Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Solving the phase problem for carbohydrate-binding proteins using selenium derivatives of their ligands: a case study involving the bacterial F17-G adhesin

Lieven Buts,[a] Remy Loris,[a] Erwin De Genst,[a] Stefan Oscarson,[b] Martina Lahmann,[b] Joris Messens,[a] Elke Brosens,[a] Lode Wyns,[a] Henri De Greve[a] and Julie Bouckaert[a]*

[a]Department of Ultrastructure, Vrije Universiteit Brussel, Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), Pleinlaan 2, 1050 Brussels, Belgium, and [b]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

Correspondence e-mail: bouckaej@vub.ac.be

Received 3 March 2003
Accepted 28 March 2003

PDB Reference: F17a-G–β-Me-SeGlcNAc complex, 1o9v, r1o9vsf.

The *Escherichia coli* adhesin F17-G is a carbohydrate-binding protein that allows the bacterium to attach to the intestinal epithelium of young ruminants. The structure of the 17 kDa lectin domain of F17-G was determined using the anomalous dispersion signal of a selenium-containing analogue of the monosaccharide ligand *N*-acetyl-D-glucosamine in which the anomeric oxygen was replaced by an Se atom. A three-wavelength MAD data set yielded good experimental phases to 2.6 Å resolution. The structure was refined to 1.75 Å resolution and was used to solve the structures of the ligand-free protein and the F17-G–*N*-acetyl-D-glucosamine complex. This selenium-carbohydrate phasing method could be of general use for determining the structures of carbohydrate-binding proteins.

1. Introduction

Bacterial adhesins mediate the attachment of bacteria to eukaryotic host cells. A large class of bacterial adhesins are found on long wire-like organelles on the surface of the cell known as fimbriae or pili (Hultgren *et al.*, 1996). These fimbrial adhesins recognize specific carbohydrate receptors on the epithelial cells of the host. The fimbrial attachment of bacterial pathogens determines their host range and tissue tropisms and is an essential step in the earliest stage of infection, preceding colonization and/or invasion of the host cells. Knowledge of the structural basis for this protein–carbohydrate recognition process could guide new approaches for preventing these harmful organisms from establishing infection.

F17 fimbriae are found on *Escherichia coli* strains causing infections of the gut epithelium in young ruminants. Their adhesin F17-G consists of an N-terminal lectin domain, which contains the carbohydrate-binding site, and a C-terminal pilin domain, which connects the protein to the body of the fimbriae. Binding of F17-G can be inhibited by the monosaccharide *N*-acetyl-D-glucosamine (GlcNAc). This specificity clearly distinguishes it from the two other fimbrial adhesins of known structure, FimH and PapGII. These two adhesin crystal structures have been solved *de novo* in complex with their carbohydrate receptors α-D-mannose (Hung *et al.*, 2002) and globotetraose (Dodson *et al.*, 2001), respectively. The FimH crystal structure was solved by multiple isomorphous replacement (Choudhury *et al.*, 1999), while for PapGII the selenomethionine-substitution method was used (Dodson *et al.*, 2001).

Here, we describe for the first time the use of a selenium derivative of a carbohydrate ligand to solve a structure. Selenium substitution of the anomeric O atom of GlcNAc was used to determine the structure of the lectin domain of the F17-G adhesin. The results indicate that chemical modification

© 2003 International Union of Crystallography
Printed in Denmark – all rights reserved

of carbohydrates may be a generally applicable approach for solving crystal structures of carbohydrate-binding proteins.

2. Experimental procedures

2.1. Synthesis of the selenium carbohydrate

The carbohydrate analogue methyl 2-acetamido-2-deoxy-1-seleno-β-D-glucopyranoside (β-Me-SeGlcNAc) was synthesized as previously described (Ogra *et al.*, 2002). The synthesis starts from the commercially available compound 2-acetamido-2-deoxy-α-D-glucopyranosyl chloride 3,4,6-triacetate (Tokyo Kasei Organic Chemicals, Japan) and provides β-Me-SeGlcNAc in two reaction steps (Fig. 1) with a 95% overall yield.

2.2. Crystallization

The lectin domain of F17-G variant a (Lintermans *et al.*, 1991) was produced as a recombinant periplasmic protein in *E. coli* strain BL21-AI and purified by affinity chromatography using *N*-acetyl-D-glucosamine-agarose (Sigma). It crystallized readily in a variety of conditions containing 10–30% PEG 4000 with a pH ranging from 4 to 8. The crystals used for data collection were grown in 30% PEG 4000, 0.1 *M* sodium acetate pH 4.6 and 0.2 *M* ammonium acetate, soaked with a cryoprotectant solution (31% PEG 8000, 10% 2-propanol, 0.1 *M* HEPES pH 7.5) containing 20 m*M* β-Me-SeGlcNAc and then flash-frozen at 100 K. Annealing by interrupting the cryostream five consecutive times was found to dramatically improve the diffraction pattern for some of the crystals.

2.3. Data collection and phasing

Data were collected at the absorption edge, inflection point and high-energy remote wavelengths as determined by a fluorescence scan for selenium (Table 1). The data sets were collected on EMBL beamline BW7A at the DESY synchrotron and processed using *DENZO*, *XDISPLAYF* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997) and *TRUNCATE* from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). The unit cell has a solvent content of approximately 33%.

An anomalous difference Patterson map revealed the presence of a set of symmetry-related peaks consistent with a single anomalous scattering site (Fig. 2). *SOLVE*/*RESOLVE* 2.02 (Terwilliger & Berendzen, 1999) was used to solve a three-wavelength MAD problem in the two possible enantiomorphic space groups $P6_122$ and $P6_522$. The scattering factors f' and f'' were refined during phase calculation. *SOLVE* identified a single site located at (0.534, 0.316, 0.050) for $P6_122$ and at (0.782, 0.318, 0.05) for $P6_522$ (fractional coordinates) using data to 2.6 Å resolution. This location could be confirmed by inspecting the anomalous difference Patterson map. The overall figure of merit for phasing was 0.65. *RESOLVE* built 130 amino-acid residues in the $P6_122$ experimental electron-density map. The map for space group $P6_522$ was not interpretable.

Using default settings and data to 1.75 Å resolution, *ARP*/*wARP* 6.0 (Perrakis *et al.*, 1999) built 155 out of 177 amino acids. Refinement with *CNS* (Brünger *et al.*, 1998) at 1.75 Å resolution and manual model building resulted in a complete model for the protein–sugar complex with good geometry and crystallographic residuals (R = 20.5%, $R_{\rm free}$ = 23.5%).

Table 1
Crystal parameters and data statistics for the MAD experiment.

Data were collected at three wavelengths from a single crystal with unit-cell parameters $a = b = 42.76$, $c = 273.7$ Å. The space group is $P6_122$ and the mosaicity of the crystal was approximately 0.6°. Values in parentheses are for the highest resolution shell.

	Absorption edge	Inflection point	High-energy remote
Wavelength (Å)	0.9788	0.9792	0.9611
Refined f'	−5.0	−8.6	−2.8
Refined f''	3.8	2.3	4.5
Resolution range (Å)	50.0–2.60 (2.69–2.60)	50.0–2.60 (2.69–2.60)	50.0–1.75 (1.81–1.75)
Total/unique reflections	587139/5337	242076/5335	959107/16346
$\langle I/\sigma(I)\rangle$	14.8	13.4	10.3 (3.9)
Reflections with $I > 3\sigma(I)$ (%)	96.7	96.6	94.0 (76.3)
Completeness (%)	99.7 (99.2)	99.5 (99.4)	97.1 (93.1)
$R_{\rm merge}$† (%)	5.8 (11.5)	6.2 (13.0)	6.6 (32.1)

† $R_{\rm merge} = \sum |I_{hkl} - \langle I_{hkl}\rangle| / \sum \langle I_{hkl}\rangle$, where I_{hkl} is a single measured intensity value for the *hkl* reflection and $\langle I_{hkl}\rangle$ is the mean of all measured reflections for the *hkl* reflection.

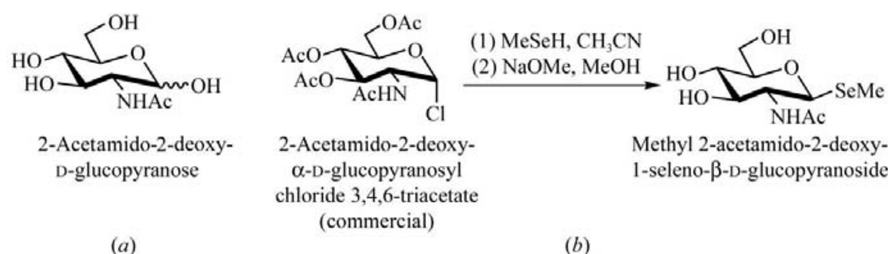

Figure 1
(*a*) Structural formula of GlcNAc (anomeric mixture). (*b*) Synthesis of the selenium derivative according to Ogra *et al.* (2002). The first step is a substitution reaction by treatment with monomethylselenol in acetonitrile. The second step is the removal of the *O*-acetyl groups with sodium methoxide in dry methanol.

3. Results and discussion

The structure of the 17 kDa F17-G lectin domain was solved by a three-wavelength MAD approach using the anomalous dispersion signal (Fig. 2) of the Se atoms in crystals soaked with β-Me-SeGlcNAc (Fig. 1). The β-Me-SeGlcNAc molecule was chosen as a selenium-containing potential ligand for the protein because the anomeric O atom (O1) was the most likely to be involved in a glycosidic link and the

least likely to be involved in direct interactions with the protein. The anomeric O atom was therefore considered to be the best candidate for substitution by selenium.

SOLVE identified a single selenium site (Z = 34) with an occupancy of 83% and a B factor of 18.2 Å^2. The experimental density in the vicinity of this site is consistent with the structure of β-Me-SeGlcNAc (Fig. 3). The asymmetric unit contains one F17-G lectin domain with a single ligand. Automatic model building using *RESOLVE* and *ARP*/*wARP* resulted in an excellent starting model for the protein structure, which reveals an immunoglobulin-like fold and will be described elsewhere.

N-acetyl-D-glucosamine (GlcNAc) has a free hydroxyl group on C1 and can interchange between the α- and β-anomeric forms in solution through opening of the saccharide ring. The modified sugar β-Me-SeGlcNAc is fixed in the β-anomeric configuration because of the substitution of the selenium with a methyl group. However, non-methylated GlcNAc bound to F17-G adopts the β configuration, as is apparent from the refined 1.65 Å resolution structure of the complex (PDB code 1o9w). Thus, the β-anomeric form is the physiologically relevant form selected upon specific binding and indeed corresponds to the GlcNAc configuration observed in all known biological glycans.

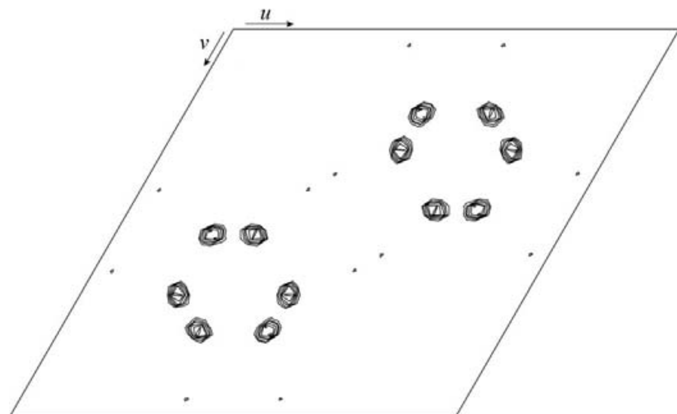

Figure 2
Anomalous difference Patterson map (for the Harker section at w = 1/6) showing the presence of a set of symmetry-related peaks consistent with a single anomalous scattering site for selenium.

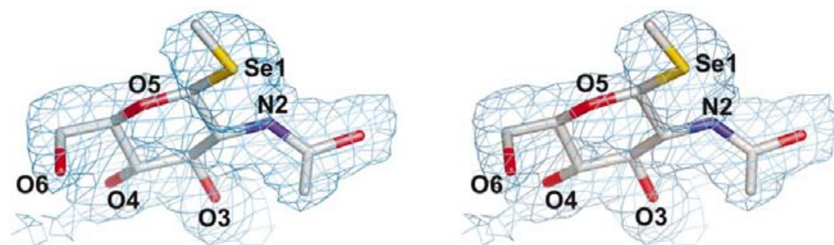

Figure 3
Stereo representation of the experimental electron density (1σ level) and the refined model for the β-Me-SeGlcNAc ligand. The sugar is presented as a ball-and-stick model with C atoms in white, O atoms in red, N atoms in blue and the Se atom in yellow.

The van der Waals radius of the Se atom (1.90 Å) is larger than that of oxygen (1.52 Å), but the binding site allows this difference. Both the Se atom and the methyl group could contribute to the fivefold better binding of β-Me-SeGlcNAc (K_d = 0.22 mM) compared with GlcNAc, which we measured with surface plasmon resonance.

The recognition by many lectins and adhesins of their receptors often depends on simple monosaccharides. This is, for instance, the basis of the classification of most of the legume lectins into monosaccharide specificity groups (Loris *et al.*, 1998). Some fimbrial adhesins also specifically recognize monosaccharides, *e.g.* FimH of uropathogenic *E. coli*, which binds α-D-mannose (Hung *et al.*, 2002), or the F17-G adhesin, which binds *N*-acetyl-β-D-glucosamine. Even in those cases where the receptor is a larger or complex oligosaccharide (*e.g. Phaseolus vulgaris* agglutinin, the PapG fimbrial adhesins), the binding can most often be inhibited with sufficiently high concentrations of a specific monosaccharide. Therefore, a relatively small repertoire of selenium sugars may prove sufficient for the initial phasing of a large number of carbohydrate-binding proteins. Alternatively, when the protein has no affinity for any monosaccharide, a specific oligosaccharide with incorporated selenium could be custom-synthesized.

4. Conclusion

Our results indicate that the incorporation of Se atoms into carbohydrates is a powerful phasing tool for carbohydrate-binding macromolecules. Moreover, the synthesis of selenium-containing analogues of carbohydrates is relatively straightforward and inexpensive. In addition, the nature of the Se substitution is non-invasive. Our case study also demonstrates that crystals of the carbohydrate-binding protein may simply be soaked with the selenium-containing saccharide in order to bind to a sugar-free crystal or to replace an already present ligand. In conclusion, the method is less laborious than selenomethionine substitution of the protein (Deacon & Ealick, 1999), which often requires reoptimization of protein expression and crystallization. A particular advantage of the use of selenium-containing carbohydrate ligands over selenomethionine substitution of the protein is the creation of a well ordered selenium-binding site by the specific binding of the ligand. An unbiased crystal structure can be obtained as well as extra information on the carbohydrate-binding site. The method is flexible in the choice of the oxygen on the monosaccharide ring to be substituted by selenium. Typically, the hydroxyl group is selected that is presumed to be least likely involved in direct hydrogen bonds with the protein: most often the reducing or anomeric group.

The authors acknowledge the use of EMBL beamline BW7A at the DESY synchrotron, Hamburg, Germany (project PX-02-370). LB is a research assistant and JB and RL are post-

doctoral fellows of the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO).

References

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, S. D. (1999). *Science*, **285**, 1061–1066.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Deacon, A. M. & Ealick, S. E. (1999). *Structure*, **7**, 161–166.

Dodson, K. W., Pinkner, J. S., Rose, T., Magnusson, G., Hultgren, S. J. & Waksman, G. (2001). *Cell*, **105**, 733–743.

Hultgren, S. J., Jones, C. H & Normark, S. (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Vol. 2, edited by F. C. Neidhardt, pp. 2730–2756. Washington, DC: ASM Press.

Hung, C.-S., Bouckaert, J., Hung, D., Pinkner, J. S., Widberg, C., DeFusco, A., Auguste, C. G., Strouse, R., Langermann, S., Waksman, G. & Hultgren, S. J. (2002). *Mol. Microbiol.* **44**, 903–915.

Lintermans, P. F., Bertels, A., Schlicker, C., Deboeck, F., Charlier, G., Pohl, P., Norgren, M., Normark, S., Van Montagu, M. & De Greve, H. (1991). *J. Bacteriol.* **173**, 3366–3373.

Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). *Biochim. Biophys. Acta*, **1383**, 9–36.

Ogra, Y., Ishiwata, K., Takayama, H., Aimi, N. & Suzuki, K. T. (2002). *J. Chromatogr. B*, **767**, 301–312.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.