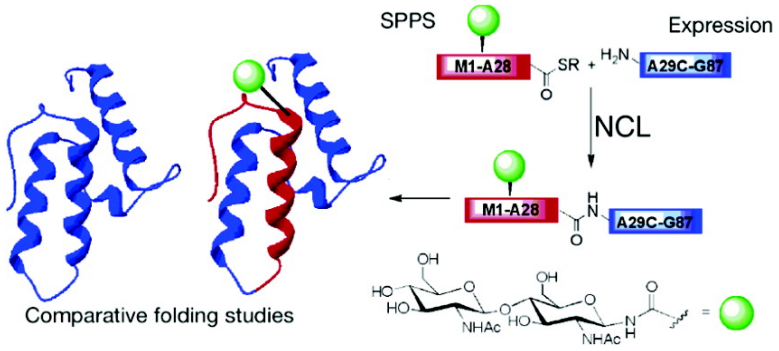
Article

# Semisynthesis of a Glycosylated Im7 Analogue for Protein Folding Studies

Christian P. R. Hackenberger, Claire T. Friel, Sheena E. Radford, and Barbara Imperiali

*J. Am. Chem. Soc.*, **2005**, 127 (37), 12882-12889• DOI: 10.1021/ja051855k • Publication Date (Web): 26 August 2005






## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 12 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 08/26/2005

# Semisynthesis of a Glycosylated Im7 Analogue for Protein Folding Studies

Christian P. R. Hackenberger,[†] Claire T. Friel,[‡] Sheena E. Radford,[‡] and Barbara Imperiali*,[†]

*Contribution from the Department of Chemistry and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and School Biochemistry and Microbiology, University of Leeds, Leeds LS2 9JT, UK*

Received March 23, 2005; E-mail: imper@mit.edu

***Abstract:*** To establish a system to address questions concerning the influence of glycosylation on protein folding pathways, we have developed a semisynthetic route toward the immunity protein Im7. This four-helix protein has been used extensively as model protein for folding studies. Native chemical ligation (NCL) affords an *N*-linked chitobiose glycoprotein analogue of Im7 with an Ala29Cys mutation. The semisynthetic approach relies on the solid-phase peptide synthesis (SPPS) of *N*-terminal thioesters (including helix I), in glycosylated or unglycosylated form, in combination with the expression of the *C*-terminal fragment of Im7 (containing helices II−IV). Detailed kinetic and thermodynamic analysis of the protein folding behavior reveals that semisynthetic Im7 analogues are well suited for protein folding studies and that the folding mechanism of the glycoprotein of this Im7 variant is not significantly altered over the unglycosylated analogue.

## Introduction

Asparagine-linked glycosylation plays a major role in determining the structure and function of numerous eukaryotic proteins.[1] This complex, enzyme-catalyzed protein modification is a cotranslational process wherein glycosyl transfer occurs as the nascent polypeptide emerges into the lumen of the endoplasmic reticulum as the protein folds to adopt a specific tertiary structure.[2] It has therefore been proposed that glycosylation may exert a critical effect on protein folding pathways.

To date, structural studies have mainly focused on understanding how *N*-linked glycosylation modulates the secondary structure of small peptide fragments.[3] However, considerably less information is available concerning the effect of glycosylation on folding mechanisms and the stability of larger protein domains or native proteins. This is a consequence of the difficulties associated with obtaining quantities of homogeneous glycoproteins for biophysical studies,[4] together with the challenge of detailed kinetic analysis of the folding of the large proteins that are typically the natural targets of *N*-linked glycosylation. Investigations addressing the folding of glycoproteins therefore have been limited to glycoprotein isolated and purified from natural sources, which include thermodynamic and stability measurements of RNAse B[5] and studies of different glycosylated variants of an invertase from *S. cerevisiae*.[6]

Herein we define a system for studying the effect of glycosylation on protein folding. We use as our model system the bacterial immunity protein, Im7, an 87-residue protein. Im7 is a member of a family of four homologous E colicin immunity proteins, the function of which is to inhibit the endonuclease domain of their specific bacterial colicin toxin.[7] The folding of Im7 has been the subject of extensive biophysical studies,[8] and the wealth of kinetic and thermodynamic data available on this protein provides an excellent basis upon which to build a detailed study of the effect of glycosylation on the entire folding landscape. Im7 and related congeners, while not naturally glycosylated, represent ideal model systems for the study of the effects of glycosylation on protein folding and stability. These proteins have a relatively simple four-helical topology, no disulfide bridges, and their small size makes the synthesis and semisynthesis of specifically elaborated glycoprotein analogues tractable. Of the family of immunity proteins, Im7 is perhaps the most interesting to study in this case, since this protein has been shown to fold to the native state via a three-state mechanism in which an on-pathway intermediate including

[†] Massachusetts Institute of Technology.
[‡] University of Leeds.

(1) (a) Varki, A. *Glycobiology* **1993**, *3*, 97−130. (b) Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683−720. (c) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357−2364. (d) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. *Science* **2001**, *291*, 2370−2376.
(2) Helenius, A.; Aebi, M. *Science* **2001**, *291*, 2364−2369.
(3) (a) O'Connor, S. E.; Imperiali, B. *J. Am. Chem. Soc.* **1997**, *119*, 2295−2296. (b) O'Connor, S. E.; Pohlmann, J.; Imperiali, B.; Saskiawan, I.; Yamamoto, K. *J. Am. Chem. Soc.* **2001**, *123*, 6187−6188. (c) Bosques, C. J.; Tschampel, S. M.; Woods, R. J.; Imperiali, B. *J. Am. Chem. Soc.* **2004**, *126*, 8421−8425.
(4) Grogan, M. J.; Pratt, M. R.; Marcaurelle, L. A.; Bertozzi, C. R. *Annu. Rev. Biochem.* **2002**, *71*, 593−634.
(5) (a) Puett, D. *J. Biol. Chem.* **1973**, *248*, 3566−3572. (b) Rudd, P. M.; Joao, H. C.; Coghill, E.; Fiten, P.; Saunders, M. R.; Opdenakker, G.; Dwek, R. A. *Biochemistry* **1994**, *33*, 17−22.
(6) Kern, G.; Kern, D.; Jaenicke, R.; Seckler, R. *Protein Sci.* **1993**, *2*, 1862−1868.
(7) James, R.; Kleanthouse, C.; Moore, G. R. *Microbiology* **1996**, *142*, 1569−1580.
(8) (a) Capaldi, A. P.; Ramachandra Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E. *Nat. Struct. Biol.* **2001**, *8*, 68−72. (b) Gorski, S. A.; Le Duff, C. S.; Capaldi, A. P.; Kalverda, A. P.; Beddard, G. S.; Moore, G. R.; Radford S. E. *J. Mol. Biol.* **2004**, *337*, 183−193. (c) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *Nat. Struct. Biol.* **2002**, *9*, 209−216. (d) Spence, G. R.; Capaldi, A. P.; Radford, S. E. *J. Mol. Biol.* **2004**, *341*, 215−226.


**10.1021/ja051855k CCC: $30.25** 

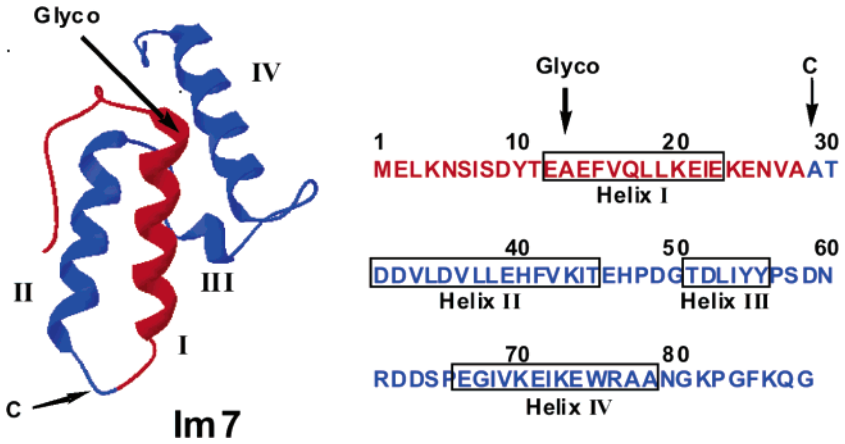



***Figure 1.*** Native sequence and structure of the immunity protein Im7. Arrows indicate sites for chitobiose modification (Glyco) and for Ala/Cys replacement for the native chemical ligation (C).

helices I, II, and IV is highly populated in the first few milliseconds of folding.[8] This intermediate has been studied in detail using hydrogen exchange[8b] and Φ-value analysis[8c] and has been shown to be stabilized by significant non-native interactions involving a number of residues from helices I, II, and IV. Importantly, the Im7 intermediate is misfolded in that residues partially solvent exposed in the native state may become transiently buried, such that non-native interhelical interfaces are formed. This raises the possibility that one or more natively solvent exposed residues could play an important role in the search for the native state while having little effect on native state stability. The type of mutation required for Φ-value analysis means that few natively solvent exposed residues have thus far been probed in the folding of Im7. The involvement of solvent exposed residues during folding, therefore, has not yet been investigated fully. The ability to incorporate carbohydrate modifications at specific positions throughout the Im7 architecture provides a unique opportunity to determine the role of solvent exposed residues in folding and to gain more information on the importance of non-native interactions at various points on the folding landscape.

To probe the influence of glycosylation on protein folding mechanisms, we now report a semisynthetic route toward an N-linked glycosylated variant of Im7. The synthetic strategy implements expressed protein ligation (EPL), which enables the connection via native chemical ligation (NCL) between a synthetic peptide with a *C*-terminal thioester, which can be derived from well-established solid-phase peptide synthesis (SPPS) methods, with an expressed protein fragment.[10] NCL has proven to be very suitable for accessing homogeneous glycoproteins, as demonstrated recently in the semisynthesis of O-linked glycoproteins.[11] Additionally, NCL allows the construction of glycoprotein analogues with unnatural carbohydrate side chains or other unnatural protein modifications. We present biophysical studies on the first glycosylated variant of Im7, which establishes the validity of this system for developing general paradigms to describe how glycosylation may affect the complex energy landscape for the folding of these model proteins.

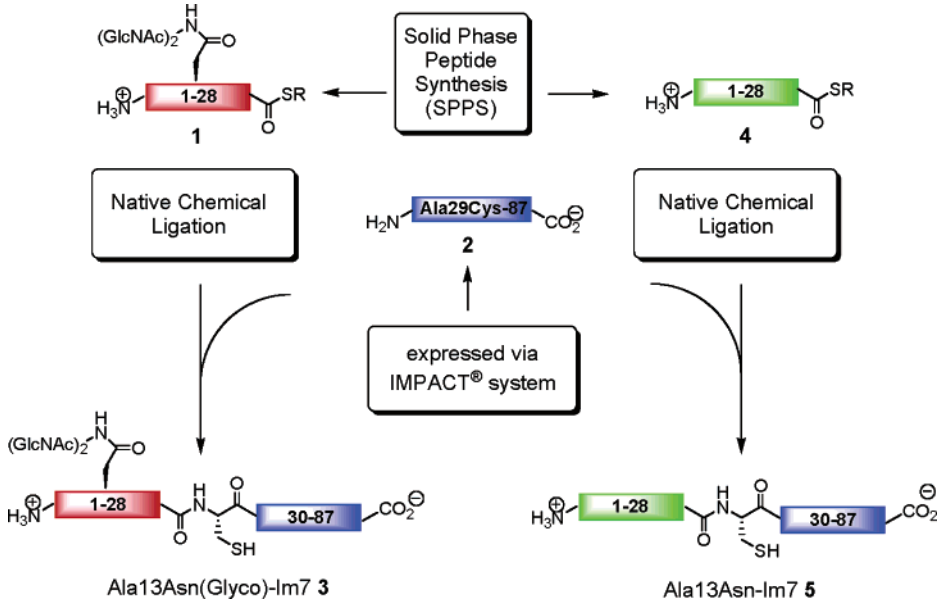
## Results and Discussion

**Synthetic Strategy.** The semisynthetic strategy for the assembly of glycosylated analogues of Im7 reported herein involves a retrosynthetic disconnection between Ala28 and Ala29 of Im7. This disconnection site was selected because it would establish the system for a favorable native ligation to the thioester of a sterically unencumbered residue at Ala28 and place the newly introduced Cys29 at an exposed site where it should not impact folding. Thus, for the assembly of Im7 analogues with glycosyl-asparagine mutations within the *N*-terminal 28 residues (Figure 1, red fragment) glycosylated thioester peptides (e.g., **1**, Scheme 1) containing residues 1−28 can be prepared by SPPS. In contrast, the larger 59-residue *C*-terminal polypeptides (e.g., **2**, Scheme 1) with an *N*-terminal Ala→Cys mutation (Figure 1, blue fragment) can be expressed. The Ala→Cys mutation is essential for the native chemical ligation step and therefore appears in all the semisynthetic proteins presented; however, it is not expected to interfere with the native Im7-fold, since the residue is placed in a loop region between helix I and II and lacks any interresidue contacts (Figure 1).[12] In addition, native chemical ligation with a *C*-terminal Ala-thioester is known to proceed with high conversion rates.[13]

As the first glycosylated variant of Im7, Ala13 was replaced with an Asn-linked chitobiose residue to yield the glycoprotein Ala13Asn(Glyco)-Im7 (**3**, Scheme 1). The chitobiose derivative was selected on the basis of previous studies, which indicated that the disaccharide has a major effect on the conformation of short peptides.[3b] Residue 13, located toward the *N*-terminus of helix I in native Im7,[14] was chosen as an initial position for the carbohydrate modification because it is solvent accessible and does not make significant interresidue contacts in the native protein. Position 13 has been studied, as part of a Φ-value analysis of Im7, and these data indicate that the peptide backbone is highly structured at the *N*-terminal end of helix I in both the intermediate and rate-limiting transition state.[8c] We predicted therefore that the Ala13Asn(Glyco) mutation would result in a glycoprotein analogue of Im7 with a native architecture similar to that of the wild-type protein and that

(9) Hofmann, R. M.; Muir, T. W. *Curr. Opin. Biotechnol*. **2002**, *13*, 297−303.
(10) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776−779.
(11) Macmillan, D.; Bertozzi, C. R. *Angew. Chem., Int. Ed.* **2004**, *43*, 1355−1359.
(12) Mutations within this loop have been shown to result in only small changes in stability of the intermediate or native state of Im7; see ref 8d.
(13) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10068−10073.
(14) Dennis, C. A.; Videler, H.; Pauptit, R. A.; Wallis, R.; James, R.; Moore, G. R.; Kleanthous, C. *Biochem. J.* **1998**, *333*, 183−191.

***Scheme 1.*** Semisynthetic Strategy for Glycosylated and Unglycosylated Im7 Analogues **3** and **5**




glycosylation should not disrupt the folding of this Im7 variant to a degree which would preclude testing the use of established techniques to determine the effect of glycosylation on the folding landscape. The unglycosylated Im7 analogue Ala13Asn-Im7 (**5**, Scheme 1) containing the native Im7-sequence with an Ala13Asn and an Ala29Cys-mutation served as a control in the folding studies, since it incorporates an Asn-residue in place of the chitobiosyl-Asn [Asn(Glyco)] in **3**. At the outset of our studies we concentrated on the semisynthesis of the unglycosylated Ala13Asn-Im7 (**5**) via thioester **4** (Scheme 1, right), as a prelude to the synthesis of the more complex glycosylated analogue.

**SPPS of the Unglycosylated and Glycosylated Thioesters 1−28 of Im7 (1 and 4).** Peptide thioesters can be generated using Fmoc-based SPPS by the alkanesulfonamide "safety-catch" strategy,[15] which requires an activation step with diazomethane or iodoacetonitrile prior to peptide cleavage with thiols to directly yield the corresponding thioester.[16] Alternatively, alkylaluminum thiolates convert peptides on HMBA (4-hydroxymethylbenzoic acid) or PAM (4-hydroxymethyl-phenylacetamidomethyl polystyrene) resins to thioesters.[17] In the current studies we used the highly acid-labile carboxy-trityl linker to the solid support (TGT-resin) to access the *C*-terminal thioester as previously introduced by our laboratory.[18] In a similar approach, other groups have employed 2-chlorotrityl linkers.[19] Highly acid-labile resin linkers allow orthogonal cleavage from the resin under mild acidic conditions without deprotection of amino acid side chains. Therefore, the *C*-terminal carboxylic acid can be transformed selectively to the thioester by 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) activation.

For the SPPS of the peptidyl thioester (**4**, Scheme 1), amino acids 1−28 (with Asn at position 13) were coupled to the commercially available Fmoc-Ala-TGT-Resin using standard Fmoc-based SPPS coupling conditions with 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate/hydroxybenzotriazole (HBTU/HOBt) activation (Scheme 2A). Test cleavages were performed throughout showing full conversion for the SPPS. For the conversion to the *C*-terminal thioester the peptide was cleaved from the resin under mild acidic conditions (0.5% TFA in $CH_2Cl_2$) and treated with benzylmercaptan, HBTU, and diisopropylethylamine (DIPEA) in THF to furnish the corresponding thioester and then globally deprotected in a single step with cleavage cocktail H, which prevents oxidation of the *N*-terminal Met residue.[20] Thioester **4** (Scheme 1) was obtained as the major product with less than 10% side products (Figure 2A).

To obtain the *N*-linked glycopeptide thioester **1** (Scheme 1), we applied the same acid labile linker strategy as in the synthesis of thioester **4**. The *N*-linked chitobiose modification was incorporated into the peptide during the SPPS using the protected Asn-linked chitobiose building block **6** (Scheme 2B).[21] Previously, glycopeptide thioesters have been synthesized either by the "safety catch" strategy[16b] or as S-alkyl-derivatives linked to a MBHA resin (*p*-methylbenzhydryl-amine).[22] We recently described a straightforward synthesis of building block **6** for the SPPS of glycopeptides based on a new carbamate

(15) (a) Kenner, G. W.; McDermott, J. R.; Sheppard, R. C. *J. Chem. Soc., Chem. Commun.* **1971**, 636−637. (b) Backes, B. J.; Virgilio, A. A.; Ellman, J. A. *J. Am. Chem. Soc.* **1996**, *118*, 3055−3056. (c) Backes, B. J.; Ellman, J. A. *J. Org. Chem.* **1999**, *64*, 2322−2330.

(16) (a) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. *J. Am. Chem. Soc.* **1999**, *121*, 11369−11374. (b) Shin, Y.; Winans, K. A.; Backes, B. J.; Kent, S. B. H.; Ellman, J. A.; Bertozzi, C. R. *J. Am. Chem. Soc.* **1999**, *121*, 11684−11689.

(17) (a) Sewing, A.; Hilvert, D. *Angew. Chem., Int. Ed.* **2001**, *40*, 3395−3396. (b) Swinnen, D.; Hilvert, D. *Org. Lett.* **2000**, *2*, 2439−2442.

(18) Mezo, A. R.; Cheng, R. P.; Imperiali, B. *J. Am. Chem. Soc.* **2001**, *123*, 3885−3891.

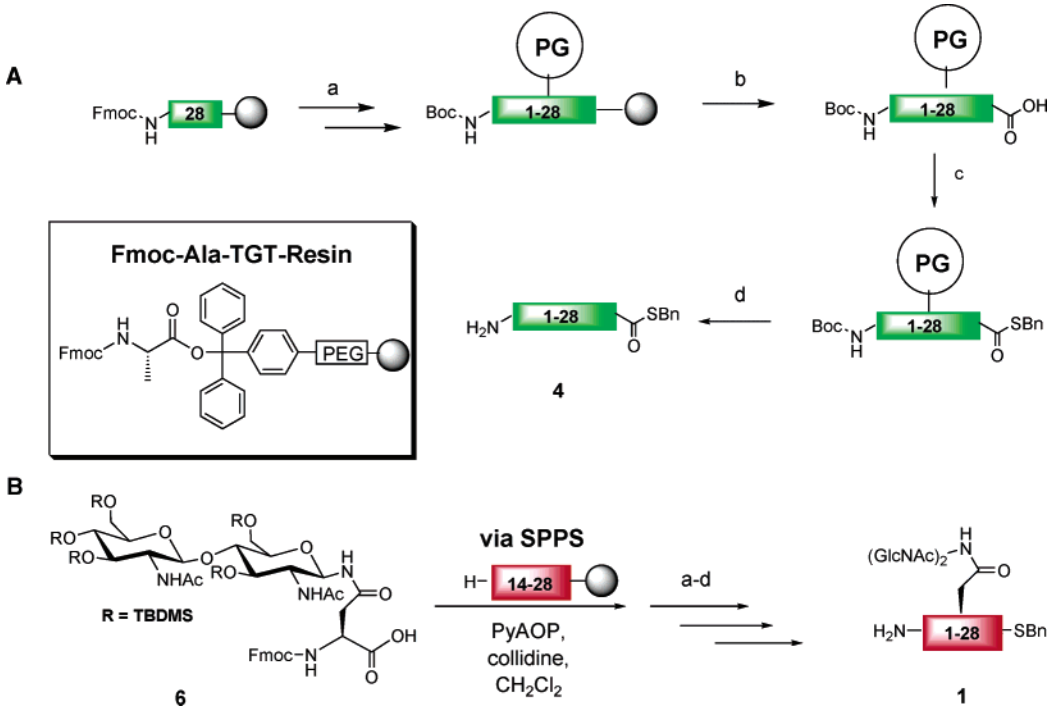
(19) (a) Futaki, S.; Sogawa, K.; Maruyama, J.; Asahara, T.; Niwa, M. *Tetrahedron Lett.* **1997**, *38*, 6237−6240. (b) Biancalana, S.; Hudson, D.; Songster, M. F.; Thompson, S. A. *Lett. Pept. Sci.* **2001**, *7*, 291−297. (c) v. Eggelkraut-Gottanka, R.; Klose, A.; Beck-Sickinger, A. G.; Beyermann, M. *Tetrahedron Lett.* **2003**, *44*, 3551−3554.

(20) Huang, H.; Rabenstein, D. L. *J. Pept. Res.* **1999**, *53*, 548−553.

(21) (a) Jobron, L.; Hummel, G. *Angew. Chem., Int. Ed.* **2000**, *39*, 1621−1624. (b) Christiansen-Brams, I.; Meldal, M.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1993**, 1461−1471. (c) Meinjohanns, E.; Meldal, M.; Paulsen, H.; Dwek, R. A.; Bock, K. *J. Chem. Soc., Perkin Trans. 1* **1998**, 549−560. (d) Holm, B.; Linse, S.; Kihlberg, J. *Tetrahedron* **1998**, *54*, 12047−12070. (e) Broddefalk, J.; Bergquist, K.; Kihlberg, J. *Tetrahedron* **1998**, *54*, 11995−12006. (f) Bosques, C. J.; Tai, V. W. F.; Imperiali, B. *Tetrahedron Lett.* **2001**, *42*, 7207−7211.

(22) (a) Hojo, H.; Aimoto, S. *Bull. Chem. Soc. Jpn.* **1991**, *64*, 111−117. (b) Hojo, H.; Watabe, J.; Nakahara, Y.; Ito, Y.; Nabeshima, K.; Toole, B. P. *Tetrahedron Lett.* **2001**, *42*, 3001−3004. (c) Hojo, H.; Haginoya, E.; Matsumoto, Y.; Nakahara, Y.; Nabeshima, K.; Toole, B. P.; Watanabe, Y. *Tetrahedron Lett.* **2003**, *44*, 2961−2964.

***Scheme 2.*** SPPS Synthesis of (A) Unglycosylated Thioester 1−28 (**4**) and (B) Glycosylated Thioester 1−28 (**1**) Using the Acid-Labile TGT Resin (PG = Protecting Group)[a]




[a] Reagents and conditions: (a) (i) 20% piperidine/DMF; (ii) Fmoc-Xaa-OH/Boc-Met-OH, HBTU, HOBt, DIPEA, NMP; (b) 0.5% TFA, $CH_2Cl_2$; (c) BnSH, HBTU, DIPEA, THF; (d) cleavage cocktail H (TFA, phenol, EDT, thioanisole, TIS, $Me_2S$, $NH_4I$, $H_2O$).

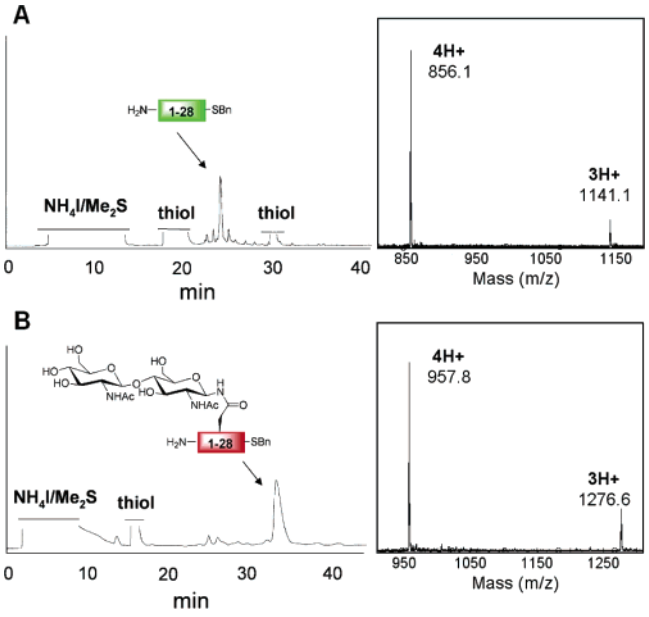



***Figure 2.*** HPLC spectra of Im7 (1−28) peptide thioesters after cleavage with cocktail H. (A) Ala13Asn thioester **4**. (B) Ala13Asn(Glyco) thioester **1**. For ESI-MS analysis, gradients, and cleavage cocktail contents see Experimental Section.

based amination protocol, which was implemented during our studies.[23]

Building block **6** (Scheme 2B) was coupled to the peptide 14−28 on the TGT-resin by activation with PyAOP/collidine in $CH_2Cl_2$, which was necessary due to the increased steric demand of the building block (Scheme 2B). A test cleavage revealed two major products, which were identified as the desired glycopeptide with, and without, one *tert*-butyldimethylsilyl (TBDMS) protecting group still attached to the carbohydrate hydroxyls. 7-Azabenzotriazol-1-yloxytris(pyrrolidino)-phosphonium hexafluorophosphate (PyAOP)/collidine coupling conditions were employed for the next two amino acids before the coupling protocol was switched back to the conventional HBTU/HOBt activation.

Transformation to the corresponding thioester **1** was performed with the same protocol as that described for the unglycosylated thioester **4** (Scheme 2B) yielding thioester **1** in an excellent conversion (>85%). In this case, peaks corresponding to products with an additional TBDMS group were not observed (Figure 2B).

**Native Chemical Ligation of Thioesters (1 and 4) with the Ala29Cys-87 Peptide (2).** Having accomplished the synthesis of the unglycosylated and glycosylated thioesters, we focused on the expression of the *C*-terminal fragment, Ala29Cys-87 (**2**, Scheme 1). The target Im7 fragment was obtained via recombinant methods using the IMPACT system, which is based on a pH-dependent intein cleavage.[24]

For the NCL of thioester **1** with the expressed protein **2** we found that standard conditions without the addition of denaturation reagents were sufficient to ensure full conversion after 24 h as monitored by analytical HPLC. However, purification of the ligated material proved to be difficult. HPLC separation using water/$CH_3CN$ (0.1% TFA) gradients was unsuccessful, and less than 5% of the desired protein material was isolated, possibly because the product aggregated in the organic solvents.

(23) Hackenberger, C. P. R.; O'Reilly, M. K.; Imperiali, B. *J. Org. Chem.* **2005**, *70*, 3574−3578.

(24) Evans, T. C., Jr.; Benner, J.; Xu, M.-Q. *J. Biol. Chem.* **1999**, *274*, 3823−3926. For a recent example, see: Pezza, J. A.; Allen, K. N.; Tolan, D. R. *Chem. Commun.* **2004**, 2412−2413.

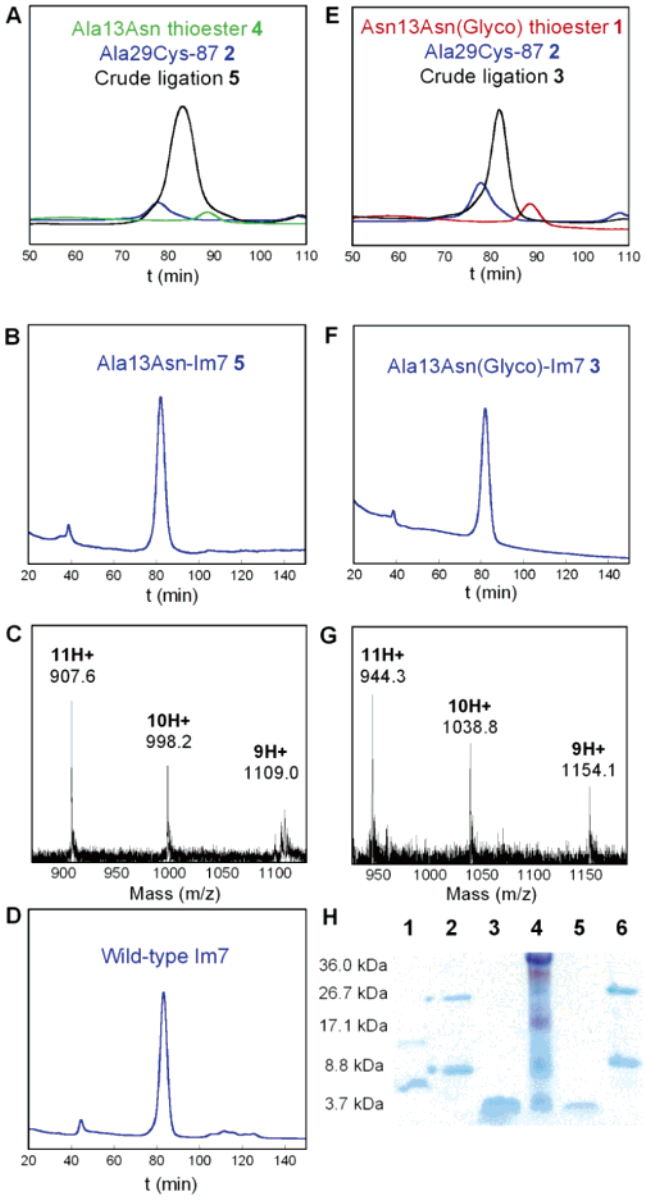



***Figure 3.*** NCL of unglycosylated and glycosylated Im7-analogues **3** and **5**: (A) SEC (see Experimental Section for exact conditions) of crude NCL of **5**. (B) SEC of purified **5**. (C) ESI-MS data for purified **5**. (D) SEC of purified wild-type Im7. (E) SEC of crude NCL of **3**. (F) SEC of purified **3**. (G) ESI-MS data for **3**. (H) Nonreducing Tris/Tricine 16.5% peptide gel. Lane 1: Ala29Cys-87 (**2**). Lane 2: Ala13Asn-Im7 (**5**). Lane 3: Ala13Asn thioester (**4**). Lane 4: Molecular weight markers. Lane 5: Ala13Asn(Glyco) thioester (**1**). Lane 6: Ala13Asn(Glyco)-Im7 (**3**).

The purification of the crude ligation mixture was ultimately accomplished by size exclusion chromatography (SEC). A typical trace of the ligation of the unglycosylated thioester **4** with the Ala29Cys-87 peptide (**2**) is shown in Figure 3A. All peaks were assigned based on ESI-MS analysis and injection of pure starting materials. The fractions which contained the ligation product **5** were pooled, dialyzed, analyzed by SEC reinjection (Figure 3B) and ESI-MS (Figure 3C), and finally compared to wild-type Im7, which was obtained and purified by established expression methods (Figure 3D).[25]

Several important observations were made during the course of the NCL. The ligation mixture must be used directly for the separation procedure without a lyophilization step to avoid oxidation products. Furthermore, we found that, upon storage, the Im7-proteins **3** and **5** oxidizes to the dimeric disulfide, which can be readily rereduced by treatment with dithiothreitol (DTT). Volatile or nonvolatile buffers can be used during SEC; however the former are preferential since this offers a direct isolation of the protein material by simple lyophilization. Finally, we noted that the expressed and reduced protein, **2**, eluted first from the SEC, despite its lower molecular mass (6.7 kDa) compared to the Im7-protein **5** (10.1 kDa). This behavior can be attributed to the hydrophobic area which is displayed by the missing helix I leading to the observed aggregation phenomenon. Considering the conversion and the recovery of the ligation product based on the amount of starting materials, a BCA assay and UV concentration determination revealed an isolated ligation yield of 87%.

The NCL for the glycosylated Im7 analogue **3** proceeded in a similar overall yield of 79% (Figure 3E-G). Further proof for the purity of protein **3** and **5** was obtained from native peptide gel analysis under nonreducing conditions (Figure 3H). Combining these results, sufficient amounts of the desired glycosylated and unglycosylated Im7-analogues **3** and **5** for the protein folding studies could be isolated.

**Protein Folding Studies.** To evaluate the differences between the glycosylated and unglycosylated analogues of the Im7 variant, fluorescence emission and far UV circular dichroism studies were performed. Characteristic tryptophan emission spectra have been reported for the folded and unfolded states of Im7.[8d] The fluorescence signal of denatured Im7 results largely from the single tryptophan residue (Trp75). This residue is packed against His47 in the native protein resulting in highly quenched fluorescence (Figure 4A). Fluorescence and CD measurements at 10 °C reveal that expressed wild-type Im7, Ala13Asn-Im7 (**5**), and Ala13Asn(Glyco)-Im7 (**3**) behave similarly under folding and unfolding conditions; all proteins adopt a highly helical native state demarked by highly quenched tryptophan fluorescence and a typical helical signal in far UV (Figure 4A and B).

A detailed analysis of the effect of glycosylation on the folding kinetics of Im7 was carried out using stopped-flow fluorescence measurements (Figure 4C). The data revealed that all three proteins (Im7, Ala13Asn-Im7 (**5**), and Ala13Asn-(Glyco)-Im7 (**3**)) fold to the native state via a three-state mechanism, as shown by the characteristic nonlinearity of the logarithm of the folding rate constants with denaturant. Also, the initial fluorescence signal of the folding transients shows the hyperfluorescence at low denaturant concentration, which is characteristic of the population of the Im7 intermediate in the dead-time of the stopped flow (~3 ms) (data not shown).[8] Fitting the data to an on-pathway three-state model (see Table 1) shows that the stability of the intermediate was not significantly effected by the incorporation of a chitobiose moiety on residue 13 ($\Delta G_{ui} \approx -13$ kJ/mol for both Ala13Asn-Im7 and Ala13Asn(Glyco)-Im7) (Table 1). In addition, the stabilities of the native states of the three proteins, determined from both kinetic analysis and temperature denaturation, were similar ($\Delta G_{un} = -24.1$ to $-26.7$ kJ/mol) (Figure 4C and D and Table 1). The slight destabilization of the unglycosylated variant Ala13Asn-Im7 (**5**) relative to wild-type Im7 ($\Delta\Delta G_{un} = +1.4$ kJ/mol) could be due to either or both of the amino acid substitutions in the semisynthetic protein. However, most im-

(25) Wallis, R.; Reilly, A.; Rowe, A.; Moore, G. R.; James, R.; Kleanthous, C. *Eur. J. Biochem.* **1992**, *297*, 687−695.

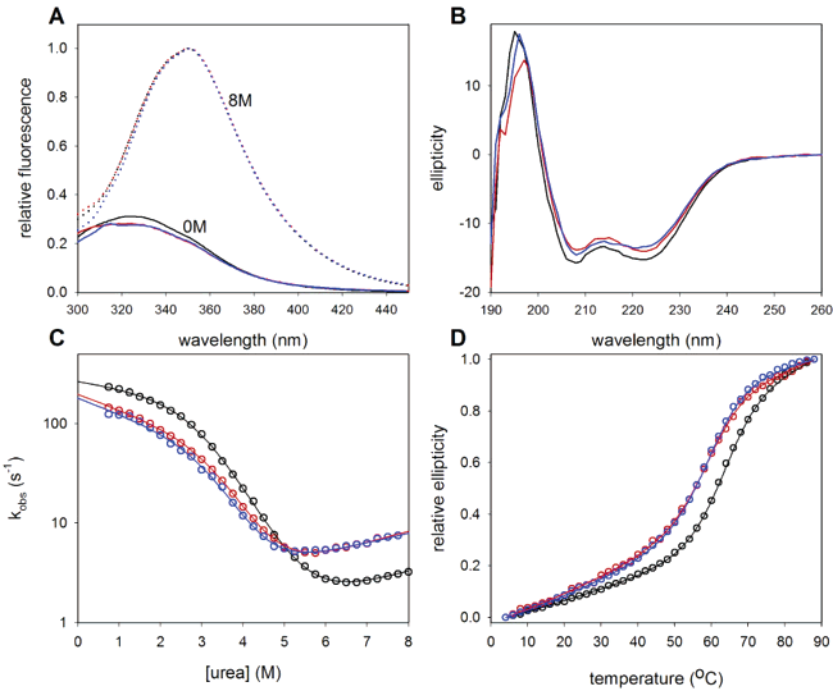



***Figure 4.*** Characterization of Im7 (black), Ala13Asn-Im7 (**5**) (red), and Ala13Asn(Glyco)-Im7 (**3**) (blue). (A) Fluorescence emission spectra under native (0 M urea) and denaturing (8 M urea) conditions. (B) Far UV CD spectra of the native proteins (0 M urea). (C) Folding and unfolding kinetics measured by stopped-flow fluorescence. (D) Thermal denaturation monitored by far UV CD (see Experimental Section for additional details).

***Table 1.*** Parameters Determined from Analysis of Thermal Denaturation Experiments ($T_m$) and Folding/Unfolding Kinetics

| protein | $T_m$ (°C) | $K_{ui}$ ($M_{ui}$)[b] | $k_{in}$[a] ($m_{in}$)[b] | $k_{ni}$[a] ($m_{ni}$)[b] | $M_{un}$[b] | $\Delta G_{ui}$[c] | $\Delta G_{un}$[c] |
|---|---|---|---|---|---|---|---|
| Im7 | 62.3 ± 0.2 | 172 ± 25 (3.89 ± 0.05) | 275 ± 10 (0.46 ± 0.07) | 0.55 ± 0.07 (0.52 ± 0.04) | 4.9 ± 0.1 | −12.1 ± 0.3 | −26.7 ± 0.5 |
| A13N | 58.0 ± 0.4 | 256 ± 68 (3.84 ± 0.11) | 198 ± 6 (0.86 ± 0.06) | 1.09 ± 0.10 (0.60 ± 0.03) | 5.3 ± 0.1 | −13.1 ± 0.6 | −25.3 ± 0.7 |
| A13NGlyco | 58.6 ± 0.2 | 222 ± 115 (3.90 ± 0.22) | 179 ± 11 (0.87 ± 0.13) | 1.40 ± 0.22 (0.51 ± 0.05) | 5.3 ± 0.3 | −12.7 ± 1.2 | −24.1 ± 1.3 |

[a] $s^{-1}$. [b] kJ $mol^{-1}$ $M^{-1}$. [c] kJ $mol^{-1}$.

portantly, the addition of a sugar at position 13 has no significant effect on the stability of the intermediate or native state or on their respective *M*-values, which report on the degree of compactness of these species relative to the denatured state (Table 1). The observation that a sugar located at the *N*-terminus of helix I does not perturb the folding pathway is in agreement with previous data which suggests that this position is structured and solvent exposed at all points on the Im7-folding landscape.[8b,c] The introduction of large polar sugar moieties at positions which are natively buried from solvent or, perhaps more interestingly, at residues, which are solvent exposed in the native state but may become buried during the course of folding would, by contrast, be predicted to have a dramatic effect on folding kinetics and will offer exciting information on the plasticity of the Im7-folding landscape and the importance of non-native interactions in the folding of this protein. While studies on further glycosylated Im7 variants are planned, the study described here demonstrates that such variants can be created in sufficient quantity and to a degree of homogeneity that allows the utilization of established techniques for the determination of their folding kinetics and thermodynamics with a high degree of accuracy.

## Conclusion

In summary, we have developed a method for semisynthesis of the immunity protein Im7, which allows access to *N*-linked glycoprotein variants as well as their unglycosylated analogues for protein folding studies. Glycosylated protein can be produced in sufficient quantities and to a degree of purity and homogeneity to allow the use of various biophysical techniques for their detailed characterization, including kinetic analysis of the folding and unfolding of the glycosylated and unglycosylated variants. These results demonstrate the power of a combination of chemical synthesis and kinetic analysis to allow the complete characterization of the role of the incorporation of individual sugar moieties, at specific positions, on the folding free energy landscape of a small protein. By using the powerful and well-characterized protein model, Im7, detailed comparisons can be made between the folding landscape of the wild-type protein and its glycosylated variants. The small size of Im7 also enhances the prospects for studies of this system using simulations, perhaps to predict locations for glycosylation that would be interesting and informative in their effect on the folding energy landscape of this protein. Further studies on the influences of glycosylation on protein folding using other glycoprotein analogues of Im7 are in progress.

## Experimental Section

**General.** All reagents, amino acids, and solvents were purchased from commercial suppliers and used without further purification. All

solvents were reagent grade and used as received. Synthetic Im7 variants Ala13Asn-Im7 (**5**) and Ala13Asn(Glyco)-Im7 (**3**) were obtained by a semisynthetic strategy as described (Scheme 1). Wild-type Im7 was expressed and purified as previously described.[25]

**Peptide Synthesis.** Unglycosylated or glycosylated peptides were synthesized on an ABI 431A peptide synthesizer using standard amide coupling conditions HBTU/HOBt. As the solid support, the TGT-Resin (Novabiochem) was used with the first amino acid (Ala) already attached to the resin. The Asn-linked chitobiose was introduced as a TBDMS-protected building block using PyAOP/collidine activation in dichloromethane.[23]

**Thioester Synthesis.** The thioester synthesis followed our previously published protocol.[18] Briefly, TGT resin containing the desired peptide was treated with 0.5% TFA in $CH_2Cl_2$ for 2 h. The resin was filtered off and washed with dichloromethane. The combined filtrates delivered the protected peptides (1−28) upon precipitation from hexanes (100 mL) and removal of the organic solvents in vacuo. The protected peptides were dissolved in THF (5 mL), and HBTU (4 equiv), diisopropylethylamine (8 equiv), and benzylmercaptan (4 equiv) were added for the conversion into the protected benzylthioesters **1** and **4**. After removal of the THF in vacuo, peptides were globally deprotected with cleavage cocktail H[20] (82.5% TFA; 2.5% ethanedithiol (EDT); 5% thioanisole; 5% $H_2O$; 5% dimethyl sulfide ($Me_2S$); with additional phenol (75 mg), $NH_4I$ (50 mg), and triispropylsilane (TIS, 100 $\mu$L) added per mL). Peptidethioesters **1** and **4** were precipitated from ether (40 mL), redissolved in $H_2O/CH_3CN$ mixtures (1:1), and purified via preparative HPLC ($C_{18}$-column, flow: 15 mL/min, water/acetonitrile (with 0.1% TFA) gradient). Electrospray ionization mass spectrometry (ESI-MS) analysis was used to confirm the identities of the unglycosylated and glycosylated thioesters.

**Glycosylated Ala13Asn Thioester 1.** The peptide thioester **1** was purified via HPLC at a gradient of 7% to 95% $CH_3CN$ over 40 min, which eluted at 34.5 min (see Figure 2B). MS (ESI-MS): 1276.6 ($[M/3]^+$, observed), 1276.7 ($[M/3]^+$, calculated); 957.8 ($[M/4]^+$, observed), 957.7 ($[M/4]^+$, calculated).

**Unglycosylated Ala13Asn Thioester 4.** The peptide thioester **4** was purified via HPLC at a gradient of 7−95% $CH_3CN$ over 35 min, which eluted at 22.8 min (see Figure 2A). MS (ESI-MS): 1141.1 ($[M/3]^+$, observed), 1141.2 ($[M/3]^+$, calculated); 856.1 ($[M/4]^+$, observed), 856.2 ($[M/4]^+$, calculated).

**Native Chemical Ligation of Ala13Asn(Glyco)-Im7 (3) and Ala13Asn-Im7 (5).** The *C*-terminal polypeptide Ala29Cys-Gly87 (**2**) was expressed using the IMPACT system, which is based on a pH-dependent intein cleavage.[24] For the native chemical ligation, a 3 mM solution (100 mM phosphate buffer, pH 7.0) of the unglycosylated and glycosylated peptide thioesters **1** and **4** was added to an equivalent volume of a 2 mM solution of expressed protein **2** (100 mM phosphate buffer, pH 7.0). 2% Benzylmercaptan and 2% thiophenol were added, and the ligation mixture was agitated for 16 h at 25 °C. A white precipitate was formed which was spun down and washed with phosphate buffer. The combined supernatants were purified directly by size exclusion chromatography [SEC, ÄKTA-FPLC, Amersham Biosciences, Superdex 75 column (optimal separation range between 10 and 70 kDa)] with either a 50 mM HEPES/150 mM NaCl buffer (pH 7.0) or a 100 mM $NH_4OAc$ buffer (pH 7.0) as eluent. Each protein was analyzed by ESI-MS, SEC-reinjection, and peptide gel electrophoresis (see Figure 3) to confirm the purity of the semisynthetic protein.

**Ala29Cys-87 (2).** The expressed protein eluted on HPLC at a gradient of 7% to 95% $CH_3CN$ over 35 min at 26.9 min. MS (ESI-MS): 1336.1 ($[M/5]^+$, observed), 1335.6 ($[M/5]^+$, calculated); 1113.6 ($[M/6]^+$, observed), 1113.2 ($[M/6]^+$, calculated); 954.6 ($[M/7]^+$, observed), 954.3 ($[M/7]^+$, calculated); 835.2 ($[M/8]^+$, observed), 835.2 ($[M/8]^+$, calculated).

**Ala13Asn(Glyco)-Im7 3.** The semisynthetic glycoprotein **3** was purified via SEC, which eluted at 80.1 min (see Figure 3F). MS (ESI-MS, see Figure 3G): 1154.1 ($[M/9]^+$, observed), 1153.9 ($[M/9]^+$, calculated); 1038.8 ($[M/10]^+$, observed), 1038.6 ($[M/10]^+$, calculated); 944.3 ($[M/11]^+$, observed), 944.3 ($[M/11]^+$, calculated).

**Ala13Asn-Im7 5.** The semisynthetic glycoprotein **3** was purified via SEC, which eluted at 80.0 min (see Figure 3B). MS (ESI-MS, see Figure 3C): 1109.0 ($[M/9]^+$, observed), 1108.8 ($[M/9]^+$, calculated); 998.2 ($[M/10]^+$, observed), 998.0 ($[M/10]^+$, calculated); 907.6 ($[M/11]^+$, observed), 907.4 ($[M/11]^+$, calculated).

**Biophysical Characterization of Wild-Type Im7, Ala13Asn-(Glyco)-Im7 3, and Ala13Asn-Im7 5.** All measurements were carried out in buffer A (50 mM sodium phosphate, 0.4 M sodium sulfate, 2 mM DTT (pH 7.0)) and at 10 °C, unless otherwise stated. DTT is added to prevent disulfide formation. This concentration of DTT is known to prevent disulfide formation in the Im7 homologue, Im9, which naturally contains a single Cys residue.[26]

**Fluorescence Spectra.** Fluorescence emission spectra ($\lambda_{ex}$ = 280 nm) of each protein were measured using a Photon Technology International fluorimeter (Ford, West Sussex, UK). For spectra of native and denatured Im7 species, proteins (~5 $\mu$M) were dissolved in buffer A or buffer A containing 8 M urea, respectively. After subtraction of the buffer, background spectra were normalized to the signal of their respective denatured states at 350 nm.

**Circular Dichroism.** Far UV CD spectra were acquired on a Jasco J715 CD spectropolarimeter (Great Dunmow, Essex, UK) using a 1 mm path length cell and a protein concentration of ~20 $\mu$M, as previously described.[8d]

**Stopped-Flow Fluorescence.** Kinetic folding and unfolding measurements were performed using an Applied Photophysics SX18.MV stopped-flow fluorimeter, as previously described.[26] Briefly, for refolding experiments the denatured protein (~40 $\mu$M) in buffer A containing 8 M urea was refolded by 1:10 (v/v) dilution into buffer A containing various concentrations of urea. For unfolding experiments folded protein (~40 $\mu$M) in buffer A without urea was diluted 1:10 (v/v) into buffer A containing different concentrations of urea.

The observed rate constants for all proteins studied were fitted to an on-pathway three-state model (eq 1).

$$\mathrm{U} \underset{k_{iu}}{\overset{k_{ui}}{\rightleftharpoons}} \mathrm{I} \underset{k_{ni}}{\overset{k_{in}}{\rightleftharpoons}} \mathrm{N} \qquad (1)$$

where U, I, and N represent the denatured, intermediate, and native states, respectively, and $k_{xy}$ is the microscopic rate constant for the conversion of $x$ to $y$. Previous analysis, using continuous-flow and stopped-flow fluorescence, has shown that the intermediate formed during Im7-folding is on-pathway.[8a] The logarithm of the observed rate constants for each data set was fitted using the analytical solution to the coupled differential equations describing eq 1[27] in Origin 7.0 (Originlab Corporation). To fit these data in the absence of measured rate constants describing the fast phase of the reaction, $k_{ui}$, and its associated $m$-value was held constant ($k_{ui}$ was set to 3000 $s^{-1}$ and $m_{ui}$ to 1.3 kJ $mol^{-1}$ $M^{-1}$, values previously determined for Im7 using continuous-flow ultra rapid mixing).[8a] The fits obtained are shown in Figure 4C as solid lines. The parameters determined from the fits are listed in Table 1. Errors quoted are the standard errors resulting from the fits. The parameters shown in Table 1 are calculated from parameters directly determined from the fits as follows: $K_{ui} = k_{ui}/k_{iu}$, $M_{ui} = m_{ui} + m_{iu}$, $M_{un} = M_{ui} + m_{in} + m_{ni}$, $\Delta G_{ui} = -RT \ln(K_{ui})$ and $\Delta G_{un} = -RT \ln(K_{ui}(k_{in}/k_{ni}))$.

**Thermal Denaturation.** Measurements were made using a Jasco-J715 spectropolarimeter with a 1 mm path length cuvette and at a protein concentration of ~20 $\mu$M. The denaturation process was monitored at

(26) Friel, C. T.; Capaldi, A. P.; Radford, S. E. *J. Mol. Biol.* **2003**, *326*, 293−305.
(27) Gorski, S. A.; Capaldi, A. P.; Kleanthous, C.; Radford, S. E. *J. Mol. Biol.* **2001**, *312*, 849−863.

220 nm. The data were fitted to a model describing a two-state transition (eq 2).

$$\text{Signal} = \frac{\left((a+bT)+\left((c+dT)\left(\exp\left(-\Delta H_m + T\left(\frac{\Delta H_m}{T_m}\right) - \Delta C_p\left(T - T_m - T\ln\left(\frac{T}{T_m}\right)\right)\right)/RT\right)\right)\right)}{\left(1+\left(\exp\left(-\Delta H_m + T\left(\frac{\Delta H_m}{T_m}\right) - \Delta C_p\left(T - T_m - T\ln\left(\frac{T}{T_m}\right)\right)\right)/RT\right)\right)} \quad (2)$$

where $T$ is the temperature in degrees Kelvin, $a + bT$ and $c + dT$ represent the pre- and post-transition baselines, respectively, $\Delta H_m$ is the enthalpy change for the transition at the midpoint, $T_m$ is the temperature at the midpoint, and $\Delta C_p$ is the specific heat change upon unfolding. The data were fitted globally sharing the $\Delta C_p$ between all data sets.

**Acknowledgment.** This work was funded by the NIH (GM 39334). C.P.R.H. acknowledges a postdoctoral fellowship from the German Academic Exchange Service (DAAD) and the German Science Foundation (DFG). C.T.F is funded by the BBSRC and S.E.R. is a BBSRC Professorial Fellow. We thank Mary O'Reilly, Beth Vogel, K. Jebrell Glover, Carlos J. Bosques, Eva Sanchez-Cobos, and Guofeng Zhang for helpful discussions and advice as well as Hector H. Hernandez and Professor Cathy Drennan for the use of their FPLC system.

JA051855K