ChemComm

This article is part of the

Glycochemistry & glycobiology web themed issue

Guest Editors are Professor Peter H. Seeberger and Dr Daniel B. Werz

Please visit the website to access the other articles in this issue:-
http://www.rsc.org/chemcomm/glycochemistry

ChemComm

Dynamic Article Links

Cite this: *Chem. Commun.*, 2012, **48**, 522–524

www.rsc.org/chemcomm

COMMUNICATION

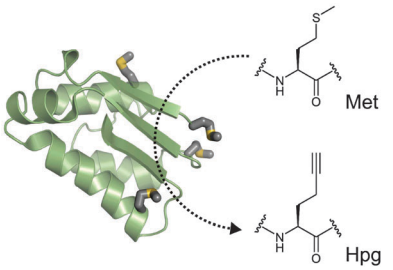
Site-selective modification of proteins for the synthesis of structurally defined multivalent scaffolds†‡

Lukas M. Artner,[a] Lars Merkel,[b] Nina Bohlke,[b] Figen Beceren-Braun,[c] Christoph Weise,[a] Jens Dernedde,[c] Nediljko Budisa*[b] and Christian P. R. Hackenberger*[a]

Received 28th September 2011, Accepted 17th October 2011
DOI: 10.1039/c1cc16039g

A combination of classical site-directed mutagenesis, genetic code engineering and bioorthogonal reactions delivered a chemically modified barstar protein with one or four carbohydrates installed at specific residues. These protein conjugates were employed in multivalent binding studies, which support the use of proteins as structurally defined scaffolds for the presentation of multivalent ligands.

Post-translational protein modifications play an important role in the regulation and organization of biological processes of living organisms and are therefore of common scientific interest.[1] In mammalians, one of the most frequent and complex modification processes is protein glycosylation, which results in the attachment of oligosaccharides as *N*- or as *O*-linked glycans.[2] Glycoproteins play a major role in recognition events, such as cell–cell interactions, and protein–antigen recognition.[3] Binding events between glycosylated cellular surfaces and carbohydrate-recognizing proteins[4]–namely lectins–often occur in a multivalent or cooperative fashion;[5] a mechanism that is also used by pathogens for infection.[6] Interestingly, analogous monovalent carbohydrates usually only bind to lectins in the low millimolar range.[7] Consequently, investigations focussing on the understanding of multivalent interactions as well as on the design of artificial multivalent binding systems have recently attracted considerable attention, in particular because carbohydrates and carbohydrate-recognizing proteins are considered as attractive medicinal targets in cancer research.[8] Over the last few years, numerous groups have explored the multivalency effect between carbohydrate-presenting scaffolds and their receptors. Commonly utilized scaffolds include polymers,[9] dextrins,[10] nanotubes[11] and nanoparticles.[12] Although the chemical access and functionalization of many polymeric systems is well developed, they often lack information about the exact number and the structural presentation of the multivalent ligands. For other scaffolds that have addressed these issues, including fullerenes[13] or viral capsids,[14] changing the rigidity and flexibility of the scaffold is often limited.



Scheme 1 The structure of ψ-b* from *Bacillus amyloliquefaciens* with mutations (K23M, E47M and K79M) indicated for ψ-b*4M. Methionine was subsequently globally replaced by Hpg during protein expression.

In this communication, we describe a systematic and modular way for the generation of multivalent binding systems by using proteins themselves as structured scaffolds, which present a defined number of carbohydrate ligands. This can be achieved by employing unnatural protein translation for the ribosomal incorporation of a specific number of unnatural functional groups into a protein, which can be chemoselectively conjugated after expression.[15,16] In particular, we used homopropargylglycine (Hpg, Scheme 1) as a non-canonical amino acid[16] in the supplementation incorporation method (SPI), which is based on *in vivo* sense codon reassignment. The use of an auxotrophic bacterial host along with a controlled protein expression allows the residue-specific replacement of a particular canonical amino acid by a non-canonical one.[2,15–17] It is important to note that this modular concept allows the positioning of multivalent ligands at preselected sites within various protein structures as well as a straightforward variation of the linker length between the scaffold and the ligands.[18]

In our model study, we attempted to engineer an artificially glycosylated protein for lectin binding studies. We chose the structurally well-defined cysteine-free "pseudo-wild-type barstar" ψ-b* from *Bacillus amyloliquefaciens* as the protein scaffold.[16,19] ψ-b* is a 10 kDa protein composed of 90 amino acids with only one methionine residue.[19] The 3D-structure of parent ψ-b*[20] revealed three solvent-exposed positions (K23, E47 and K79)

[a] *Freie Universität Berlin, Institut für Chemie und Biochemie, Takustr. 3, 14195 Berlin, Germany. E-mail: hackenbe@chemie.fu-berlin.de; Fax: +49 30 838 52551*
[b] *TU Berlin, Institut für Chemie, Franklinstraße 29, D-10587 Berlin, Germany. E-mail: budisa@biocat.tu-berlin.de; Fax: +49 30 314 28 279*
[c] *Charité—Universitätsmedizin Berlin, Institut für Laboratoriumsmedizin, Klinische Chemie und Pathobiochemie, CBF, Hindenburgdamm 30, 12200 Berlin*

† This article is part of the *ChemComm* 'Glycochemistry and glycobiology' web themed issue.

‡ Electronic supplementary information (ESI) available: Brief outline of the synthesis, MALDI-spectra. See DOI: 10.1039/c1cc16039g

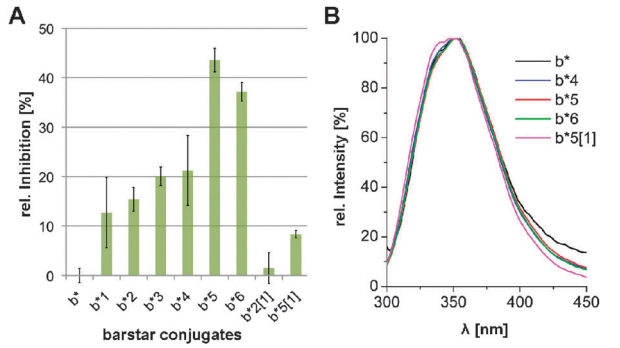
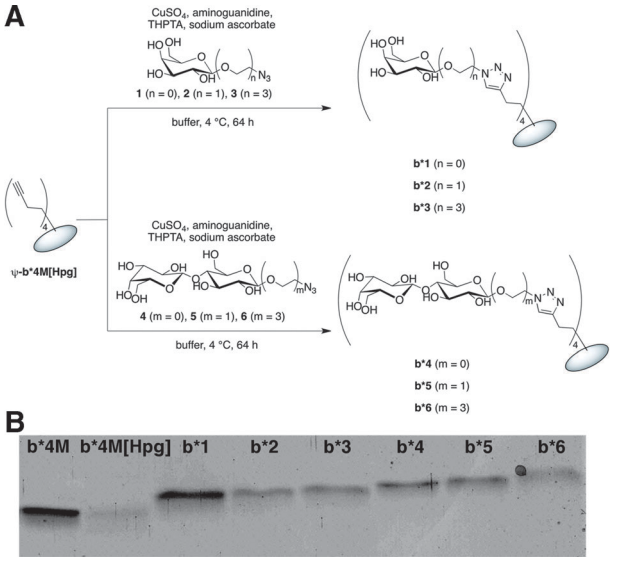


Fig. 1 (A) Relative inhibition of PNA binding measured by SPR in competition with ψ-b*4M[Hpg] (denoted as b*) and protein conjugates **b*1–b*6**, **b*2[1]** and **b*5[1]**. The SPR data for the conjugates were normalized against the unfunctionalized protein ψ-b*4M[Hpg]; (B) fluorescence emission spectra of ψ-b*4M[Hpg] (denoted as b*) and multivalent lactose glycoprotein conjugates **b*4–b*6** and **b*5[1]** excited at 280 nm. For experimental details see ESI.‡



Scheme 2 (A) Functionalization of ψ-b*4M[Hpg] with azido-sugars **1–6** by CuAAC. (B) Sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE; 12%) of parent ψ-b*4M[Met] (left) along with ψ-b*4M[Hpg] (second from left) and related protein conjugates **b*1–b*6**. Expectedly, migration times of glycoprotein conjugates are shifted when compared with non-conjugated species. For experimental details see ESI.‡

which were subsequently exchanged to methionine *via* site-specific mutagenesis, giving rise to ψ-b*4M (Scheme 1). In this way, four alkyne-containing Hpg amino acids were efficiently introduced to ψ-b*4M by SPI yielding the congener denoted by ψ-b*4M[Hpg], which can be used to conjugate carbohydrates to barstar to four unnatural residues.[16,21] In addition, ψ-b*1M[Hpg] was expressed containing a single Hpg at the N-terminus for comparative lectin binding studies with mono-glycosylated barstar proteins.

In our studies, we chose the well-established copper-catalyzed 1,3-dipolar azide–alkyne cycloaddition (CuAAC),[22] commonly termed as "click" reaction,[23] in particular because of the straightforward accessibility of different azide-functionalized carbohydrates[24] with three linker lengths as well as the high expression rates of proteins with alkyne-bearing amino acids.[15,25]

For lectin binding studies of the glycosylated protein scaffolds, peanut agglutinin (PNA) was chosen, for which weak inhibition with galactose and stronger inhibition with lactose are known. Consequently, six β-linked azido derivatives of galactose (**1–3**) and lactose (**4–6**) with different linker lengths were probed, which were synthesized from known protocols (see ESI‡).[9–12,24] The CuAAC-reaction of ψ-b*4M[Hpg] with azido-sugars **1–6** was performed under optimized conditions recently reported by Finn and coworkers, using tris-(hydroxypropyltriazolylmethyl)amine (THPTA) as a Cu(I) stabilizing ligand.[26] Non-commercially available THPTA was synthesized in 65% yield by using $Cu(MeCN)_4PF_6$ as catalyst. Strict implementation of the previously mentioned CuAAC protocol did not lead to fully functionalized protein, but with an extended reaction time and lower temperature we managed to get full functionalization of ψ-b*4M[Hpg] with all six azido-sugars to yield protein conjugates **b*1–b*6** (Scheme 2A), as verified by the corresponding gel shifts in gel electrophoresis and MALDI-ToF MS (see Scheme 2B and ESI‡). Additionally, the single alkyne-containing ψ-b*1M[Hpg] showed full conversion with galactose **2** and lactose **5** to mono-glycosylated **b*2[1]** and **b*5[1]** (see ESI‡).

After purification by dialysis, the inhibitory effect of the artificial glycoproteins was probed *via* a competitive surface plasmon resonance (SPR) binding assay, where the inhibition of sugar-functionalized barstar proteins on PNA binding to the immobilized Thomsen–Friedenreich (TF) antigen was analyzed. In brief, the TF antigen coupled multivalently to a polyacrylamide backbone was coated to a Biacore chip surface. Binding signals of the analytes passed over the chip were recorded as resonance units (RU). The respective RU for PNA binding alone was set to 100% and served as positive control. Potential inhibitory compounds preincubated with PNA should then demonstrate reduced binding (*X*% of control). Measurements were performed with unfunctionalized ψ-b*4M[Hpg] as control as well as quadruple functionalized barstar proteins **b*1–b*6** and the mono-functionalized proteins **b*2[1]** and **b*5[1]** (Fig. 1A). Unspecific inhibition of PNA by ψ-b*4M[Hpg] appeared to be very low (approximately 0.2%). In addition, it was found that the lactose conjugated proteins **b*4**, **b*5** and **b*6** inhibited PNA-binding between 21% and 44%, in which the ethyl-spaced lactose conjugate **b*5** showed the strongest inhibitory effect of all proteins (Fig. 1A). As expected, the inhibitory effect was significantly lower for galactose conjugates **b*1**, **b*2** and **b*3**, which can be rationalized by the previously mentioned lower carbohydrate specificity of PNA. Finally, both mono-glycosylated proteins **b*2[1]** and **b*5[1]** showed a significantly reduced inhibitory potency, thereby pointing towards increased binding due to the presentation of several lactose ligands. Free ligands in contrast showed no relevant inhibition at all (ESI‡).

In the next step, we focussed on the structural analysis of the lactose scaffolds, which showed the strongest inhibitory effect. Specifically, we measured the fluorescence of the parent protein and its conjugates as changes in the emission spectra and quantum yields represent a specific fingerprint of the protein tertiary structure.[27] Barstar contains two solvent-exposed Trp residues (positions 38 and 44) and buried Trp53.[19] Trp-residues not only dominate its absorbance and fluorescence profiles but also play a crucial role in ψ-b* structural integrity.[28] In ψ-b*

a network of cooperative interactions exists around Trp53, which is sandwiched between Phe56 and Phe74.[19] These residues play a central role in maintaining the stability of barstar and all those features are reflected in the fluorescence emission maxima.

As shown in Fig. 1B the fluorescence emission spectrum of ψ-b*4M[Hpg] is composed of a typical spectral shoulder between 330–340 nm (contribution of buried Trp53) and a peak maximum ~350 nm (from solvent-exposed residues Trp38/Trp44). The profiles of its glycoconjugates **b*4–b*6** and **b*5[1]** are essentially identical. Therefore, it can be reasonably assumed that the structural integrity is not significantly compromised by the conjugation reaction, as no hypsochromic or bathochromic shift of the fluorescence maximum could be observed although the shoulder in the b*5[1] spectrum is slightly more pronounced.

In summary, we have presented a strategy for engineering an artificial protein scaffold for multivalent binding studies by conjugation with carbohydrate moieties. We demonstrated this by introducing one or four galactose and lactose residues at preselected solvent-exposed sites in a barstar protein by a combination of unnatural protein translation and bioorthogonal functionalization. The SPR experiments show that this artificial protein scaffold acquired the highest capacity to inhibit PNA binding upon conjugation with four ethylene-glyco spaced lactose residues as opposed to the mono-glycosylated barstar protein without any detrimental effect on native tertiary structure. We anticipate that the high-resolution crystal and solution structures available for many proteins will offer almost ideal platforms for decorating those biopolymers with multivalent ligands, including carbohydrate or even peptide motifs. This will include proteins with different structural features and various levels of rigidity and will provide a very promising route for the design of other protein-based multivalent systems. We believe that the strategy presented here provides a solid basis for the further development and design of molecular systems for multivalent binding studies.

The authors acknowledge financial support from the German Science Foundation (Emmy-Noether program, HA 4468/2-1), the SFB 765, the Boehringer-Ingelheim Foundation and the Fonds der chemischen Industrie (FCI).

Notes and references

1 C. T. Walsh, S. Garneau-Tsodikova and G. J. Gatto, Jr, *Angew. Chem., Int. Ed.*, 2005, **44**, 7342.
2 D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131.
3 D. A. Tirrell, *Nature*, 2004, **430**, 837; H. Lis and N. Sharon, *Chem. Rev.*, 1998, **98**, 637; Y. C. Lee and R. T. Lee, *Acc. Chem. Res.*, 1995, **28**, 321.
4 N. Sharon, *J. Biol. Chem.*, 2007, **282**, 2753; C. R. Bertozzi, *Chem. Biol.*, 1995, **2**, 703.
5 L. L. Kiessling, J. E. Gestwicki and L. E. Strong, *Angew. Chem., Int. Ed.*, 2006, **45**, 2348; M. Mammen, S. K. Choi and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 1998, **37**, 2755.
6 C. R. Bertozzi and L. L. Kiessling, *Science*, 2001, **291**, 2357.
7 D. Deniaud, K. Julienne and S. G. Gouin, *Org. Biomol. Chem.*, 2011, **9**, 966; B. E. Collins and J. C. Paulson, *Curr. Opin. Chem. Biol.*, 2004, **8**, 1.
8 D. H. Dube and C. R. Bertozzi, *Nat. Rev. Drug Discovery*, 2005, **4**, 477.
9 I. Papp, J. Dernedde, S. Enders and R. Haag, *Chem. Commun.*, 2008, 5851; X.-L. Sun, K. M. Faucher, M. Houston, D. Grande and E. L. Chaikof, *J. Am. Chem. Soc.*, 2002, **124**, 7258.
10 R. V. Vico, J. Voskuhl and B. J. Ravoo, *Langmuir*, 2011, **27**, 1391; A. Mazzaglia, D. Forde, D. Garozzo, P. Malvagna, B. J. Ravoo and R. Darcy, *Org. Biomol. Chem.*, 2004, **2**, 957; A. García-Barrientos, J. J. García-López, J. Isac-García, F. Ortega-Caballero, C. Uriel, A. Vargas-Berenguel and F. Santoyo-González, *Synthesis*, 2001, **7**, 1057.
11 R. Su, L. Li, X. Chen, J. Han and S. Han, *Org. Biomol. Chem.*, 2009, **7**, 2040; L. Gu, P. G. Luo, H. Wang, M. J. Meziani, Y. Lin, L. M. Veca, L. Cao, F. Lu, X. Wang, R. A. Quinn, W. Wang, P. Zhang, S. Lacher and Y.-P. Sun, *Biomacromolecules*, 2008, **9**, 2408.
12 E. Mahon, T. Aastrup and M. Barboiu, *Chem. Commun.*, 2010, **46**, 5491; P. Wu, X. Chen, N. Hu, U. C. Tam, O. Blixt, A. Zettl and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2008, **47**, 5022.
13 S. Cecioni, V. Oerthel, J. Iehl, M. Holler, D. Goyard, J.-P. Praly, A. Imberty, J.-F. Nierengarten and S. Vidal, *Chem.–Eur. J.*, 2011, **17**, 3252.
14 A. Kussrow, E. Kaltgrad, M. L. Wolfenden, M. J. Cloninger, M. G. Finn and D. J. Bornhop, *Anal. Chem.*, 2009, **81**, 4889; D. Banerjee, A. P. Liu, N. R. Voss, S. L. Schmid and M. G. Finn, *ChemBioChem*, 2010, **11**, 1273; Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless and M. G. Finn, *J. Am. Chem. Soc.*, 2003, **125**, 3192. For a recent highlight, see: A. K. Udit, C. P. R. Hackenberger and M. K. O'Reilly, *ChemBioChem*, 2010, **11**, 481.
15 N. Budisa, *Angew. Chem., Int. Ed.*, 2004, **43**, 6426; L. Wang and P. G. Schultz, *Angew. Chem., Int. Ed.*, 2005, **44**, 34.
16 L. Merkel, H. S. G. Beckmann, V. Wittmann and N. Budisa, *ChemBioChem*, 2008, **9**, 1220.
17 M. G. Hoesl and N. Budisa, *ChemBioChem*, 2011, **12**, 552.
18 For a related approach, in which Lys-residues of β-lactoglobulin were conjugated to carbohydrates, see: R. S. Loka, C. M. Sadek, N. A. Romaniuk and C. W. Cairo, *Bioconjugate Chem.*, 2010, **21**, 1842.
19 S. Lepthien, L. Merkel and N. Budisa, *Angew. Chem., Int. Ed.*, 2010, **49**, 5446.
20 M. Rubini, S. Lepthien, R. Golbik and N. Budisa, *Biochim. Biophys. Acta, Proteins Proteomics*, 2006, **1764**, 1147.
21 For reviews see: E. M. Sletten and C. R. Bertozzi, *Angew. Chem., Int. Ed.*, 2009, **48**, 6974; C. P. R. Hackenberger and D. Schwarzer, *Angew. Chem., Int. Ed.*, 2008, **47**, 10030. For recent applications of CuAAC, see: S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony and B. G. Davis, *Nature*, 2007, **446**, 1105; E. Kaya, K. Gutsmiedl, M. Vrabel, M. Müller, P. Thumbs and T. Carell, *ChemBioChem*, 2009, **10**, 2858; H. Lin and C. T. Walsh, *J. Am. Chem. Soc.*, 2004, **126**, 13998.
22 V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2596; C. W. Tornoe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057. For seminal contributions, see: R. Huisgen, G. Szeimies and L. Moebius, *Chem. Ber.*, 1967, **100**, 2494.
23 S. K. Mamidyala and M. G. Finn, *Chem. Soc. Rev.*, 2010, **39**, 1252; H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2001, **40**, 2004.
24 R. Kumar, P. Tiwari, P. R. Maulik and A. K. Misra, *Eur. J. Org. Chem.*, 2006, **1**, 74; R. Kumar, P. R. Maulik and A. K. Misra, *Glycoconjugate J.*, 2008, **25**, 595; R. R. Schmidt, *Angew. Chem.*, 1986, **98**, 213.
25 S. Dong, L. Merkel, L. Moroder and N. Budisa, *J. Pept. Sci.*, 2008, **14**, 1148.
26 D. Cantillo, M. Ávalos, R. Babiano, P. Cintas, J. L. Jiménez and J. C. Palacios, *Org. Biomol. Chem.*, 2011, **9**, 2952; V. Hong, A. K. Udit, R. A. Evans and M. G. Finn, *ChemBioChem*, 2008, **9**, 1481; V. Hong, S. I. Presolski, C. Ma and M. G. Finn, *Angew. Chem., Int. Ed.*, 2009, **48**, 9879; S. I. Presolski, V. Hong, S.-H. Cho and M. G. Finn, *J. Am. Chem. Soc.*, 2010, **132**, 14570.
27 J. R. Lakowitz, in *Protein fluorescence*, Kluwer Academic/Plenum Publisher, 2nd edn, 1999.
28 N. Budisa, S. Alefelder, J. H. Bae, R. Golbik, C. Minks, R. Huber and L. Moroder, *Protein Sci.*, 2001, **10**, 1281.

