N-PEGylation of a Reverse Turn is Stabilizing in Multiple Sequence Contexts unlike N-GlcNAcylation

Joshua L. Price,[a,b] Evan T. Powers,[a,b,1] and Jeffery W. Kelly[a,b,c,1]

[a]Department of Chemistry, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California 92037, United States. [b]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California 92037, United States. [c]Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Rd., La Jolla, California 92037, United States.

[1]To whom correspondence should be addressed. jkelly@scripps.edu, epowers@scripps.edu

Supporting Information

Synthesis of PEGylated Fmoc-protected Asparagine

Allyl N^2*-fluorenylmethyoxycarbonyl*-N^4*-{11-methoxy-3,6,9-trioxaundecyl}-L-asparaginate* ***1*** *(Fmoc-Asn(PEG)-OAll)*

Procedure

Fmoc-Asn(PEG)-OAll **1** was synthesized following the procedure of Herzner and Kunz[1]: to a solution of Allyl *N*^{2}-fluorenylmethyoxycarbonyl-L-aspartate (Fmoc-Asp-OAll, 2.28 g, 5.77 mmol) in dry dichloromethane (50 mL) was added isobutyl (2-isobutoxy)-1,2-dihydroquinoline-1-carboxylate (IIDQ, 2.57 mL, 8.66 mmol), and the resulting mixture was stirred for 15 min at room temperature under an argon atmosphere. Then, 11-methoxy-3,6,9-trioxaundecylamine (1g, 4.81 mmol) was added, and stirring was continued for 24 h. After dilution with dichloromethane (100 mL), and washing with brine (50 mL) and water (50 mL), the organic solution was dried with $MgSO_4$, and concentrated by rotary evaporation to give a yellow oil. The desired product was purified by flash chromatography over silica gel using ethyl acetate/hexanes (1:9 for ~800 ml, then 3:7 for ~800 mL), then acetic acid/ethyl acetate (1:100 for ~1200 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 (2.61 g, 4.4 mmol, 93% yield). R_f = 0.15 (1:100 acetic acid/ethyl acetate).

Analytical Data

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.66 (d, *J* = 7.5 Hz, 2H, Fmoc aryl C-***H***), 7.53 (t, *J* = 7.0 Hz, 2H, Fmoc aryl C-***H***), 7.30 (t, *J* = 7.4 Hz, 2H, Fmoc aryl C-***H***), 7.21 (t, *J* = 7.4 Hz, 2H, Fmoc aryl C-***H***), 6.71 (t, *J* = 4.9 Hz, 1H, -$C_\beta H_2$-CON***H***-CH_2-CH_2-O-), 6.27 (d, *J* = 8.6 Hz, 1H, -CON***H***-C_αH(COOAll)-), 5.81 (m, 1H, -O-CH_2-C***H***=CH_2), 5.24 (apparent d, *J* = 17.2 Hz, 1H, -O-CH_2-CH=C(***H***$_{trans}$)H_{cis}), 5.13 (apparent dd, *J* = 10.5, 0.8 Hz, 1H, -O-CH_2-CH=C(H_{trans})***H***$_{cis}$), 4.57

(m, 3H, -NH-C_α***H***(COOAll)-$C_\beta H_2$-, -O-C***H*****$_2$**-CH=CH_2), 4.38 (dd, ^2J = 10.5, ^3J = 7.3 Hz, 1H, Fmoc Ar_2CH-C(***H*****$_a$**)H_b-O-), 4.26 (dd, ^2J = 10.3, ^3J = 7.6 Hz, 1H, Fmoc Ar_2CH-C(H_a)***H*****$_b$**-O-), 4.14 (apparent t, J = 7.1 Hz, 1H, Fmoc Ar_2C***H***-CH_2-O-), 3.59 – 3.40 (m, 14H, -C***H*****$_2$**-O-), 3.38-3.30 (m, 2H, -CONH-C***H*****$_2$**-CH_2-O-), 3.26 (s, 3H, -O-C***H*****$_3$**), 2.87 (dd, J = 15.8, 5.2 Hz, 1H, -C_αH(COOAll)-C_β(***H*****$_a$**)H_b-CONH-), 2.68 (dd, J = 15.8, 4.3 Hz, 1H, -C_αH(COOAll)-C_β(H_a)***H*****$_b$**-CONH-). The full ^{1}H NMR spectrum for **1** is shown in Fig. S1.

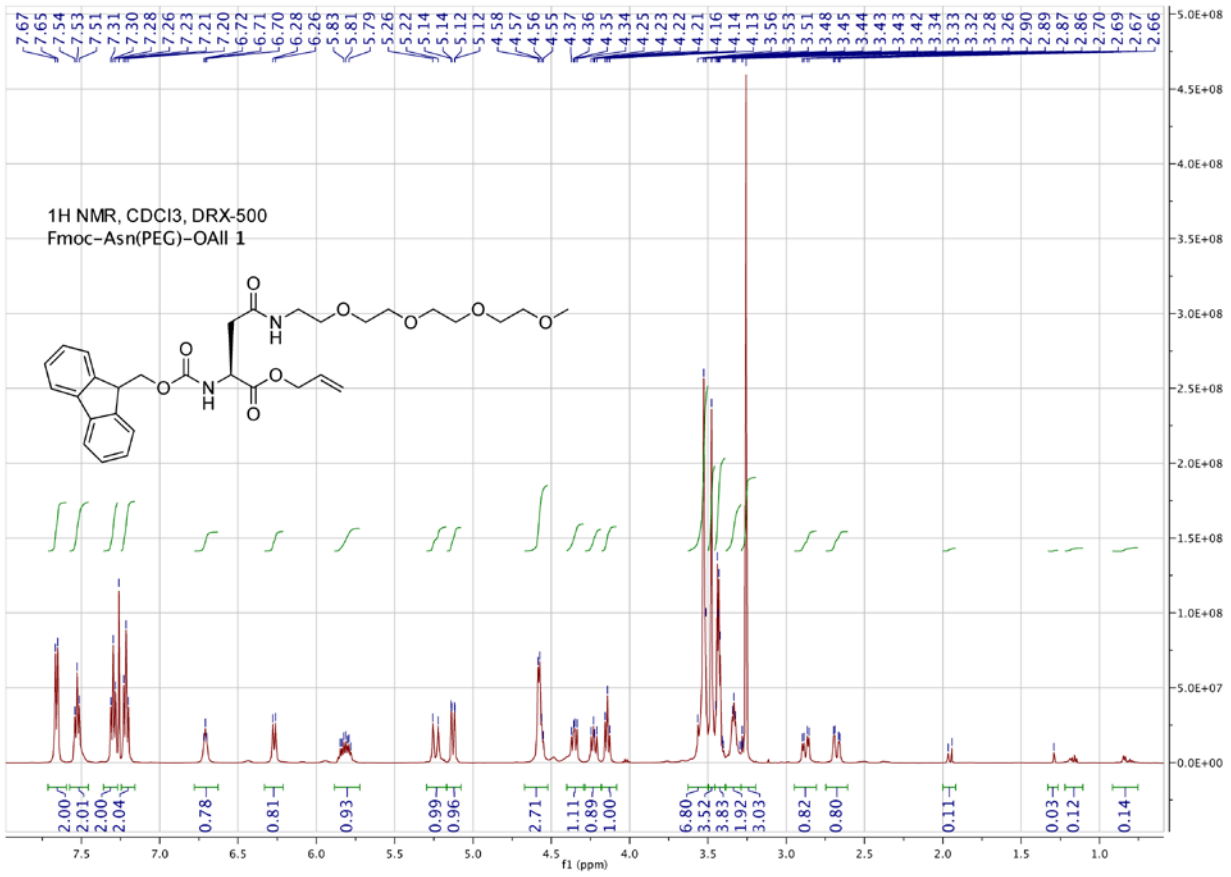


Figure S1. ^{1}H NMR spectrum for Fmoc-Asn(PEG)-OAll **1**.

^{13}C NMR (126 MHz, $CDCl_3$) δ 170.93, 169.78 (-$C_\beta H_2$-***C***ONH-, -C_αH(***C***OOAll)-$C_\beta H_2$-), 156.12 (Fmoc Ar_2-CH-CH_2-O-***C***ONH-), 143.80, 143.66, 141.10, 141.08 (four Fmoc aryl *ipso C*'s), 131.63 (-O-CH_2-***C***H=CH_2), 128.17 (benzene), 127.54, 126.94, 126.93, 125.10, 125.01, 119.79 (Fmoc aryl ***C***-H), 118.21 (-O-CH_2-CH=**C**H_2), 77.42, 77.16, 76.91 ($CDCl_3$), 71.68, 70.28,

70.22, 70.19, 70.02, 69.46 (-***C***H_2-O-), 66.91 (Fmoc Ar_2CH-***C***H_2-O-), 65.94 (-O-***C***H_2-CH=CH_2), 58.73 (-O-***C***H_3), 50.89 (-NH-***C***$_\alpha$H(COOAll)-C$_\beta H_2$-), 46.95 (Fmoc Ar_2***C***H-CH_2-O-), 39.23 (-CONH-***C***H_2-CH_2-O-), 37.26 (-C$_\alpha$H(COOAll)-***C***$_\beta H_2$-CONH-). The full ^{13}C NMR spectrum for **1** is shown in Fig. S2.

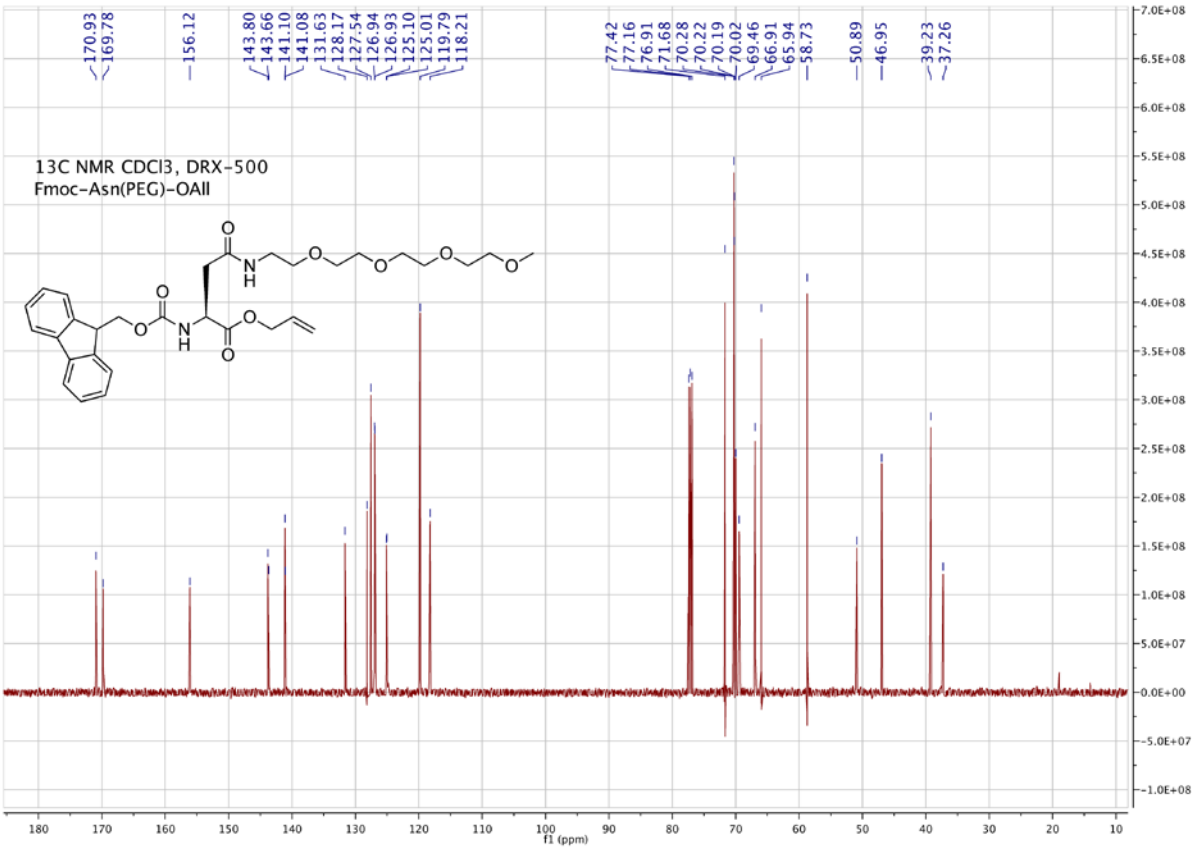


Figure S2. ^{13}C NMR spectrum for Fmoc-Asn(PEG)-OAll **1**.

Assignments of the ^{1}H and ^{13}C NMR spectra for **1** were made by analogy with published spectral datalle for related compounds,[1] and using a two-dimensional HMQC experiment (see Fig. S3) to identify the one-bond C-H correlations shown in Table S1.

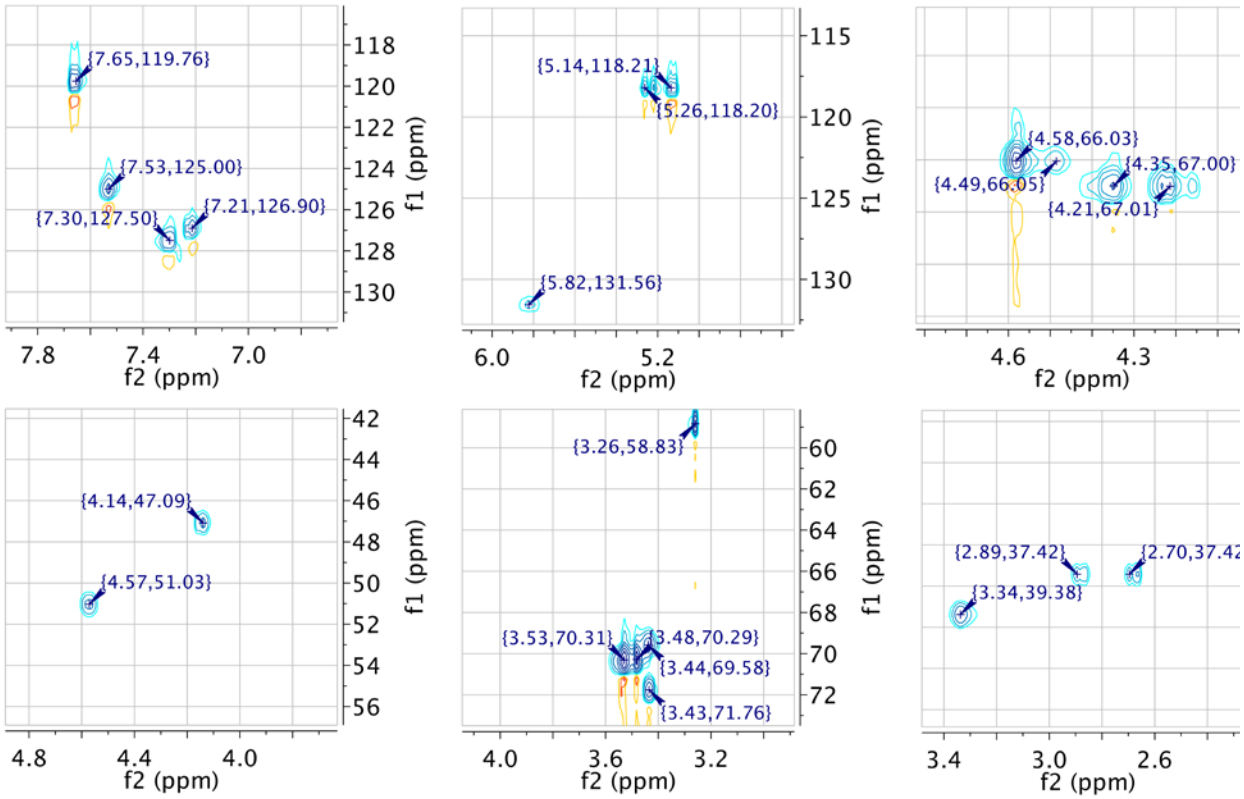


Figure S3. Two-dimensional HMQC spectrum of Fmoc-Asn(PEG)-OAll **1**.

Table S1. One-bond C-H correlations identified from HMQC experiment on Fmoc-Asn(PEG)-OAll **1**

^{1}H δ	^{13}C δ	Assignment
7.65	119.76	Fmoc aryl ***C-H***
7.53	125.00	Fmoc aryl ***C-H***
7.30	127.50	Fmoc aryl ***C-H***
7.21	126.90	Fmoc aryl ***C-H***
5.82	131.56	-O-CH_2-***CH***=CH_2
5.26	118.20	-O-CH_2-CH=***C(H$_{trans}$)H***$_{cis}$
5.14	118.21	-O-CH_2-CH=***C***(H_{trans})***H$_{cis}$***
4.58	66.03	-O-***C(H$_a$)***H_b-CH=CH_2
4.49	66.05	-O-***C***(H_a)***H$_b$***-CH=CH_2
4.57	51.03	-NH-***C$_α$H***(COOAll)-$C_βH_2$-
4.35	67.00	Fmoc Ar_2CH-***C(H$_a$)***H_b-O-
4.21	67.01	Fmoc Ar_2CH-***C***(H_a)***H$_b$***-O-
4.14	47.09	Fmoc Ar_2***CH***-CH_2-O-
3.53	70.31	-O-***CH$_2$***-
3.48	70.29	-O-***CH$_2$***-
3.44	69.58	-O-***CH$_2$***-
3.43	71.76	-O-***CH$_2$***-
3.34	39.38	-CONH-***CH$_2$***-CH_2-O-
3.26	58.83	-O-***CH$_3$***
2.89	37.42	-$C_α$H(COOAll)-***C$_β$(H$_a$)***H_b-CONH-
2.70	37.42	-$C_α$H(COOAll)-***C$_β$***(H_a)***H$_b$***-CONH-

High-resolution electrospray ionization time-of-flight mass spectrometry (ESI-TOF): calculated *m*/*z* for $C_{31}H_{41}N_2O_9$ $(M+H)^+$ is 585.2812, found 585.2835 (see Fig. S4).

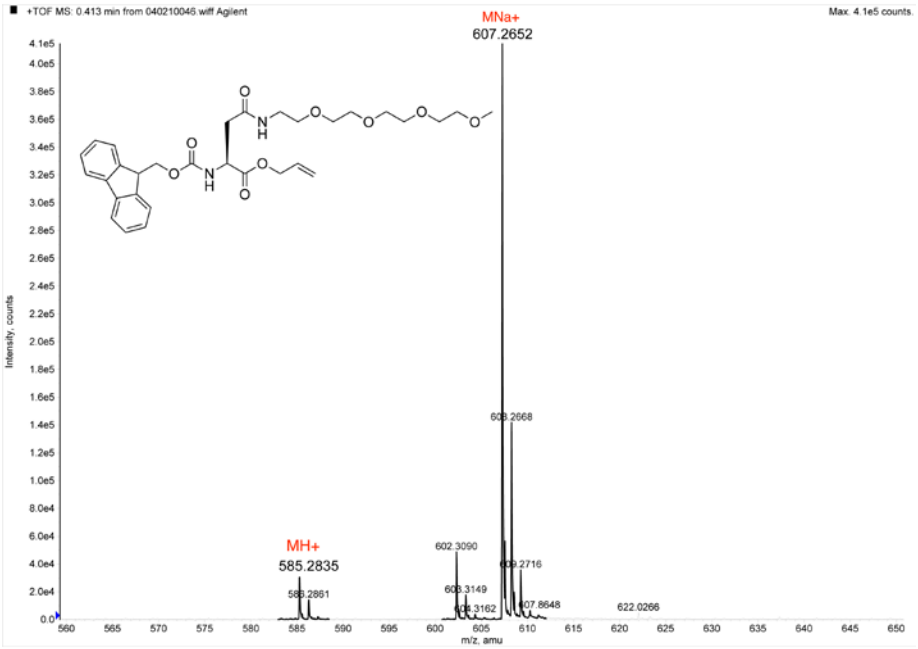


Figure S4. ESI-TOF data for Fmoc-Asn(PEG)-OAll **1**.

N^2-*fluorenylmethyoxycarbonyl*-N^4-*{11-methoxy-3,6,9-trioxaundecyl}*-L-*asparagine* ***2*** *(Fmoc-Asn(4-PEG)-OH)*

Procedure

Fmoc-Asn(4-PEG)-OH **2** was synthesized following the procedure of Herzner and Kunz[1]: to a solution of Fmoc-Asn(4-PEG)-OAll **1** (2.61 g, 4.46 mmol) in dry tetrahydrofuran (90 mL) was added N-methylaniline (725 µL, 0.717 g, 6.70 mmol) and tetrakis(triphenylphosphine)palladium(0) (62 mg, 1.2 mol %) under an argon atmosphere. The resulting mixture was stirred at room temperature for 22 h, with the flask covered in foil to exclude light. The solution was concentrated by rotary evaporation, and the resulting residue was diluted with dichloromethane (150 mL), and extracted with 0.25 N HCl (50 mL) and water (50 mL). The organic solution was dried with $MgSO_4$ and concentrated by rotary evaporation. The desired product was purified by flash chromatography over silica gel using hexanes (~500 mL), then ethyl acetate/hexanes (3:7 for ~500 ml, then 7:3 for ~400 mL), then acetic acid/ethyl

acetate/hexanes (1:30:70 for ~400 mL), then acetic acid/ethyl acetate (1:100, ~3000 mL) as eluents. The product was concentrated via rotary evaporation (chloroform was added to the concentrated product and then removed by rotary evaporation to remove residual ethyl acetate; benzene was subsequently added to the concentrated product and removed via rotary evaporation to remove residual ethyl acetate and acetic acid) and dried *in vacuo* to give a yellow oil (1.08 g, 1.98 mmol, 44% yield). $R_f = 0$ (a baseline TLC spot in 1:100 acetic acid/ethyl acetate).

Analytical Data

^{1}H NMR (500 MHz, $CDCl_3$) δ 9.14 (s, 1H, -NH-C_αH(COO***H***)-$C_\beta H_2$-), 7.68 (d, *J* = 7.5 Hz, 2H, Fmoc aryl C-***H***), 7.56 (t, *J* = 7.5 Hz, 2H, Fmoc aryl C-***H***), 7.39 – 7.30 (m, 2H, Fmoc aryl C-***H***), 7.29 – 7.15 (m, 2H, Fmoc aryl C-***H***), 6.44 (s, 1H, -CON***H***-C_αH(COOH)-), 4.58 (broad apparent s, 1H, -CONH-C_α***H***(COOH)-$C_\beta H_2$-), 4.33 (dd, ^2J = 10.2 Hz, ^3J = 8.1 Hz, 1H, Fmoc Ar_2CH-C(***H*****$_a$**)H_b-O-CONH-), 4.28 (dd, ^2J = 10.2 Hz, ^3J = 7.9 Hz, 1H, Fmoc Ar_2CH-C(H_a)***H*****$_b$**-O-CONH-), 4.15 (apparent t, *J* = 6.9 Hz, 1H, Fmoc Ar_2C***H***-CH_2-O-), 3.65-3.42 (m, 14H, -C***H*****$_2$**-O-), 3.41-3.32 (m, 2H, -CONH-C***H*****$_2$**-CH_2-O-), 3.28 (s, 3H, O-C***H*****$_3$**), 2.91 (apparent d, *J* = 11.7 Hz, 1H, -C_αH(COOH)-C_β(***H*****$_a$**)H_b-CONH-), 2.77 (apparent d, *J* = 8.7, 1H, -C_αH(COOH)-C_β(H_a)***H*****$_b$**-CONH-). Note the absence of a resonance for the Asn side-chain amide proton (-$C_\beta H_2$-CON***H***-CH_2-CH_2-O-), possibly due to fast exchange (ESI-TOF data for **2** shown below indicate that this proton is nonetheless present). Resonances for the following contaminants[2] were also observed: ethyl acetate [4.07 (q, *J* = 7.1 Hz, CH_3-C***H*****$_2$**-O-), 1.99 (s, C***H*****$_3$**-CO-), 1.21 (t, C***H*****$_3$**-CH_2-O)], hexanes [1.33 – 1.24 (m), 0.89 (m)], unidentified [1.73 – 1.60 (m), 1.45 – 1.35 (m)]. The full ^{1}H NMR spectrum for **2** is shown in Fig. S5.

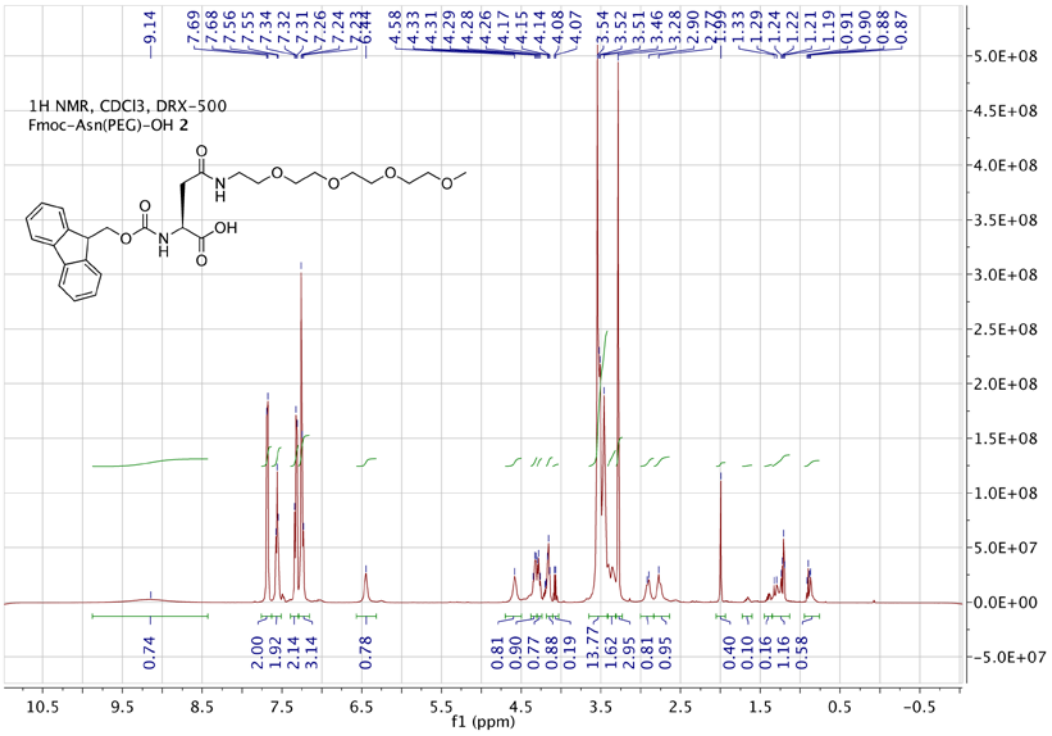


Figure S5. ^{1}H NMR spectrum for Fmoc-Asn(PEG)-OH **2**.

^{13}C NMR (126 MHz, $CDCl_3$) δ 173.16, 167.71 (-NH-C_αH(***C***OOH)-$C_\beta H_2$-, -$C_\beta H_2$-***C***ONH-CH_2-), 156.14 (Fmoc-O-***C***ONH-), 143.85, 143.72, 141.16, 141.14 (four Fmoc aryl *ipso* C's), 127.63, 127.05, 125.18, 125.12, 119.85 (Fmoc Ar ***C***-H), 77.42, 77.16, 76.91 ($CDCl_3$), 71.64, 70.31, 70.17, 70.13, 69.87, 69.32, (-***C***H_2-O-), 67.04 (Fmoc Ar_2CH-***C***H_2-O-), 58.77 (-O-***C***H_3), 50.89 (-NH-***C***$_\alpha$H(COOH)-$C_\beta H_2$-), 46.99 (Fmoc Ar_2***C***H-CH_2-O-), 39.45 (-NH-***C***H_2-CH_2-O-), 37.68 (-C_αH(COOH)-***C***$_\beta H_2$-CONH-). Resonances for the following contaminants[2] were also observed: benzene (128.26), ethyl acetate [171.10, (-CH_2-CH_2-OO***C***-CH_3), 60.35 (CH_3-***C***H_2-O-), 20.97 (***C***H_3-CO-), 14.13 (***C***H_3-CH_2-O-)], hexanes (30.30, 28.85, 23.69, 22.91, 14.00), unidentified [130.86, 128.74, 68.08, 38.66, 10.92]. The full ^{13}C NMR spectrum for **2** is shown in Fig. S6.

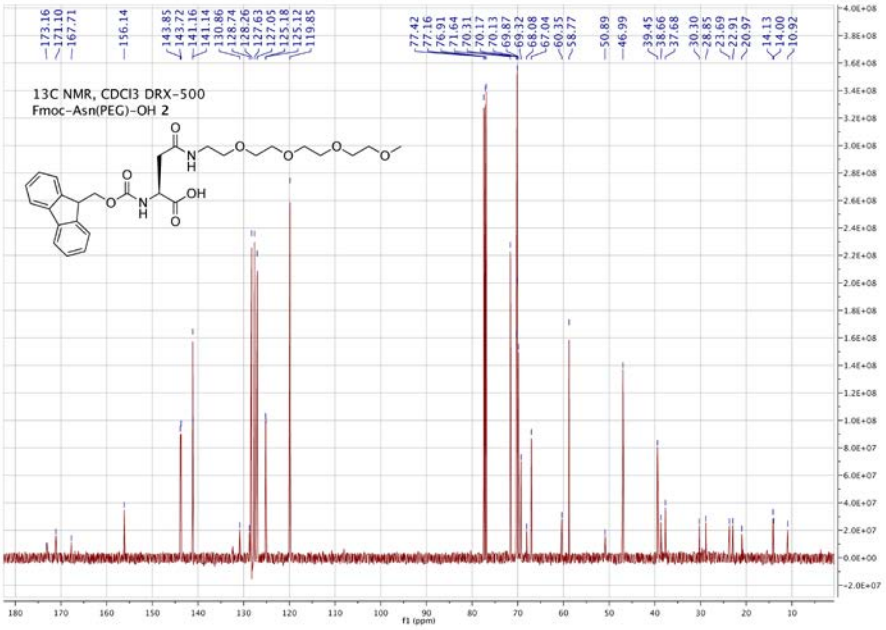


Figure S6. ^{13}C NMR spectrum for Fmoc-Asn(PEG)-OH **2**.

Assignments of the ^{1}H and ^{13}C NMR spectra for **2** were made by analogy with published spectral data for related compounds,[1] and using a two-dimensional HMQC experiment (see Fig. S7) to identify the one-bond C-H correlations shown in Table S2.

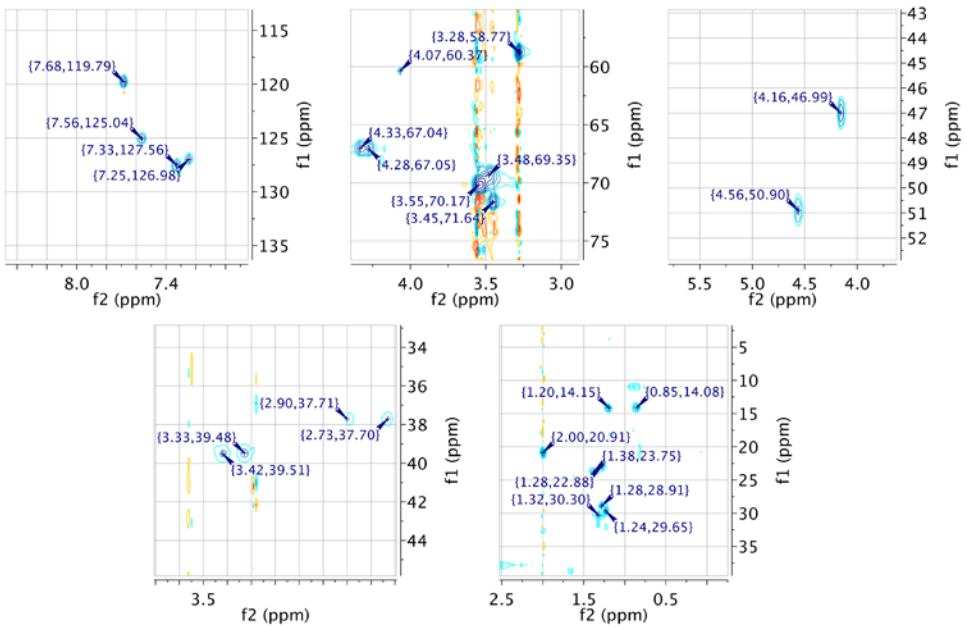


Figure S7. Two-dimensional HMQC spectrum of Fmoc-Asn(PEG)-OH **2**.

Table S2. One-bond C-H correlations identified from HMQC experiment on Fmoc-Asn(PEG)-OH **2**

^{1}H δ	^{13}C δ	Assignment
7.68	119.79	Fmoc aryl ***C-H***
7.56	125.04	Fmoc aryl ***C-H***
7.33	127.56	Fmoc Ar ***C-H***
7.25	126.98	Fmoc Ar ***C-H***
4.56	50.90	-NH-$\boldsymbol{C_{\alpha}H}$(COOH)-$C_{\beta}H_2$-
4.33	67.04	Fmoc Ar_2CH-$\boldsymbol{C(H_a)}H_b$-OOC-NH-
4.28	67.04	Fmoc Ar_2CH-$C(H_a)\boldsymbol{H_b}$-OOC-NH-
4.16	46.99	Fmoc $Ar_2\boldsymbol{CH}$-CH_2-O-
3.55	70.17	-O-$\boldsymbol{CH_2}$-
3.48	69.35	-O-$\boldsymbol{CH_2}$-
3.45	71.64	-O-$\boldsymbol{CH_2}$-
3.42	39.51	-CONH-$\boldsymbol{C(H_a)}H_b$-CH_2-O-
3.33	39.48	-CONH-$C(H_a)\boldsymbol{H_b}$-CH_2-O-
3.28	58.77	-O-$\boldsymbol{CH_3}$
2.90	37.71	-$C_{\alpha}H$-$\boldsymbol{C_{\beta}(H_a)}H_b$-CONH-
2.73	37.70	-$C_{\alpha}H$-$C_{\beta}(H_a)\boldsymbol{H_b}$-CONH-
4.07	60.45	EtOAc CH_3-$\boldsymbol{CH_2}$-OOC-CH_3
2.00	20.91	EtOAc CH_3-CH_2-OOC-$\boldsymbol{CH_3}$
1.38	23.75	Hexanes
1.32	30.30	Hexanes
1.28	28.91	Hexanes
1.28	22.88	Hexanes
1.24	29.65	Hexanes
1.20	14.15	EtOAc $\boldsymbol{CH_3}$-CH_2-OOC-CH_3
0.85	14.08	Hexanes

Small amounts of impurities (including benzene, ethyl acetate, and hexanes) were detected in **2** (as indicated by the ^{1}H and ^{13}C NMR spectra). Because these do not interfere with the use of **2** in subsequent peptide synthesis efforts (see below), no further purification was attempted.

ESI-TOF: *m*/*z* for $C_{28}H_{37}N_2O_9$ $(M+H)^+$; calc. 545.2499, found 545.25 (see Fig. S8).

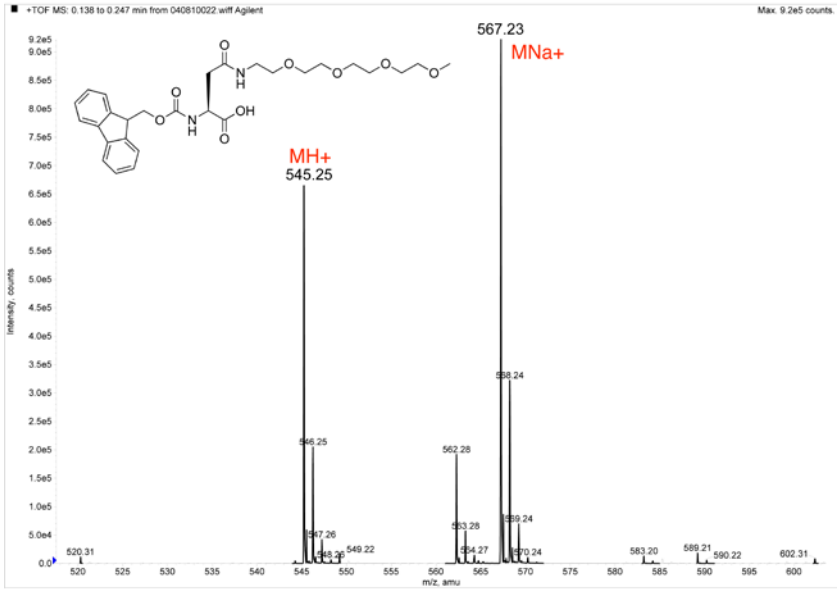


Figure S8. ESI-TOF data for Fmoc-Asn(PEG)-OH **2**.

Protein Synthesis

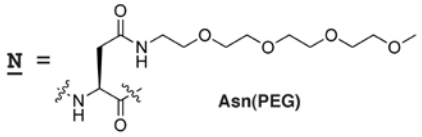
General

Proteins **6**, **6-F**, **6-T**, and **6-F,T**, were synthesized previously.[3] PEGylated proteins **6PEG**, **6PEG-F**, **6PEG-T**, and **6PEG-F,T** were synthesized as C-terminal acids, employing a solid phase peptide synthesis approach using a standard Fmoc Nα protecting group strategy either manually (**6PEG**, **6PEG-F**, **6PEG-T**) or via a combination of manual and automated peptide synthesis (**6PEG-F,T**) using an Applied Biosystems 433A automated peptide synthesizer). The complete amino acid sequences of **6PEG**, **6PEG-F**, **6PEG-T**, and **6PEG-F,T** are given in Table S3. 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 TGT resin and all Fmoc-protected α-amino acids (with acid-labile side-chain protecting groups) were purchased from EMD Biosciences, except for Fmoc-Asn(PEG)-OH, which was

synthesized as described above. Piperidine and N,N-diisopropylethylamine (DIEA) were obtained from Aldrich; N-methyl pyrrolidinone (NMP) was procured from Applied Biosystems.

Table S3. PEGylated WW Protein Sequences and MALDI-TOF Data

Protein	Sequence	Calc. $[M+H]^+$ (g/mol)	Found $[M+H]^+$ (g/mol)
6PEG	H_2N-KLPPGWEKRM**S**RS**N**G**R**VYYFNHITNASQFERPSG-COOH	4200.1	4199.3
6PEG-F	H_2N-KLPPGWEKRM**F**RS**N**G**R**VYYFNHITNASQFERPSG-COOH	4260.2	4261.2
6PEG-T	H_2N-KLPPGWEKRM**S**RS**N**G**T**VYYFNHITNASQFERPSG-COOH	4145.1	4145.8
6PEG-F,T	H_2N-KLPPGWEKRM**F**RS**N**G**T**VYYFNHITNASQFERPSG-COOH	4205.1	4205.0
	6 10 15 20 25 30 35 39		



A general protocol for manual solid phase peptide synthesis follows: Fmoc-Gly-loaded NovaSyn TGT resin (109 mg, 25 μmol at 0.23 mmol/g resin loading) was aliquotted into a fritted polypropylene syringe and allowed to swell in CH_2Cl_2 and 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 stirred at room temperature for 5 minutes. The deprotection solution was 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 resulting mixture was stirred at room temperature for 15 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 to a newly deprotected amine on resin, the desired Fmoc-protected amino acid (125 μmol, 5 eq) and HBTU (125 μmol, 5 eq) were dissolved by vortexing in 1.25 mL 0.1 M HOBt (125 μmol, 5 eq) in NMP. To the dissolved amino acid solution was added 44 μL DIEA (250 μmol, 10 eq). 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 resulting mixture was stirred at room temperature for at least 1 h. Selected amino acids (including Fmoc-Asn(PEG)-OH) were double coupled as needed to allow the coupling reaction to proceed to completion. 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 ~4 h 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 (~45 mL). Following centrifugation, the ether was decanted, and the pellet (containing the crude protein) was stored at -20 °C until purification.

HPLC Purification and MS Characterization of Pin WW variants

Immediately prior to purification, the crude proteins were dissolved in 1:1 water:acetonitrile. Proteins were purified by reverse-phase HPLC on a C18 column using a linear gradient of water in acetonitrile with 0.2% v/v TFA. HPLC fractions containing the desired protein product were pooled, frozen, and lyophilized. Proteins were identified by matrix-assisted laser desorption/ionization time-of-flight spectrometry (MALDI-TOF, Table S3 and Figs. S9-12), and purity was established by analytical HPLC (Figs. S13-S16).

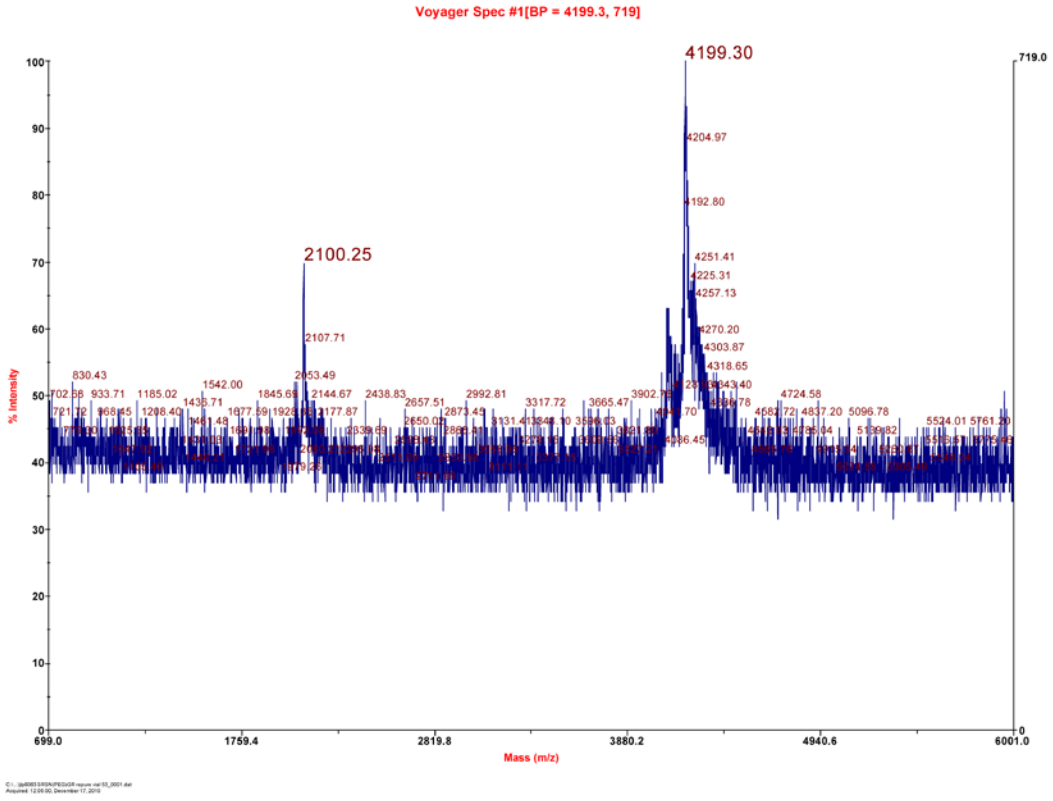


Figure S9. MALDI-TOF data for PEGylated protein **6PEG**.

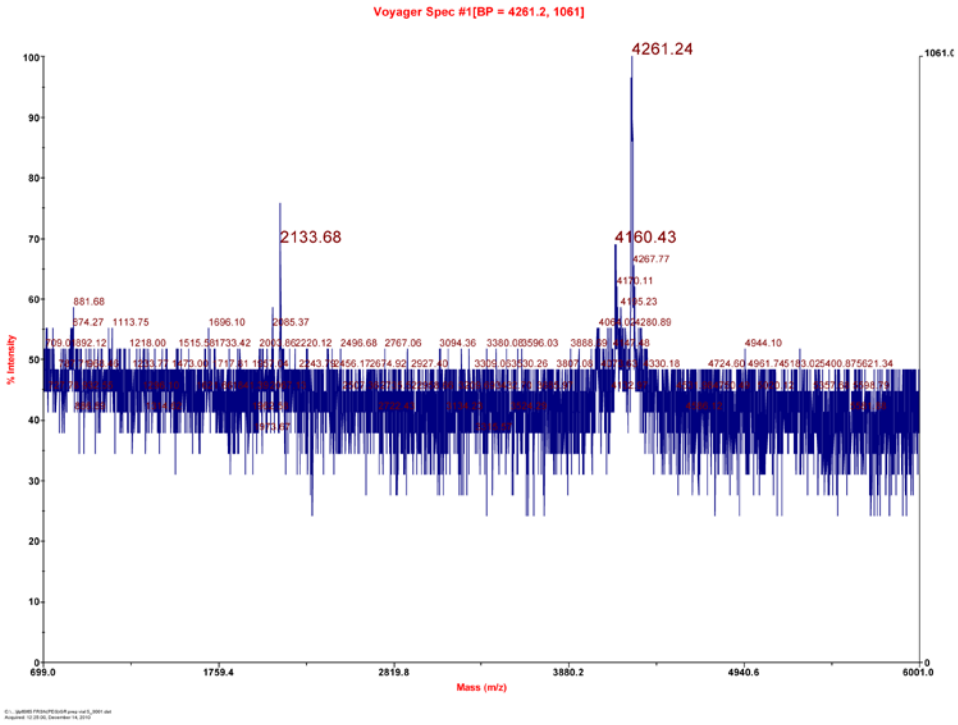


Figure S10. MALDI-TOF data for PEGylated protein **6PEG-F**.

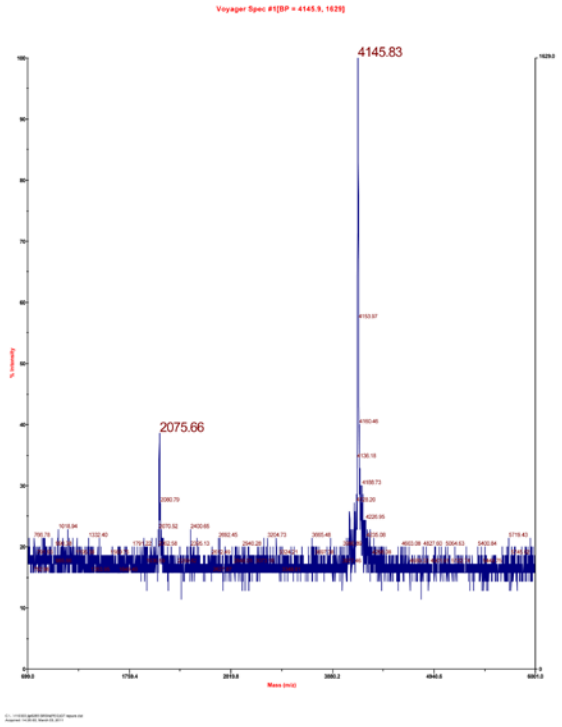


Figure S11. MALDI-TOF data for PEGylated protein **6PEG-T**.

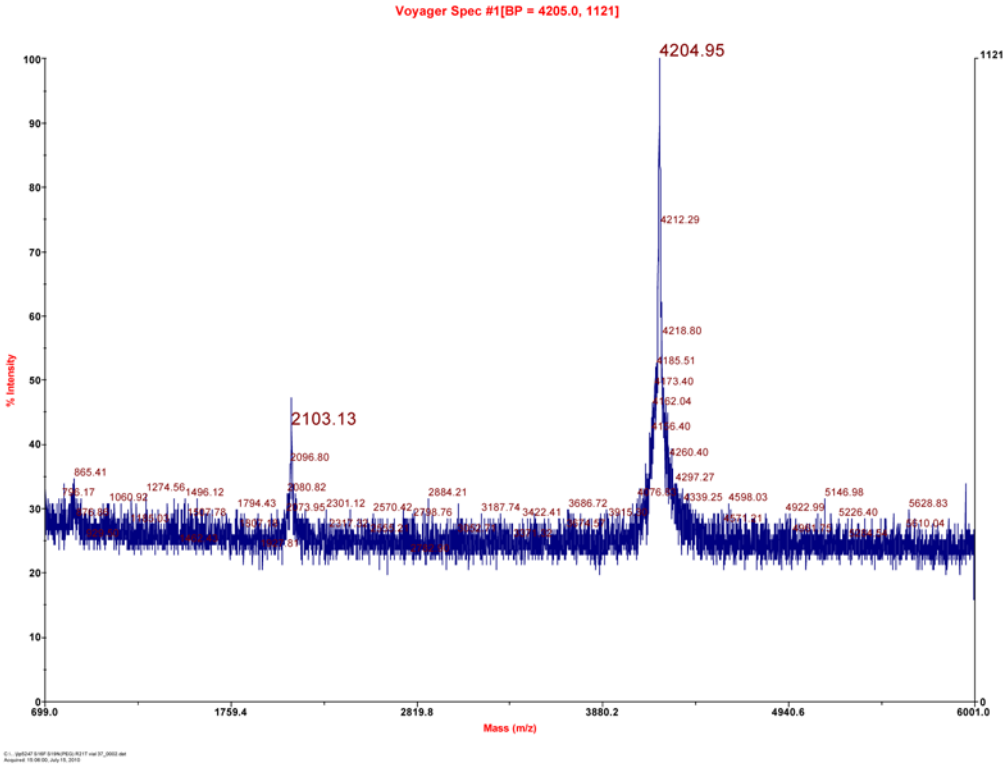


Figure S12. MALDI-TOF data for PEGylated protein **6PEG-F,T**.

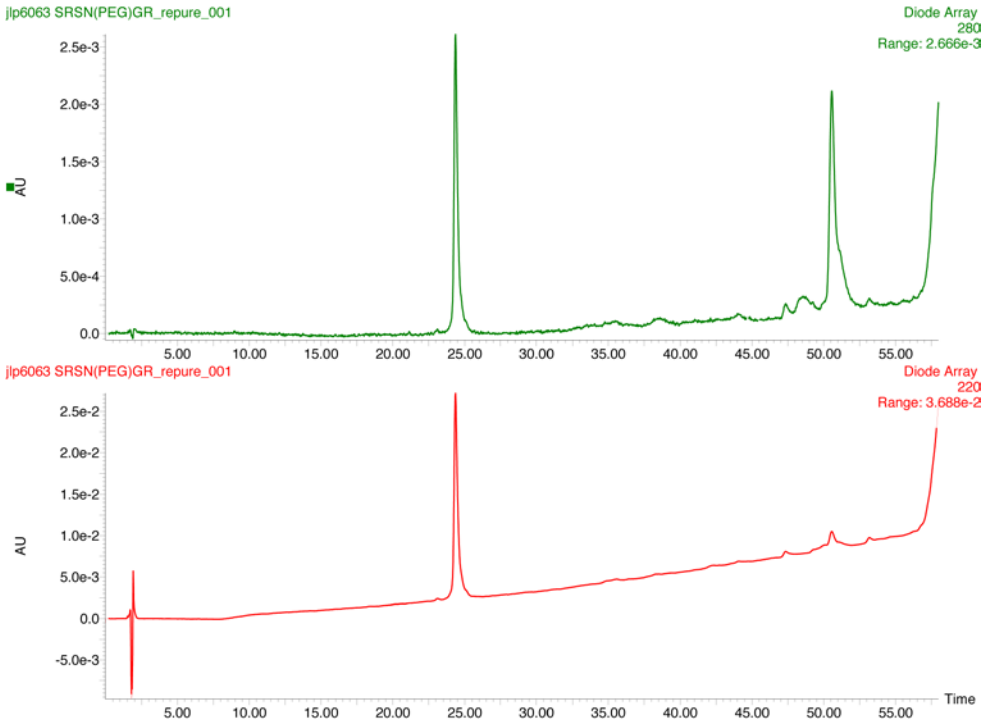


Figure S13. Analytical HPLC data for PEGylated protein **6PEG** after two successive rounds of preparative purification. Sample was injected onto a C18 analytical column (4.6 mm × 10 cm) and eluted using a linear gradient of 10-60% acetonitrile in water (constant 0.2% TFA) over 50 minutes. Desired product elutes at ~24.5 minutes. Peak at ~50 minutes is likely a contaminant from previous uses of the analytical column.

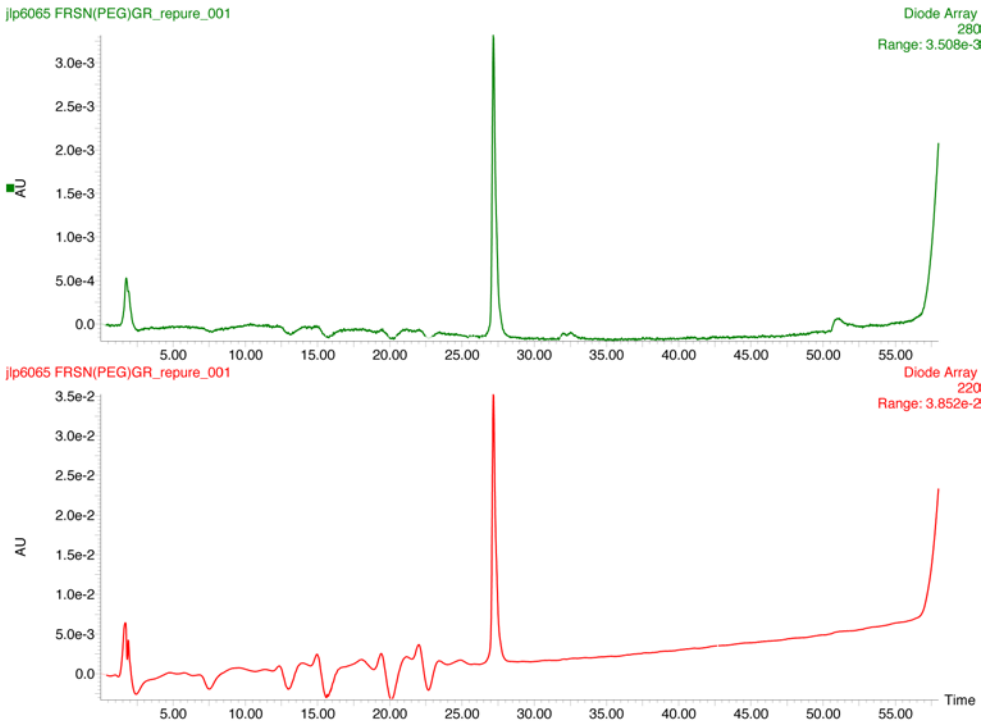


Figure S14. Analytical HPLC data for PEGylated protein **6PEG-F** after two successive rounds of preparative purification. Sample was injected onto a C18 analytical column (4.6 mm × 10 cm) and eluted using a linear gradient of 10-60% acetonitrile in water (constant 0.2% TFA) over 50 minutes. Desired product elutes at ~27.5 minutes.

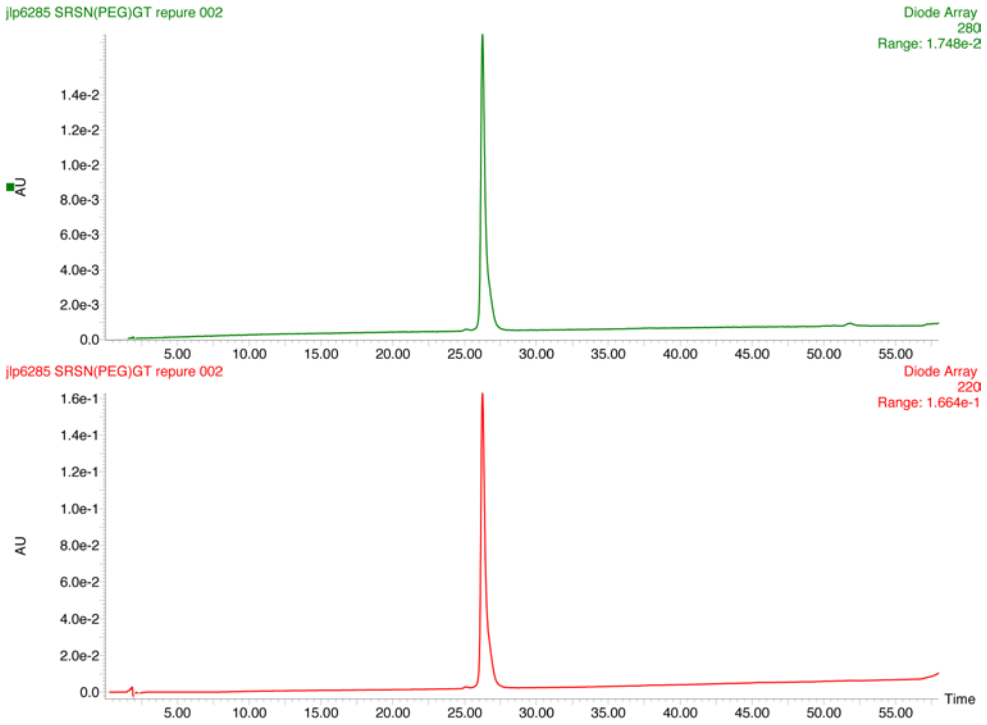


Figure S15. Analytical HPLC data for PEGylated protein **6PEG-T** after two successive rounds of preparative purification. Sample was injected onto a C18 analytical column (4.6 mm × 10 cm) and eluted using a linear gradient of 10-60% acetonitrile in water (constant 0.2% TFA) over 50 minutes. Desired product elutes at ~26.5 minutes.

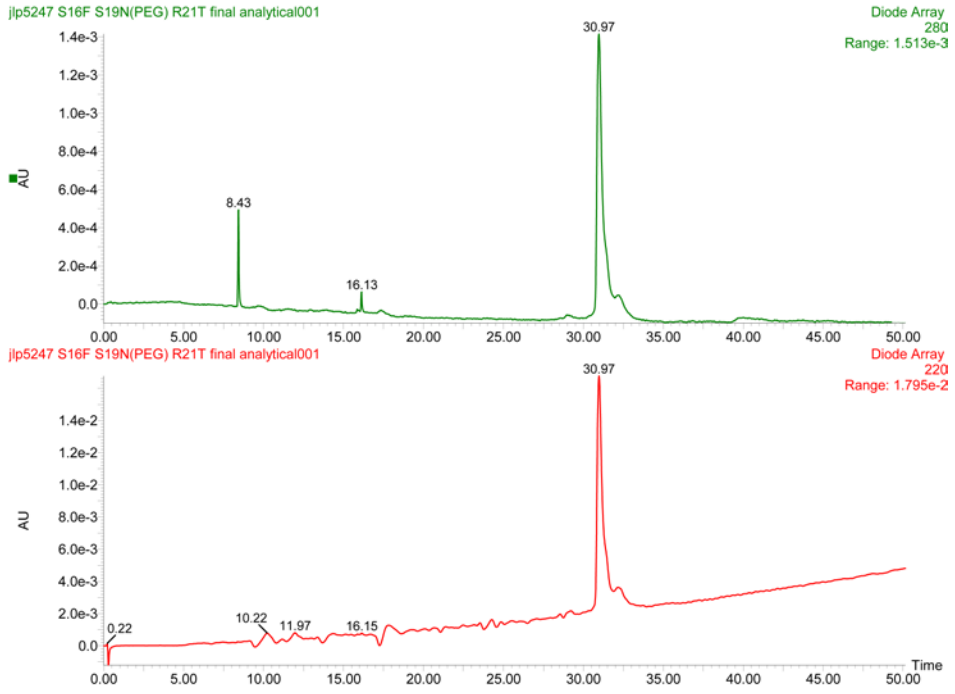


Figure S16. Analytical HPLC data for PEGylated protein **6PEG-F,T** after one round of preparative purification. Sample was injected onto a C18 analytical column (4.6 mm × 5 cm) and eluted using a linear gradient of 10-60% acetonitrile in water (constant 0.2% TFA) over 50 minutes. Desired product elutes at ~31 minutes.

Circular Dichroism Spectropolarimetry

Measurements were made with an Aviv 62A DS Circular Dichroism Spectropolarimeter, using quartz cuvettes with a path length of 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 $M^{-1}cm^{-1}$).[4] CD spectra were obtained by monitoring molar ellipticity from 340 to 200 nm, with 5 second averaging times. Variable temperature CD data were obtained by monitoring molar ellipticity at 227 nm from 0.2 to 98.2 °C at 2 °C intervals, with 90 s equilibration time between data points and 30 s averaging times.

Variable temperature CD data were fit to the following model for two-state thermally induced unfolding transitions:

$$[\theta] = \frac{(D_0 + D_1 \cdot T) + K_f(N_0 + N_1 \cdot T)}{1 + K_f} \quad \text{(S1)}$$

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 $\Delta G_f(T)$ according to the following equation:

$$K_f = \exp\left[\frac{-\Delta G_f(T)}{RT}\right] \quad \text{(S2)}$$

where R is the universal gas constant (0.0019872 kcal/mol/K). The midpoint of the thermal unfolding transition (or melting temperature T_m) was calculated by fitting $\Delta G_f(T)$ to either of two equations. The first equation is derived from the van't Hoff relationship:

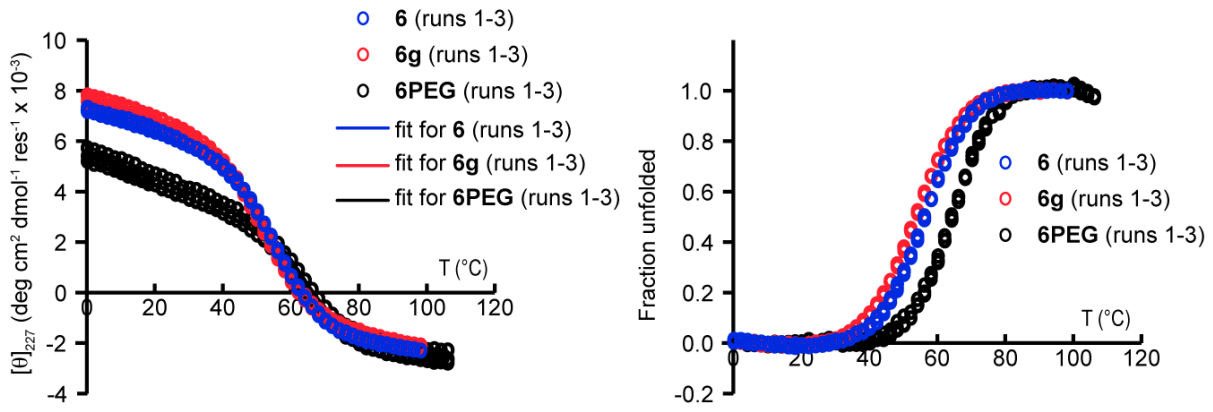
$$\Delta G_f(T) = \frac{\Delta H(T_m)}{T_m}(T_m - T) + \Delta C_p\left[T - T_m - T\ln\left(\frac{T}{T_m}\right)\right] \quad \text{(S3)}$$

where $\Delta H(T_m)$ is the enthalpy of folding at the melting temperature and ΔC_p is the heat capacity of folding ($\Delta H(T_m)$, ΔC_p, and T_m are parameters of the the fit). The second equation represents $\Delta G_f(T)$ as a Taylor series expansion about the melting temperature:

$$\Delta G_f(T) = \Delta G_0 + \Delta G_1(T - T_m) + \Delta G_2(T - T_m)^2 \quad \text{(S4)}$$

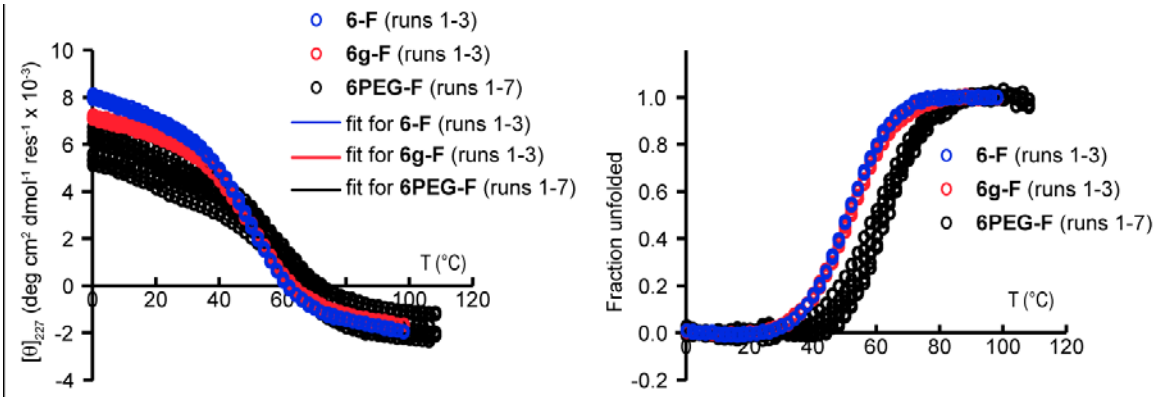
where ΔG_0, ΔG_1, and ΔG_2 are parameters of the fit and T_m is a constant obtained from the van't Hoff fit (in equation S3). The ΔG_f values displayed in the main text for each Pin WW protein were obtained by averaging the ΔG_f values (calculated at 338.15 K using equation 3) from each of three or more replicate variable temperature CD experiments on the same protein.

CD spectra for Pin WW proteins **6PEG**, **6PEG-F**, **6PEG-T**, and **6PEG-F,T** appear in Fig. S17. Variable temperature CD data for proteins **6PEG**, **6PEG-F**, **6PEG-T**, and **6PEG-F,T** appear in Figs. S18-S21 (previously reported[3,5] variable temperature CD data for non-glycosylated proteins **6**, **6-F**, **6-T**, and **6-F,T** and for GlcNAc-containing glycoproteins **6g**, **6g-F**, **6g-T**, and **6g-F,T** are included in Figs. S18-S21 for comparison). Figs. S18-S21 also shows values for the parameters from equations S1−S4 that were used to fit the variable temperature CD data. The standard error for each fitted parameter is also shown. These standard parameter errors were used to estimate the uncertainty in the average ΔG_f values given in the main text by propagation of error.



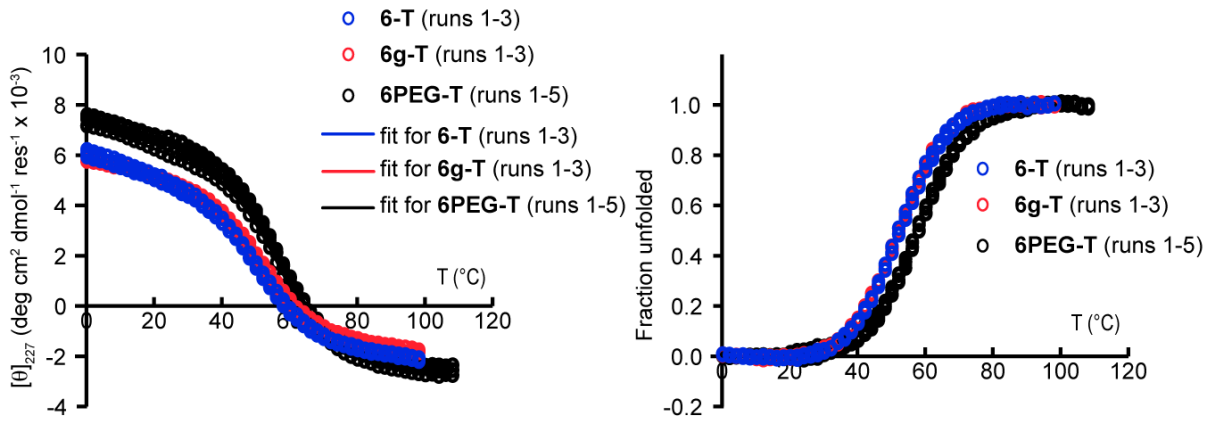
	ΔH(Tm) (kcal/mol)	ΔCp (kcal/mol/K)	Tm (K)	ΔG0 (kcal/mol)	ΔG1 (kcal/mol/K)	ΔG2 (kcal/mol/K²)	N0 ([θ])	N1 ([θ]/K)	D0 ([θ])	D1 ([θ]/K)
6PEG										
run 1	-40.5 (1.2)	-0.4 (0.3)	337.4 (0.4)	-0.0005 (0.0445)	0.120 (0.004)	0.0004 (0.0005)	20.7	-0.054	6.8	-0.024
run 2	-40.6 (1.2)	-0.3 (0.3)	337.7 (0.4)	-0.0004 (0.0422)	0.120 (0.004)	0.0004 (0.0005)	21.0	-0.057	6.8	-0.026
run 3	-38.7 (1.0)	-0.5 (0.3)	337.2 (0.4)	-0.0008 (0.0460)	0.115 (0.003)	0.0007 (0.0004)	18.9	-0.050	5.6	-0.022

Figure S18. Variable temperature CD data and fit parameters (see equations S1-S4) for PEGylated protein **6PEG** (in which position 19 is Asn-PEG) at 10 μM protein concentration in 20 mM sodium phosphate buffer, pH 7. Also included are previously reported (ref. 5) variable temperature CD data for non-glycosylated protein **6** (in which position 19 is Asn), and for GlcNAc-containing glycoprotein **6g** (in which position 19 is Asn-GlcNAc).



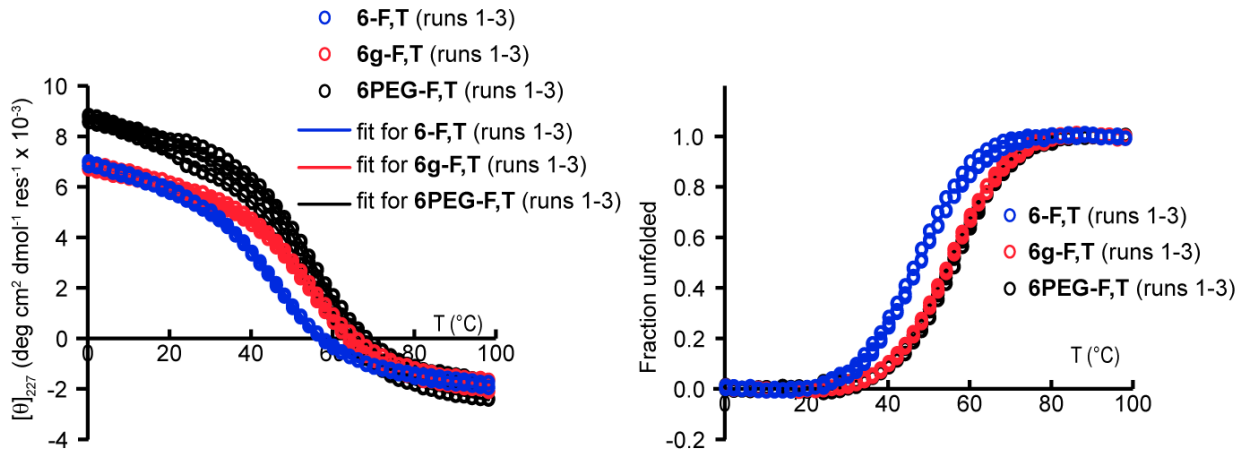
	ΔH(Tm) (kcal/mol)	ΔCp (kcal/mol/K)	Tm (K)	ΔG0 (kcal/mol)	ΔG1 (kcal/mol/K)	ΔG2 (kcal/mol/K²)	N0 ([θ])	N1 ([θ]/K)	D0 ([θ])	D1 ([θ]/K)
6PEG-F										
run 1	-33.8 (1.1)	-0.1 (0.2)	334.6 (0.5)	-0.0003 (0.0497)	0.101 (0.003)	0.0001 (0.0003)	21.8	-0.05	5.4	-0.02
run 2	-38.8 (2.9)	0.3 (0.5)	336.4 (0.7)	-0.0004 (0.0823)	0.115 (0.008)	-0.0005 (0.0008)	25.9	-0.07	3.9	-0.02
run 3	-38.4 (4.4)	0.4 (0.8)	336.9 (1.1)	-0.0005 (0.1250)	0.114 (0.013)	-0.0006 (0.0012)	27.1	-0.07	3.2	-0.01
run 4	-32.3 (1.0)	-0.1 (0.2)	333.6 (0.5)	-0.0003 (0.0494)	0.097 (0.003)	0.0001 (0.0003)	17.1	-0.04	5.7	-0.02
run 5	-36.5 (1.5)	0.1 (0.3)	335.7 (0.5)	-0.0002 (0.0533)	0.109 (0.005)	-0.0001 (0.0005)	20.4	-0.06	3.0	-0.01
run 6	-36.5 (1.4)	0.1 (0.3)	335.6 (0.4)	-0.0003 (0.0486)	0.109 (0.004)	-0.0002 (0.0004)	21.1	-0.05	5.0	-0.02
run 7	-29.9 (1.2)	-0.2 (0.3)	332.3 (1.0)	-0.0006 (0.0890)	0.090 (0.004)	0.0003 (0.0005)	12.6	-0.03	5.9	-0.02

Figure S19. Variable temperature CD data and fit parameters (see equations S1-S4) for PEGylated protein **6PEG-F** (in which position 19 is Asn-PEG) at 10 μM protein concentration in 20 mM sodium phosphate buffer, pH 7. Also included are previously reported (ref. 5) variable temperature CD data for non-glycosylated protein **6-F** (in which position 19 is Asn), and for GlcNAc-containing glycoprotein **6g-F** (in which position 19 is Asn-GlcNAc).



	ΔH(T$_m$) (kcal/mol)	ΔC$_p$ (kcal/mol/K)	T$_m$ (K)	ΔG$_0$ (kcal/mol)	ΔG$_1$ (kcal/mol/K)	ΔG$_2$ (kcal/mol/K^2)	N$_0$ ([θ])	N$_1$ ([θ]/K)	D$_0$ ([θ])	D$_1$ ([θ]/K)
6PEG-T										
run 1	-30.9 (0.7)	-0.1 (0.2)	330.5 (0.4)	-0.0002 (0.0399)	0.094 (0.002)	0.0001 (0.0002)	17.3	-0.04	4.6	-0.02
run 2	-30.2 (0.5)	-0.1 (0.1)	331.3 (0.3)	-0.0001 (0.0296)	0.091 (0.001)	0.0001 (0.0002)	20.0	-0.05	3.1	-0.01
run 3	-29.0 (0.8)	-0.3 (0.2)	330.9 (0.6)	-0.0008 (0.0529)	0.088 (0.002)	0.0005 (0.0003)	19.0	-0.04	3.2	-0.01
run 4	-31.3 (0.5)	-0.2 (0.1)	331.5 (0.3)	-0.0001 (0.0302)	0.094 (0.002)	0.0003 (0.0002)	20.4	-0.05	3.4	-0.02
run 5	-30.6 (0.5)	-0.2 (0.1)	330.5 (0.4)	-0.0003 (0.0346)	0.093 (0.002)	0.0003 (0.0002)	18.7	-0.04	5.2	-0.02

Figure S20. Variable temperature CD data and fit parameters (see equations S1-S4) for PEGylated protein **6PEG-T** (in which position 19 is Asn-PEG) at 10 μM protein concentration in 20 mM sodium phosphate buffer, pH 7. Also included are previously reported (ref. 5) variable temperature CD data for non-glycosylated protein **6-T** (in which position 19 is Asn), and for GlcNAc-containing glycoprotein **6g-T** (in which position 19 is Asn-GlcNAc).



	ΔH(T$_m$) (kcal/mol)	ΔC$_p$ (kcal/mol/K)	T$_m$ (K)	ΔG$_0$ (kcal/mol)	ΔG$_1$ (kcal/mol/K)	ΔG$_2$ (kcal/mol/K^2)	N$_0$ ([θ])	N$_1$ ([θ]/K)	D$_0$ ([θ])	D$_1$ ([θ]/K)
6PEG-F,T										
run 1	-32.0 (1.2)	-0.2 (0.3)	328.7 (0.8)	-0.0022 (0.2506)	0.091 (0.011)	0.0008 (0.0014)	21.5	-0.05	8.7	-0.03
run 2	-32.1 (1.0)	-0.1 (0.3)	328.9 (0.6)	0.0370 (0.0592)	0.097 (0.003)	0.0003 (0.0004)	23.4	-0.05	9.3	-0.03
run 3	-32.2 (0.9)	-0.3 (0.3)	329.6 (0.7)	0.0129 (0.0645)	0.099 (0.003)	0.0005 (0.0004)	27.2	-0.07	7.6	-0.03

Figure S21. Variable temperature CD data and fit parameters (see equations S1-S4) for PEGylated protein **6PEG-F,T** (in which position 19 is Asn-PEG) at 10 μM protein concentration in 20 mM sodium phosphate buffer, pH 7. Also included are previously reported (ref. 5) variable temperature CD data for non-glycosylated protein **6-F,T** (in which position 19 is Asn), and for GlcNAc-containing glycoprotein **6g-F,T** (in which position 19 is Asn-GlcNAc).

References

(1) Herzner, H.; Kunz, H. *Carbohydrate Res.* **2007**, *342*, 541-557.
(2) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62*, 7512-7515.
(3) 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*, 571-575.
(4) Edelhoch, H. *Biochemistry* **1967**, *6*, 1948–1954.
(5) Price, J. L.; Powers, D. L.; Powers, E. T.; Kelly, J. W. *Proc. Natl. Acad. Sci. USA* **2011**, 108,14127-14132.