CORRESPONDENCE

Building meaningful models of glycoproteins

To the Editor:
The building of crystallographic models of proteins is guided by understanding of the primary structure of proteins and by well-established and rigorously applied stereochemical principles. However, a cursory survey of Protein Data Bank entries containing oligosaccharides suggests that of the order of one-third of entries contain significant errors in carbohydrate stereochemistry, nomenclature or even consistency with the electron density maps. Many of the stereochemical errors can be detected by reference to conformational studies of glycans[1,2] and to publicly available resources (http://www.glycosciences.de/tools/). However, these errors also indicate that there is a wide discrepancy in the sophistication of building and validation tools available for protein and carbohydrate models.

An example of the difficulties that can be encountered when building crystallographic models of glycoproteins is the recent model proposed by Szakonyi *et al.*[3] of the Epstein-Barr virus major envelope glycoprotein, EBV gp350. EBV gp350 was expressed in *Spodoptera frugiperda* Sf9 cells, and Szakonyi *et al.*[3] report that they observed electron density corresponding to the oligosaccharide chains of fourteen N-linked glycosylation sites. The crystallization of such a heavily glycosylated glycoprotein is a notable achievement. However, the proposed model contains not only systematic errors in carbohydrate stereochemistry, presumably resulting from inadequate parameter files, but also hitherto unreported motifs in the primary structures of the glycans.

The previously undescribed glycosidic linkages and motifs that Szakonyi *et al.*[3] propose include Man-(1→3)-GlcNAc and GlcNAc-(1→3)-GlcNAc linkages (of indeterminate anomericity) within the trimannosyl core, hybrid-type glycans containing a terminal Man-(1→3)-GlcNAc linkage on the 3-antennae, and β-galactosyl motifs capping oligomannose-type glycans. We suggest that, in the absence of supporting evidence, electron density at 3.5-Å resolution should not be used to support linkages incompatible

with the known biosynthetic routes of N-glycan processing. One of the advantages of the expression of EBV gp350 in Sf9 cells is the restricted range of carbohydrate structures that are likely to be observed. Such information can guide model building. Sf9 cell glycosylation is dominated by oligomannose-type ($Man_{5–9}GlcNAc_2$) and paucimannose-type glycans ($Man_{2–3}GlcNAc_2$ with or without core fucose)[4]. The highly unusual $Gal_2Man_7GlcNAc_2$ glycans proposed by Szakonyi *et al.*[3] could conceivably correspond to the underprocessed α-glucosylated oligomannose glycans derived from the highly conserved starting structure of glycosylation, $Glc_3Man_9GlcNAc_2$ (ref. 2), but these are rarely observed on mature secreted glycoproteins[5]. Finally, the building of such large glycans seems to be inadequately supported by the deposited structure factor data.

In the wake of advances in the production and crystallization of glycoproteins, we believe there should now be a wider discussion about the application and control of refinement standards applied to glycosylation. This requires more robust validation procedures for newly reported structures as well as a systematic evaluation of those previously described (scrutiny from which the authors of this letter are not exempt). We suggest that carbohydrate-specific building and validation tools, capable of guiding the construction of biologically relevant and stereochemically accurate models, should be integrated into popular crystallographic software. Rigorous treatment of the structural biology of glycosylation can only enhance the structural analysis of glycoproteins and our understanding of their functions.

Max Crispin, David I. Stuart & E. Yvonne Jones
Division of Structural Biology, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, UK.
e-mail: max@strubi.ox.ac.uk

COMPETING INTERESTS STATEMENT
The authors declare no competing financial interests.

1. Petrescu, A.J., Petrescu, S.M., Dwek, R.A. & Wormald, M.R. *Glycobiology* **9**, 343–352 (1999).
2. Wormald, M.R. *et al. Chem. Rev.* **102**, 371–86 (2002).
3. Szakonyi, G. *et al. Nat. Struct. Mol. Biol.* **13**, 996–1001 (2006).
4. Altmann, F., Staudacher, E., Wilson, I.B. & Marz, L. *Glycoconj. J.* **16**, 109–123. (1999).
5. Parodi, A. *J. Annu. Rev. Biochem.* **69**, 69–93 (2000).

Berman, Henrick, Nakamura & Markley respond:
The Worldwide Protein Data Bank (wwPDB, http://www.wwpdb.org) maintains the archive of macromolecular structural data determined experimentally by crystallography, NMR and cryo-EM.

The wwPDB welcomes the above comments on the quality of the models of oligosaccharides contained in the PDB. We are well aware of the problems encountered, such as the common error in N-linked N-acetyl-D-glucosamine attached to asparagine; there are approximately two hundred cases for which we have had to assign, by stereochemistry matching, the incorrect 2-(acetylamino)-2-deoxy-α-D-glucopyranose (NDG) rather than the correct 2-(acetylamino)-2-deoxy-β-D-glucopyranose (NAG).

The authors state that about one-third of the entries containing oligosaccharides have significant errors in carbohydrate stereochemistry. Lüttke *et al.*[1] carried out a detailed analysis of the carbohydrate structures in the PDB and reported that about 30% of all PDB entries containing carbohydrates have one or several errors in glycan description that are mainly due to wrong assignment of saccharide units. Lüttke and von der Lieth have made available software called pdb-care (http://www.glycosciences.de/tools/pdb-care/) that can identify and assign carbohydrate structures using only atom types and their three-dimensional atom coordinates. This software provides a valuable check on carbohydrates built into electron density; systematic names and the respective PDB residues are compared and inconsistencies reported. Additionally, the reliability of

reported and calculated connectivities for molecules listed in HETATOM records is checked and unusual values are reported.

Some of the problems with, for example, O-linked carbohydrates in the PDB arise when electron density is modeled without enough regard to known chemistry. As reviewed in an article on bacterial glycoproteomics[2], the bacterial pilin proteins of the pathogenic bacteria *Neisseria meningitidis*, *Neisseria gonorrhoeae* and *Pseudomonas aeruginosa* are all O-glycosylated[3,4]. The pilin glycans from *Neisseria* species share a common structure, in particular with respect to the unusual O-linked sugar residue 2,4-diacetamido-2,4,6-trideoxyhexose (DATDH)[4]. Like that of the neisseriae, the pseudomonas pilin *O*-glycan is a short-chain oligosaccharide, although the sugar components differ. However, in the PDB, the pilin structure from *Neisseria gonorrhoeae* has been reported with galactose-α-1,3-*N*-acetylglucosamine-serine in entries 1AY2 (ref. 5) and 2PIL (ref. 6). These two PDB entries were published before the work of Hegge *et al.*[4]. In later PDB entries (2HI2 and 2HIL, ref. 7), the correct sugar, 2,4-bis(acetylamino)-1,5-anhydro-2,4-dideoxy-*d*-glucitol, is reported O-linked to serine.

The building of glycosidic linkages and motifs into electron density depends on the availability of known structures. The databases and tools being worked on under several glycobiology initiatives will go a long way toward bringing the primary structure of glycans to the attention of nonspecialists and enabling researchers to build three-dimensional models of complex carbohydrates for fitting into electron density. These include EuroCarbDB (http://www.eurocarbdb.org), the Consortium for Functional Glycomics (http://www.functionalglycomics.org/fg) in the US, and the Japanese Consortium for Glycobiology and Glycotechnology (http://www.jcgg.jp/E/). Through the efforts of these consortia, the wwPDB is seeking advice and tools for validation of carbohydrate stereochemistry and primary structure.

We agree with the comments of Crispin *et al.* that, in the absence of supporting evidence, incorrect glycans may be built into low-resolution maps; however, with knowledge of the correct glycan structure, it should be possible to fit these structures into even low-resolution maps. This does not solve the problem for PDB entries such as 2H6O (ref. 8), in which electron density was found for the N-linked glycans but the nature of the linkages and motifs was unknown.

With the evolution of experimental methods, functional knowledge of proteins, and methods used to process PDB data, inconsistencies have been introduced into the archive. The members of the wwPDB have worked together to remediate these data to ensure the uniformity of archived entries: see http://remediation.wwpdb.org/. As part of the ligand remediation project, we have identified structures that most probably contain incorrect Asn-NDG linkages. In the future, we will include more validation procedures designed to detect likely errors in carbohydrate models.

Helen M. Berman
RCSB Protein Data Bank, Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, USA.
e-mail: berman@rcsb.rutgers.edu

Kim Henrick
Macrosmolecular Structural Database, European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK.

Haruki Nakamura
Protein Data Bank Japan, Research Center for Structural Biology, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Osaka 565-0871, Japan

John Markley
BioMagResBank, Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706-1544, USA

1. Lütteke, T., Frank, M. & von der Lieth, C.-W. *Carbohydr. Res.* **339**, 1010–1020 (2004).
2. Hitchen, P.G. & Dell, A. *Microbiology* **152**, 1575–1580 (2006).
3. Castric, P., Cassels, F.J. & Carlson, R.W. *J. Biol. Chem.* **276**, 26479–26485 (2001).
4. Hegge, F.T. *et al. Proc. Natl. Acad. Sci. USA* **101**, 10798–10803 (2004).
5. Parge, H.E. *et al. Nature* **378**, 32 (1995).
6. Forest, K.T., Dunham, S.A., Koomey, M. & Tainer, J.A. *Mol. Microbiol.* **31**, 743–752 (1999).
7. Craig, L. *et al. Mol. Cell* **23**, 651–662 (2006).
8. Szakonyi, G. *et al. Nat. Struct. Mol. Biol.* **13**, 996–1001 (2006).