Supporting Information

Chen et al. 10.1073/pnas.1015356107

SI Methods

General HPLC Purification and Characterization Procedure. All final peptide and protein purification were carried out on a preparative reverse phase HPLC (C_{18}) with a gradient of increasing eluent A (acetonitrile, 0.1% TFA) in eluent B (water, 0.1% TFA). The standard gradient proceeds from 5:95 A:B to 95:5 A:B. The length of the gradient was adjusted based on the reaction to allow for adequate separation of peaks, from 30 to 50 min. Elution from the column was monitored by UV absorbance at 280 and 228 nm.

All purified Im7 variants were characterized by analytical reverse phase HPLC (C_{18}) with the same solvent gradient as described to assess their level of purity.

Synthesis and Purification of the Im7(C59-G87) Glycopeptides. The Im7(C59-G87) glycopeptides (Fig. S1) were synthesized on Fmoc-Gly-NovaSyn TGT resin (NovaBiochem) on an ABI 431A automated peptide synthesizer (Applied Biosystems) using standard Fmoc (9-fluoenylmethoxycarbonyl) coupling procedures. Fmoc deprotection was carried out using 20% 4-methylpiperidine in NMP (1-methyl-2-pyrrolidinone). Double coupling procedure was carried out for each amino acid using 4 eq Fmoc amino acid per eq resin, 4 eq *N*-hydroxybenzotriazole, and 4 eq 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate as activating agents, with DIPEA (diisopropylethylamine) in NMP, for 1 h at room temperature. Acetic anhydride capping was carried out to terminate unreacted free amines after each amino acid coupling. The coupling of the Fmoc-Asn (Chitobiose-$TBDMS_5$)-OH building block (1), as well as the two subsequent amino acids, were carried out by hand using 4 eq of the Fmoc amino acid, and 4 eq 7-azabenzotriazol-1-yloxy-tris (pyrrolidino)phosphonium hexafluorophosphate as the coupling agent, with DIPEA in DMF (dimethylformamide). The completion of each reaction was verified by using 2,4,6-trinitrobenzenesulfonic acid to check for the presence of a free amino group.

Cleavage of the peptide from the resin and global deprotection was carried out using cleavage cocktail K [82% TFA, 3% EDT (1,2 ethanedithiol), 5% thioanisol, 5% phenol, and 5% water], shaking for 2 h at room temperature. The desired glycopeptide was precipitated in cold diethyl ether, purified by preparative reverse phase C_{18} HPLC, confirmed by electrospray ionization mass spectrometry (ESI-MS) to be within 1 Da of expected mass, analyzed by RP-HPLC and shown to be ≥95% pure, and lyophilized for storage.

Synthesis and Purification of the Im7(MEH_6E2-A28) Glycopeptide Thioesters. The N-terminal His-tagged Im7(MEH_6E2-A28) glycopeptides (Fig. S1) were synthesized on Fmoc-Ala-NovaSyn TGT resin (NovaBiochem) using the same protocol as described for the Im7(C59-G87) peptide (above). Upon completion of the synthesis, the glycopeptide was cleaved from the resin using cleavage cocktail A (0.5% TFA, 99.5% dichloromethane), shaking for 2 h at room temperature. These conditions resulted in a peptide with a free carboxyl terminus and fully protected side-chain functionality. This peptide was dried under reduced pressure, washed with diethyl ether, redissolved in anhydrous DMF, and treated with 10 eq of benzyl mercaptan, 3 eq of benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate, followed by 3 eq of DIPEA. The mixture was stirred for 8 h under N_2. The DMF solvent was removed under high vacuum, and the resulting yellow oil was treated for 2 h at room temperature with cleavage cocktail K for global deprotection. The desired glycopeptide thioesters were precipitated in cold diethyl ether, purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, shown by RP-HPLC analysis to be ≥95% pure, and lyophilized for storage.

General Information for All Protein Procedures. All buffer solutions were filtered before use (Millipore 0.2 μM). Final protein concentrations were determined by absorbance at 280 nm (Im7 ε_{280} nm $= 9{,}700\ M^{-1}\,cm^{-1}$).

Mutagenesis of Im7 Nonglycosylated Pseudo-Wild-Type Variants. Introduction of site specific mutations into Im7 was carried out by using the Quikchange mutagenesis kit (Stratagene) and the pTrc(Im7)-99A vector as the template. The pTrc(Im7)-99A vector encodes the native Im7 protein with an N-terminal hexahistidine tag, described previously (2). All variants were sequenced to ensure that the gene contained the desired change and no others.

Expression and Purification of Nonglycosylated Im7 Variants. Starting from a 5 mL overnight culture, BL21(DE3) cells (Stratagene) encoding the gene for the Im7 nonglycosylated variants were grown at 37 °C in 1 L of LB broth containing carbenicillin antibiotic (100 μg/mL) to an OD_{600} of 0.6–0.8. At that point protein production was induced by the addition of 1 mM IPTG. After 3 h, the cells were harvested by centrifugation (7,500 × *g*) for 30 min, washed once with a 0.9% NaCl solution, centrifuged again (7,500 × *g*) for 30 min, and the cell pellet was frozen at −80 °C until needed.

All purification steps were performed at 4 °C. Frozen cell pellets from the Im7 expression were thawed and resuspended in 5% of the original culture volume in buffer A (50 mM Tris-acetate, 5 mM imidazole, pH 8). The cells were lysed by sonication, followed by centrifugation (142,400 × *g*) for 1 h to remove cellular debris and membrane fractions. The resulting supernatant was slowly applied to a column of Ni-nitrilotriacetate (Ni-NTA) agarose equilibrated with buffer A. After washing with 5 column volumes of buffer A, the purified protein was eluted with buffer B (50 mM Tris-acetate, 250 mM imidazole, pH 8). Fractions containing a significant amount of desired protein were combined and purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, shown by RP-HPLC to be ≥95% pure, and lyophilized for storage. Typical yield was approximately 15 mg/L of expression volume.

Cloning of the Intein Im7(C29-G87) Construct [Chitin Binding Domain-Intein-Im7(C29-G87)]. Starting from the pTrc($Im7_{A29C}$)-99A vector, the Im7(C29-G87) gene fragment was amplified using primers which contained 5′ Sap1 and 3′ BamH1 restriction sites and the Vent polymerase (New England Biolabs) under standard conditions described by the manufacturer for 30 cycles. Both the amplicons and the IMPACT pTWIN1 vector were then doubly digested with Sap1 and BamH1 restriction enzymes (New England Biolabs), separated by agarose gel electrophoresis, and purified with the QIAquick gel extraction kit (Qiagen). Ligations of the amplified fragment into the cut vector were conducted with the T4 DNA ligase kit (Promega) for 1.5 h at 16 °C. The ligation product was transformed into a competent DH5α strain of *Escherichia coli* (Invitrogen), spread on plates containing LB media and carbenicillin. The desired pTWIN1(Im7C29-G87) plasmid from colonies formed on the plates was isolated, verified by sequencing and transformed into a competent BL21(DE3) strain of *E. coli* for protein expression.

Expression and Purification of Im7(C29-G87) Peptide. Starting from a 5 mL overnight culture, BL21(DE3) cells (Stratagene) expressing pTWIN1(Im7(C29-G87) (Fig. S2*A*) were grown at 37 °C in LB broth containing carbenicillin antibiotic to an OD_{600} of 0.6–0.8. At that point protein production was induced by the addition of 1 mM IPTG. Expression at this temperature affords the overexpressed protein as inclusion bodies. After 3 h, the cells were harvested by centrifugation (7,500 × *g*) for 30 min, washed once with a 0.9% NaCl solution, centrifuged again (7,500 × *g*) for 30 min, and the cell pellet was frozen at −80 °C until needed.

All purification steps were performed at 4 °C. Cell pellets containing the $Im7_{C29-G87}$ construct were thawed and resuspended in 5% of the original culture volume in buffer (20 mM Tris • HCl, 500 mM NaCl, pH 8.5). The cells were disrupted by sonication and centrifuged (142,400 × *g*) for 1 h to spin down the cellular debris and inclusion bodies containing the desired protein product. The isolated pellet was then homogenized into 10% of the original culture volume in buffer (20 mM Tris • HCl, 500 mM NaCl, 7 M guanidinium chloride, pH 8) and stirred for 1 h. The cellular debris was removed from the dissolved protein by centrifugation (142,400 × *g*) for 1 h. The supernatant was dialyzed into refolding buffers R-1/2/3/4/5/6 (20 mM Tris • HCl, 500 mM NaCl, 8/6/4/2/0/0 M urea, 1 mM DTT, pH 8.5), which contained sequentially diluted concentrations of urea, for 12 h in each buffer. The dialyzed product was then added to a column of chitin resin (New England Biolabs), of 1% the original culture volume and prewashed with buffer R-6, at a flow rate of 0.5 mL/ min. The column was washed with 15 column volumes of buffer R-6 and then incubated in buffer C (20 mM Tris • HCl, 1 mM DTT, pH 6.0) at room temperature for 16 h to induce intein cleavage. The column was then washed with buffer C until all desired protein had been eluted. The isolated Im7(C29-G87) peptide was then further purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass, shown by RP-HPLC ≥95% pure, and lyophilized for storage. A 1 L expression typically yielded approximately 4–5 mg of pure desired product.

Cloning of the Im7(MEH_6E2-S58) Intein CBD Construct (Im7(MEH_6E2-S58)-Intein-CBD). Starting from the pTrc(Im7)-99A vector, the Im7(MEH_6E2-S58) peptide was amplified using primers with engineered 5′ Nde1 and 3′ Sap1 restriction sites using the pFu Turbo polymerase (Stratagene) under standard conditions described by the manufacturer. Both the amplicons and the IMPACT pTWIN1 vector were doubly digested with Nde1 and Sap1 restriction enzymes (New England Biolabs), fractionated by agarose gel electrophoresis, and purified with the QIAquick gel extraction kit (Qiagen). Ligations were conducted with the T4 DNA ligase kit (Promega) for 15 min at room temperature. The ligation product was transformed into a competent DH5α strain of *E. coli* (Invitrogen), spread on plates containing LB media and carbenicillin. The desired TWIN1[Im79 (MEH_6E2-S580] plasmid from colonies formed on the plates was isolated, verified by sequencing, and transformed into a competent BL21(DE3) strain of *E. coli* for protein overexpression.

Expression and Purification of Im7(MEH_6E2-S58) Peptide Thioester. Starting from a 5 mL overnight culture, BL21(DE3) cells (Stratagene) expressing pTWIN1(Im7MEH_6E2-S58) (Fig. S2*B*) were grown at 37 °C in LB broth containing carbenicillin antibiotic to an OD_{600} of 0.6–0.8. At that point the temperature was lowered to 16 °C and protein production was induced by the addition of 1 mM IPTG for 16 h. The cells were harvested by centrifugation (7,500 × *g*) for 30 min, washed once with a 0.9% NaCl solution, centrifuged again (7,500 × *g*) for 30 min, and the cell pellet was frozen at −80°C until needed.

All purification steps were performed at 4 °C. Cell pellets containing the $Im7_{M1-S58}$ construct were thawed and resuspended in 2.5% of the original culture volume in buffer (20 mM Hepes, pH 8). The cells were lysed by sonication, followed by centrifugation (142,400 × *g*) for 1 h to remove cellular debris and membrane proteins. The supernatant was slowly applied to a column containing Ni-NTA agarose equilibrated with buffer A. After washing with 4 column volumes of buffer (20 mM Hepes, pH 6), the purified construct was eluted with 25 mL of buffer (20 mM Hepes, 200 mM imidazole, pH 6). To the elution was added 2 g sodium 2-mercaptoethanesulfonate and incubated for 8 h at 37 °C followed by 8 h at 4 °C to cleave the intein and release the Im7(M1-S58) peptide thioester. To remove the CBD-intein, the mixture was then applied to a column of chitin resin of 0.5% the original culture volume. The flow-through containing the desired Im7(MEH_6E2-S58) peptide thioester was collected (Fig. S4*A*) and purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass (Fig. S4*C*), shown by RP-HPLC to be ≥95% pure (Fig. S4*B*), and lyophilized for storage. A 1 L expression typically yields approximately 10 mg of pure desired product.

Native Chemical Ligation. Native chemical ligation methods are based on those reported previously (3). Specifically, the appropriate lyophilized peptide pairs were dissolved in buffer (100 mM sodium phosphate, pH 7) and quantified by UV using the derived extinction coefficient based on the peptide sequences. The synthetic glycopeptides were brought to a concentration of 2 mM, whereas the expressed recombinant protein fragment was used in excess at 3 mM. To initiate the ligation reaction, an equal volume of the appropriate peptide and recombinant fragment was mixed, followed by the addition of 2% benzyl mercaptan and 2% benzenethiol. After 16 h of gentle shaking at room temperature, the precipitated thiols were removed by centrifugation (40,000 × *g*) and the full-length Im7 variant was purified by preparative reverse phase C_{18} HPLC, confirmed by ESI-MS to be within 1 Da of expected mass (Fig. S5 and Table S2), checked by RP-HPLC to be ≥95% pure (Fig. S6), and lyophilized for storage. By HPLC peak area analysis, the typical yield for the ligation reaction was 80% (Fig. S5*A*).

Stopped-Flow Fluorescence Studies of Im7 Folding. Experiments were performed at 10 °C in buffer A (50 mM sodium phosphate, 400 mM sodium sulfate, 1 mM EDTA, pH 7). Refolding experiments were performed by 1:10 dilution of approximately 0.35 mg/mL protein in buffer A containing 8 M urea into buffered solutions with final urea concentrations in the range 0.75–8.0 M. Unfolding experiments were measured in the same way but the initial protein solution did not contain urea. The final urea concentration ranged from 3.0–8.0 M for unfolding experiments. At each urea concentration, at least seven kinetic traces were averaged and fitted to a single exponential function using the manufacturer's software in order to obtain the observed rate constant (k_{obs}) as well as the initial and final fluorescence signals. After subtracting the buffer blanks, both the initial and final fluorescence signals were normalized to the fluorescence signal of the 7.75 M urea sample. The k_{obs} and normalized endpoint and initial fluorescence signals for each variant were fitted to the analytical solution for an on-pathway three-state model using Igor Pro 6.0 (Wavemetrics) as previously described. The k_{ui} and m_{ui} were fixed to the values previously obtained for WT Im7 from ultrarapid mixing continuous-flow measurements ($k_{ui} = 1573\ s^{-1}$, $m_{ui} = 1.23\ kJ\,mol^{-1}\,M^{-1}$). Intermediate stability was determined by allowing k_{iu} to vary.

Replica Exchange Molecular Dynamics (REMD) Simulations of Glycosylation Site Peptides. REMD simulations of the seven glycosylation site peptides in both glycosylated and nonglycosylated forms were performed using CHARMM 33b2 (4) with the GBSW (Generalized Born model with a simple smoothed SWitching function) force field (5) and a salt concentration of 140 mM. Each peptide

was seven residues in length (with the glycosylation site residue at the center) and was built in an extended conformation with an acetyl group at the N terminus and a methylamino (NHMe) group at the C terminus to mimic the backbone in the context of the full-length protein. Parameters for the chitobiose glycan were based on similar chemical groups in the existing CHARMM22 (6) and carbohydrate solution force fields (7) whenever possible, and others were derived from a previous computational study on O-linked glycosylation (8).

Each peptide was replicated, energy minimized using 200 steps of adopted basis Newton–Raphson minimization, and heated to 16 exponentially spaced temperatures ranging from 280 to 700 K (9) over a period of 50 ps. Following heating, the replicas were then equilibrated at each temperature for additional 450 ps using a Nosé–Hoover thermostat (10). During the exchange portion of the REMD simulation, swaps were attempted between (alternating) neighboring pairs of replicas every 5 ps. Using atom-based cutoffs, electrostatic and van der Waals interactions were switched to zero at 16 Å (5). The nonbonded list was truncated at 20 Å (5). Bond lengths involving hydrogen atoms were fixed with SHAKE/Roll and RATTLE/Roll, and a time step of 2 fs was used with a velocity Verlet integrator (11).

To evaluate the convergence of each REMD simulation, block averages of the conformational probabilities were calculated every 5 or 10 ns (see below). Because these probabilities depend entirely on the ϕ/ψ distribution of the central residue, they are an appropriate metric for convergence of this distribution. REMD simulations were first run for 25 ns, using the last 20 ns of each simulation for analysis. If the standard errors of the mean probabilities were greater than 0.05, then the simulation was extended for an additional 10 ns (35 ns total time) and statistics were recalculated. If the standard errors of the means were still greater than 0.05 after 35 ns of simulation, then the simulation was extended to 50 ns, using the last 40 ns for analysis with 10 ns block averages. Any simulations not converged after 50 ns were run for an additional 20 ns (70 ns total time).

Analysis of Conformational Preferences at Each Glycosylation Site. The ϕ/ψ distribution of the central residue of each glycosylation site peptide was calculated using the structures generated by the 280 K replicas of the REMD simulations. These ϕ/ψ values were then classified as belonging to the α-helical conformation if ($-180° < \phi < 0°$, $-100° < \psi < 45°$), the β- sheet conformation if ($-180° < \psi < -45°$, $45° < \psi < 225°$), or the turn/α_L conformation if ($0° < \psi < 180°$, $-90° < \psi < 90°$). These definitions of α-helix, β-sheet, and turn/α_L conformations were taken from Hovmöller et al. (12). Structures with ϕ/ψ angles not belonging to one of these three categories comprised less than 0–4% of the conformational ensembles.

The probabilities for sampling these conformations were calculated by dividing the total number of structures with ϕ/ψ angles corresponding to each conformation by the total number of structures. Uncertainty estimates were determined by calculating the standard error of the 5 or 10 ns block averages, with the block length determined by the total simulation length, as described in the previous section.

1. Hackenberger CP, O'Reilly MK, Imperiali B (2005) Improving glycopeptide synthesis: A convenient protocol for the preparation of beta-glycosylamines and the synthesis of glycopeptides. *J Org Chem* 70:3574–3578.
2. Capaldi AP, Kleanthous C, Radford SE (2002) Im7 folding mechanism: Misfolding on a path to the native state. *Nat Struct Biol* 9:209–216.
3. Hackenberger CP, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J Am Chem Soc* 127:12882–12889.
4. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30:1545–1614.
5. Chen J, Im W, Brooks CL, III (2006) Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J Am Chem Soc* 128:3728–3736.
6. MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
7. Kuttel M, Brady JW, Naidoo KJ (2002) Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J Comput Chem* 23:1236–1243.
8. Spiriti J, Bogani F, van der Vaart A, Ghirlanda G (2008) Modulation of protein stability by O-glycosylation in a designed Gc-MAF analog. *Biophys Chem* 134:157–167.
9. Ho BK, Dill KA (2006) Folding very short peptides using molecular dynamics. *PLoS Comput Biol* 2:e27.
10. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys Rev A* 31:1695–1697.
11. Lamoureux G, Roux B (2003) Modeling induced polarization with classical Drude oscillatots: Theory and molecular synamics simulation algorithm. *J Chem Phys* 119:3025–3039.
12. Hovmoller S, Zhou T, Ohlson T (2002) Conformations of amino acids in proteins. *Acta Crystallogr D Biol Crystallogr* 58:768–776.

A

Solid phase peptide synthesis

Expressed via IMPACT system

(GlcNAc-GlcNAc)-NH

$H_3\overset{+}{N}$–MEH$_6$E2-A28–SR

+

$H_3\overset{+}{N}$–T30-G87–CO_2^- (SH)

EPL

(GlcNAc-GlcNAc)-NH

$H_3\overset{+}{N}$–MEH$_6$E2-A28–T30-G87–CO_2^- (SH)

Full-length Im7 A29C glycomutant

B

Solid phase peptide synthesis

Expressed via IMPACT system

(GlcNAc-GlcNAc)-NH

$H_3\overset{+}{N}$–N60-G87–CO_2^- (SH)

+

$H_3\overset{+}{N}$–MEH$_6$E2-S58–SR

EPL

(GlcNAc-GlcNAc)-NH

$H_3\overset{+}{N}$–MEH$_6$E2-S58–N60-G87–CO_2^- (SH)

Full-length Im7 D59C glycomutant

Fig. S1. Semisynthesis of Im7 glycosylated variants. Each variant is assembled from a glycopeptide made by solid phase peptide synthesis and a larger peptide produced by recombinant expression methods. EPL, expressed protein ligation. (*A*) Ligation between residues 28 and 29 for semisynthesis of helix I and helix I–II loop glycomutants, and (*B*) ligation between 59 and 60 for semisynthesis of helix IV and helix III–IV loop glycomutants.

Fig. S2. (*A*) Intein cleavage mechanism and the production of the Im7(C29-G87) recombinant protein fragment. (*B*) Intein cleavage mechanism and the production of the Im7(MEH_6E2-S58) recombinant fragment. Note that this recombinant fragment contains a C-terminal thioester.

Fig. S3. (*A*) Fluorescence emission spectra of WT Im7 and Im7 variants in 0 M (solid line) or 8 M urea (dashed line). (*B*) Far-UV CD spectrum of WT Im7 and Im7 variants. Ellipticity units, $(\theta \times 10^{-3})$mdeg·cm^2·dmol^{-1}.

Fig. S4. Preparation and characterization of Im7(MEH$_6$E2-S58). (*A*) SDS-PAGE gel of Im7(MEH$_6$E2-S58) purification. Lanes: 1, gel ladder; 2, cell lysate; 3, flow through Ni-NTA column; 4, Ni-NTA column wash; 5, Ni-NTA elution; 6, sodium 2-mercaptoethanesulfonate 8 h incubation; 7, flow-through chitin column. (*B*) Reverse phase preparative C_{18} RP-HPLC trace for purification of Im7(MEH$_6$E2-S58) from lane 7 of gel. (*C*) ESI-MS of Im7(MEH$_6$E2-S58) showing protein of expected mass (observed mass, *x*; expected mass, *y*).

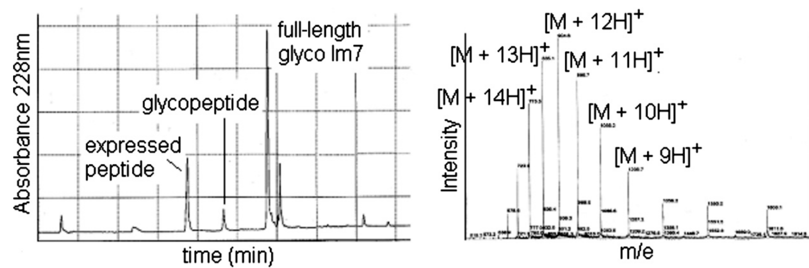


Fig. S5. Example of an expressed protein ligation reaction for the synthesis of the Im7(A29C/V27NGlyco) variant. (*A*) A typical ligation reaction proceeds to approximately 80% completion relative to the limiting glycopeptide starting material. The reaction was monitored by reverse phase C_{18} HPLC at 228 nm. (*B*) The full-length Im7(A29C/V27NGlyco) glycoprotein product was characterized by ESI-MS.

A29C N5

A29C N5-Glyco

A29C K20N

A29C K20N-Glyco

A29C V27N

A29C V27N-Glyco

D59C N60

D59C N60-Glyco

D59C K73N

D59C K73N-Glyco

D59C A78N

D59C A78N-Glyco

Fig. S6. Analytical RP-HPLC of purified Im7 variants, showing absorbance at 280 nm wavelength over elution time (minutes). RP-HPLC conditions were described in the *SI Methods*. All Im7 variants eluted between 20 and 25 min. All Im7 variants used in folding experiments were confirmed to be ≥95% pure. Characterization of the Im7(A29C/A13N) variants was previously published.

A29C N5

Average ϕ/ψ pmf of N5

A29C N5-Glyco

Average ϕ/ψ pmf of N5+glycan

A29C A13N

Average ϕ/ψ pmf of N13

A29C A13N-Glyco

Average ϕ/ψ pmf of N13+glycan

A29C K20N

Average ϕ/ψ pmf of N20

A29C K20N-Glyco

Average ϕ/ψ pmf of N20+glycan

A29C V27N

Average ϕ/ψ pmf of N27

A29C V27N-Glyco

Average ϕ/ψ pmf of N27+glycan

D59C N60

Average ϕ/ψ pmf of N60

D59C N60-Glyco

Average ϕ/ψ pmf of N60+glycan

D59C K73N

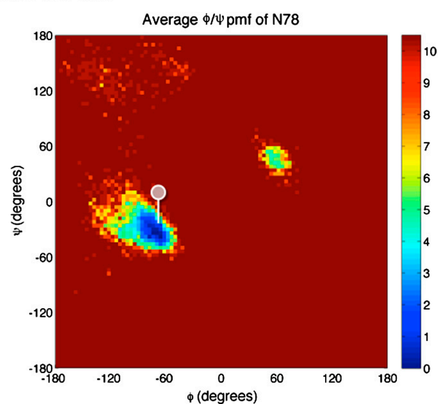
Average ϕ/ψ pmf of N73

D59C K73N-Glyco

Average ϕ/ψ pmf of N73+glycan

ψ (degrees)

ϕ (degrees)



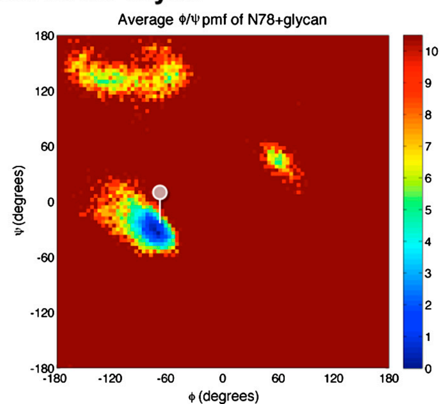


Fig. S7. Relative free energies with respect to backbone dihedral angles ϕ/ψ for nonglycosylated and glycosylated variants of the central residue at the experimentally introduced Im7 glycosylation site: N5, N13, N20, N27, N60, N73, and N78. The relative free energies are calculated from a potential of mean force (pmf): $W(\phi,\psi) = -RT \log p(\phi,\psi)$, where $p(\phi,\psi)$ is the ϕ/ψ probability distribution of the central residue obtained from REMD simulations of the relevant Im7 heptapeptide subsequence. The "push-pin" symbol designates the ϕ/ψ of the original amino acid at this position in native Im7.

Table S1. Values of β_1 and β_{TS2} for Im7 variants

Variant	β_1	β_{TS2}
WT Im7	0.76 ± 0.02	0.90 ± 0.01
A29C N5	0.78 ± 0.01	0.89 ± 0.01
A29C N5-Glyco	0.76 ± 0.01	0.90 ± 0.01
A29C A13N	0.73 ± 0.02	0.89 ± 0.01
A29C A13N-Glyco	0.74 ± 0.02	0.90 ± 0.01
A29C K20N	0.75 ± 0.01	0.91 ± 0.01
A29C K20N-Glyco	0.82 ± 0.03	0.91 ± 0.01
A29C V27N	0.79 ± 0.01	0.90 ± 0.004
A29C V27N-Glyco	0.82 ± 0.01	0.93 ± 0.01
D59C N60	0.80 ± 0.01	0.93 ± 0.004
D59C N60-Glyco	0.77 ± 0.01	0.92 ± 0.01
D59C K73N	0.77 ± 0.01	0.92 ± 0.004
D59C K73N-Glyco	0.74 ± 0.05	0.92 ± 0.02
D59C A78N	0.79 ± 0.01	0.92 ± 0.003
D59C A78N-Glyco	0.82 ± 0.01	0.91 ± 0.005

Table S2. ESI-MS characterization of Im7 variants. The masses listed represent the dominant peaks from the mass spectra

Im7 Variant	Charge	Calculated mass	Observed mass
A29C N5	$[M+9H]^+$	1,209.7	1,209.9
	$[M+10H]^+$	1,088.9	1,089.0
	$[M+11H]^+$	990.0	990.1
A29C N5-Glyco	$[M+9H]^+$	1,255.0	1,255.3
	$[M+10H]^+$	1,129.6	1,129.9
	$[M+11H]^+$	1,027.0	1,027.6
A29C K20N	$[M+9H]^+$	1,208.2	1,207.7
	$[M+10H]^+$	1,087.5	1,087.0
	$[M+11H]^+$	988.7	988.3
A29C K20N-Glyco	$[M+9H]^+$	1,253.4	1,254.1
	$[M+10H]^+$	1,128.2	1,128.9
	$[M+11H]^+$	1,025.7	1,026.4
A29C V27N	$[M+9H]^+$	1,211.4	1,090.0
	$[M+10H]^+$	1,090.4	991.1
	$[M+11H]^+$	991.3	908.5
A29C V27N-Glyco	$[M+9H]^+$	1,256.6	1,255.9
	$[M+10H]^+$	1,131.1	1,130.5
	$[M+11H]^+$	1,028.3	1,027.8
D59C N60	$[M+9H]^+$	1,204.8	1,204.4
	$[M+10H]^+$	1,084.5	1,084.0
	$[M+11H]^+$	986.0	985.5
D59C N60-Glyco	$[M+9H]^+$	1,250.1	1,249.3
	$[M+10H]^+$	1,125.2	1,124.6
	$[M+11H]^+$	1,023.0	1,022.4
D59C K73N	$[M+9H]^+$	1,203.3	1,202.7
	$[M+10H]^+$	1,083.1	1,082.7
	$[M+11H]^+$	984.7	984.3
D59C K73N-Glyco	$[M+9H]^+$	1,248.5	1,247.9
	$[M+10H]^+$	1,123.8	1,123.2
	$[M+11H]^+$	1,021.7	1,021.2
D59C A78N	$[M+9H]^+$	1,209.6	1,209.1
	$[M+10H]^+$	1,088.8	1,088.3
	$[M+11H]^+$	989.9	989.5
D59C A78N-Glyco	$[M+9H]^+$	1,254.9	1,254.1
	$[M+10H]^+$	1,129.5	1,128.9
	$[M+11H]^+$	1,026.9	1,026.4

Calculated masses were derived using Peptide Mass Calculator v3.2 (http://rna.rega.kuleuven.ac.be/masspec/pepcalc.htm) under electrospray series positive mode setting.