Analysis of the dispersity in carbohydrate loading of synthetic glycoproteins using MALDI-TOF mass spectrometry†

Mitul K. Patel, Balakumar Vijayakrishnan, Julia R. Koeppe, Justin M. Chalker, Katie J. Doores and Benjamin G. Davis*

Received 23rd August 2010, Accepted 29th September 2010
DOI: 10.1039/c0cc03420g

Statistical correlation of mass spectrum peak broadening with product dispersity in protein conjugation reactions allows more detailed characterization of putative therapeutic conjugates.

The presentation of carbohydrates on proteins and cell surfaces has far reaching biological significance.[1] Carbohydrates act as markers for critical signalling events and interact with sugar-binding proteins using a high surface density of carbohydrates to improve affinity.[2] Synthetic glycoproteins that mimic this surface coverage can interact with these proteins and are useful for clinical applications[3] such as vaccines,[4] enzyme replacement therapy,[5] targeted drug delivery[6] and gene therapy.[7]

Currently, most glycoprotein therapeutics are heterogeneous, formed from the indiscriminate conjugation of glycans, for instance at primary amines on native lysine residues and the *N*-terminus.[8,9] Since lysine has a high natural abundance, multiple modifications occur. A striking drawback is the formation of a statistical mixture of products and the challenge of measuring and controlling the product dispersity.

Conjugates can be characterized with matrix-assisted laser desorption/ionization (MALDI) mass spectrometry (MS), but for large molecular weights, the resolving power is insufficient to distinguish each product.[10] Hence, a broad peak is observed, where the centre is the mean copy number of ligands per protein. Information on the dispersity of the sample is typically neglected; this is surprising since such proteins may be used clinically. Fuller characterization would allow better batch reproducibility[11] and understanding of reactivity and therapeutic structure–activity relationships.

At present, there are no simple methods to determine sample dispersity. Top–down proteomics may reveal possible modification sites, but product distribution is not readily apparent.[12] Fourier transform-ion cyclotron resonance MS offers high resolution and may be capable of directly measuring distributions,[13] but requires extensive computation[14] and expensive instrumentation. Broadened peaks of MALDI-MS can, in principle, reveal information on the underlying distribution. The parent peak of the unmodified protein can be approximated as a Gaussian peak which can be mathematically expressed from its full width at half maximum (FWHM). The peak shape is given by (1), where x is the measured mass and δ^2 is the variance, obtained from the FWHM.‡ The broadened peak after modification is also approximately Gaussian and can be represented as (2), where λ^2 is the variance of this peak.‡

$$\frac{1}{\sqrt{2\pi\delta^2}}e^{\frac{-x^2}{2\delta^2}} \quad (1)$$

$$f(x) = \frac{1}{\sqrt{2\pi\lambda^2}}e^{\frac{-(x-\mu)^2}{2\lambda^2}} \quad (2)$$

Indiscriminate modification produces a mixture of protein products, each of which generates its own Gaussian "peaklet". Inadequate resolution leads to peaklet coalescence and a widened peak (Fig. 1). Therefore, the spectrum of the modified protein can be represented as a sum of peaklets, weighted by a function representing the product distribution. The theoretical product distribution can be calculated by modelling the reaction as a series of consecutive competitive second order events.[15] This distribution and its analysis are sensitive to the reaction conditions. For instance, at very low reagent concentration, a Poisson distribution is likely most appropriate, whereas for a reaction purely defined by statistics (where only the number of reactive groups is considered to affect the rate constants), a binomial distribution offers an exact analytical solution. However, these probability functions are only valid over a narrow range of reaction criteria. A Gaussian distribution is more mathematically flexible and valid over a broader range of conditions.† Assuming the variance of each peaklet stays constant as δ^2 and ionization efficiency is unaltered by successive modifications, we can write (with a continuity correction):

$$f(x) = \sum_{0}^{n}\left[\frac{1}{\sqrt{2\pi\delta^2}}e^{\frac{-(x-mn)^2}{2\delta^2}}\int_{n-0.5}^{n+0.5}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx\right] \quad (3)$$

where m is mass change per modification, n is the total number of reactive sites per protein, μ is the relative mean mass of conjugates and σ^2 is the variance of distribution.

Eqn (2) and (3) represent different analyses of the same function: that of the broadened peak of the synthetic conjugate mixture. By equating the two formulae, σ can be

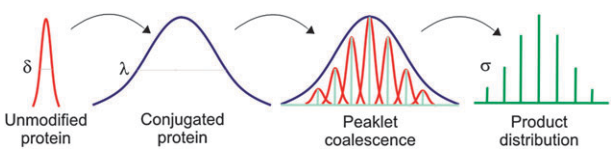

Fig. 1 Peaklet coalescence leads to peak broadening. By analysis of peak widths, information on the product distribution can be obtained.

Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Mansfield Road, Oxford, OX1 3TA, UK. E-mail: Ben.Davis@chem.ox.ac.uk; Fax: +44 (0)1865 285002; Tel: +44 (0)1865 285001

† Electronic supplementary information (ESI) available: Extended discussion on the Gaussian approximation and peak tailing, experimental details and MATLAB M-file scripts. See DOI: 10.1039/c0cc03420g

determined. Thus, by simply measuring FWHM of modified and unmodified proteins, the product distribution variance can be calculated. Furthermore, since the area between $\pm 2\sigma$ equates to 95% of the total, $\mu \pm 2\sigma$ represents the range within which 95% of adducts exist thus giving a direct, physically intuitive measure of dispersity.

To demonstrate the versatility of the approach, we synthesized various glycoconjugates and determined their dispersity using MALDI with time of flight (TOF) detection.§ Mannose-bearing synthetic glycoproteins mimic the glycosylation on envelope protein gp120 and are potential HIV vaccines.[3,4] As a model system, we linked mannose residues to immunogenic carrier proteins using thiophosgene conjugation chemistry, a strategy previously explored for synthetic vaccine constructs.[8,16] Such glycoconjugates are representative of a type of vaccine design, which unlike many existing vaccines, contain pure glycan components. Initially, we conjugated monosaccharide **1** with bovine serum albumin (BSA, 59 lysines + *N*-terminus), a common model protein for vaccines (Fig. 2). MALDI peaks were used to determine the extent of modification and sample dispersity for increasing ratios of **1** : lysine. For instance, BSA modified with a mean of 15 copies of **1** has a dispersity (2σ) of 8 (95% of proteins incorporate 7–23 copies of **1**)—a more informative description of the actual product composition than simply the mean.

We then investigated coupling reactions of BSA with a mannose disaccharide (Fig. 3). The greater steric demand of **2** required a larger excess of reagent to achieve similar levels of modification as **1**. Interestingly, glycoconjugates of **2** showed more dispersity than the corresponding conjugates of **1** (Fig. 4). The dispersity is determined by the ratios of the consecutive rate constants (the distribution constants).[15] The average rate constant of each successive modification is expected to decrease due to the reducing number of available lysines and increasing steric crowding from the protein structure. Thus, the rate of this decrease determines the dispersity of the sample, with a sharper fall resulting in a less dispersed

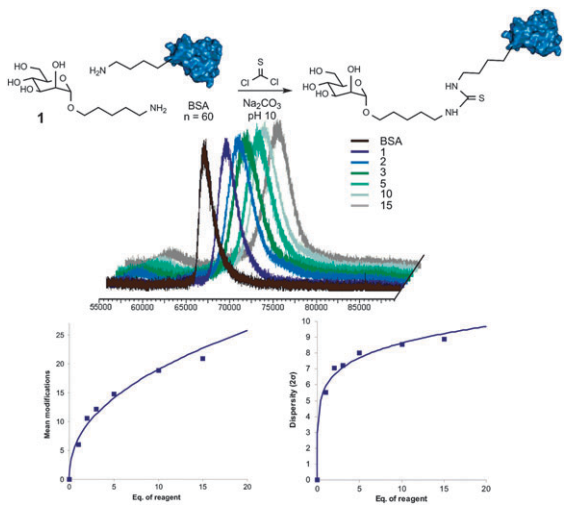

Fig. 2 Top: reaction scheme for conjugation to BSA; middle: mass spectrum of glycoconjugates with increasing equivalents of **1** per lysine; bottom: mean extent of modification and dispersity *versus* reagent excess.

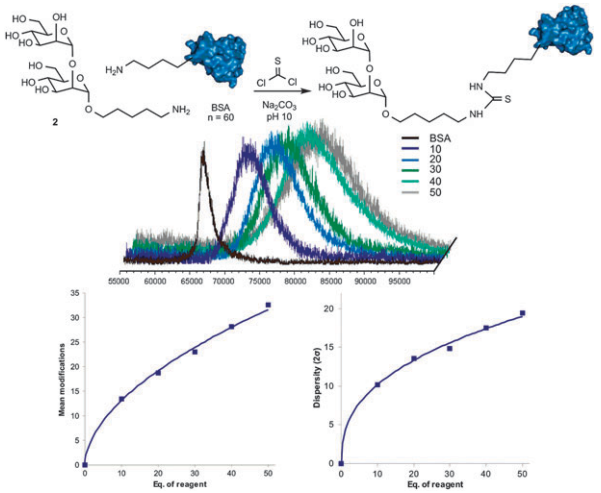

Fig. 3 Top: reaction scheme for conjugation to BSA; middle: mass spectrum of glycoconjugates with increasing equivalents of **2** per lysine; bottom: mean extent of modification and dispersity *versus* reagent excess.

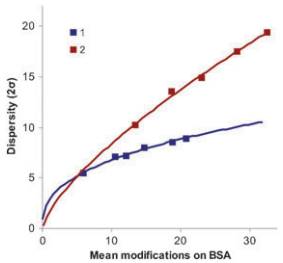

Fig. 4 Dispersity of glycoconjugates of BSA with **1** or **2**.

product distribution. Since **2** is bulkier than **1**, a slower reaction rate might have been expected. However, consecutive rate constants are the critical determinant and may not be affected to the same extent. If earlier rates in the sequence are retarded to a greater extent than later ones, the progressive decrease in reaction rates is much slower leading to a more dispersed product distribution. The overall outcome revealed that use of reagent **1** as compared to **2** has a greater effect on earlier lysine couplings than with later, less solvent-exposed residues which are already less reactive.

Applicability to other proteins was demonstrated by conjugating **2** with the cross reacting material fragment of diphtheria toxin (CRM-197), another common protein for vaccine constructs (Fig. 5).[17] Compared to the reaction of **2** with BSA, the extent of modification with CRM was reduced due to the lower number of reactive sites (only 40 amines). The dispersity of CRM conjugates is also lower compared with BSA, since the statistical fall in sequential rate constants is exaggerated when there are fewer reactive groups.

Having established the method on model systems, we applied the analysis to a prospective HIV vaccine where a mannose tetrasaccharide (D1 arm) was conjugated to a carrier protein to mimic the glycan display on gp120. We have recently evaluated such Qβ glycoproteins for immunogenic activity.[18] Since the Qβ monomer is a low molecular weight protein (14 kDa, 8 amines), the individual peaklets are well

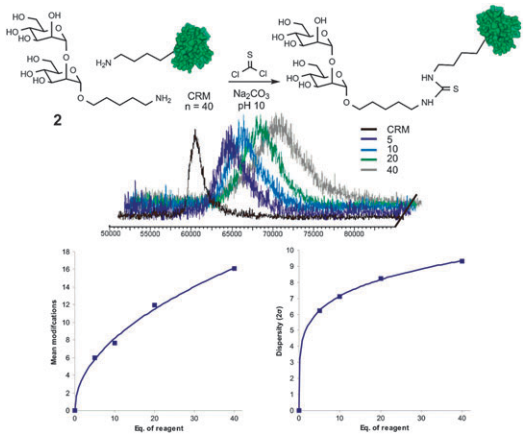

Fig. 5 Top: reaction scheme for conjugation to CRM; middle: mass spectrum of glycoconjugates with increasing equivalents of **2** per lysine; bottom: mean extent of modification and dispersity *versus* reagent excess.

resolved by MALDI (Fig. 6) and provided direct validation of the assumptions underpinning our dispersity analysis. Individual peaklets have identical FWHM, and distribution profiles match theoretical expectations indicating similar ionisation of the different protein products.[19] The product distribution is directly evident, showing an average of 1.5 modifications and a dispersity of 1.5. This allowed consistent levels of modification between two constructs carrying different glycans, which was vital for comparing biological efficacy. Most common carrier proteins are larger than Qβ and peaklet coalescence occurs, as for the conjugation of the same D1 arm with BSA. Loading-dispersity analysis reveals 10.8 ± 8.8 modifications; similar BSA conjugate with **1** has 10.6 ± 7.2 modifications, consistent with the trend of larger sugars producing higher dispersity. Without this dispersity parameter, the two glycoconjugates would be considered similarly modified. A dispersity analysis is therefore vital to understand the true nature of the products.

The simple mathematical method we describe allows dispersity in copy-number to be determined and expressed intuitively, even for non-ideal cases exhibiting high-mass peak tailing.[20]† Comparisons can be made between reagents: bulkier sugars can increase dispersity, while proteins with fewer reactive groups cause a decrease. Glycoconjugates can be better characterized without additional experiments than is routine,¶ allowing control of loading and distribution. Given that very few synthetic vaccines in mainstream use are well defined,[4] our method enables better reproducibility and more logical comparisons. The dispersity measure proposed here may find use akin to the polydispersity index commonly used in polymer science. Other applications of therapeutic relevance can be envisaged. Indiscriminate *in vivo* protein glycation of lysines is implicated in several diseases and measuring the dispersity of glycated proteins would assist in disease diagnostics.[21]

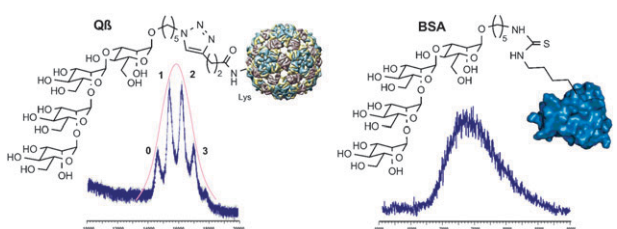

Fig. 6 Glycoconjugates of the D1 arm with Qβ (left) and BSA (right) and MALDI mass spectra of the corresponding glycoproteins.

We thank the International AIDS Vaccine Initiative (MKP, BV, JRK, KJD), the Rhodes Trust and NSF (JMC) for financial support. BGD is a Royal Society-Wolfson Research Merit Award recipient and supported by an EPSRC LSI Platform Grant. We also thank V. Hong and Prof. M. G. Finn for their assistance in the preparation of the Qβ conjugates.

Notes and references

‡ FWHM = $\sqrt{\text{variance}} \times 2\sqrt{2 \ln 2}$

§ MALDI-TOF detector settings, laser irradiance and matrix preparation were identical for each set of comparisons to ensure consistency.

¶ Other possible methods beyond MS that might aid analysis include those based on differential electrophoresis such as isoelectric focussing.

1 R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683.
2 N. Jayaraman, *Chem. Soc. Rev.*, 2009, **38**, 3463; J. J. Lundquist and E. J. Toone, *Chem. Rev.*, 2002, **102**, 555.
3 K. J. Doores, D. P. Gamblin and B. G. Davis, *Chem.–Eur. J.*, 2006, **12**, 656.
4 R. Roy, *Carbohydrate Based Vaccines*, ACS Symposium Series, 2008, vol. 989; M. Hecht, P. Stallforth, D. V. Silva, A. Adibekian and P. H. Seeberger, *Curr. Opin. Chem. Biol.*, 2009, **13**, 354; R. Roy, *Drug Discovery Today: Technol.*, 2004, **1**, 327.
5 Y. Zhu, X. Li, A. McVie-Wylie, C. Jiang, B. L. Thurberg, N. Raben, R. J. Mattaliano and S. H. Cheng, *Biochem. J.*, 2005, **389**, 619.
6 M. A. Robinson, S. T. Charlton, P. Garnier, X.-T. Wang, S. S. Davis, A. C. Perkins, M. Frier, R. Duncan, T. J. Savage, D. A. Wyatt, S. A. Watson and B. G. Davis, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14527; B. G. Davis and M. A. Robinson, *Curr. Opin. Drug Discovery Dev.*, 2002, **5**, 279.
7 O. M. T. Pearce, K. D. Fisher, J. Humphries, L. W. Seymour, A. Smith and B. G. Davis, *Angew. Chem., Int. Ed.*, 2005, **44**, 1057.
8 G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, 2nd edn, 1996; M. Izumi, S. Okumura, H. Yuasa and H. Hashimoto, *J. Carbohydr. Chem.*, 2003, **22**, 317.
9 For well defined glycoproteins see: B. G. Davis, *Chem. Rev.*, 2002, **102**, 579; D. P. Gamblin, E. M. Scanlan and B. G. Davis, *Chem. Rev.*, 2009, **109**, 131; B. G. Davis, *Pure Appl. Chem.*, 2009, **81**, 285.
10 D. J. Harvey, *Int. J. Mass Spectrom.*, 2003, **226**, 1.
11 C. Jones, *Carbohydrate Based Vaccines*, ACS Symposium Series, 2008, vol. 989, ch. 2, pp. 21–35.
12 B. A. Garcia, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 193.
13 A. G. Marshall, C. L. Hendrickson and G. S. Jackson, *Mass Spectrom. Rev.*, 1998, **17**, 1; N. L. Kelleher, M. W. Senko, M. M. Siegel and F. W. McLafferty, *J. Am. Soc. Mass Spectrom.*, 1997, **8**, 380; K. Nagy, K. Vékey, T. Imre, K. Lundányi, M. P. Barrow and P. J. Derrick, *Anal. Chem.*, 2004, **76**, 4998.
14 D. M. Horn, R. A. Zubarev and F. W. McLafferty, *J. Am. Soc. Mass Spectrom.*, 2000, **11**, 320.
15 H. J. R. Maget, *J. Polym. Sci., Part A: Polym. Chem.*, 1964, **2**, 1281.
16 R. D. Astronomo, H. Lee, C. N. Scanlan, R. Pantophlet, C. Huang, I. A. Wilson, O. Blixt, R. A. Dwek, C. Wong and D. R. Burton, *J. Virol.*, 2008, **82**, 6359; D. H. Buss and I. J. Goldstein, *J. Chem. Soc. C*, 1968, 1457.
17 G. Giannini, R. Rappuoli and G. Ratti, *Nucleic Acids Res.*, 1984, **12**, 4063.
18 K. J. Doores, Z. Fulton, V. Hong, M. K. Patel, C. N. Scanlan, M. R. Wormald, M. G. Finn, D. R. Burton, I. A. Wilson and B. G. Davis, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17107.
19 F. K. Yeboah and V. A. Yaylayan, *Nahrung*, 2001, **45**, 164.
20 E. Sachon, G. Clodic, T. Blasco and G. Bolbach, *J. Am. Soc. Mass Spectrom.*, 2007, **18**, 1880.
21 T. Niwa, *Mass Spectrom. Rev.*, 2006, **25**, 713.