Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation

Mark M. Chen^a, Alice I. Bartlett^b, Paul S. Nerenberg^c, Claire T. Friel^b, Christian P. R. Hackenberger^a, Collin M. Stultz^c, Sheena E. Radford^b, and Barbara Imperiali^{a,1}

^aDepartment of Chemistry and Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139; ^bAstbury Centre for Structural Molecular Biology, Institute of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK; and ^cDepartment of Electrical Engineering and Computer Science and the Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Contributed by Barbara Imperiali, October 20, 2010 (sent for review September 13, 2010)

N-linked glycosylation modulates protein folding and stability through a variety of mechanisms. As such there is considerable interest in the development of general rules to predict the structural consequences of site-specific glycosylation and to understand how these effects can be exploited in the design and development of modified proteins with advantageous properties. In this study, expressed protein ligation is used to create site-specifically glycosylated variants of the bacterial immunity protein Im7 modified with the chitobiose disaccharide (GlcNAc-GlcNAc). Glycans were introduced at seven solvent exposed sites within the Im7 sequence and the kinetic and thermodynamic consequences of N-linked glycosylation analyzed. The $\Delta\Delta G^\circ$ values for glycan incorporation were found to range from +5.2 to -3.8 kJ·mol⁻¹. In several cases, glycosylation influences folding by modulating the local conformational preferences of the glycosylated sequence. These locally mediated effects are most prominent in the center of α -helices where glycosylation negatively effects folding and in compact turn motifs between segments of ordered secondary structure where glycosylation promotes folding and enhances the overall stability of the native protein. The studies also provide insight into why glycosylation is commonly identified at the transition between different types of secondary structure and when glycosylation may be used to elaborate protein structure to protect disordered sequences from proteolysis or immune system recognition.

Covalent protein modifications modulate and diversify the structures and functions of the naïve protein products that are encoded by the genome (1). Amongst the many varied transformations, N-linked glycosylation is perhaps the most chemically complex and ubiquitous occurring throughout all domains of life (2, 3). The large hydrophilic glycans that are appended to proteins have been implicated in a myriad of biological processes (4), including modulation of protein stability, oligomerization, and aggregation (5, 6), endoplasmic reticulum (ER) quality control and protein trafficking (7), host cell-surface interactions (8), and to modulate enzyme activity (9). In particular, the cotranslational timing of eukaryotic N-glycosylation and the profound impact of the modification on protein folding has inspired the application of experimental (10), computational (11), and bioinformatic (12) approaches targeted at developing general paradigms that can be implemented to define and manipulate glycosylation-induced effects on protein structure and, further, to establish a glycosylation code (13) as a predictive tool. These approaches are of significant fundamental importance as protein modifications, such as glycosylation, take center stage in our understanding of complex systems from both a fundamental viewpoint as well as in the application of glycoprotein therapeutics in modern medicine (14).

There are major impediments to the detailed analysis of the effects of glycosylation on protein folding and stability. Most natively glycosylated proteins are large, multisubunit, and/or

membrane-associated complexes that are currently immensely challenging to study at the level of detail that would provide discrete information on the site-specific effects of glycosylation. Furthermore, biophysical studies are hampered by the limited availability of chemically defined materials for analysis, which is largely due to the intrinsic heterogeneity both of the glycan structure and the specific site occupancy of the glycan in natively or heterologously expressed glycoproteins. For these reasons, there has been a considerable focus on the conformational analysis of defined glycopeptides, which can be prepared via chemical synthesis (15). Such studies have revealed that N-linked glycans can directly influence local peptide conformation, for example, by promoting the formation of more compact structures (16) and modulating disulfide bond formation (17). There is significantly less known, however, about the effects of glycosylation with specific sugar moieties on the stability and folding kinetics of full-length proteins. Recent advances in methods for expressed protein ligation (EPL) have begun to provide opportunities for the assembly of site-specifically modified glycoproteins (18), which in turn affords discretely glycosylated proteins for in-depth analysis.

As part of an initiative to derive general paradigms for dissecting and predicting the physical effects of N-linked glycosylation on protein folding landscapes, we present a detailed biophysical analysis of the effect of site-specific modification of the four helical protein Im7 with the N-glycan chitobiose (Fig. 1). This small, globular 87-residue protein, which lacks disulfides or proline *cis*-imides, is an ideal target for assembly via EPL to prepare chemically homogeneous, site-specifically glycosylated proteins for analysis of folding and stability. Moreover previous analyses have provided an excellent understanding of the folding landscape of Im7 in atomistic detail (19–21). Im7 is known to fold via a three-state mechanism (Fig. 2) involving a rapid desolvation and collapse of the unfolded protein to an on-pathway three helical intermediate, followed by a slower transition in which a specifically packed hydrophobic core is adopted (19, 21). This characteristic is particularly relevant for the current studies, because glycosylation has been proposed to exert an effect on folding via destabilization of the unfolded state, rather than a stabilization of the folded state (11, 22). Therefore, the ability to dissect the folding landscape of Im7 offers the opportunity to derive insight into the influence of glycans on the rate and mechanism of adoption of native and nonnative compact folded states.

Author contributions: C.M.S., S.E.R., and B.I. designed research; M.M.C., A.I.B., P.S.N., C.T.F., and C.P.R.H. performed research; M.M.C., A.I.B., P.S.N., C.T.F., C.P.R.H., C.M.S., S.E.R., and B.I. analyzed data; and M.M.C., A.I.B., P.S.N., C.T.F., C.P.R.H., C.M.S., S.E.R., and B.I. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed. E-mail: imper@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1015356107/-DCSupplemental.

Fig. 1. Im7 sequence and EPL strategies. (A) Amino acid sequence of Im7 showing helices I–IV and the seven glycosylation sites. (B) Structure of chitobiose-Asn. (C) Ribbon diagram of Im7 (1AYI) (32) illustrating positions 29 and 59 (cyan markers), which were mutated to cysteine enabling the ligation strategies that allow introduction of glyco-Asn. Also illustrated in space-filling rendition (blue) are native Im7 residues at each of the selected glycosylation sites.

In these studies, two EPL strategies are used to create seven glycosylated Im7 variants with Asn-chitobiose (Asn-GlcNAc₂), at positions 5, 13, 20, 27, 60, 73, or 78 (Fig. 1). The sites were selected based on the known structure of Im7 to represent surface sites where glycosylation would not interfere with the interior core of the protein and to encompass different structural regions. These included sites in extended regions or loops (5, 27, and 60), sites at the N or C termini of helices (13 and 78), and sites within α -helices (20 and 73) (Table 1). Additionally, the sites vary in the mobility of the α -carbon of the modified residues as shown by their crystallographic *B* factors (Table 1). For each glycosylation site, we analyzed the conformational properties, stability, and folding kinetics of the semisynthetic glycoproteins and compared these properties with those of the corresponding pseudo-WT variants, which differed only by the absence of the appended glycan. Although N-linked glycoproteins are generally modified with considerably larger glycans (3), the truncated chitobiose disaccharide is a useful and relevant model because it is common to all eukaryotic N-linked glycoproteins and has been shown to exhibit a strong effect on the local peptide structure, even without the rest of the glycan structure (10, 17). In addition, whereas the mechanism of N-linked glycosylation requires a conserved Asn-Xaa-Ser/Thr motif, here we have focused exclusively on defining the effect of the glycan on folding and stability and therefore have not included the auxiliary hydroxyamino acid in the Im7 mutations.

In parallel with the experimental studies, a series of replica exchange molecular dynamics (REMD) simulations (23) were performed on a family of relevant glycosylated and nonglycosylated heptapeptides in order to determine how the dihedral angle

preferences of the target asparagine are altered by glycosylation. Together the results reveal that N-linked glycosylation alters the folding rate and stability of Im7 in a site-specific manner that depends critically on the secondary structure at the site of modification. Overall the studies provide general guidelines to inform where the introduction of N-linked glycosylation could be used to modulate the properties of the folding polypeptide chain without compromising the overall structure of the modified protein.

Results

Semisynthesis of Homogeneously Glycosylated Im7 Variants. EPL was used for the semisynthesis of milligram quantities of seven discretely glycosylated Im7 variants in which a natural or introduced Asn was modified with chitobiose at positions 5, 13, 20, 27, 60, 73, or 78 (Fig. 1 and *SI Text*). The glycosylated A29C A13N variant was prepared by ligation between synthetic (glyco-N13) Im7(M1-A28) C-terminal thioester and the expressed Im7(C29-G87) peptide (24). An analogous ligation between synthetic Im7(MEH₆E2-A28) glycopeptide thioesters and the expressed Im7(C29-G87) peptide (*SI Text*) was used to prepare N-terminal His-tagged analogs with glycosylation at residues 5, 20, and 27. Variants with glycosylation at residues 60, 73, and 78 were prepared using a strategy with disconnection at C59. This strategy involved ligation between an Im7(MEH₆E2-S58) peptide thioester (*SI Text*), prepared by recombinant expression via intein chemistry (25), and a synthetic Im7(C59-G87) glycopeptide. The ligation sites (A29C and D59C) are solvent exposed and lack interresidue contacts, and therefore are unlikely to effect folding and stability.

Characterization of the Glycosylated Im7 Variants. The seven glycosylation sites were chosen to lie in different regions of the Im7 structure and to be solvent exposed, such that each variant should still fold to the native structure. The conformational properties of each new variant were assessed by fluorescence and CD spectroscopy and compared with WT Im7 (*SI Text*). The fluorescence of the single tryptophan (W75) is a sensitive probe of the Im7 native structure because W75 is highly fluorescent in the unfolded state and only weakly fluorescent in the native state as a result of close packing against H47 (26). Fluorescence measurements under

Fig. 2. Folding mechanism of Im7. Representation of the three-state folding mechanism of Im7 showing the on-pathway intermediate (*I*). Also shown is the β_T -value, which represents the relative solvent-accessible surface area buried in each species normalized to burial in the native state. The values quoted are those for WT Im7, the β_T -value for TS1 is taken from ref. 19.

Table 1. Structural parameters for residues 5, 13, 20, 27, 60, 73, and 78 in native Im7 derived from Protein Data Bank entry 1AYI (32)

	ϕ/ψ dihedral angles	Secondary structure	SASA *, $\text{\AA}^2$	<i>B</i> -factor C_α , $\text{\AA}^2$
N5	(–)62.4 (–)34.3	N term extended	143 (260)	28.79
A13	(–)66.4 (–)39.9	α -helix (N term)	67 (209)	13.37
K20	(–)73.0 (–)34.2	α -helix (middle)	100 (303)	14.16
V27	(–)101.3 (+)15.5	1 of 5-aa loop	128 (250)	22.12
N60	(–)86.8 (+)7.6	7 of 11-aa loop	171 (260)	45.63
K73	(–)55.7 (–)47.8	α -helix (middle)	83 (303)	13.82
A78	(–)69.9 (–)22.4	α -helix (C term)	85 (209)	15.21

*Value in brackets represents maximum solvent-accessible surface area (SASA) for specific residue.

Table 2. Kinetic and thermodynamic parameters for the folding and unfolding kinetics of the Im7 variants

Variant	K_{UI} (M_{UI} [kJ·mol ⁻¹ ·M ⁻¹])	k_{IN} (s ⁻¹) (m_{IN} [kJ·mol ⁻¹ ·M ⁻¹])	k_{NI} (s ⁻¹) (m_{NI} [kJ·mol ⁻¹ ·M ⁻¹])	ΔG°_{UI} kJ·mol ⁻¹	ΔG°_{UN} kJ·mol ⁻¹
WT Im7	125.2 ± 30 (3.9 ± 0.1)	276 ± 13 (0.74 ± 0.10)	1.31 ± 0.13 (0.51 ± 0.04)	-11.4 ± 0.6	-23.9 ± 0.6
A29C N5	302 ± 58 (3.66 ± 0.08)	151 ± 3 (0.54 ± 0.04)	1.09 ± 0.11 (0.49 ± 0.03)	-13.4 ± 0.5	-25.0 ± 0.5
A29C N5 Glyco	352 ± 66 (3.70 ± 0.07)	207 ± 4.5 (0.69 ± 0.04)	1.07 ± 0.11 (0.46 ± 0.03)	-13.8 ± 0.4	-26.2 ± 0.5
A29C A13N	253 ± 103 (3.86 ± 0.17)	197 ± 10 (0.85 ± 0.10)	1.08 ± 0.15 (0.60 ± 0.05)	-13.0 ± 1.0	-25.3 ± 1.0
A29C A13N Glyco	218 ± 91 (3.92 ± 0.18)	178 ± 9 (0.86 ± 0.11)	1.38 ± 0.17 (0.52 ± 0.04)	-12.7 ± 1.0	-24.1 ± 1.0
A29C K20N	67.6 ± 12.7 (3.96 ± 0.07)	191 ± 7 (0.81 ± 0.10)	1.43 ± 0.08 (0.49 ± 0.02)	-9.9 ± 0.4	-21.4 ± 0.5
A29C K20N Glyco	14.7 ± 8.6 (4.43 ± 0.11)	169 ± 11 (0.47 ± 0.23)	2.58 ± 0.08 (0.51 ± 0.01)	-6.3 ± 0.7	-16.2 ± 0.68
A29C V27N	67.9 ± 7.8 (4.09 ± 0.05)	166 ± 4 (0.56 ± 0.07)	1.33 ± 0.05 (0.52 ± 0.01)	-9.9 ± 0.3	-21.3 ± 0.3
A29C V27N Glyco	432 ± 93 (4.35 ± 0.10)	140 ± 4 (0.57 ± 0.05)	1.43 ± 0.13 (0.37 ± 0.03)	-14.3 ± 0.5	-25.1 ± 0.6
D59C N60	267 ± 35 (4.13 ± 0.06)	159 ± 3 (0.64 ± 0.04)	1.55 ± 0.09 (0.37 ± 0.02)	-13.2 ± 0.3	-24.1 ± 0.3
D59C N60 Glyco	296 ± 51 (4.14 ± 0.07)	199 ± 5 (0.79 ± 0.05)	1.25 ± 0.09 (0.42 ± 0.02)	-13.4 ± 0.4	-25.3 ± 0.4
D59C K73N	112 ± 16 (3.82 ± 0.06)	134 ± 2 (0.70 ± 0.05)	2.50 ± 0.10 (0.42 ± 0.01)	-11.1 ± 0.3	-20.4 ± 0.4
D59C K73N Glyco	15.8 ± 3.9 (3.77 ± 0.16)	132 ± 8 (0.92 ± 0.34)	2.61 ± 0.10 (0.43 ± 0.01)	-6.5 ± 0.7	-15.7 ± 0.7
D59C A78N	92.2 ± 8.5 (4.06 ± 0.04)	170 ± 3 (0.66 ± 0.05)	1.62 ± 0.05 (0.40 ± 0.01)	-10.6 ± 0.2	-21.6 ± 0.23
D59C A78N Glyco	63.7 ± 8.9 (4.05 ± 0.06)	133 ± 4 (0.42 ± 0.08)	1.28 ± 0.07 (0.46 ± 0.02)	-9.8 ± 0.3	-20.7 ± 0.4

All rate constants are expressed in s⁻¹, all m values are expressed in kJ·mol⁻¹·M⁻¹ and all free energies are expressed in kJ·mol⁻¹. Data were acquired at 10 °C in 50 mM sodium phosphate (pH 7.0) and 0.4 M Na₂SO₄. Errors were calculated as described previously (19).

cause local distortion of the helix. Significantly, for both K20N and K73N glycosylation led to destabilization of the protein with $\Delta\Delta G^{\circ}_{UN}$ values of 5.2 ± 0.7 kJ/mol for K20N-glyco and 4.8 ± 0.6 kJ/mol for K73N-glyco (Fig. 4 and Table 2). This destabilization is predominantly due to the glycan influencing early steps in folding, as shown by the decreased K_{UI} , however glycosylation of the K20N variant also increases the rate constant for unfolding (k_{NI}) (Table 2). The experimental observation that glycosylation negatively impacts the collapse of the unfolded to the intermediate state is supported by the REMD simulations (23) which enabled us to map the free energy landscape of each peptide representing the K20N and K73N variants, with and without the glycan, at 280 K. To quantify the change in local conformational preferences upon glycosylation, the ϕ/ψ -propensities of the central Asn residue in each peptide in both the glycosylated and nonglycosylated forms were compared (Table 3). The simulations show that even in the short heptapeptide sequences corresponding to the glycosylation site in the K20N and K73N variants there is a significant reduction in α -helical propensity upon introduction of the glycan moiety.

Glycosylation is Well Tolerated at the Termini of α -Helices. Residues 13 and 78 are located near the N and C termini of helices I and IV, respectively (Fig. 1). In previous studies, the substitutions A13G and A78G were shown to be destabilizing ($\Delta\Delta G^{\circ}_{UN} \sim 2$ –5 kJ·mol⁻¹; ref. 27) consistent with entropic stabilization of the unfolded state. For the variants in this study, the glyco-Asn is preceded or succeeded by a turn or loop structure and, therefore, unlike the K20N and K73N variants, there is space to accommodate the introduced glycan. REMD simulations of the heptapep-

tides including these sites suggest that glycosylation may alter the conformational preferences of the local peptide sequences (Table 3). However, it appears that introduction of glycosylation at these sites in the protein does not have a significant effect on stability or the folding–unfolding rate constants perhaps because in this case the local structure can adapt to accommodate the modification without perturbing the overall structure.

The minimal effect of glycosylation at the termini of α -helices in these studies underscores a commonly observed principle, which is well documented in statistical analyses of glycoprotein structures, that there is an elevated probability of finding glycosylation sites where there is a change in secondary structure (12). Additionally, when glyco-Asn is found associated with α -helical structures, the most common placement is at the helix termini. As noted above, of the 88 nonredundant N-glycosylation sites documented to be in α -helices, 50% are at helix termini and an additional 14% are in single-turn helices (28).

The Effect of Glycosylation in Loops and Turns is Dependent on the Residue Location and Conformational Flexibility. Residues 5, 27, and 60 are in regions of WT Im7 that do not adopt regular secondary structure (Fig. 1 and Table 1), and the three Im7 variants (N5, V27N, and N60) studied in these regions have moderate to high conformational mobility as defined by B factors derived from the structure analysis of WT Im7 (Table 1). V27N is unique within the set of Im7 analogs because it is the only variant in which glycosylation clearly promotes the rate of folding and enhances stability. Whereas previous studies have reported only small changes in $\Delta\Delta G^{\circ}_{UN}$ upon mutation at position 27 (21, 27), in this case, the presence of the glycan led to stabilization of the protein with $\Delta\Delta G^{\circ}_{UN}$ value of -3.8 ± 0.9 kJ/mol (Fig. 4 and Table 2). Examination of the kinetic data reveals that this effect results predominantly from stabilization of I by the glycan (increased K_{UI} , Table 2). Such a dramatic stabilization is unusual for substitution of only a single amino acid at a surface exposed site (29).

Analysis of the structures of nonredundant glycosylation sites established that the prevalence of N-linked glycosylation in ordered turn or coil structures which link two elements of secondary structure is more favorable than predicted from sequence alone (12). Additionally, fluorescence (16) and NMR (30) studies of peptides and their corresponding glycopeptides have revealed that glycosylation promotes the adoption of compact turn structures. Significantly, recent studies on the adhesion domain of the human immune cell receptor cluster of differentiation 2 (hCD2ad), a small domain with β -sandwich topology that is natively glycosylated at the ($i + 2$) residue of a type I β -turn, revealed that glycosylation has a dramatic effect on protein folding

Fig. 4. Effect of glycosylation on protein stability. Comparison of the experimentally measured difference in the free energy of folding ($\Delta\Delta G^{\circ}_{UI}$ and $\Delta\Delta G^{\circ}_{UN}$) between the glycosylated variants and the corresponding nonglycosylated pseudo-WTs.

Table 3. Conformational preferences of Asn and glyco-Asn residues at the (experimentally introduced) glycosylation sites in heptapeptide sequences, depicted as percent of conformers in α -helix, β -sheet, or turn/ α_L conformations (see [S1 Text](#) for definitions)

	Pseudo-WT			Glycosylated pseudo-WT		
	α -Helix	β -Sheet	Turn/ α_L	α -Helix	β -Sheet	Turn/ α_L
N5	0.87 $\pm$ 0.03	0.06 $\pm$ 0.01	0.07 $\pm$ 0.02	0.66 $\pm$ 0.05	0.19 $\pm$ 0.03	0.14 $\pm$ 0.04
A13N	0.39 $\pm$ 0.03	0.06 $\pm$ 0.02	0.55 $\pm$ 0.01	0.34 $\pm$ 0.08	0.43 $\pm$ 0.07	0.19 $\pm$ 0.08
K20N	0.84 $\pm$ 0.05	0.07 $\pm$ 0.02	0.09 $\pm$ 0.05	0.59 $\pm$ 0.07	0.20 $\pm$ 0.04	0.19 $\pm$ 0.04
V27N	0.69 $\pm$ 0.04	0.07 $\pm$ 0.02	0.24 $\pm$ 0.02	0.43 $\pm$ 0.05	0.32 $\pm$ 0.03	0.23 $\pm$ 0.06
N60	0.88 $\pm$ 0.02	0.05 $\pm$ 0.01	0.07 $\pm$ 0.02	0.90 $\pm$ 0.02	0.06 $\pm$ 0.01	0.03 $\pm$ 0.02
K73N	0.81 $\pm$ 0.03	0.11 $\pm$ 0.03	0.08 $\pm$ 0.05	0.34 $\pm$ 0.04	0.64 $\pm$ 0.04	0.00 $\pm$ 0.00
A78N	0.90 $\pm$ 0.04	0.04 $\pm$ 0.01	0.06 $\pm$ 0.03	0.69 $\pm$ 0.05	0.25 $\pm$ 0.03	0.05 $\pm$ 0.02

Structures with dihedral angles not in one of these categories comprised <0–4% of the conformational ensembles. The data were obtained from a backbone dihedral angle analysis of structures generated using REMD simulations of 1m7 heptapeptide subsequences.

rate and stability (10). In these cases, the observed effects of glycosylation may result from a shift in backbone dihedral angle preferences of the modified residue, a reduction in the conformational space available to the peptide in the vicinity of the introduced glycan that results in a loss of conformational entropy (22), or the adoption of new, favorable noncovalent interactions between the glycan and the local peptide sequence. Because the major effect of the glycan on folding and stability of the V27N Im7 variant is manifested early in folding, during formation of the intermediate state, we focus on consideration of the first two effects. In native Im7, V27 is designated as adopting a coil conformation (ϕ $(-)$ 101.34°, ψ $(+)$ 15.50°) (Table 1) and falls in the region of the Ramachandran plot between typical α -helix and β -strand ϕ/ψ dihedral angle space. Examination of the REMD simulation results reveals that glycosylation reduces the propensity of asparagine to adopt right-handed α -helical structures (Table 3 and *SI Text*) and appears to increase the population with β -sheet dihedral angles. Therefore, part of the effect of glycosylation may be to shift the ϕ/ψ angle preferences in a way that promotes the early folding process, which subsequently results in adoption of the collapsed intermediate state. However, because the molecular dynamics simulations on the simple hepta-peptides do not directly account for the observed effect on Im7 folding, it is likely that for V27N that there must be involvement of a more extended stretch of the protein sequence and therefore more detailed simulations may be needed to provide a better prediction of the effects. For the V27N variant, an additional outcome of glycosylation is likely the promotion of the native structure because this conformation would relieve unfavorable steric clashes. To provide an indication of the steric encumbrance associated with glycosylation, we have modeled the Asn-chitobiose into position 27 of Im7 and depict a composite figure with the glycan in the three preferred orientations (Fig. 5) (31).

Finally, glycosylation of N5 and N60, shows only minor stabilizing effects ($\sim 1 \text{ kJ}\cdot\text{mol}^{-1}$) on the overall stability of these Im7 variants (Fig 4 and Table 2). These analogs share selected properties, which may account for the absence of significant effects. In both cases, the B factors are relatively large, 28.79 and 45.63 $\text{\AA}^2$ for N5 and N60, respectively, in WT Im7, and the Asn side chains in the native protein are fully surface exposed with minimal interactions with the rest of the protein structure as evidenced by the solvent-accessible surface area parameters for these two residues (Table 1). Additionally, both N5 and N60 are near disordered residues in the native structure. Specifically, L3 and K4 as well as S58 and D59 are disordered in the X-ray analysis of Im7. Therefore, although the REMD simulations suggest that glycosylation may have an effect on the local peptide conformational preferences, such changes could be easily tolerated both during folding and within the native fold. From a practical perspective, even when not affecting folding or stability, N-linked glycosylation could potentially be used to advantage in relatively disordered sites such as N5 and N60 because the glycan could protect the disordered sequences of the protein from

proteolysis or immune system recognition, which are well-documented consequences of N-linked glycosylation (4).

Conclusions

N-linked glycosylation presents opportunities to modulate protein folding by mechanisms that are distinct from modifications that are designed to alter the packed core of the protein structure. In addition, there is considerable interest in the development of general rules to predict the specific structural, kinetic, and thermodynamic consequences of site-specific glycosylation and to understand how these rules can be exploited in the design and development of modified proteins with advantageous properties.

Detailed biophysical analysis of the effect of N-linked glycosylation on small proteins and protein domains provides valuable information that can be applied toward understanding and predicting the potential effects of glycosylation in larger systems, which may be intractable to analysis in atomistic detail because of the complexity of the system or the inability to produce suitable quantities of homogeneously glycosylated proteins for analysis. Recent analysis of the effect of N-linked glycosylation on hCD2ad revealed that the modification has a dramatic effect on the protein folding rate and stability (10). Furthermore, *in silico* folding studies of glycosylated variants of the β -sheet-rich SH3 domain support the proposal that N-linked glycosylation influences the protein folding landscape by modulating the collapse of the unfolded state to afford intermediates along the folding pathway.

Fig. 5. Representation of glycosylated Im7 with chitobiose-Asn at position 27. The illustration depicts Im7 in a ribbon representation with mesh space-filling surface. The sugar was treated as a rigid body. The χ_1 angle set to each of the known preferred orientations: (-60° (pale blue), ($+180^\circ$ (purple), and ($+60^\circ$ (turquoise), and χ_2 is set to 0° (32). The illustration was created using the Im7 structure (1AYI) (32) in which residue 27 was replaced with chitobiose-Asn.

(11). In the current study, using Im7 as a model all α -helical protein, we have shown that N-linked glycosylation effects protein folding rates and stability in a tunable manner that is predictable based on knowledge of the native protein structure and the effect of glycosylation on local sequence preferences. This information provides the framework for further fundamental studies of the effects of glycosylation on protein behavior, as well as the route toward practical applications in the production of proteins with customized properties.

Methods

Structure, Sequence, and Numbering of Im7. The Im7 Protein Data Bank code is 1AY1 (32). Im7 figures were rendered using Chimera (University of California, San Francisco). The residue numbers listed throughout this paper are those of the untagged protein, where (E²) corresponds to the second residue of the protein (27).

Expression and Purification of Im7 Variants. Cloning, expression and purification of Im7 variants is described in the *SI Text*.

Semisynthesis of Glycosylated Im7 Variants. Fmoc-Asn(chitobiose-TBDMS)₂-OH was prepared as described previously (33). The glycosylated Im7 variants were prepared via ligation of recombinant peptides and synthetic glycopeptides as summarized in Fig. 1 and detailed in the *SI Text*. The synthesis and purification of the Im7(C59-G87) glycopeptides and the Im7(MEH₆E2-A28) glycopeptide thioesters, the cloning of Intein Im7(C29-G87) construct, the expression and purification of Im7(C29-G87) peptide, and the conditions for EPL are described in the *SI Text*. The EPL reaction was carried out under native conditions, and the full-length glycoprotein product was isolated using preparative RP-HPLC, shown by analytical RP-HPLC to be greater than 95% in purity, and characterized by electrospray ionization-MS (*SI Text*).

Circular Dichroism of Im7 Variants. Far-UV CD spectra were acquired on an Aviv Model 202 spectropolarimeter in buffer A (50 mM sodium phosphate, 400 mM sodium sulfate, 1 mM EDTA, pH 7) at 25 °C. Im7 concentrations were

standardized using UV absorbance at 280 nm under denaturing conditions (Im7 ϵ_{280} nm = 9,700 M⁻¹ cm⁻¹).

Fluorescence Spectroscopy of Im7 Variants. Fluorescence emission spectra of the Im7 variants were measured using a Photon Technology International Fluorimeter as described previously (21). Spectra of all denatured states were assumed to have the same maximum intensity at 350 nm. The spectrum of each native protein was normalized to the intensity of the respective denatured state, allowing direct comparison of the fluorescence intensity between variants.

Stopped-Flow Fluorescence Studies of Im7 Variant Folding. Kinetic measurements and analysis of the Im7 variants were based on approaches previously reported for Im7 (19, 21). Folding–unfolding measurements were carried out on an Applied-Photophysics SX1.8MV stopped-flow fluorimeter. Details of the experimental protocols and analysis can be found in the *SI Text*.

Replica Exchange Molecular Dynamics Simulations. REMD simulations of heptapeptides corresponding to local sequence at the seven glycosylation site, in both glycosylated and nonglycosylated forms, were performed using CHARMM 33b2 (34) with the GBSW (Generalized Born with a smooth Switching function) force field (35). Parameters for the chitobiose glycan were based on similar chemical groups in the existing CHARMM22 (36) force field and carbohydrate solution force field (37) carbohydrate parameter set whenever possible, and others were derived from previous work on O-linked glycosylation (38). For each peptide, these simulations were carried out at 16 exponentially spaced temperatures (ranging from 280 to 700 K) for 25–70 ns, depending on the convergence of the simulations as calculated from 5 or 10 ns block averages (see *SI Text*).

ACKNOWLEDGMENTS. The authors thank Prof. Andrei Petrescu for providing access to the Structural Assessment of Glycosylation Sites database. This work was supported by the National Institutes of Health GM039334 (to B.I.), the National Science Foundation 0821391 (to C.M.S.), and the Biotechnology and Biological Sciences Research Council Grants BB/526502/1 (to A.I.B) and 24/B17145 (to C.T.F.).

- Walsh CT, Garneau-Tsodikova S, Gatto GJ (2005) Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew Chem Int Ed Engl* 44:7342–7372.
- Abu-Qarrn M, Eichler J, Sharon N (2008) Not just for eukarya anymore: Protein glycosylation in bacteria and archaea. *Curr Opin Struct Biol* 18:544–550.
- Weerapana E, Imperiali B (2006) Asparagine-linked protein glycosylation: From eukaryotic to prokaryotic systems. *Glycobiology* 16:91R–101R.
- Varki A (1993) Biological roles of oligosaccharides: All of the theories are correct. *Glycobiology* 3:97–130.
- Mitra N, Sinha S, Ramya TNC, Suroliya A (2006) N-linked oligosaccharides as outfitters for glycoprotein folding, form and function. *Trends Biochem Sci* 31:156–163.
- Bosques CJ, Imperiali B (2003) The interplay of glycosylation and disulfide formation influences fibrillization in a prion protein fragment. *Proc Natl Acad Sci USA* 100:7593–7598.
- Lederkremer GZ (2009) Glycoprotein folding, quality control and ER-associated degradation. *Curr Opin Struct Biol* 19:515–523.
- Imberty A, Varrot A (2008) Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* 18:567–576.
- Skropeta D (2009) The effect of individual N-glycans on enzyme activity. *Bioorg Med Chem* 17:2645–2653.
- Hanson SR, et al. (2009) The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc Natl Acad Sci USA* 106:3131–3136.
- Shental-Bechor D, Levy Y (2008) Effect of glycosylation on protein folding: A close look at thermodynamic stabilization. *Proc Natl Acad Sci USA* 105:8256–8261.
- Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14:103–114.
- Shental-Bechor D, Levy Y (2009) Folding of glycoproteins: Toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol* 19:524–533.
- Sola RJ, Griebenow K (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci* 98:1223–1245.
- Buskas T, Ingale S, Boons GJ (2006) Glycopeptides as versatile tools for glycobiology. *Glycobiology* 16:113R–136R.
- Imperiali B, Rickert KW (1995) Conformational implications of asparagine-linked glycosylation. *Proc Natl Acad Sci USA* 92:97–101.
- O'Connor SE, Pohlmann J, Imperiali B, Saskiawan I, Yamamoto K (2001) Probing the effect of the outer saccharide residues of N-linked glycans on peptide conformation. *J Am Chem Soc* 123:6187–6188.
- Payne RJ, Wong CH (2010) Advances in chemical ligation strategies for the synthesis of glycopeptides and glycoproteins. *Chem Commun (Cambridge, UK)* 46:21–43.
- Friel CT, Smith DA, Vendruscolo M, Gsponer J, Radford SE (2009) The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat Struct Mol Biol* 16:318–324.
- Friel CT, Capaldi AP, Radford SE (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J Mol Biol* 326:293–305.
- Bartlett AL, Radford SE (2010) Desolvation and development of specific hydrophobic core packing during Im7 folding. *J Mol Biol* 396:1329–1345.
- Hoffmann D, Florke H (1998) A structural role for glycosylation: lessons from the hp model. *Folding Des* 3:337–343.
- Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. *Chem Phys Lett* 314:141–151.
- Hackenberger CPR, Friel CT, Radford SE, Imperiali B (2005) Semisynthesis of a glycosylated Im7 analogue for protein folding studies. *J Am Chem Soc* 127:12882–12889.
- Evans TC, Jr, Benner J, Xu MQ (1999) The in vitro ligation of bacterially expressed proteins using an intein from *Methanobacterium thermoautotrophicum*. *J Biol Chem* 274:3923–3926.
- Spence GR, Capaldi AP, Radford SE (2004) Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* 341:215–226.
- Capaldi AP, Kleinhous C, Radford SE (2002) Im7 folding mechanism: Misfolding on a path to the native state. *Nat Struct Biol* 9:209–216.
- Petrescu A-J (2009) SAGS database. *Structural Assessment of Glycosylation Sites* <http://sags.biochim.ro/>.
- Foit L, et al. (2009) Optimizing protein stability in vivo. *Mol Cell* 36:861–871.
- O'Connor SE, Imperiali B (1998) A molecular basis for glycosylation-induced conformational switching. *Chem Biol* 5:427–437.
- Imberty A, Perez S (1995) Stereochemistry of the N-glycosylation sites in glycoproteins. *Protein Eng* 8:699–709.
- Dennis CA, et al. (1998) A structural comparison of the colicin immunity proteins Im7 and Im9 gives new insights into the molecular determinants of immunity-protein specificity. *Biochem J* 333(Pt 1):183–191.
- Hackenberger CPR, 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.
- Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30:1545–1614.
- Chen J, Im W, Brooks CL, 3rd (2006) Balancing solvation and intramolecular interactions: Toward a consistent generalized Born force field. *J Am Chem Soc* 128:3728–3736.
- MacKerell AD, et al. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616.
- 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.
- 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.