

Efficient N-Terminal Glycoconjugation of Proteins by the N-End Rule

Lars Merkel,^[a] Henning S. G. Beckmann,^[b] Valentin Wittmann,^{*,[b]} and Nediljko Budisa^{*,[a]}

The importance of protein N terminus sequence composition for cell physiology was recognized more than two decades ago.^[1] However, its relevance for chemical protein engineering through an expanded genetic code was demonstrated only very recently.^[2] Nature changes the chemistry of the N terminus by posttranslational modifications (PTMs) such as long-chain alkylation, acetylation, myristoylation, glycosylation, etc.^[3,4] This, in turn, influences the lifetimes and general metabolic fates of tagged proteins in different ways according to the N-end rules.^[5] Although Met is the first amino acid in a newly synthesized protein, it is usually enzymatically removed from the mature protein when the second position is occupied by a non-bulky residue (for example, Ala, Cys, Gly).^[6] On the other hand, bulky amino acids—such as Lys, Arg, Leu, Phe, and Ile—in this position protect the N-terminal Met from being processed.^[6] The extension of these rules to noncanonical amino acids occupying the protein N terminus should enable the *in vivo* generation of stable artificial N-terminal handles. Their subsequent chemical derivatization might generate new specific functions, especially if carbohydrates are attached.

Protein glycosylation is the most complex form of PTM, and the development of methods for the preparation of homogeneously glycosylated proteins or related mimics is of utmost importance for unraveling the biological roles of glycans.^[7,8] Traditional approaches for chemical posttranslational modification by derivatization of solvent-accessible reactive side chains such as those of lysine and cysteine often lead either to heterogeneous or to hardly reproducible mixtures of modified proteins,^[9] and so selective methods that will allow absolute control of the type and positions of glycans within a protein are required. Walsh and co-workers applied the copper(I)-catalyzed^[10,11] Huisgen [3+2] cycloadditions^[12] (CCHCs) of azides and alkynes (examples of click chemistry reactions^[13]) in order to introduce carbohydrate residues into synthetic cyclopeptide antibiotics containing propargylglycine residues.^[14] The potential to translate a noncanonical functional group—such as ketone, azide, or alkyne—that can be chemoselectively modified into a protein structure during ribosomal synthesis^[15–17] opens a way to generate homogeneous, structurally defined

glycoproteins with the carbohydrate installed at a preselected site.

In the context of suppressor-based methodology, site-specific incorporation of *p*-acetyl-phenylalanine was first performed. Aminoxy derivatives of various sugars were subsequently coupled to the keto group in the protein.^[15] A step further was the introduction of monoglycosylated amino acids into proteins in response to an amber stop codon,^[18,19] to which additional sugar molecules were attached by the enzymatic action of glycosyltransferases.^[18] While this work provided a proof-of-principle for the use of an expanded genetic code in glycobiology, suppression-based methodology is generally limited by low production yields and a technically expensive experimental setup. Very recently, Davis and co-workers used a more efficient auxotrophy-based residue-specific method to introduce azidohomoalanine (Aha) and homopropargylglycine (Hpg) into engineered TIM barrel and LacZ proteins.^[20] Their substituted proteins (Met→Aha and Met→Hpg) were subjected to CCHC reactions with alkyne- or azide-substituted carbohydrates to provide homogeneous protein glycoconjugates.

Only recently, we have shown that excision of N-terminal Met analogues can be effectively prevented by the presence of bulky amino acids such as Arg or Lys at the second and even third sequence positions.^[2] Here we demonstrate that the presence of two bulky Lys residues in sequence positions two and three of the protein barstar indeed prevents excision of N-terminal Aha. This approach artificially creates a new handle exclusively at the protein's N terminus, which allows the use of the engineered azido function for subsequent N-terminal conjugation with synthetic alkyne-derivatized carbohydrates.

Several methods for site-selective modification of the N termini of proteins have been developed. To some extent, direct acylation of N-terminal amino groups can be achieved because of their lower pK_a values (in relation to side-chain amino groups) by careful control of the reaction pH.^[22] The selectivity of this reaction is limited, however, because of the large number of lysine residues contained in most proteins. Reactive carbonyl groups can be produced through periodate oxidation of N-terminal serine and threonine residues^[23] (delivering aldehydes) or through reactions between N termini and pyridoxal-5-phosphate, leading in a two-step sequence to the formation of ketones.^[24] These carbonyl groups can subsequently be ligated with alkoxyamines to form oximes. Oxime formation, however, while proceeding with high chemoselectivity, is potentially reversible, which can lead to loss of the conjugated moiety. Alternative methods include treatment of N-terminal cysteine residues with thioesters to produce peptide bonds through native chemical ligation^[25–27] and with aldehydes to form thiazolidines.^[28] Beside these chemical methods, several approaches based on *in vitro* translation with pre-charged initiator tRNAs have been reported.^[29,30]

[a] L. Merkel, Dr. N. Budisa

Max Planck Institute of Biochemistry, Molecular Biotechnology
Am Klopferspitz 18, 82152 Martinsried (Germany)
Fax: (+49) 89-8578-3557
E-mail: budisa@biochem.mpg.de

[b] H. S. G. Beckmann, Prof. Dr. V. Wittmann

Fachbereich Chemie, Universität Konstanz
78457 Konstanz (Germany)
Fax: (+49) 7531-88-4573
E-mail: mail@valentin-wittmann.de

Supporting information for this article is available on the WWW under <http://www.chembiochem.org> or from the author.

Barstar is a small recombinant protein consisting of 90 amino acids and is widely used for folding studies. In our study, engineered cysteine-free "pseudo-wild-type barstar" (ψ -b*), Pro28Ala/Cys41Ala/Cys83Ala with only one Met residue at the N terminus (Met1), was employed (Scheme 1 A).^[31] N-termi-

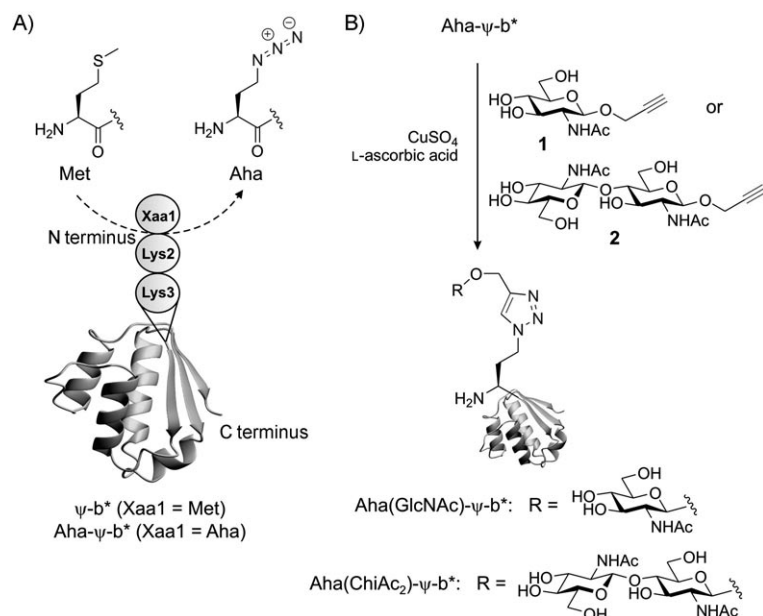
disaccharide at the beginning of the core structure common to all *N*-glycoproteins. *O*-GlcNAc modification of proteins is a ubiquitous form of PTM predominantly found in nuclear and cytoplasmic proteins and is involved in many cellular regulation processes.^[32] Both carbohydrate structures are known to

be recognized by wheat germ agglutinin (WGA), a lectin that is used to detect natural *O*-GlcNAc PTM.

The CCHC reactions between Aha- ψ -b* and the alkyne-containing glycosides **1** and **2** were each carried out in aqueous buffer in the presence of CuSO₄ and L-ascorbic acid (Scheme 1 B). Mass spectrometric analyses of the obtained glycoconjugates Aha-(GlcNAc)- ψ -b* and Aha(ChiAc₂)- ψ -b* revealed full agreement between expected and found masses, as shown in Figure 1 and Table 1. Interestingly, after the CCHC reaction, the Aha- ψ -b* species was no longer detectable by ESI-MS, indicating virtually quantitative glycoconjugation. Furthermore, sugar conjugation caused a small but detectable shift in SDS-PAGE of the corresponding protein samples (Figure S2).

Far-UV CD analyses of the native protein, Aha- ψ -b*, and related glycoproteins revealed almost superimposable spectra (Figure 2A). The unchanged spectral shapes strongly suggest identical overall secondary structures, at least within the limits of this spectroscopic technique. Small variations in CD intensities between ψ -b* and Aha- ψ -b* are most probably due to minor differences in sample concentrations, whereas the dichroic intensity increase around 220 nm in the glycoconjugated variants suggests an increase in ordered structure.

On the other hand, replacement of the N-terminal Met of ψ -b* with Aha and subsequent glycoconjugation leads to slightly less stable proteins in terms of *T*_m values that are lowered by 2–4 °C as revealed in thermal unfolding experiments (Table 1, (Figure S3)). Interestingly, the chemical exchange from thioether (Met) to azide (Aha) at the



Scheme 1. A) Three-dimensional structure (ribbon plot) of pseudo-wild-type barstar (ψ -b*; Pro28Ala/Cys41Ala/Cys83Ala)^[21] with marked N and C termini. The overall protein structure of ψ -b* consists of a $\beta\alpha\beta$ motif which is characteristic for some nucleic acid binding proteins. ψ -b* contains a single N-terminal Met, which was replaced with azido-homoalanine (Aha) by the auxotrophy-based residue-specific method to give Aha- ψ -b*. The second and third sequence positions are occupied by bulky Lys residues, which protect the N-terminal amino acid from cotranslational cleavage. B) CCHC reactions between azide-containing Aha- ψ -b* and propargyl glycosides **1** and **2** resulted in the formation of triazole-linked glycoconjugates Aha(GlcNAc)- ψ -b* and Aha(ChiAc₂)- ψ -b*, respectively.

nal glycoconjugation was expected to deliver a novel functional feature such as the ability to bind lectins. In the process, its original function (that is, ribonuclease inhibition) should be retained.

For successful Met \rightarrow Aha replacement in recombinant ψ -b*, *Escherichia coli* B834(DE3) host cells were grown in new minimal medium (NMM) with Met (0.025 mM) as natural substrate until its exhaustion, followed by simultaneous addition of Aha and target gene induction with isopropyl- β -D-thiogalactopyranoside (Supporting Information). Under these conditions, azide-labeled protein Aha- ψ -b* (Scheme 1 A) was expressed in yields of about 50%. From one liter of culture we purified 5 mg, in comparison with 10 mg L⁻¹ for ψ -b*. Electrospray ionization mass-spectrometric analysis (ESI-MS) of ψ -b* and Aha- ψ -b* clearly revealed a high level of replacement (ca. 90–95%), with small amounts of the parent protein as contaminant (Figure S1).

For N-terminal modification of the protein variant Aha- ψ -b* by CCHC reactions, the unprotected propargyl glycosides of *N*-acetylglucosamine (GlcNAc β -O-CH₂-CCH, **1**) and *N,N'*-diacetylchitobiose (GlcNAc β (1,4)-GlcNAc β -O-CH₂-CCH, **2**) were selected (Scheme 1 B). *N,N'*-Diacetylchitobiose (ChiAc₂) represents the

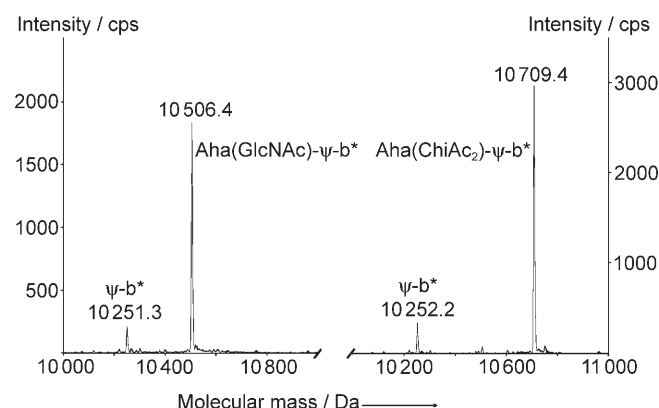


Figure 1. Deconvoluted ESI mass spectra after modification of Aha- ψ -b* with **1** and **2**, respectively. The smaller peaks in both spectra correspond to the mass of parent ψ -b* (see Table 1 for details). The absence of Aha- ψ -b* in the spectra confirms virtually quantitative glycoconjugation at the protein N terminus.

Protein	Mass [Da]		T_m [°C] ^[a]	λ_{max} [nm] ^[b]
	expected	found		
ψ -b*	10252.5	10252.3	67.89	345
Aha- ψ -b*	10247.4	10247.3	65.87	333
Aha(GlcNAc)- ψ -b*	10506.7	10506.4	64.17	333
Aha(ChiAc ₂)- ψ -b*	10709.9	10709.4	63.86	333

[a] The T_m value is the midpoint of denaturation (melting temperature).
 [b] λ_{max} is the fluorescence emission maximum determined by protein sample excitation at 280 nm (slit 5.0 nm).

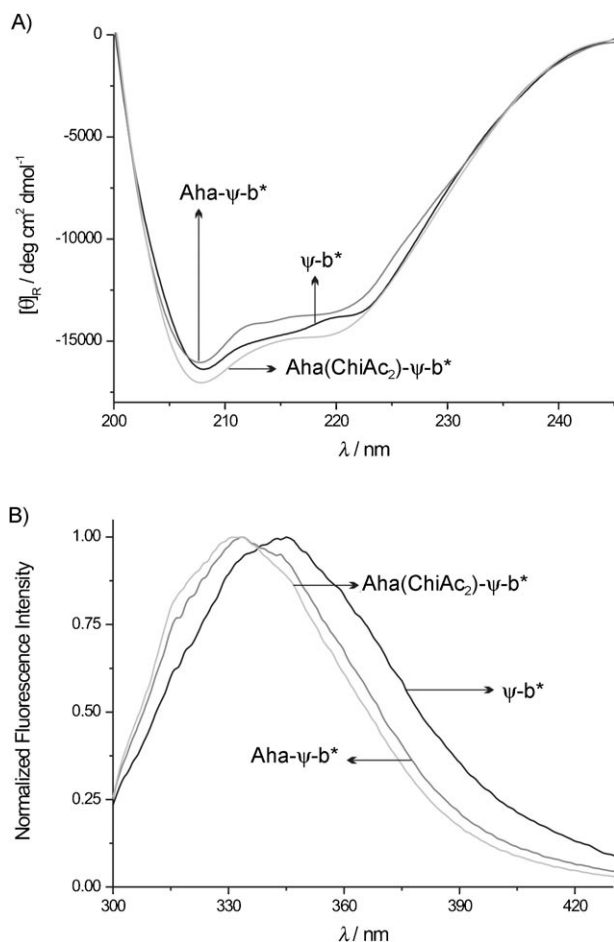


Figure 2. Effect of glycoconjugation on barstar secondary (A) and tertiary (B) structure upon introduction of Aha and subsequent conjugation with *N,N'*-diacetylchitobiose (ChiAc₂) through CCHC reactions. Changes in secondary structure were monitored by UV and CD spectroscopy in the 200–250 nm range. Fluorescence emission spectra were recorded in the 300–450 nm range with excitation at 280 nm (Table 1). Both the fluorescence and the CD spectra of Aha(GlcNAc)- ψ -b* are almost identical to those of Aha(ChiAc₂)- ψ -b* (Supporting Information).

N terminus generates an increase of about 20% in the curve steepness. Subsequent carbohydrate attachment through CCHC does not affect this value significantly. Evidently, both Aha incorporation and sugar conjugation are responsible for the higher cooperativity of the unfolding processes. We have previously already observed that isosteric replacement of Met

in proteins with related analogues is accompanied by rather large differences in van't Hoff enthalpies, despite the moderate shifts in the T_m values.^[33] The effects detected by CD spectroscopy, such as increased α -helical content and folding cooperativity of the protein, upon Aha incorporation and glycoconjugation are fully confirmed by fluorescence measurements. As shown in Figure 2B, the fluorescence emission maxima of the Aha- ψ -b* and Aha(ChiAc₂)- ψ -b* are blue-shifted by 12 nm (cf. Table 1), supporting a more stable tertiary structure. These blue spectral shifts are obviously associated with strong influences of the azido side chain or triazole ring of the attached sugars on the local environment of the barstar N terminus.

The carbohydrate-modified barstar variants inhibit the RNase activity of barnase in standard inhibition assays on RNA digestion,^[34] as does the parent variant. To demonstrate that WGA specifically binds to glycosylated barstar mutants, surface plasmon resonance (SPR) studies were performed. Aha- ψ -b*, Aha(GlcNAc)- ψ -b*, and Aha(ChiAc₂)- ψ -b* were immobilized on a CM5 sensor chip activated with carbodiimide/*N*-hydroxysuccinimide (NHS). The steady-state binding of WGA to the chip surface at different WGA concentrations is shown in Figure 3.

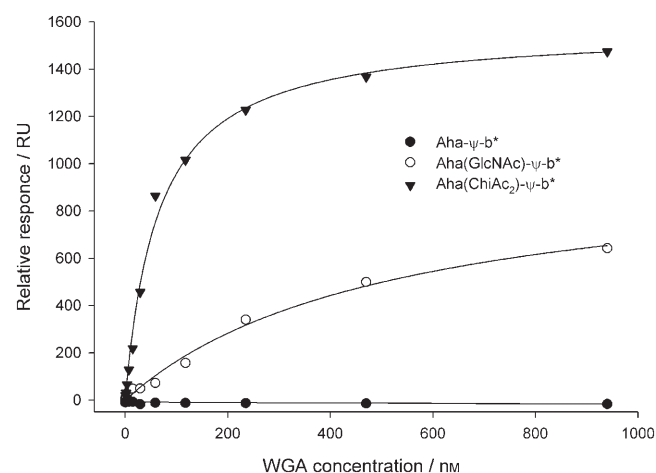


Figure 3. Equilibrium responses from SPR experiments with immobilized barstar variants as a function of WGA concentration. Experimental data were fitted by use of the steady-state binding model to obtain K_d values. While WGA shows no binding to Aha- ψ -b*, it binds to Aha(GlcNAc)- ψ -b* with $K_d = 517 \pm 78$ nM and to Aha(ChiAc₂)- ψ -b* with $K_d = 64 \pm 6$ nM.

While WGA does not bind to Aha- ψ -b*, it shows high affinity towards Aha(GlcNAc)- ψ -b* ($K_d = 517 \pm 78$ nM) and even higher affinity towards Aha(ChiAc₂)- ψ -b* ($K_d = 64 \pm 6$ nM). The K_d values are in good agreement with those determined for immobilized GlcNAc derivatives under comparable conditions.^[35] These results indicate that our approach allows engineering of new lectin affinities whereas the natural activity of barstar is conserved.

In summary, we have presented a strategy for N-terminal labeling of proteins making use of the N-end rules. Bulky amino acids such as Lys at positions two and three in barstar protect not only N-terminal Met but also the noncanonical amino acid

Aha from being excised. This offers the potential for N-terminal labeling of proteins by simple engineering of a bulky amino acid at position two even if the protein usually does not contain any Met. The method gives higher production yields than existing approaches for N-terminal labeling based on in vitro translation with pre-charged initiator tRNAs. Since the residue-specific incorporation method used here is statistical, all Met residues in a protein sequence are substituted by the analogue during the replacement. This problem, however, can be easily circumvented by use of site-directed mutagenesis, because Met is generally very rare (only about 1.5% in all known protein sequences).^[36] We were able to demonstrate that different propargyl glycosides can be ligated to the azide-containing barstar mutant Aha- ψ -b* through CCHC reactions in almost quantitative yields, giving access to stable and active homogeneous glycoforms. Carbohydrate epitopes incorporated by this strategy can serve as recognition motifs for lectins, as was shown by SPR experiments. Important applications of attached carbohydrate labels include protein purification by lectin affinity chromatography and lectin-directed cell-type-specific protein targeting.

Experimental Section

CCHC reactions: A typical reaction mixture consisted of Aha- ψ -b* (400 μ L, 0.24 mM [2.5 mg mL⁻¹] in 50 mM Tris-HCl pH 8, final concentration: 1 mg mL⁻¹), Tris-HCl pH 8 (100 μ L, 1 M), CuSO₄ (50 μ L, 50 mM in H₂O), L-ascorbic acid (50 μ L, 50 mM in H₂O), H₂O (300 μ L), and an aqueous solution of the propargyl glycoside (100 μ L, 40 mM). The reaction mixture was shaken for 24 h at 4 °C. As a negative control, Aha- ψ -b* was incubated under the same conditions but without CuSO₄/L-ascorbic acid. Afterwards all samples were dialyzed three times against PBS at 4 °C. The reaction yield was ~80% (1 mL of 0.8 mg mL⁻¹ of glycoconjugated protein).

SPR experiments: SPR measurements were carried out on a Biacore T100 instrument (Biacore, Uppsala, Sweden) with use of CM5 sensor chips and PBS as running buffer. Barstar samples were dissolved in acetate buffer pH 4.0 (10 mM, 50 μ g mL⁻¹) and immobilized by amide coupling to carboxyl groups in the chips' matrix, activated with *N*-ethyl-*N'*-dimethylaminopropylcarbodiimide hydrochloride/NHS by the standard procedure recommended by Biacore. The reference flow cell was not treated.

To determine WGA binding, solutions of WGA in running buffer of different concentrations (0.92–940 nM) were injected consecutively at 30 μ L min⁻¹ over the active surface. After each WGA injection, the surface was regenerated by two injections of GlcNAc (0.5 M, 15 s, 30 μ L min⁻¹). The signal of the reference flow cell was subtracted. Sensorgrams were analyzed by use of the steady-state affinity program of the Biacore T100 Evaluation Software.

Acknowledgements

Prof. Luis Moroder is gratefully acknowledged for critical reading of the manuscript. We are grateful to Waltraud Wenger and Elisabeth Weyher-Stingl for their excellent technical assistance. H.S.G.B. acknowledges a stipend according to the Landesgraduiertenförderungsgesetz. N.B. and L.M. are grateful for the support from the BioFuture Program of the Federal Ministry of

Education and Research of Germany and the Munich Center of Integrated Protein Sciences (CIPS-M).

Keywords: barstar • cycloadditions • glycoproteins • N-end rules • N-terminal methionine excision

- [1] A. Bachmair, D. Finley, A. Varshavsky, *Science* **1986**, 234, 179–186.
- [2] L. Merkel, Y. Cheburkin, B. Wiltshi, N. Budisa, *ChemBioChem* **2007**, 8, 2227–2232; A. Wang, N. W. Nairn, R. S. Johnson, D. A. Tirrell, K. Grabstein, *ChemBioChem* **2008**, 9, 324–330.
- [3] C. Giglione, A. Boularot, T. Meinel, *Cell. Mol. Life Sci.* **2004**, 61, 1455–1474.
- [4] C. T. Walsh, *Posttranslational Modifications of Proteins: Expanding Nature's Inventory*, Roberts, Englewood, **2006**.
- [5] A. Varshavsky, *Cell* **1992**, 69, 725–735.
- [6] A. Benbassat, K. Bauer, S. Y. Chang, K. Myambo, A. Boosman, S. Chang, *J. Bacteriol.* **1987**, 169, 751–757.
- [7] *Topics in Current Chemistry, Vol. 267: Glycopeptides and Glycoproteins: Synthesis, Structure, and Application* (Ed.: V. Wittmann), Springer, Heidelberg, **2007**.
- [8] B. G. Davis, *Chem. Rev.* **2002**, 102, 579–601.
- [9] G. de Santis, J. B. Jones in *Methods in Molecular Biology, Vol. 182: In Vitro Mutagenesis Protocols* (Ed.: J. Braman), Humana, Totowa, **2002**, pp. 55–65.
- [10] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem.* **2002**, 114, 2708–2711; *Angew. Chem. Int. Ed.* **2002**, 41, 2596–2599.
- [11] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, 67, 3057–3064.
- [12] R. Huisgen, *Angew. Chem.* **1963**, 75, 604–637; *Angew. Chem. Int. Ed. Engl.* **1963**, 2, 565–598.
- [13] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem.* **2001**, 113, 2056–2075; *Angew. Chem. Int. Ed.* **2001**, 40, 2004–2021.
- [14] H. N. Lin, C. T. Walsh, *J. Am. Chem. Soc.* **2004**, 126, 13998–14003.
- [15] H. T. Liu, L. Wang, A. Brock, C. H. Wong, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, 125, 1702–1703.
- [16] J. C. M. van Hest, K. L. Kiick, D. A. Tirrell, *J. Am. Chem. Soc.* **2000**, 122, 1282–1288.
- [17] K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi, *Proc. Natl. Acad. Sci. USA* **2002**, 99, 19–24.
- [18] Z. W. Zhang, J. Gildersleeve, Y. Y. Yang, R. Xu, J. A. Loo, S. Uryu, C. H. Wong, P. G. Schultz, *Science* **2004**, 303, 371–373.
- [19] C. H. Röhrig, O. A. Retz, L. Hareng, T. Hartung, R. R. Schmidt, *ChemBioChem* **2005**, 6, 1805–1816.
- [20] S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis, *Nature* **2007**, 446, 1105–1109.
- [21] M. Rubini, S. Lepthien, R. Golbik, N. Budisa, *Biochim. Biophys. Acta Proteins Proteomics* **2006**, 1764, 1147–1158.
- [22] D. P. Baker, E. Y. Lin, K. Lin, M. Pellegrini, R. C. Petter, L. L. Chen, R. M. Arduini, M. Brickelmaier, D. Wen, D. M. Hess, L. Chen, D. Grant, A. Whitty, A. Gill, D. J. Lindner, R. B. Pepinsky, *Bioconjugate Chem.* **2006**, 17, 179–188.
- [23] K. F. Geoghegan, J. G. Stroh, *Bioconjugate Chem.* **1992**, 3, 138–146.
- [24] J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, M. B. Francis, *Angew. Chem.* **2006**, 118, 5433–5437; *Angew. Chem. Int. Ed.* **2006**, 45, 5307–5311.
- [25] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, *Justus Liebig's Ann. Chem.* **1953**, 583, 129–149.
- [26] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, 266, 776–779.
- [27] T. W. Muir, *Annu. Rev. Biochem.* **2003**, 72, 249–289.
- [28] J. P. Tam, Q. Yu, Z. Miao, *Biopolymers* **1999**, 51, 311–332.
- [29] N. Muranaka, M. Miura, H. Taira, T. Hohsaka, *ChemBioChem* **2007**, 8, 1650–1653.
- [30] S. Mamaev, J. Olejnik, E. K. Olejnik, K. J. Rothschild, *Anal. Biochem.* **2004**, 326, 25–32.
- [31] R. Golbik, G. Fischer, A. R. Fersht, *Protein Sci.* **1999**, 8, 1505–1514.
- [32] N. E. Zachara, G. W. Hart, *Chem. Rev.* **2002**, 102, 431–438.

- [33] N. Budisa, R. Huber, R. Golbik, C. Minks, E. Weyher, L. Moroder, *Eur. J. Biochem.* **1998**, *253*, 1–9.
- [34] R. W. Hartley, *Methods Enzymol.* **2001**, *341*, 599–611.
- [35] Y. Shinohara, Y. Hasegawa, H. Kaku, N. Shibuya, *Glycobiology* **1997**, *7*, 1201–1208.

- [36] N. Budisa, *Angew. Chem.* **2004**, *116*, 6586–6624; *Angew. Chem. Int. Ed.* **2004**, *43*, 6426–6463.
-

Received: January 25, 2008
Published online on April 2, 2008
