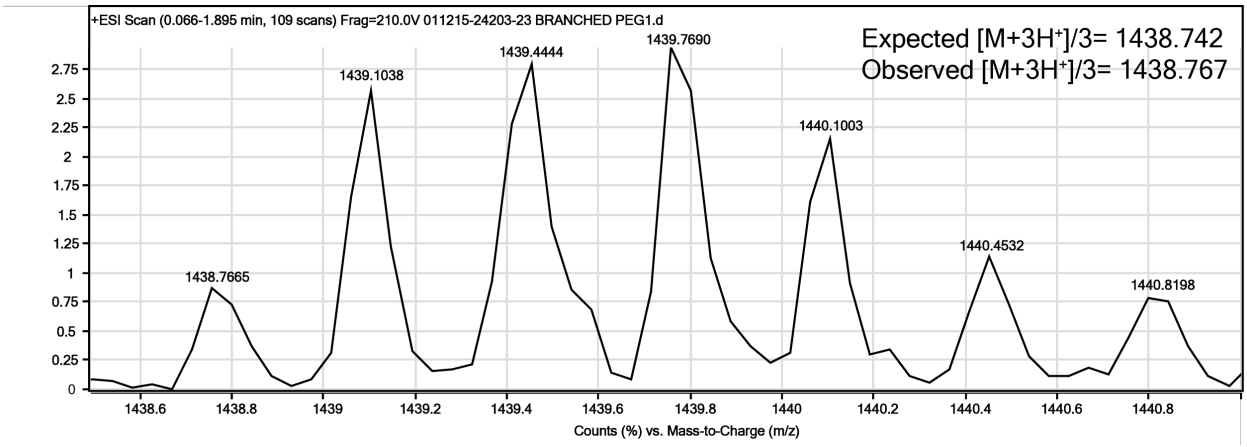


Figure 4-30. ESI-TOF spectrum for WW variant **23BP**.

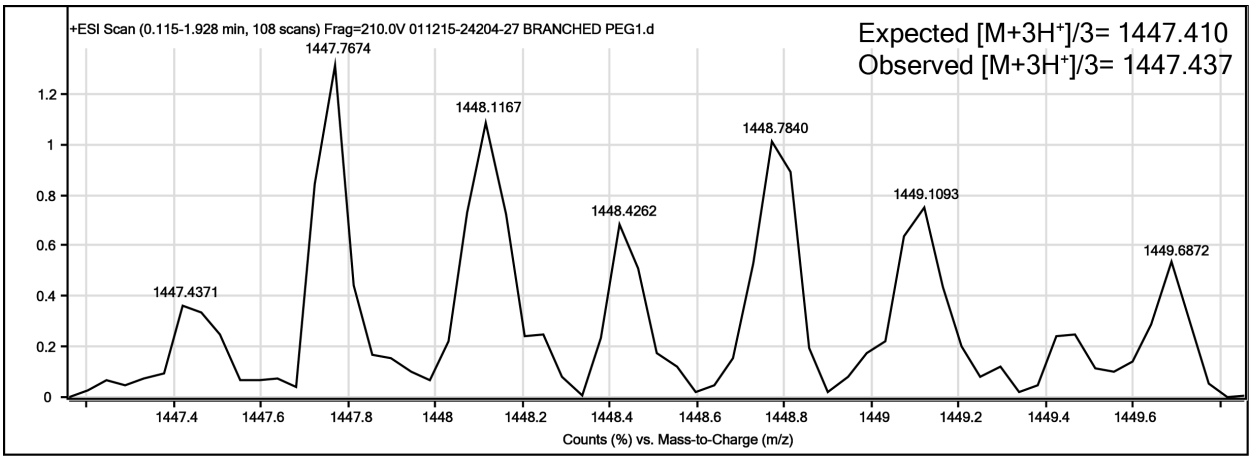


Figure 4-31. ESI-TOF spectrum for WW variant **27BP**.

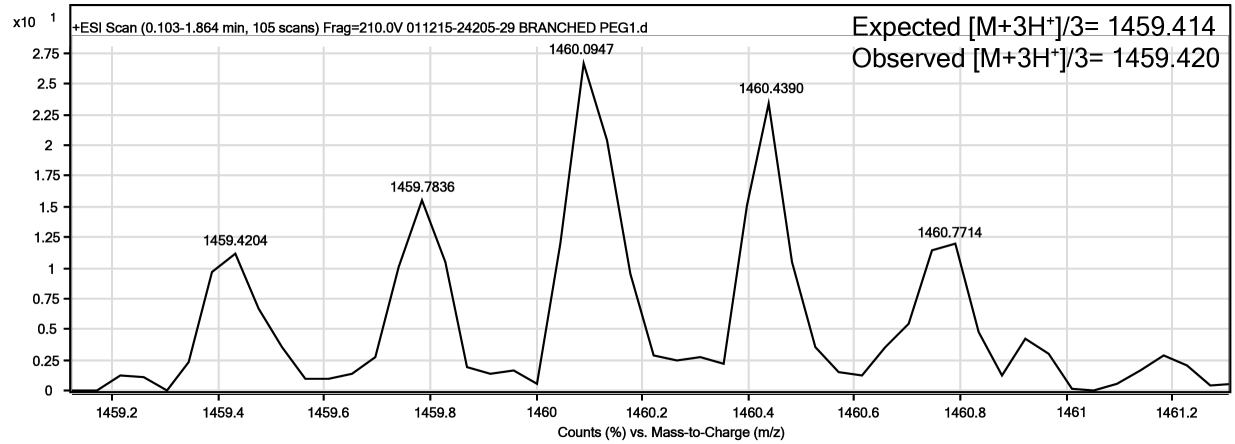


Figure 4-32. ESI-TOF spectrum for WW variant **29BP**.

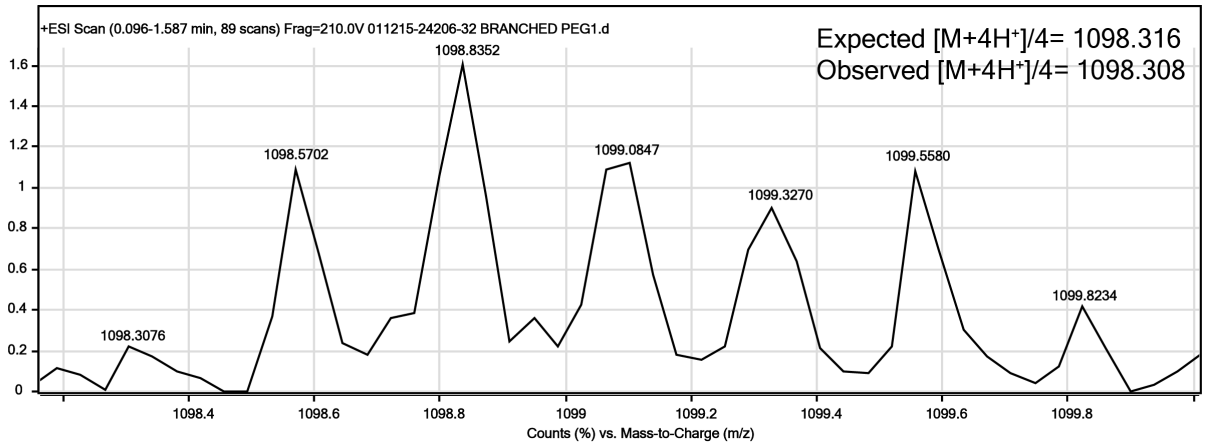


Figure 4-33. ESI-TOF spectrum for WW variant **32BP**.

4.4.2 HPLC

HPLC traces for proteins **16Aha, 16AhaP, 18Aha, 18AhaP, 23Aha, 23AhaP, 27Aha, 27AhaP, 29Aha, 29AhaP, 32Aha, 32AhaP, 16PrF, 16PrFp, 18PrF, 18PrFp, 23PrF, 23PrFp, 27PrF, 27PrFp, 29PrF, 29PrFp, 32PrF, 32PrFp, 16BP, 18BP, 23BP, 27BP, 29BP, and 32BP** are shown in Figures 4-34 through 4-63.

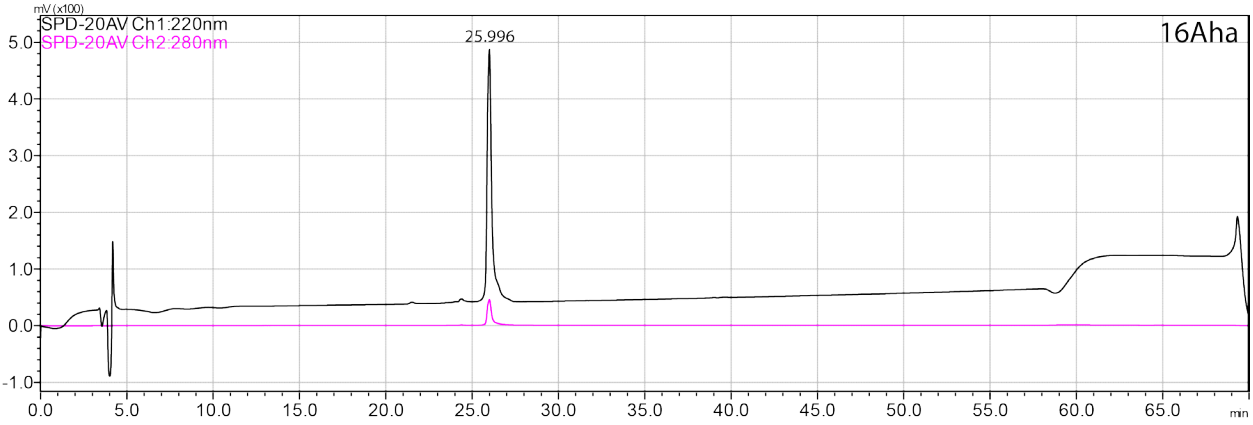


Figure 4-34. Analytical HPLC Data for WW variant **16Aha**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

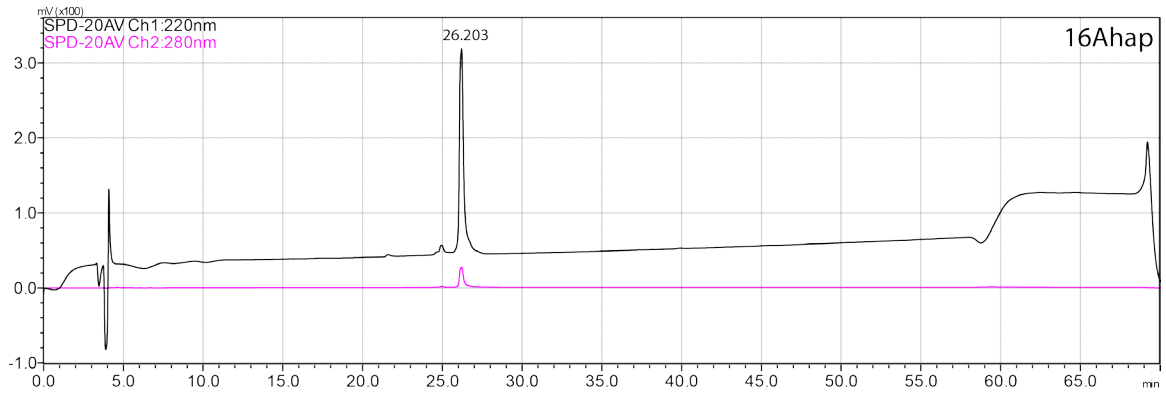


Figure 4-35. Analytical HPLC Data for WW variant **16AhaP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

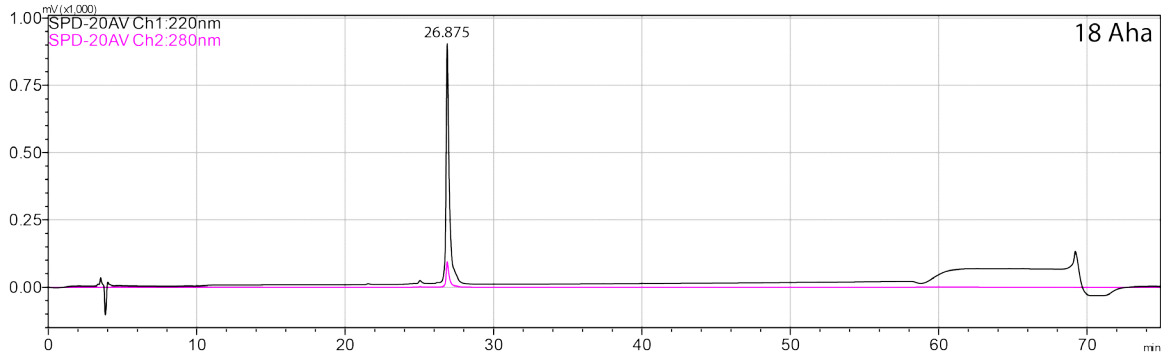


Figure 4-36. Analytical HPLC Data for WW variant **18Aha**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

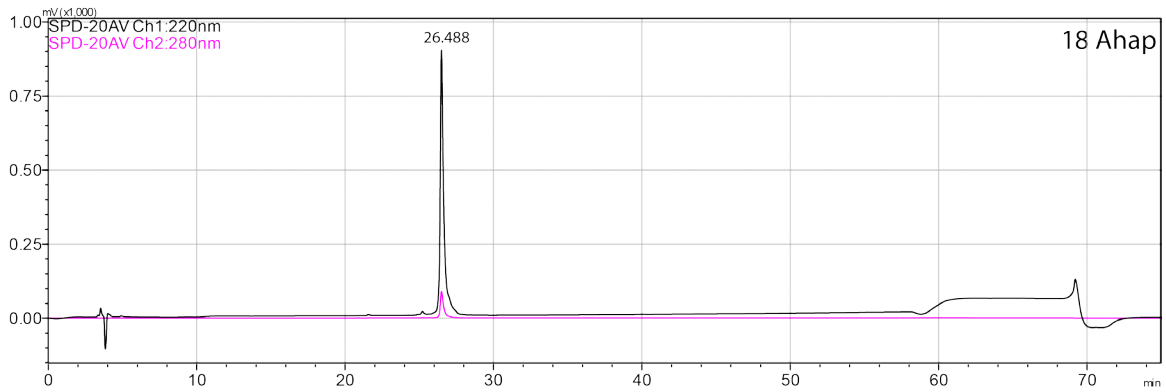


Figure 4-37. Analytical HPLC Data for WW variant **18AhaP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

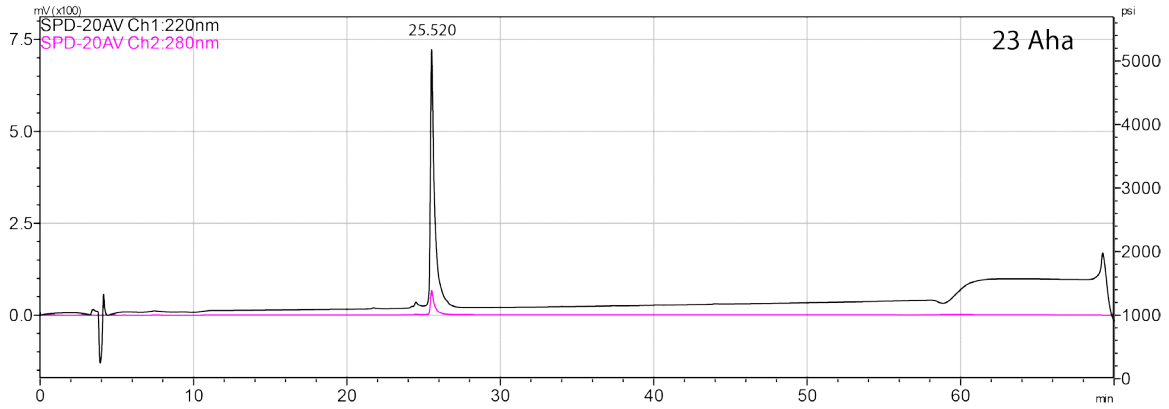


Figure 4-38. Analytical HPLC Data for WW variant **23Aha**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

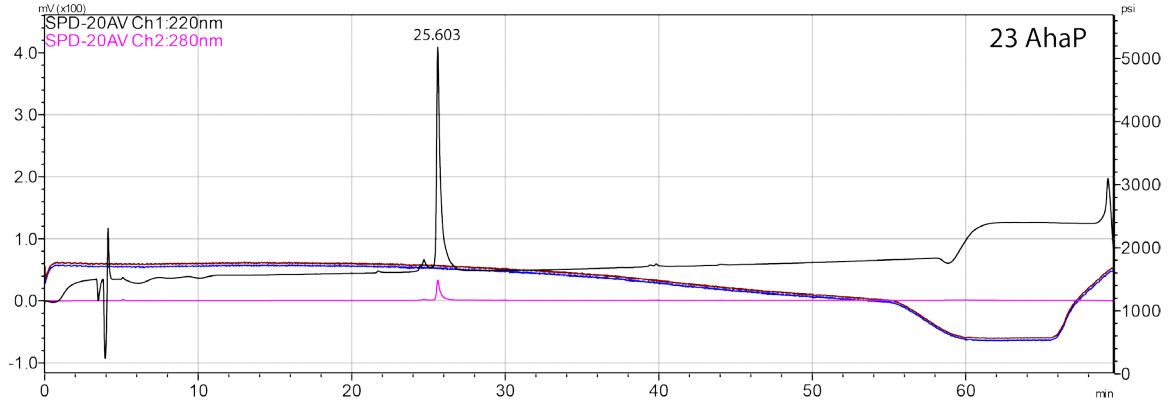


Figure 4-39. Analytical HPLC Data for WW variant **23AhaP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

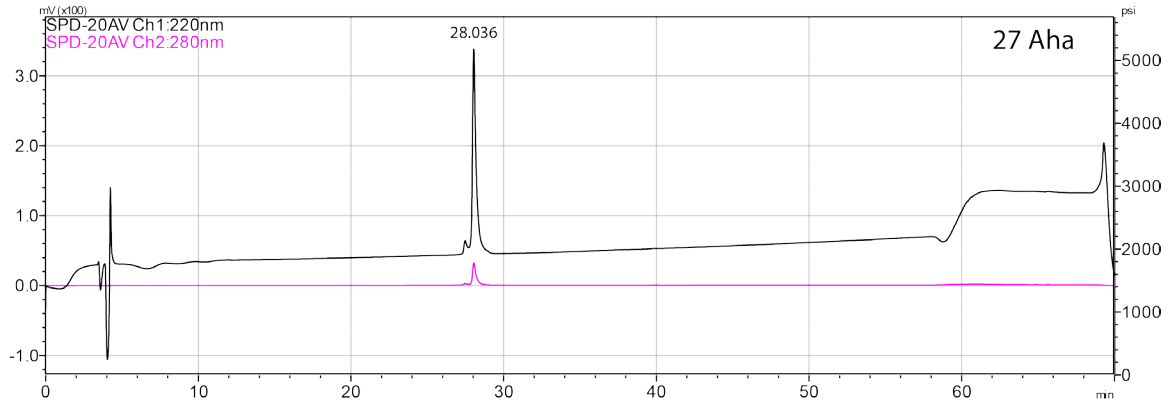


Figure 4-40. Analytical HPLC Data for WW variant **27Aha**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

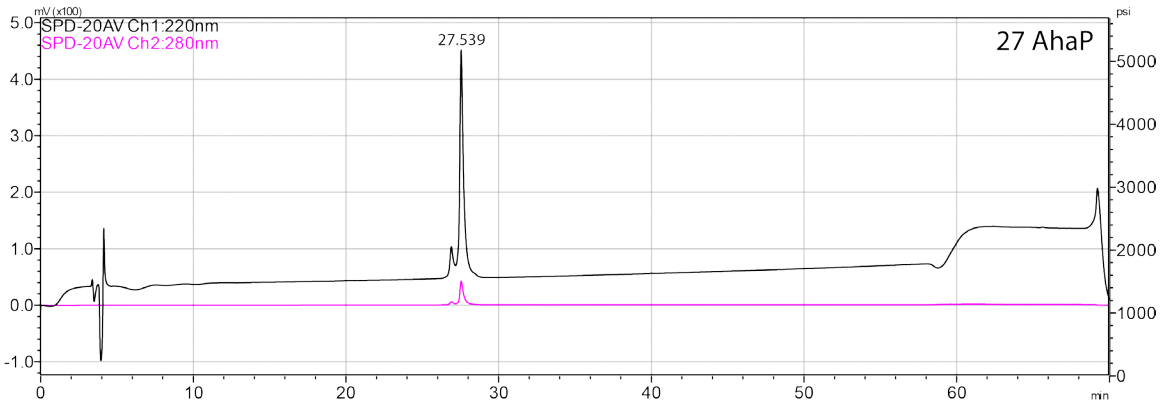


Figure 4-41. Analytical HPLC Data for WW variant **27AhaP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

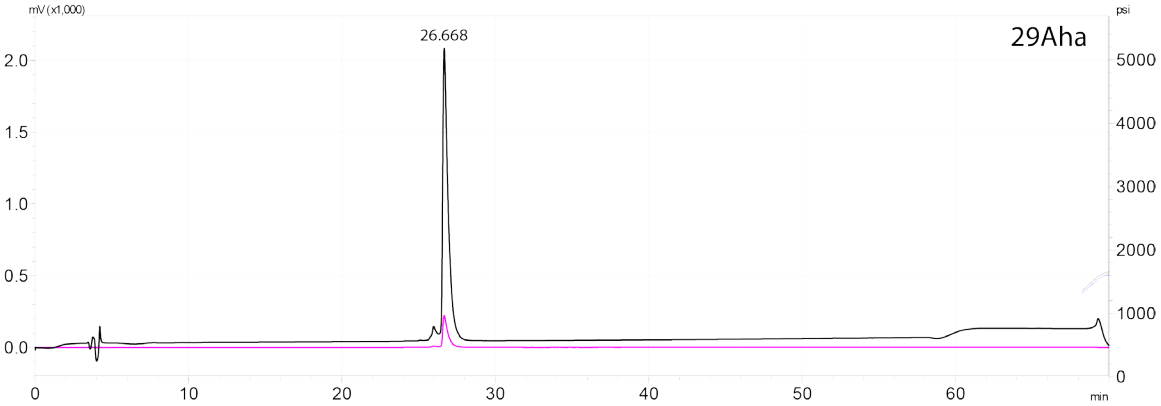


Figure 4-42. Analytical HPLC Data for WW variant **29Aha**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

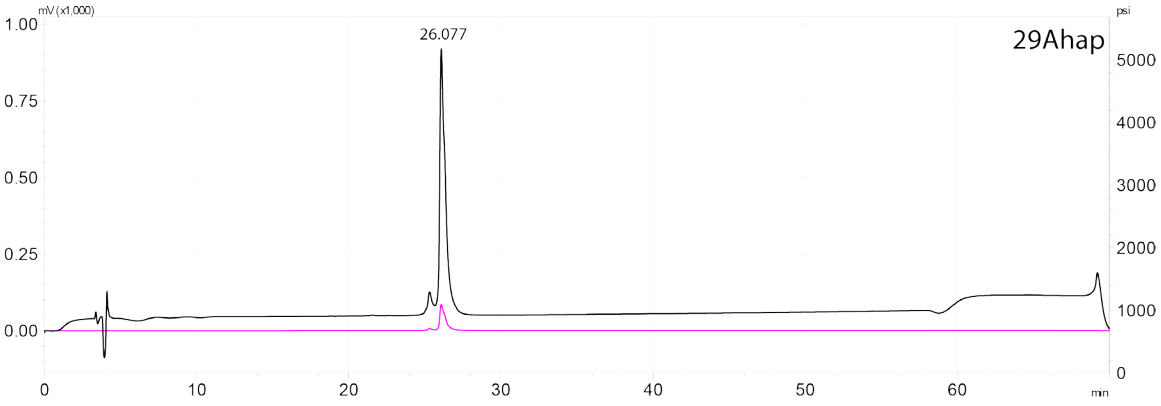


Figure 4-43. Analytical HPLC Data for WW variant **29AhaP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

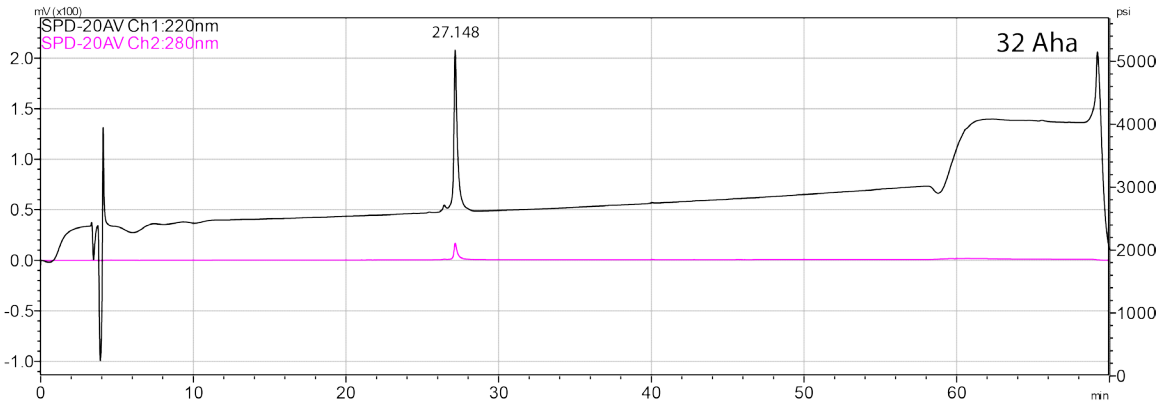


Figure 4-44. Analytical HPLC Data for WW variant **32Aha**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

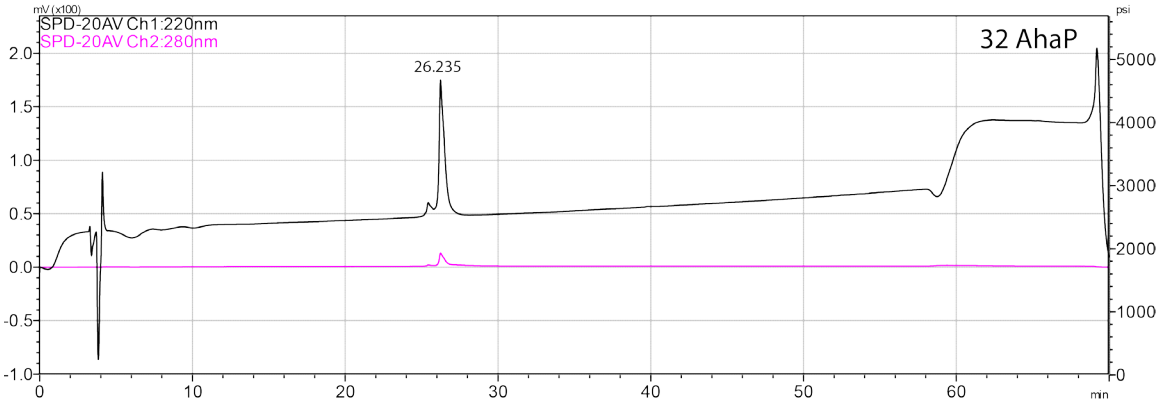


Figure 4-45. Analytical HPLC Data for WW variant **32AhaP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

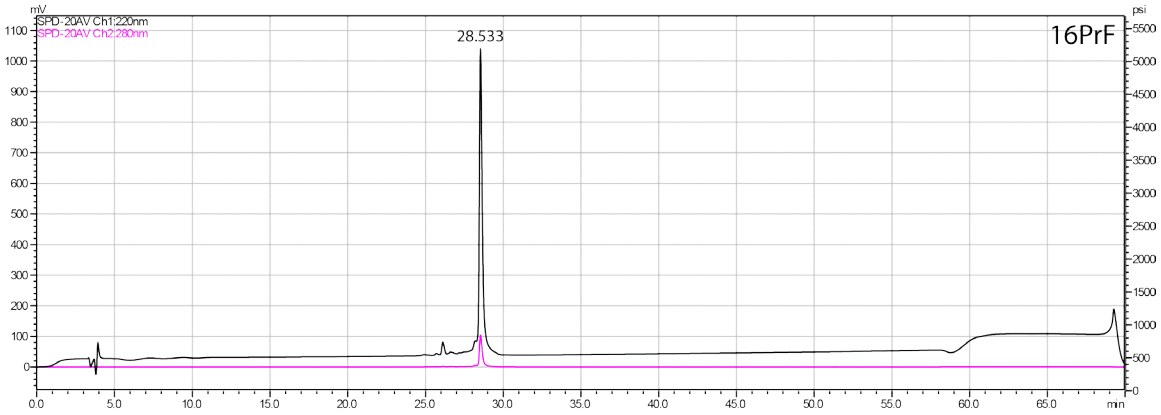


Figure 4-46. Analytical HPLC Data for WW variant **16PrF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min

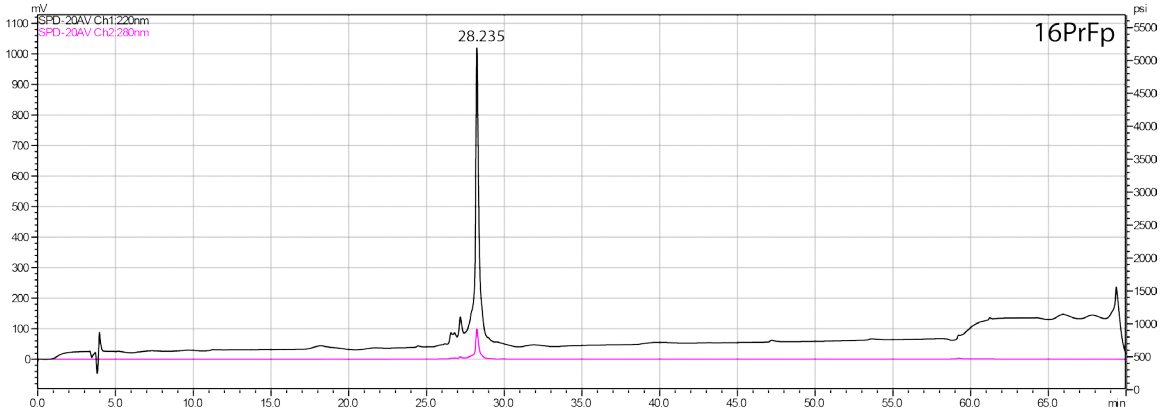


Figure 4-47. Analytical HPLC Data for WW variant **16PrFp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

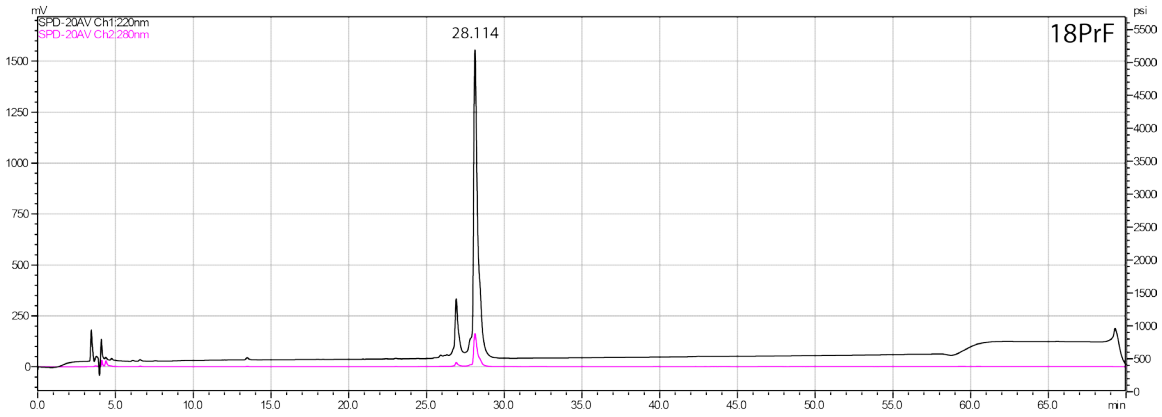


Figure 4-48. Analytical HPLC Data for WW variant **18PrF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

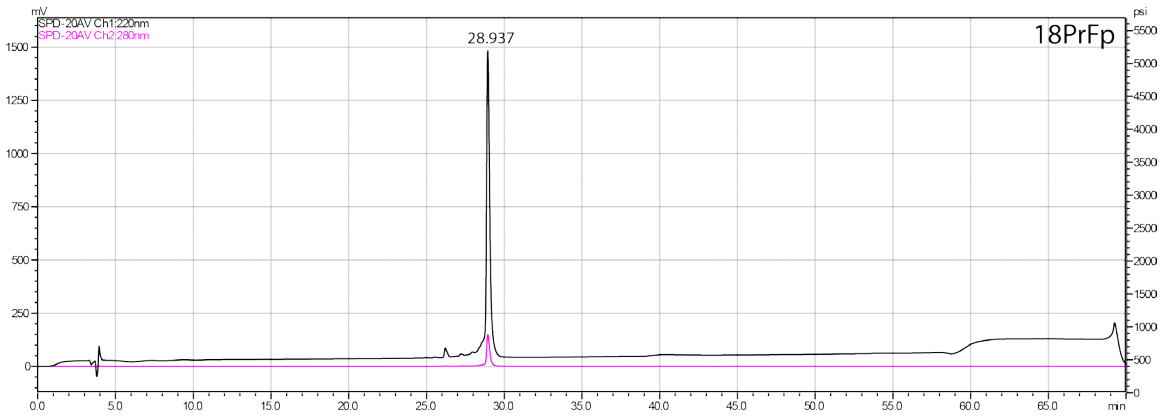


Figure 4-49. Analytical HPLC Data for WW variant **18PrFp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

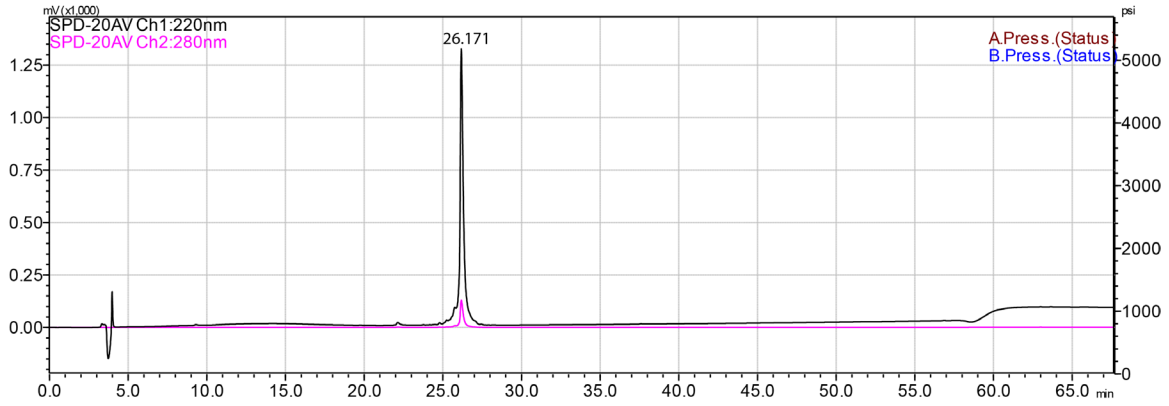


Figure 4-50. Analytical HPLC Data for WW variant **23PrF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

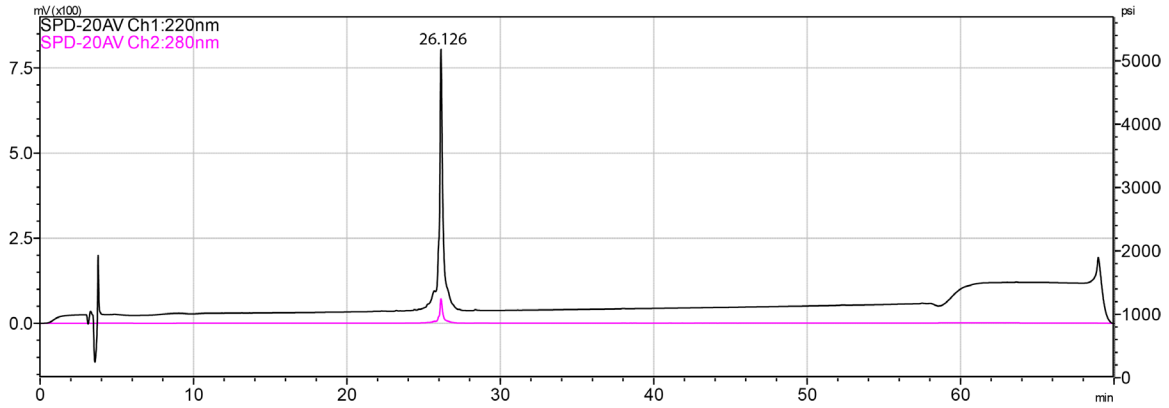


Figure 4-51. Analytical HPLC Data for WW variant **23PrFp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

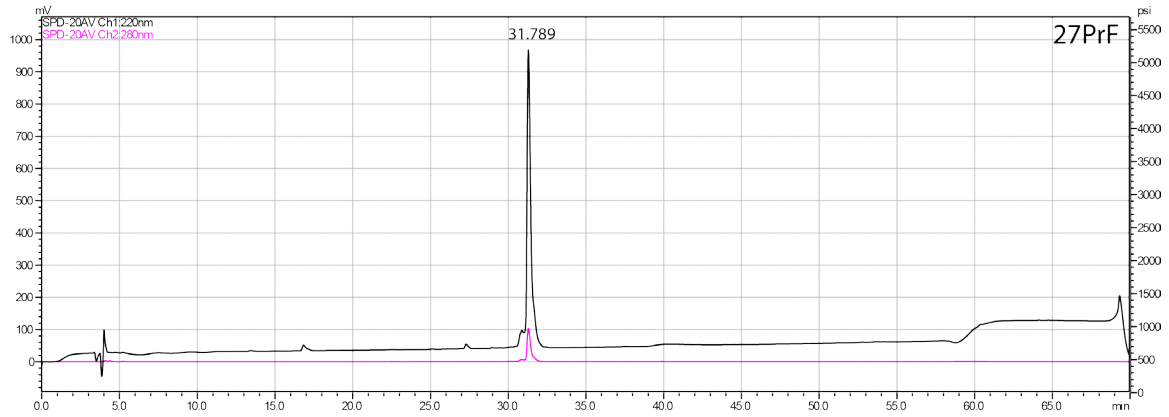


Figure 4-52. Analytical HPLC Data for WW variant **27PrF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

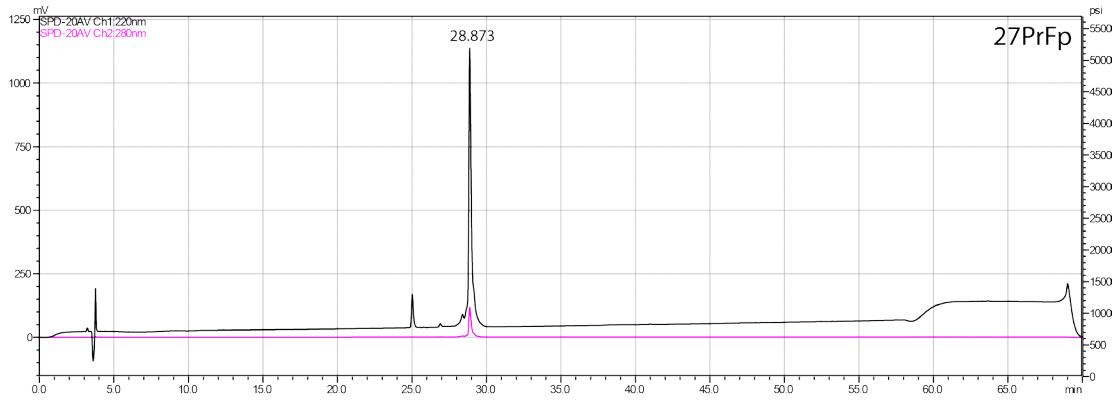


Figure 4-53. Analytical HPLC Data for WW variant **27PrFp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

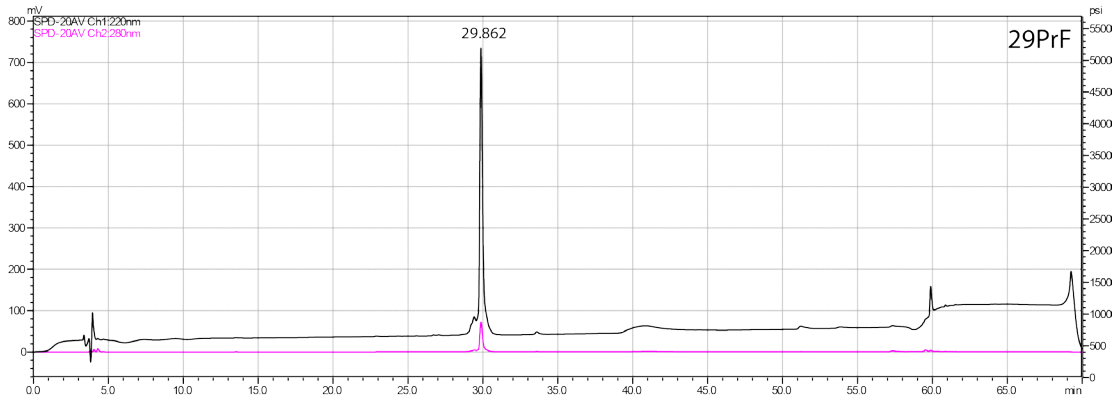


Figure 4-54. Analytical HPLC Data for WW variant **29PrF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

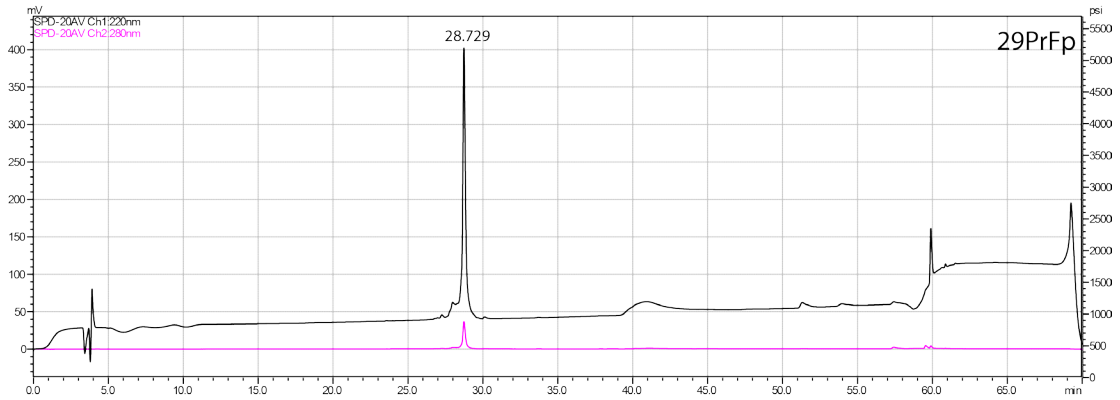


Figure 4-55. Analytical HPLC Data for WW variant **29PrFp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

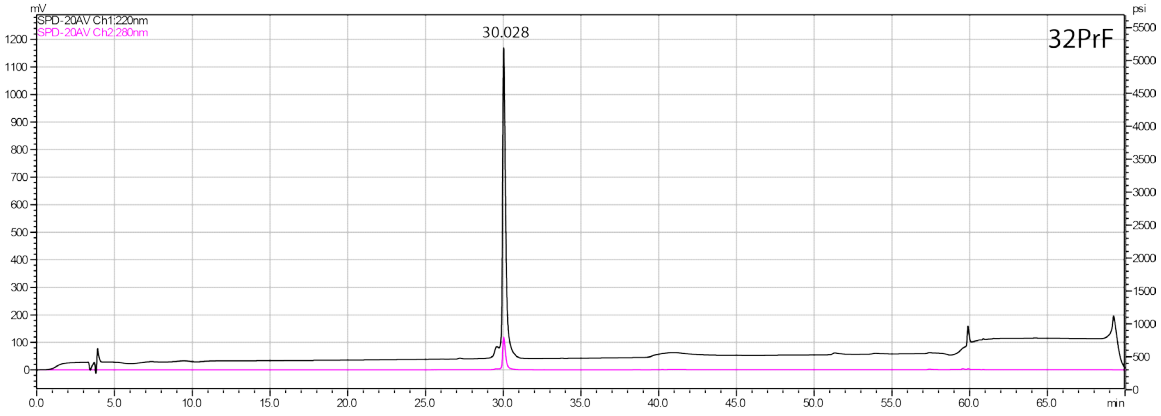


Figure 4-56. Analytical HPLC Data for WW variant **32PrF**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

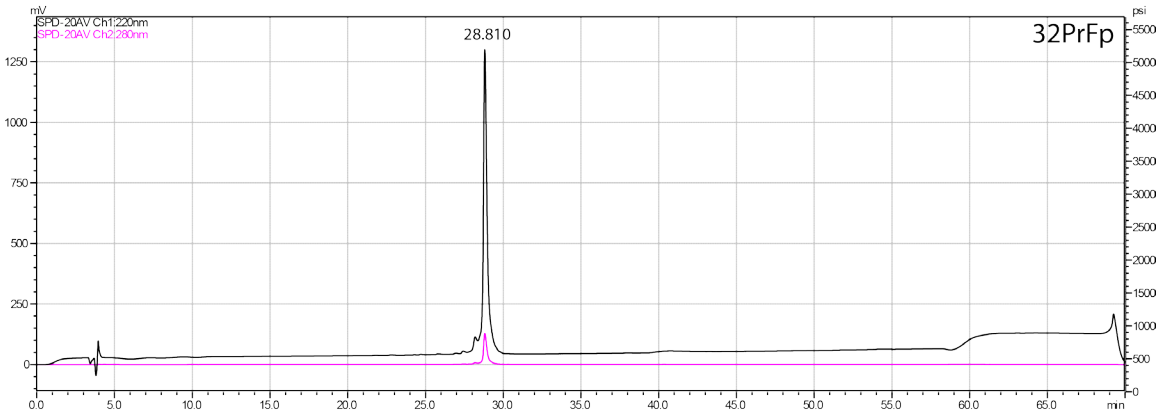


Figure 4-57. Analytical HPLC Data for WW variant **32PrFp**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

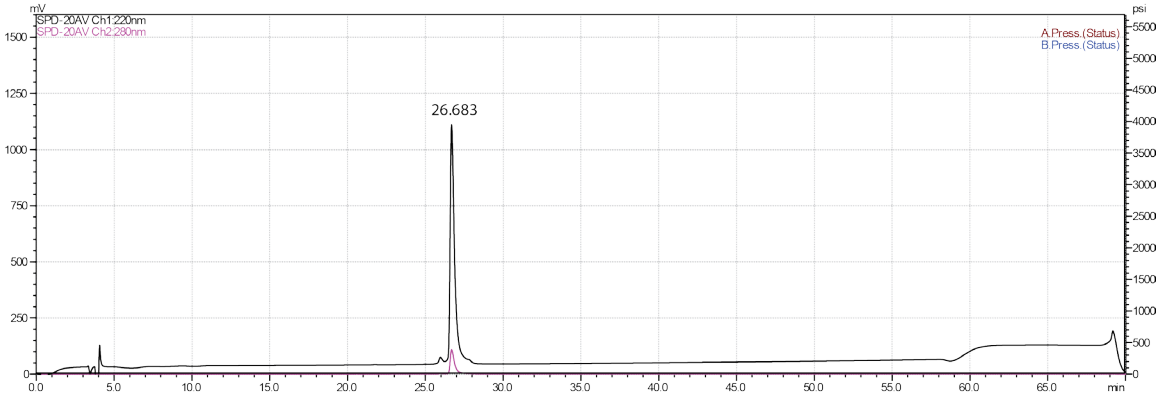


Figure 4-58. Analytical HPLC Data for WW variant **16BP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

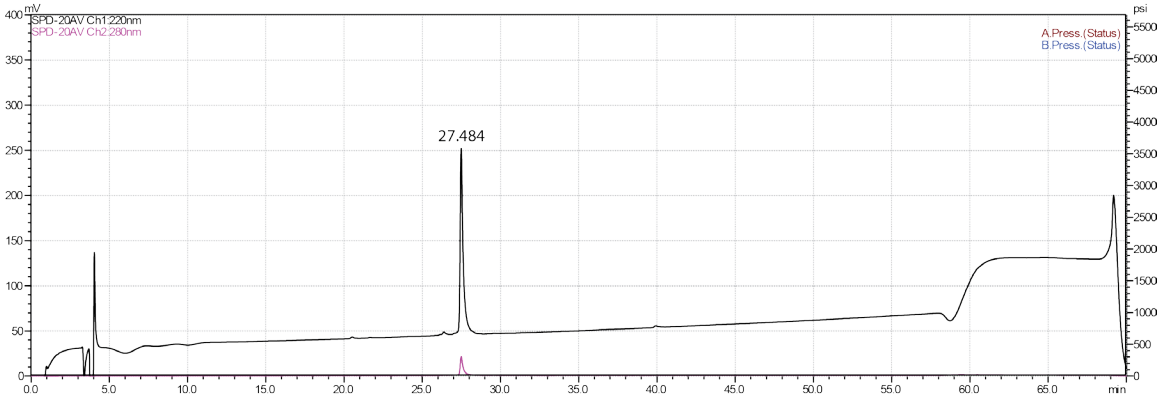


Figure 4-59. Analytical HPLC Data for WW variant **18BP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

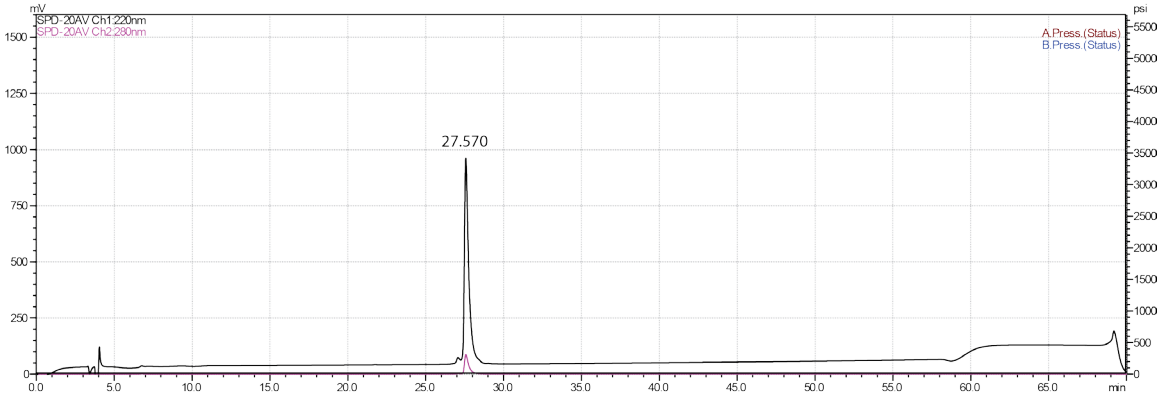


Figure 4-60. Analytical HPLC Data for WW variant **23BP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

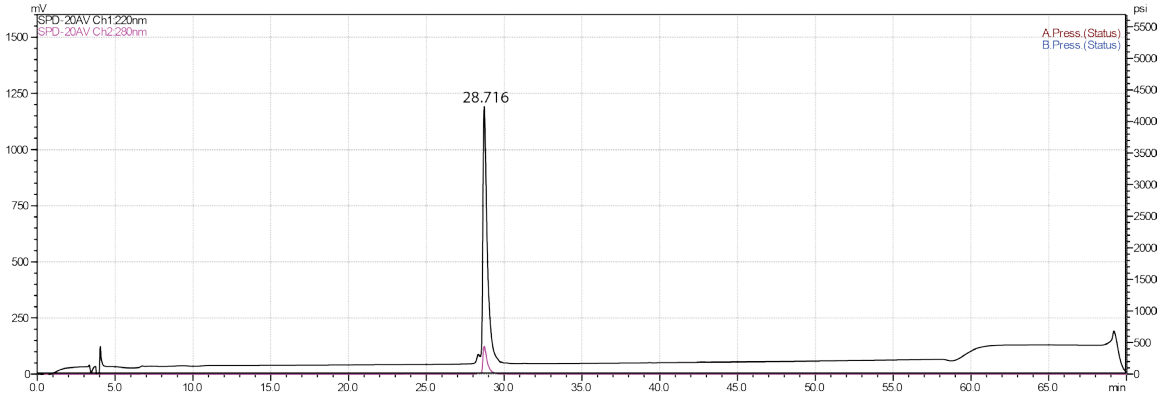


Figure 4-61. Analytical HPLC Data for WW variant **27BP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

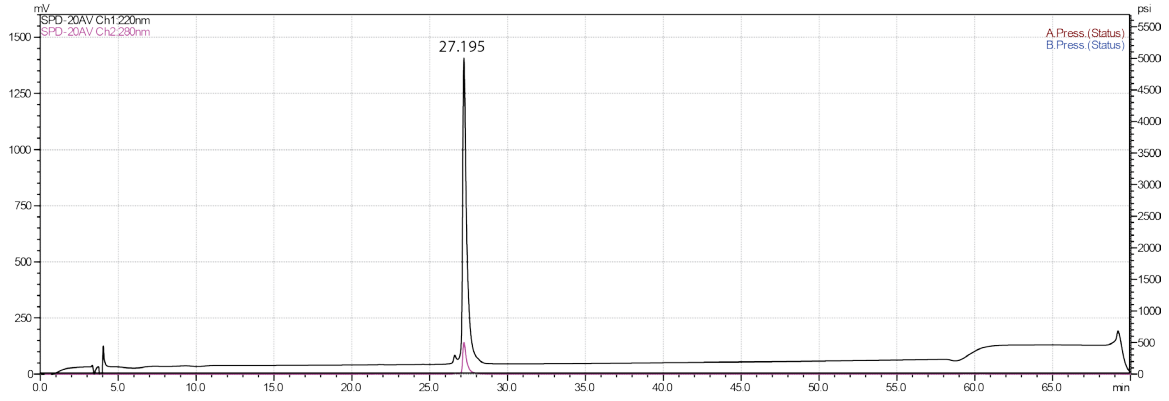


Figure 4-62. Analytical HPLC Data for WW variant **29BP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

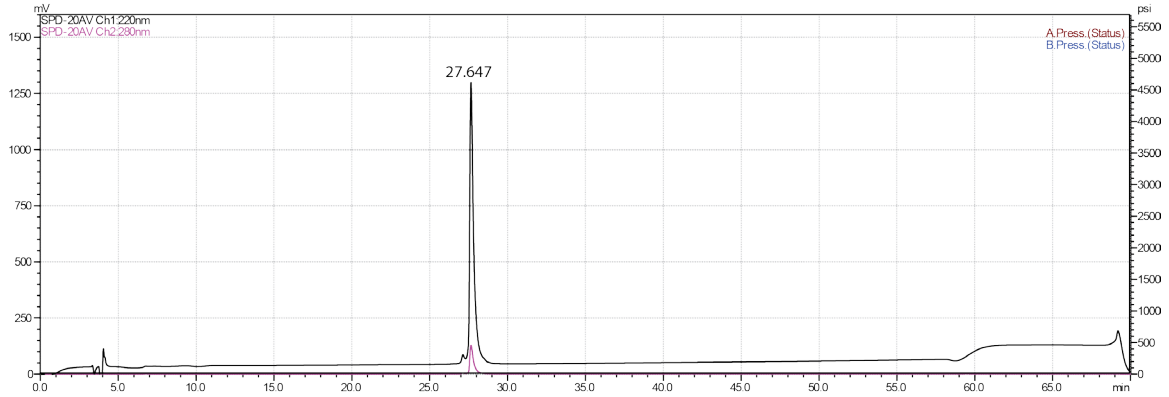
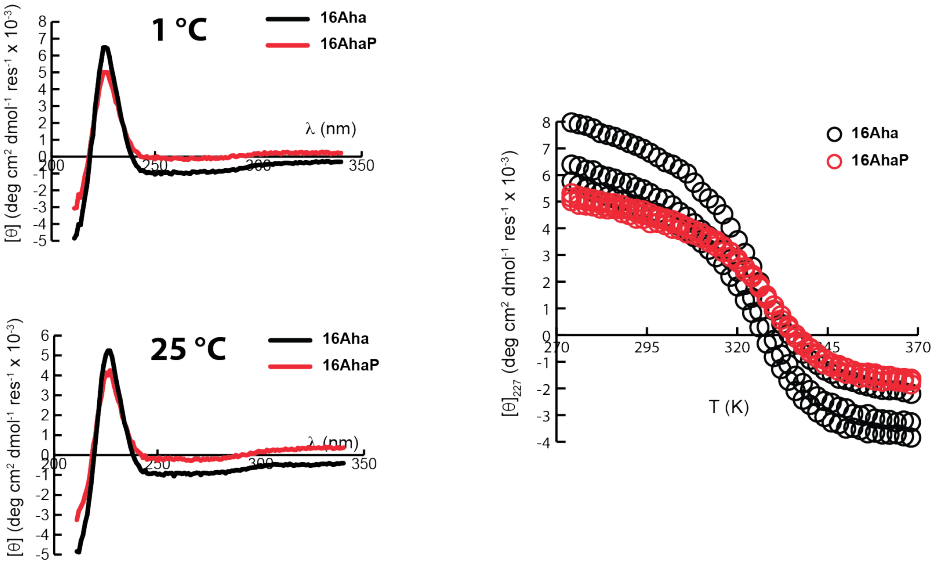


Figure 4-63. Analytical HPLC Data for WW variant **32BP**. Protein solution was injected onto a C18 analytical column and eluted using a linear gradient of 10-60% B (A=H2O, 0.1% TFA; B= MeCN, 0.1% TFA) over 50 minutes, followed by a 10 minute rinse (95% B), and a 10 minute column re-equilibration (10% B) with a flow rate of 1 mL/min.

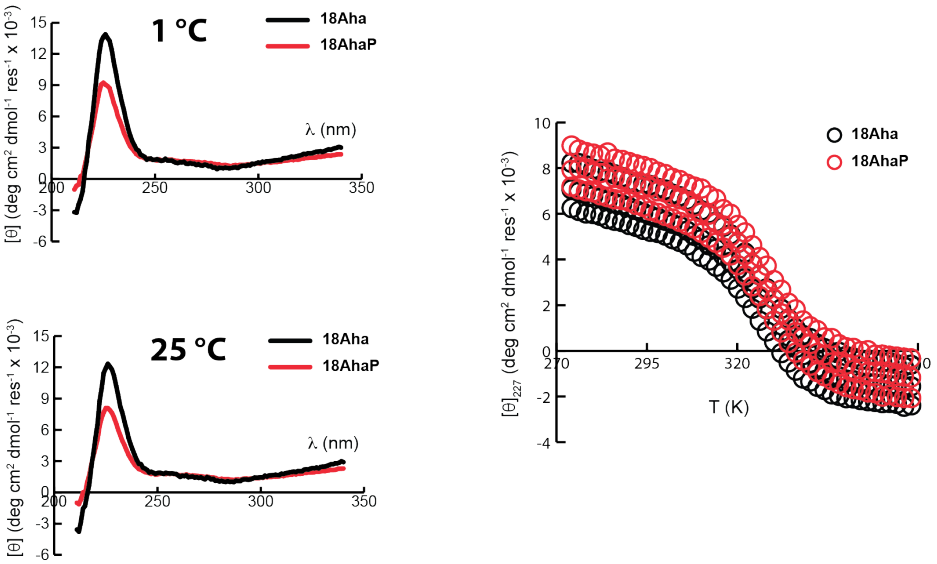
4.4.3 CD Spectra and Thermal Denaturation Plots

CD spectra and thermal denaturation plots for **16Aha, 16AhaP, 18Aha, 18AhaP, 23Aha, 23AhaP, 27Aha, 27AhaP, 29Aha, 29AhaP, 32Aha, 32AhaP, 16PrF, 16PrFp, 18PrF, 18PrFp, 23PrF, 23PrFp, 27PrF, 27PrFp, 29PrF, 29PrFp, 32PrF, 32PrFp, 16BP, 18BP, 23BP, 27BP, 29BP, and 32BP** are shown in Figure**s** 4-64 through 4-63.



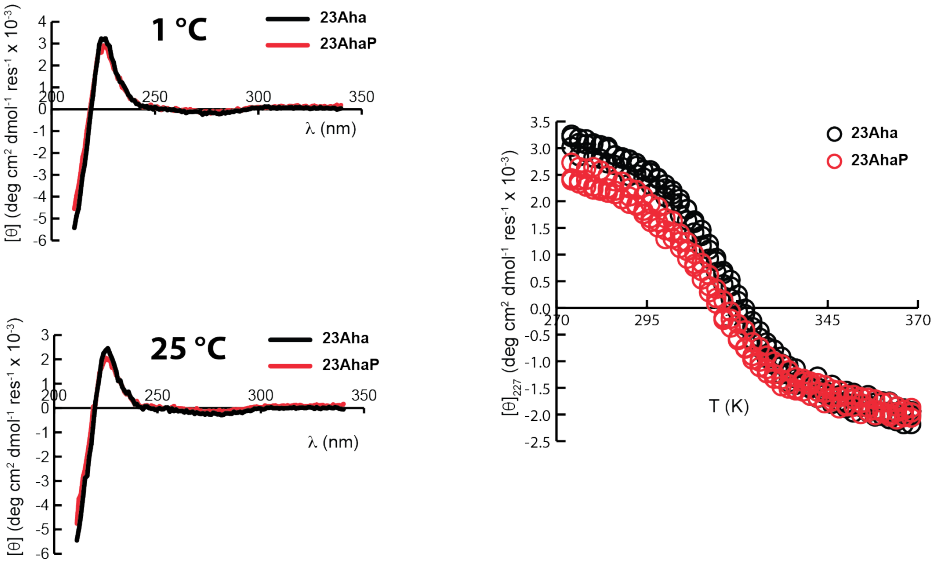
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16Aha	325.9 ± 0.3	-29.0 ± 0.5	-0.52 ± 0.10	0.9999	0.048
16AhaP	330.0 ± 0.3	-31.8 ± 0.4	-0.36 ± 0.12	0.9999	0.030

Figure 4-64. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **16Aha** (black) and **16AhaP** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



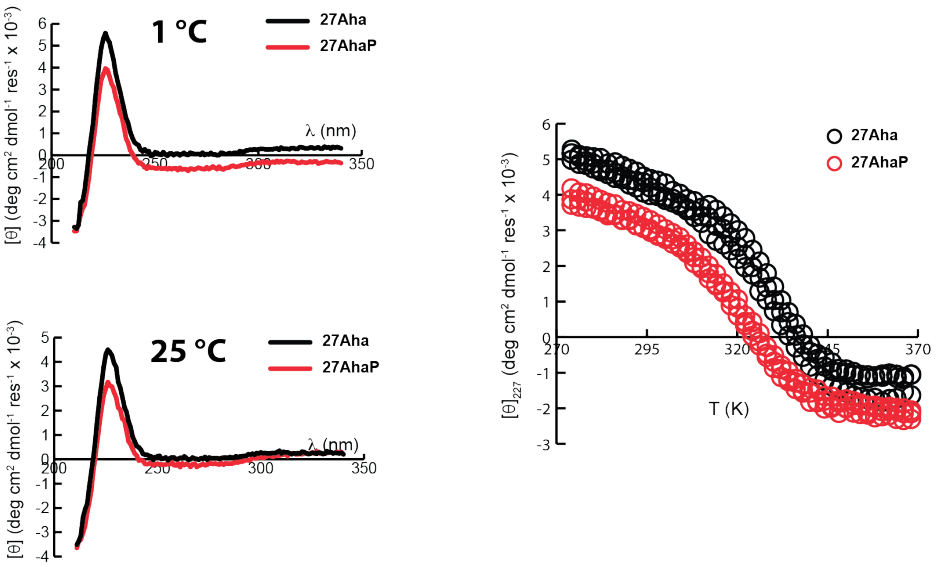
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
18Aha	327.6 ± 0.4	-31.5 ± 0.6	-0.48 ± 0.15	0.9998	0.045
18AhaP	329.2 ± 0.2	-30.9 ± 0.5		0.9998	0.053

Figure 4-65. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **18Aha** (black) and **18AhaP** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



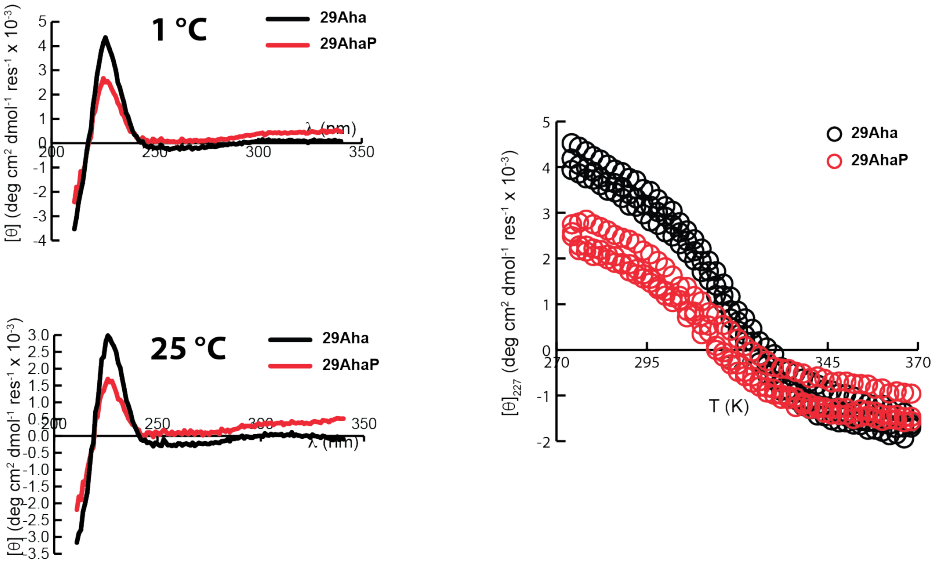
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
23Aha	317.8 ± 0.3	-25.3 ± 0.8		0.9993	0.052
23AhaP	314.9 ± 0.4	-27.2 ± 1.2		0.9988	0.061

Figure 4-66. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **23Aha** (black) and **23AhaP** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



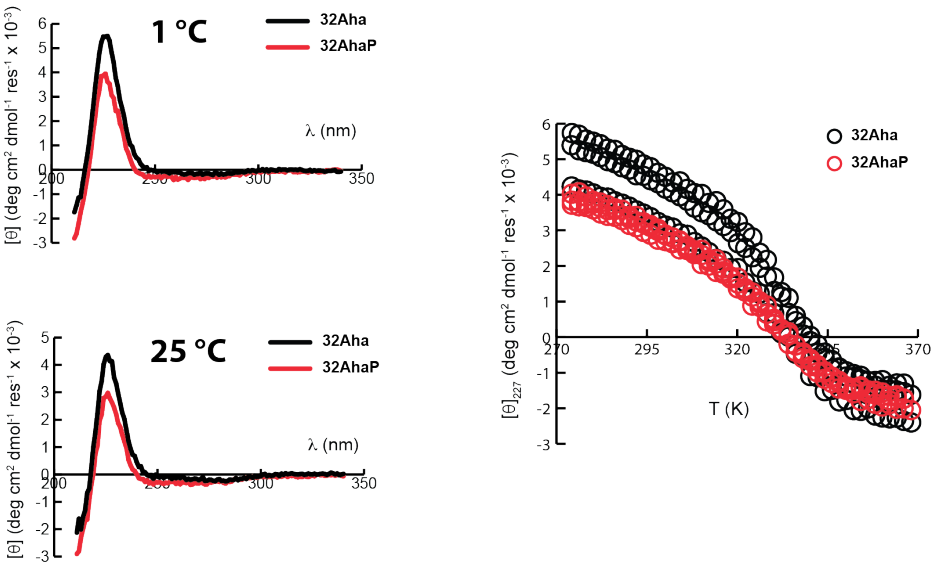
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
27Aha	331.7 ± 0.2	-32.5 ± 0.8	-0.67 ± 0.07	0.9992	0.071
27AhaP	325.3 ± 0.2	-25.6 ± 0.4		0.9994	0.058

Figure 4-67. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **27Aha** (black) and **27AhaP** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



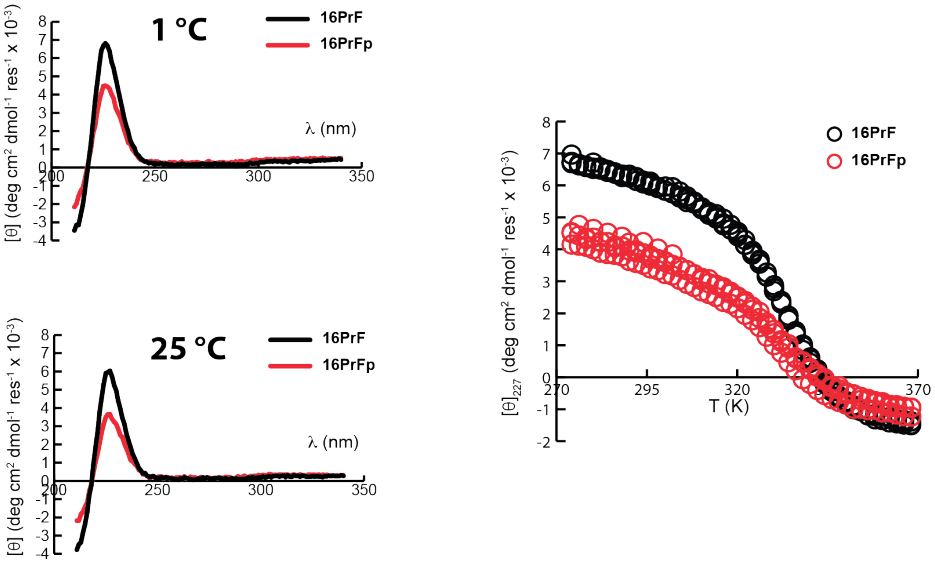
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
29Aha	320.7 ± 0.1	-26.1 ± 0.3		0.9996	0.046
29AhaP	315.7 ± 0.6	-24.8 ± 1.4		0.9986	0.056

Figure 4-68. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **29Aha** (black) and **29AhaP** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



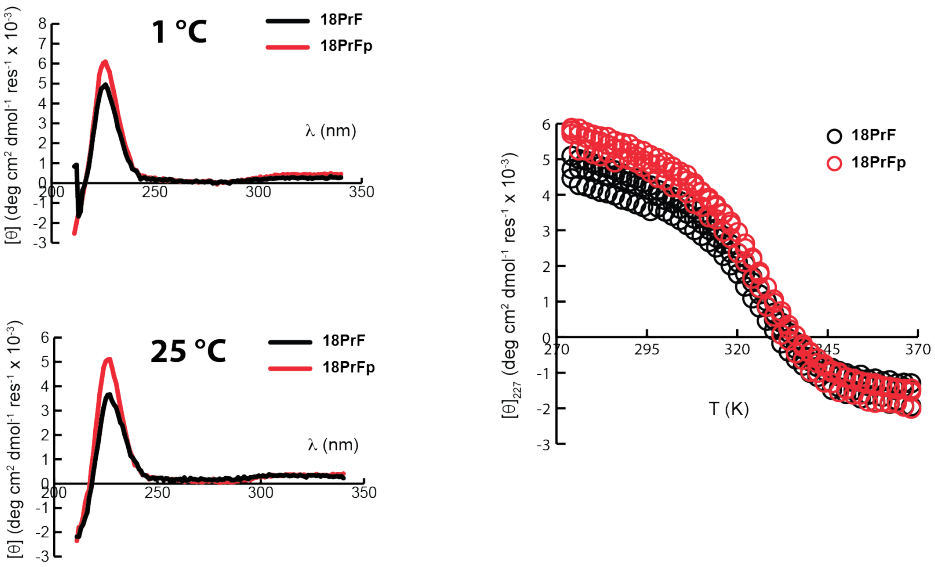
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
32Aha	333.4 ± 0.3	-31.9 ± 1.3	-0.69 ± 0.07	0.9995	0.063
32AhaP	333.0 ± 0.5	-28.0 ± 1.7	-0.66 0.07	0.9995	0.057

Figure 4-69. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **32Aha** (black) and **32AhaP** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



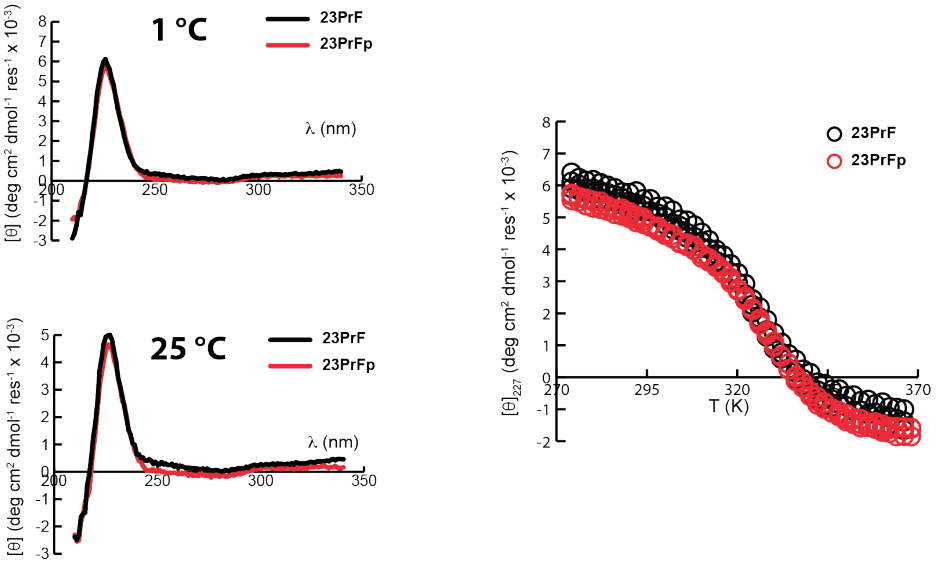
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
16PrF	333.4 ± 0.2	-31.8 ± 0.6	-0.69 ± 0.04	0.9998	0.045
16PrFp	332.8 ± 0.5	-33.2 ± 2.1	-0.82 ± 0.08	0.9992	0.072

Figure 4-70. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **16PrF** (black) and **16PrFp** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



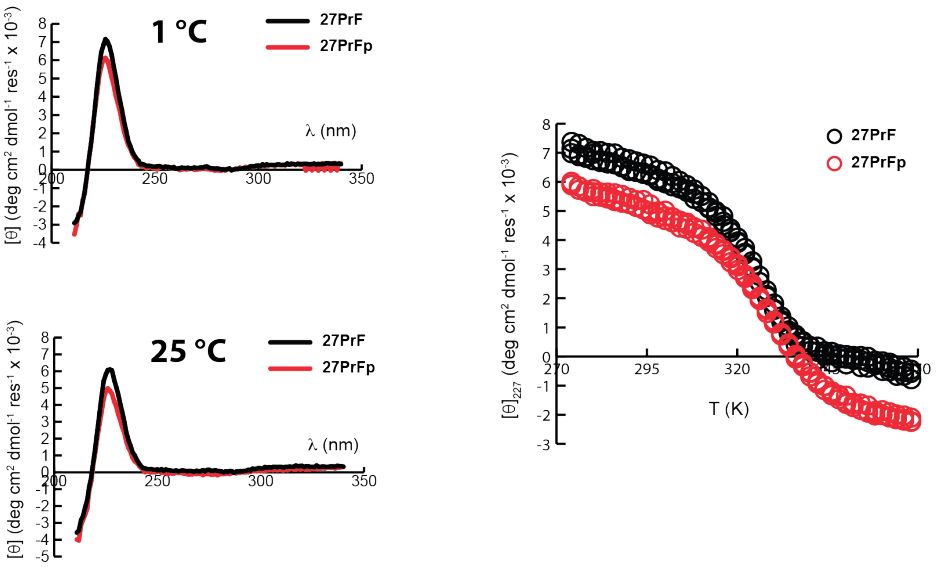
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
18PrF	326.0 ± 0.2	-28.3 ± 0.7	-0.63 ± 0.07	0.9997	0.046
18PrFp	326.6 ± 0.4	-27.7 ± 0.7	-0.53 ± 0.12	0.9996	0.058

Figure 4-71. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **18PrF** (black) and **18PrFp** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



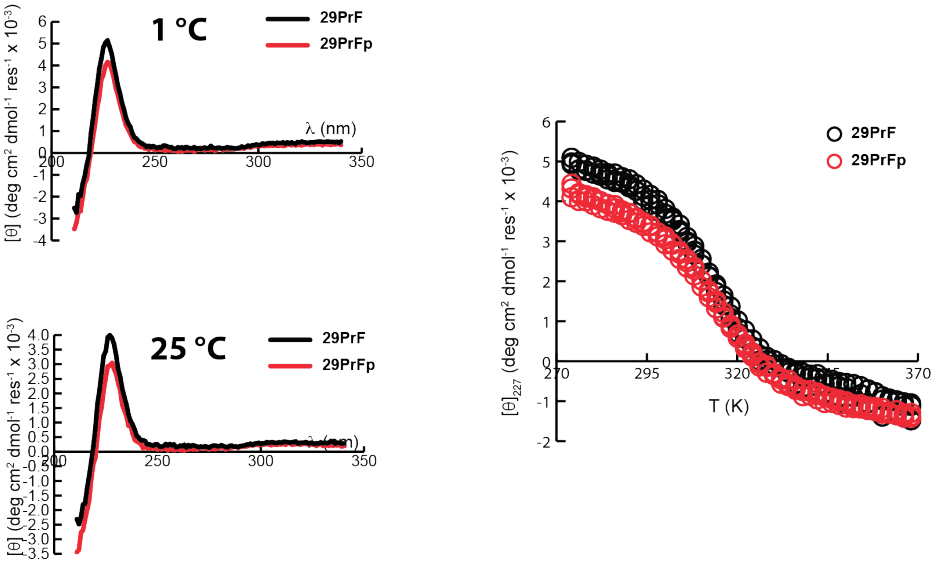
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
23PrF	325.2 ± 0.2	-26.6 ± 0.6	-0.59 ± 0.06	0.9998	0.042
23PrFp	328.7 ± 0.2	-28.6 ± 0.7	-0.65 ± 0.04	0.9998	0.042

Figure 4-72. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **23PrF** (black) and **23PrFp** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



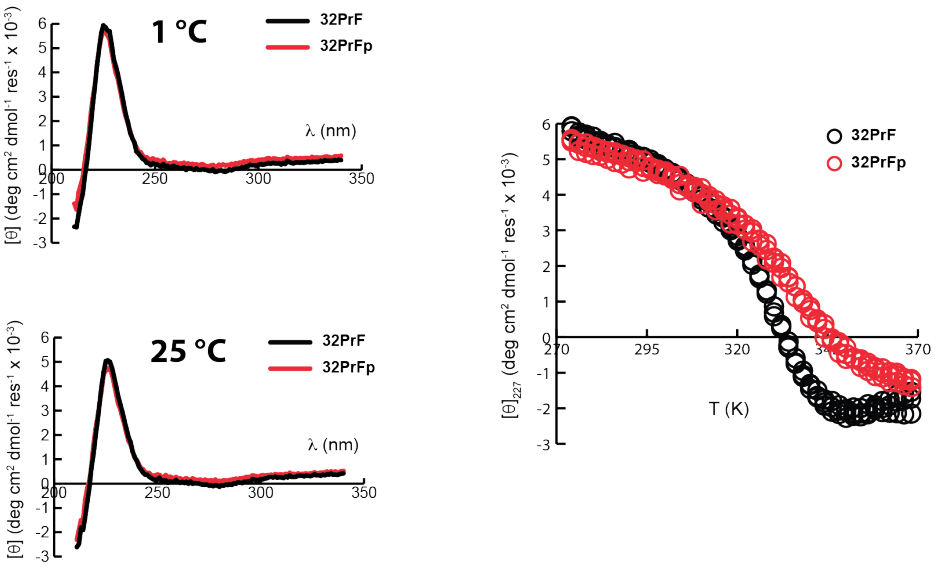
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
27PrF	325.1 ± 0.2	-37.5 ± 0.9	-0.98 ± 0.06	0.9997	0.058
27PrFp	331.6 ± 0.1	-26.7 ± 0.3	-	0.9997	0.057

Figure 4-73. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **27PrF** (black) and **27PrFp** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



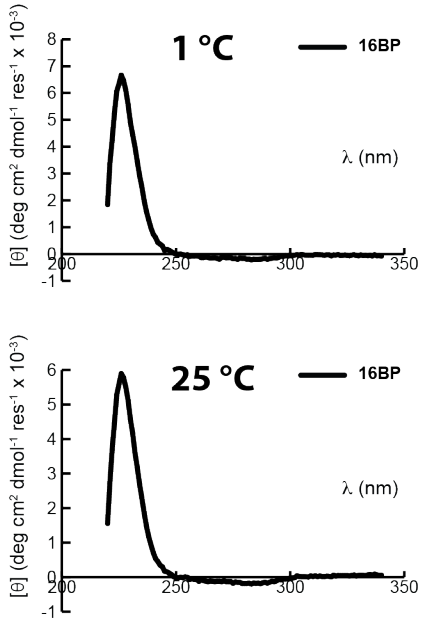
Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
29PrF	312.5 ± 0.4	-27.8 ± 1.1	-0.92 ± 0.08	0.9996	0.052
29PrFp	316.3 ± 0.2	-26.2 ± 0.5	-	0.9998	0.033

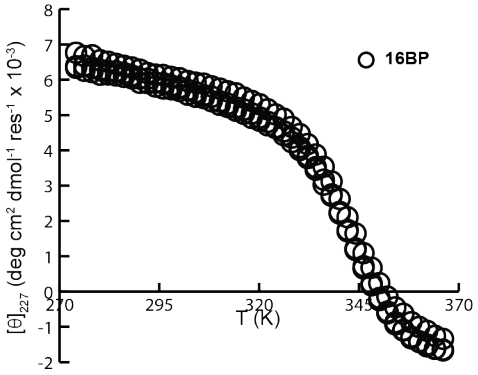
Figure 4-74. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **29PrF** (black) and **29PrFp** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



Peptide	Tm/ K	$\Delta H(T_m)$ / kcal mol^{-1}	ΔC_p / kcal mol^{-1} K^{-1}	R^2	rmsd error
32PrF	331.9 ± 0.2	-31.5 ± 0.8	-0.66 ± 0.05	0.9996	0.070
32PrFp	336.6 ± 0.3	-23.5 ± 0.6	-0.41 ± 0.05	0.9993	0.062

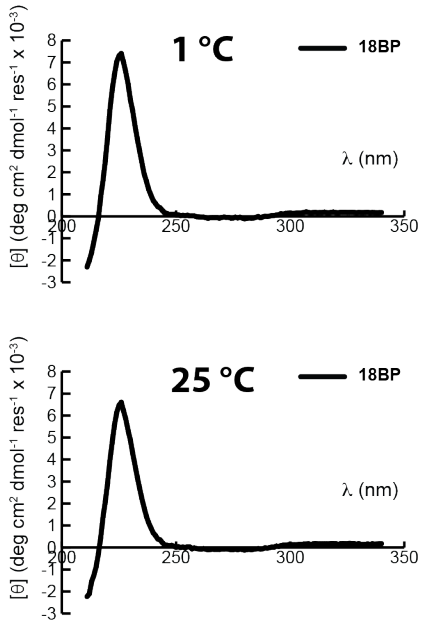
Figure 4-75. CD spectra (50 µM) and variable temperature CD data (50 µM) for WW variants **32PrF** (black) and **32PrFp** (red) in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.

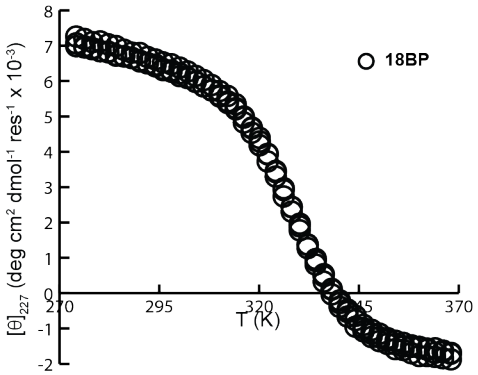




Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error
16BP	343.40 ± 0.05	-36.13 ± 0.27	-0.64 ± 0.03	0.99989	0.02993

Figure 4-76. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **16BP** in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.





Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error
18BP	330.44 ± 0.11	-29.05 ± 0.21	0.09 ± 0.05	0.99987	0.04050

Figure 4-77. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **18BP** in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.

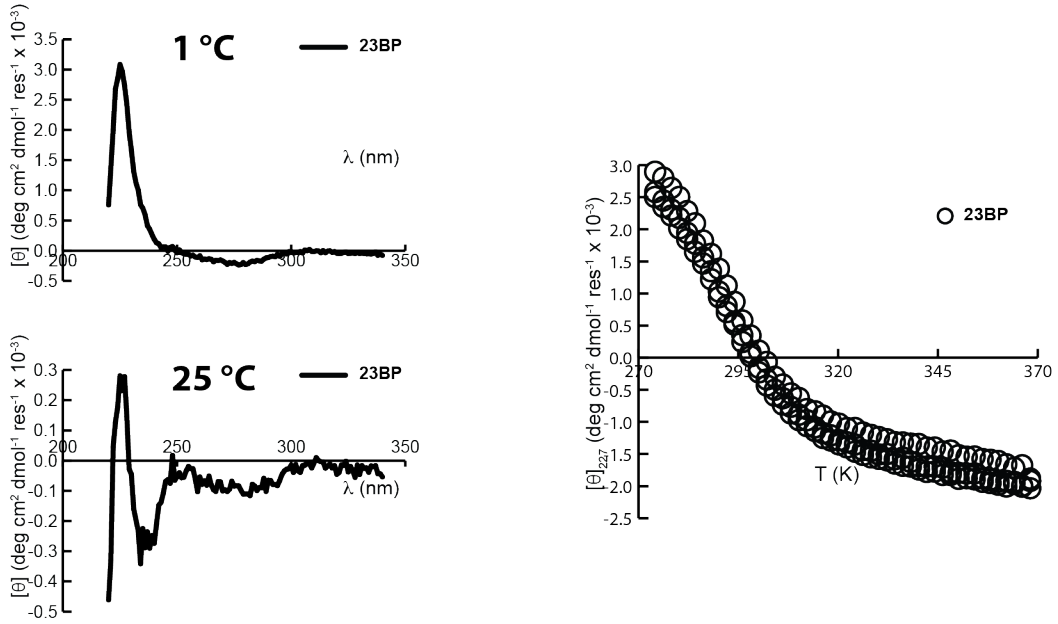
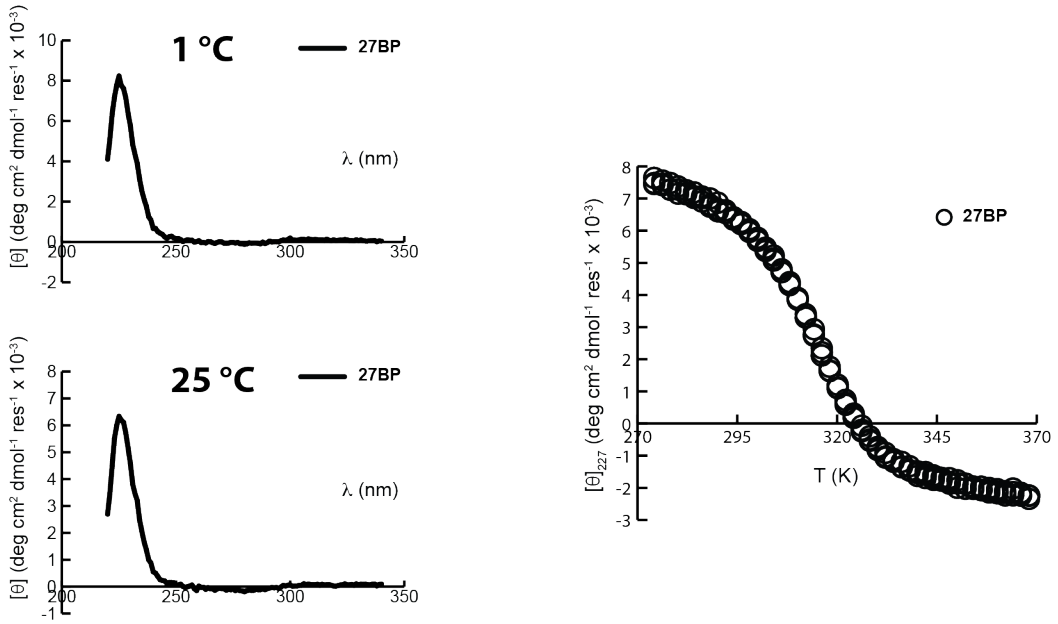
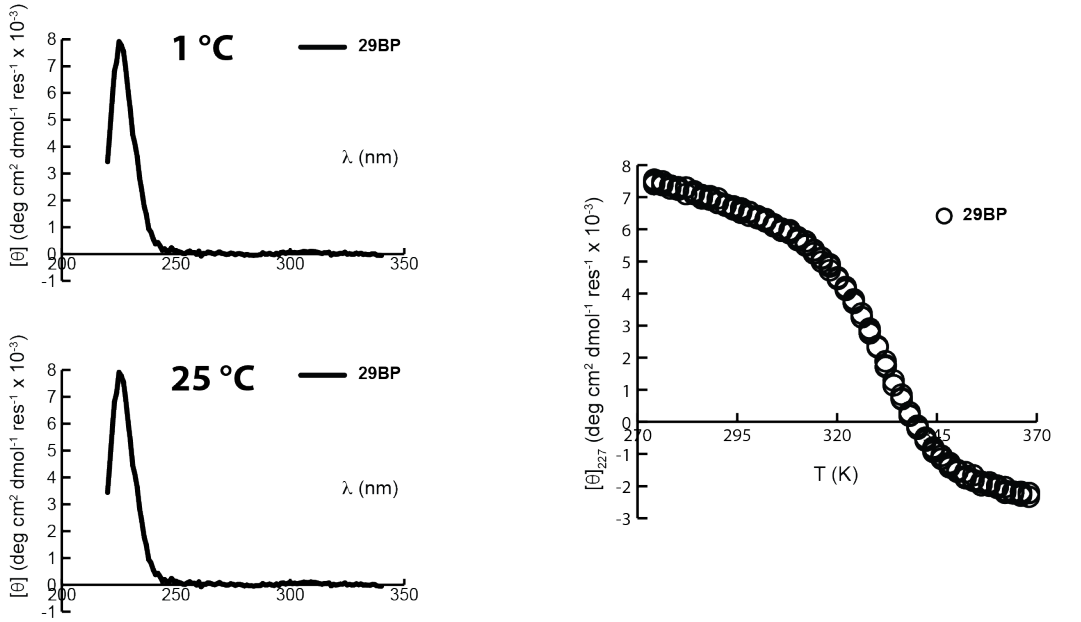


Figure 4-78. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **23BP** in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



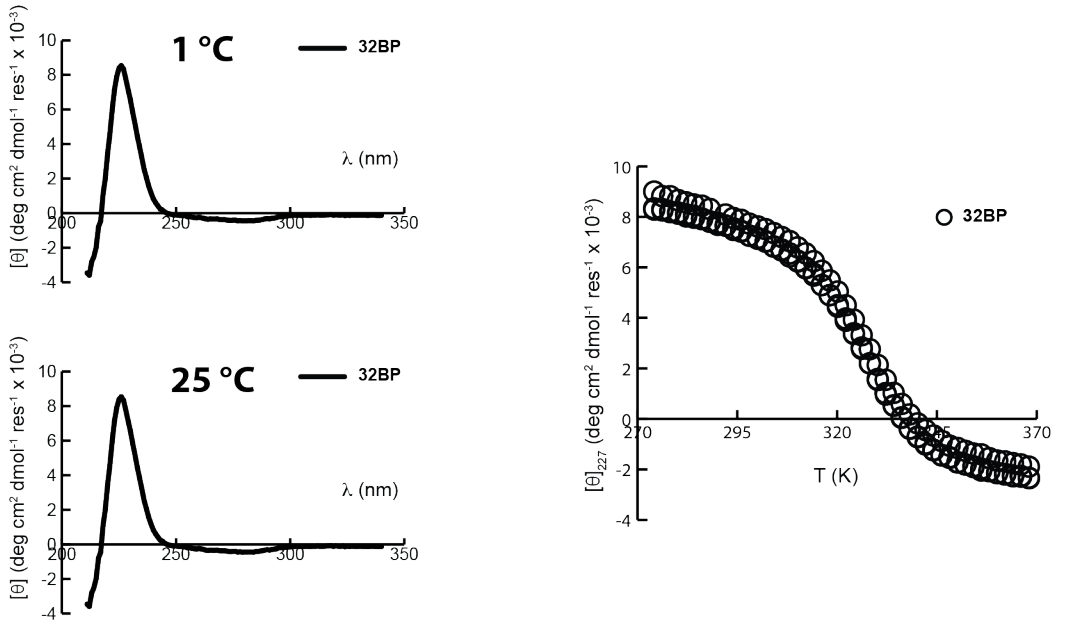
Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error
27BP	314.65 ± 0.35	-26.04 ± 0.66	-0.53 ± 0.08	0.99993	0.03582

Figure 4-79. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **27BP** in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error
29BP	331.19 ± 0.14	-30.06 ± 0.34	-0.59 ± 0.04	0.99993	0.03476

Figure 4-80. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **29BP** in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.



Peptide	T_m	$\Delta H(T_m)$	ΔC_p	R^2	rmsd error
32BP	327.37 ± 0.39	-30.35 ± 0.42	-0.32 ± 0.12	0.99988	0.05042

Figure 4-81. CD spectra (50 μM) and variable temperature CD data (50 μM) for WW variants **32BP** in 20 mM sodium phosphate, pH 7. Fit parameters appear in the table, along with standard errors.

4.5 References

(1) Lawrence, P. B.; Gavrilov, Y.; Matthews, S. S.; Langlois, M. I.; Shental-Bechor, D.; Greenblatt, H. M.; Pandey, B. K.; Smith, M. S.; Paxman, R.; Torgerson, C. D.; Merrell, J. P.; Ritz, C. C.; Prigozhin, M. B.; Levy, Y.; Price, J. L. *J. Am. Chem. Soc.* **2014**, *136*, 17547.

(2) 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. *Bioconjugate Chem.* **2013**, *24*, 796.