Structural context for protein *N*-glycosylation in bacteria: The structure of PEB3, an adhesin from *Campylobacter jejuni*

ERUMBI S. RANGARAJAN,¹ SMITA BHATIA,² DAVID C. WATSON,²
CHRISTINE MUNGER,¹ MIROSLAW CYGLER,^{1,3} ALLAN MATTE,³
AND N. MARTIN YOUNG²

¹Department of Biochemistry, McGill University, Montreal, Quebec H3G 1Y6, Canada

²Institute for Biological Sciences, NRC, Ottawa, Ontario K1A 0R6, Canada

³Biotechnology Research Institute, NRC, Montreal, Quebec H4P 2R2, Canada

(RECEIVED December 19, 2006; FINAL REVISION February 16, 2007; ACCEPTED February 16, 2007)

Abstract

Campylobacter jejuni is unusual among bacteria in possessing a eukaryotic-like system for *N*-linked protein glycosylation at Asn residues in sequons of the type Asp/Glu-Xaa-Asn-Xaa-Ser/Thr. However, little is known about the structural context of the glycosylated sequons, limiting the design of novel recombinant glycoproteins. To obtain more information on sequon structure, we have determined the crystal structure of the PEB3 (Cj0289c) dimer. PEB3 has the class II periplasmic-binding protein fold, with each monomer having two domains with a ligand-binding site containing citrate located between them, and overall resembles molybdate- and sulfate-binding proteins. The sequon around Asn90 is located within a surface-exposed loop joining two structural elements. The three key residues are well exposed on the surface; hence, they may be accessible to the PglB oligosaccharyltransferase in the folded state.

Keywords: adhesin; *Campylobacter jejuni*; *N*-glycosylation; PEB3; sequon structure

Campylobacter jejuni is remarkable among bacteria in possessing an *N*-glycosylation system that resembles, in several respects, that of eukaryotic organisms (Szymanski and Wren 2005). The central enzyme of this *pgl* system is PglB, an oligosaccharyltransferase that is homologous to the eukaryotic oligosaccharyltransferase Stt3 in yeast (Wacker et al. 2002). It transfers the heptasaccharide GalNAc- α 1,4-GalNAc- α 1,4-(Glc- β 1,3)-GalNAc- α 1,4-GalNAc- α 1,4-GalNAc- α 1,3-Bac, where Bac is 2,4-diacetamido-2,4,6-trideoxy-D-Glc (Young et al. 2002) from an undecaprenylpyrophosphate carrier to specific Asn residues

on acceptor proteins (Young et al. 2002; Kowarik et al. 2006a). PglB forms *N*-glycosidic bonds between an Asn side chain and the modified GlcNAc, bacillosamine, or other acetamido sugars (Feldman et al. 2005), unlike the GlcNAc specificity of the eukaryotic system (Wacker et al. 2006).

The protein-acceptor characteristics of PglB are less well defined, and it is not clear whether *N*-glycosylation is carried out before or after the acceptor protein is folded, or in both states. Though the Asn sequons for PglB were thought to be the same as the eukaryotic one, Asn-Xaa-Ser/Thr (Young et al. 2002; Wacker et al. 2006), it was recently shown that PglB utilizes a more extended sequon, Asp/Glu-Xaa-Asn-Xaa-Ser/Thr (Kowarik et al. 2006a). However, additional features of the acceptor proteins appear necessary for the *N*-glycosylation. The structural context of the sequon is an obvious possibility,

Reprint requests to: Martin Young, Institute for Biological Sciences, NRC, 100 Sussex Drive, Ottawa, Ontario K1A 0R6, Canada, e-mail: Martin.Young@nrc-cnrc.gc.ca; fax: (613) 941-1327.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062737507>.

particularly given the structural biases of eukaryotic glycosylation sites (Petrescu et al. 2004, 2006) and the greater length of this prokaryotic sequon.

Almost no structural information is available on *C. jejuni* glycoproteins. Though protein glycosylation is essential for cell adhesion, colonization, and complementation (Linton et al. 2002; Szymanski et al. 2002; Larsen et al. 2004), nearly all *C. jejuni* glycoproteins are of unknown function and lack homologs in the PDB. One exception is the protein AcrA, which contains two glycosylation sites (Kowarik et al. 2006a).

One of the most abundant glycoproteins, PEB3 (Linton et al. 2002; Wacker et al. 2002) has a single sequon around Asn90 that is ~50% glycosylated. A strong antigen (Pei et al. 1991), it is a surface protein (Linton et al. 2002), with 51.8% sequence identity to Paa, an *Escherichia coli* adhesin (Batisson et al. 2003), and 49% identity to AcfC, an accessory colonization factor from *Vibrio cholera* (Peterson and Mekalanos 1988). PEB3's adhesin properties have been confirmed in experiments with Hep-2 cells (Drs. B.W. Wren and R. Langdon, pers. comm.). To investigate the structural context of PEB3's sequon, we determined its three-dimensional crystal structure. The five-residue sequon is in a well-exposed conformation on the protein surface, which will be of value for the designed introduction of sequons into carrier proteins.

Results

Both unlabeled and SeMet-substituted PEB3 behaved as dimers in dynamic light scattering and gel-filtration chromatography experiments (results not shown). The crystal structure was solved using a Se SAD experiment and refined at 1.6 Å resolution (Table 1). The two molecules of PEB3 in the asymmetric unit can be superimposed with a root-mean-squares deviation of 0.22 Å for 230 Cα atoms. Each monomer consists of nine α-helices and nine β-strands assembled into two domains (Fig. 1). The domains are discontinuous, with domain 1 consisting of residues Asp21–Phe99 and Arg213–Thr251, and domain 2 of residues Arg100–Tyr212. Each α/β domain has a central five-stranded β-sheet against which α-helices are packed, and are structurally similar, superposing with an RMSD of 1.6 Å for 112 Cα pairs. The two domains are connected on one side by two extended β-strands (β4, β9) that together form a β-ribbon, with strand β9 contributing to the β-sheets in both domains. A cleft is formed between the domains on the face opposite the connecting β-strands. In domain I, the strand order is β2–β1–β3–β9–β4, with strand β9 running antiparallel to the other four strands. In domain II, the topology of the mixed β-sheet is β7–β6–β8–β5–β9, with strand β5 running antiparallel to the other four strands. This chain

Table 1. X-ray crystallographic data

Data collection	
Space Group	<i>P</i> 2 ₁
Unit Cell	
a(Å)	46.5
b(Å)	101.7
c(Å)	59.1
β(°)	109.0
Z ^a	4
Resolution (Å) (highest resolution shell)	50–1.6 (1.65–1.60)
Wavelength (Å)	0.9792
Observed <i>hkl</i>	503501 ^b
Unique <i>hkl</i>	137826 ^c
Redundancy	3.7 (3.4)
Completeness	99.1 (95.5)
<i>R</i> _{sym} ^d	0.056 (0.332)
<i>I</i> /σ(<i>I</i>)	19.1 (4.1)
Wilson B (Å ²)	20.1
Refinement	
Resolution (Å)	50–1.60
<i>R</i> _{work} (# <i>hkl</i>) ^e	0.189 (61,232)
<i>R</i> _{free} (# <i>hkl</i>) ^f	0.215 (3373)
Average B-factors (Å ²) (# atoms)	
Protein	20.5 (3643)
Solvent	28.3 (349)
Citrate	18.5 (26)
Ramachandran plot	
Allowed (%)	100
Generous (%)	0
Disallowed (%)	0
rms Bonds (Å)	0.009
rms Angles (°)	1.1
PDB code	2HXW

^aNumber of molecules in the unit cell.

^{b,c}Bijvoet pairs unmerged.

^d $R_{\text{sym}} = (\sum |I_{\text{obs}} - I_{\text{avg}}|) / \sum I_{\text{avg}}$.

^e $R_{\text{work}} = (\sum |F_{\text{obs}} - F_{\text{calc}}|) / \sum F_{\text{obs}}$ (merged data).

^f $R_{\text{free}} = R_{\text{work}}$ but for 5% of reflections not included in refinement (merged data).

topology is that of the classic type II periplasmic-binding protein family (Dwyer and Hellinga 2004).

Dimer formation buries ~1845 Å² (15.8%) of the solvent-accessible surface of each monomer, and there are extensive intersubunit interactions. Residues in helix α3 in one subunit interact with residues from helices α5 and α6 in the other subunit. The β4 strand of one monomer interacts with the loop between β4 and α6 of the other monomer. Additionally, the loop between β8 and β9 in one subunit interacts with α5 in the second subunit. Helix α5 also interacts with residues of β9 in the other monomer.

PEB3 crystallized in the presence of 200 mM sodium citrate revealed clear density (Fig. 2) that was interpreted as a citrate ion bound within the cleft of each subunit of the dimer (Fig. 1). This cleft is formed mainly by loop regions that are situated between the secondary structure elements β1–α1, β2–α2, β5–α5, and α6–β6.

Figure 1. Structure of PEB3. (A) The PEB3 monomer, colored according to domain structure, with domain 1 colored sky-blue and domain 2 colored tan. Elements of secondary structure are labeled. (B) The PEB3 dimer, with monomer A colored according to secondary structure, and monomer B colored according to domain structure. The citrate ligands are depicted in stick representation, and the sequons to which N-linked glycans can be attached are colored magenta. Figures were prepared with PyMol (<http://pymol.sourceforge.net/>).

The organization of the ligand-binding sites is related by pseudo twofold symmetry, such that they are on the same face of the dimer. The interactions between the three ionizable carboxyl groups of the citrate and the protein involve both direct and water-mediated hydrogen bonds. Residues interacting with citrate include Thr138^{OG}, Ser139^{OG1}, Ser173^{OG}, the main chain amides of Ser139 and Ser173, and the carbonyl O of Asn137, Asn170, and Ser171 (Fig. 2). Citrate binding results in a hydrogen-bonding network that spans domains I and II.

The five-residue sequon Asp-Phe-Asn-Val-Ser around Asn90 occurs in a loop region connecting $\alpha 3$ and $\beta 4$ that is well exposed on the protein surface (Fig. 2). The Asp and Ser side chains are solvent exposed alongside the Asn, while those of the intervening Phe and Val point toward the interior of the protein.

Discussion

The structure indicates PEB3 may function as both a transporter protein and an adhesin, resembling PEB1a, which is an Asp/Glu transporter and an adhesin (Leon-Kempis Mdel et al. 2006). While its citrate binding

suggests a function related to dicarboxylic acid or tricarboxylic acid transport, there is another possible citrate transporter in *C. jejuni*, Cj0203. PEB3's sequence identities with the *E. coli* Paa adhesin and *V. cholera* AacF (Batisson et al. 2003) include the citrate-binding residues, suggesting binding of a common ligand.

A search for related protein structures, including use of the DALI server (<http://www.ebi.ac.uk/dali/>) revealed PEB3's highest similarity was with several class II periplasmic-binding proteins. All have similarly located binding sites for doubly or triply charged ions, but they are monomeric. They include molybdate-binding proteins from *E. coli* (ModA, PDB 1WOD, RMSD of 1.4 Å for 135 C α pairs) and *Azotobacter vinelandii* (PDB 1ATG, RMSD of 1.7 Å for 134 C α pairs), sulfate-binding protein from *Salmonella typhimurium* (PDB 1SBP, RMSD of 1.6 Å for 143 C α pairs), and ferric iron-binding protein from *Haemophilus influenzae* (PDB 1MRP, RMSD of 1.7 Å for 113 C α pairs). Structural similarity was also observed with a putative phosphate-binding protein (PDB 1TWY) with a RMSD of 1.3 Å for 96 C α pairs.

N-glycosylation sites in eukaryotic glycoproteins often occur between elements of secondary structure (Petrescu

Figure 2. PEB3-binding site and sequon structure. (A) Citrate binding-site density, 2F₀-F_C (omit) electron density map contoured at a level of 3σ. (B) Hydrogen-bonding interactions between PEB3 and citrate. (C) The PEB3 sequon ⁸⁸Asp-Phe-Asn-Val-Ser⁹², with the key surface-exposed residues labeled. (D) The region of the sequon ¹¹⁸Asp-Ser-Asn-Ile-Thr¹²² in Cj0982c (PDB 1XT8).

et al. 2004, 2006) and, similarly, the PEB3 site is in a connecting loop between α3 and β4. The exposed nature of the site (Figs. 1, 2) suggests that the *pgl* heptasaccharide could be attached to the Asn without rearrangement of the local protein structure. Another natural substrate for PglB, the Cys transport protein, Cj0982c, which also shares the class II periplasmic-binding protein fold (Muller et al. 2005), has a local structure around its single sequon ¹¹⁸Asp-Ser-Asn-Ile-Thr¹²² (Kowarik et al. 2006a) resembling that of PEB3 in its side-chain arrangement (Fig. 2) and in its position between two structural elements, the domain-connecting strand β4 and the short helix, α5.

However, in PEB3 the sequon is part of domain 1, while in Cj0982c it is in domain 2.

Two other *C. jejuni* glycoproteins, Cj0143c and Cj0734c, have homologs in the PDB (1PQ4 and 1HSL, respectively) whose structured regions do not include the N-terminal segments equivalent to the glycosylation sites. In these cases, glycosylation of loosely structured regions seems to be occurring in the manner of short peptides in vitro (Glover et al. 2005).

These examples indicate PglB can modify both folded and unfolded proteins in vivo and that no specific structural context or presentation of the sequon is required.

However, predicting the behavior of a particular eukaryotic substrate from these findings is not straightforward. A ribonuclease mutant (Kowarik et al. 2006b) was more reactive with PglB when unfolded, while a green fluorescent protein with a sequon introduced into an exposed loop was modifiable in the folded state. The PEB3 sequon presentation described here should provide a useful template for the design of glycosylation sites into other acceptor proteins in order to create novel glycoconjugate species.

Materials and Methods

Cloning, overexpression, and purification of PEB3

A construct consisting of the open reading frame without the periplasmic leader sequence, i.e., residues 21–251 (Pei et al. 1991), with a C-terminal His₆ sequence, was cloned into the expression plasmid pCWori⁺ and maintained in *E. coli* AD202. Cells were grown at 37°C in 1 L of 2YT medium supplemented with 150 µg/mL⁻¹ ampicillin to A₆₀₀ = 0.5, followed by induction with 1 mM isopropylthiogalactoside, and growth continued for 6 h. For preparation of SeMet-labeled protein, *E. coli* DL41(DE3) cells (*metA*⁻) were grown in LeMaster medium containing 25 mg/L⁻¹ SeMet (Hendrickson et al. 1990) for 18 h at 20°C. Cells were lysed in 10 mM Hepes buffer (pH 7.5) containing protease inhibitors (Roche) by mechanical disruption. After centrifugation, the soluble fraction was adjusted to 500 mM NaCl and 50 mM imidazole, and loaded onto a 5-mL HisTrap column (GE Healthcare). PEB3 was eluted with a linear gradient of 50–500 mM imidazole and the PEB3 fractions were dialyzed against 10 mM Hepes (pH 7.5), 250 mM NaCl.

Crystallization, structure determination, and refinement

PEB3 crystals were grown by hanging-drop vapor diffusion by equilibrating 1 µL of protein (16 mg/mL with 5 mM DTT) mixed with 1 µL of reservoir solution, 18% (w/v) polyethylene-glycol 3350, 0.2 M di-ammonium hydrogen citrate. Crystals appeared in 1 d at 20°C and were cryoprotected with 5% (v/v) 2-methyl-2,4-pentanediol in reservoir solution prior to flash cooling in a N₂ cold stream at 100 K (Oxford Cryosystems). Diffraction data were collected at beamline X12B, National Synchrotron Light Source, using a ADSC Quantum-4 CCD detector and processed using HKL2000 (Otwinowski and Minor 1997).

The structure of SeMet-labeled PEB3 was determined by single-wavelength anomalous dispersion at the Se peak energy and refined using data collected to 1.6 Å resolution (Table 1). The Se substructure was solved using SHELXD (Schneider and Sheldrick 2002) and used to compute phases using SHELXE to 2 Å resolution, having a figure-of-merit 0.68 following density modification. These phases were extended to 1.6 Å resolution and density modified using the program RESOLVE (Terwilliger 2000), yielding a figure-of-merit of 0.78. Approximately 90% of the model was built into the 1.6 Å map using ARP/wARP (Morris et al. 2003), and the remainder with COOT (Emsley and Cowtan 2004). The model was refined using REFMAC5 (Murshudov et al. 2003), with no reflection σ -cutoff applied.

The final model contains two PEB3 monomers in the asymmetric unit, residues 21–174 and 177–251 of monomer A, residues 22–251 of monomer B, two bound citrates, and 531 water molecules. Coordinates and structure factors have been deposited in the RCSB PDB under accession code 2HXW.

Acknowledgments

X-ray diffraction data were measured at beamline X12B of the National Synchrotron Light Source. Financial support comes principally from the Office of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy, and the National Center for Research Resources of the National Institutes of Health. We thank Alexei Soares and Dieter Schneider for assistance in data collection. This work was supported by the NRC's Genomics and Health Initiative and by a grant from the Canadian Institutes of Health Research, 200103GSP-90094-GMX-CFAA-19924 to M.C.

References

- Batisson, I., Guimond, M.-P., Girard, F., An, H., Zhu, C., Oswald, E., Fairbrother, J.M., Jacques, M., and Harel, J. 2003. Characterization of the novel factor Paa involved in the early steps of the adhesion mechanism of attaching and effacing *Escherichia coli*. *Infect. Immun.* **71**: 4516–4525.
- Dwyer, M.A. and Hellinga, H.W. 2004. Periplasmic binding proteins: A versatile superfamily for protein engineering. *Curr. Opin. Struct. Biol.* **14**: 495–504.
- Emsley, P. and Cowtan, K. 2004. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**: 2126–2132.
- Feldman, M.F., Wacker, M., Hernandez, M., Hitchen, P.G., Marolda, C.L., Kowarik, M., Morris, H.R., Dell, A., Valvano, M.A., and Aebi, M. 2005. Engineering N-linked protein glycosylation with diverse O-antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **102**: 3016–3021.
- Glover, K.J., Weerapana, E., Numao, S., and Imperiali, B. 2005. Chemo-enzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl-transferase from *Campylobacter jejuni*. *Chem. Biol.* **12**: 1311–1315.
- Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**: 1665–1672.
- Kowarik, M., Young, N.M., Numao, S., Schulz, B.L., Hug, I., Callewaert, N., Mills, D.C., Watson, D.C., Hernandez, M., Kelly, J.F., et al. 2006a. Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25**: 1957–1966.
- Kowarik, M., Numao, S., Feldman, M.F., Schultz, B.L., Callewaert, N., Keirmaier, E., Catrein, I., and Aebi, M. 2006b. N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**: 1148–1150.
- Larsen, J.C., Szymanski, C., and Guerry, P. 2004. N-linked protein glycosylation is required for full competence in *Campylobacter jejuni* 81–176. *J. Bacteriol.* **186**: 6508–6514.
- Leon-Kempis Mdel, R., Guccione, E., Mulholland, F., Williamson, M.P., and Kelly, D.J. 2006. The *Campylobacter jejuni* PEB1a adhesion is an aspartate/ glutamate-binding protein of an ABC transporter essential for microaerobic growth on dicarboxylic acids. *Mol. Microbiol.* **60**: 1262–1275.
- Linton, D., Allan, E., Karlyshev, A.V., Cronshaw, A.D., and Wren, B.W. 2002. Identification of N-acetylglucosamine-containing glycoproteins PEB3 and CgpA in *Campylobacter jejuni*. *Mol. Microbiol.* **43**: 497–508.
- Morris, R.J., Parrakis, A., and Lamzin, V.S. 2003. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**: 229–244.
- Muller, A., Thomas, G.H., Horler, R., Brannigan, J.A., Blagova, E., Levnikov, V.M., Fogg, M.J., Wilson, K.S., and Wilkinson, A.J. 2005. An ATP-binding cassette-type cysteine transporter in *Campylobacter jejuni* inferred from the structure of an extracytoplasmic solute receptor protein. *Mol. Microbiol.* **57**: 143–155.
- Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. 2003. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 247–255.

- Otwinowski, Z. and Minor, W. 1997. Processing of x-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Pei, Z., Ellison III, R.T., and Blaser, M.J. 1991. Identification, purification, and characterization of major antigenic proteins of *Campylobacter jejuni*. *J. Biol. Chem.* **266**: 16363–16369.
- Peterson, K.M. and Mekalanos, J.J. 1988. Characterization of the *Vibrio cholerae* ToxR regulon: Identification of novel genes involved in intestinal colonization. *Infect. Immun.* **56**: 2822–2829.
- Petrescu, A.-J., Milac, A.-L., Petrescu, S.M., Dwek, R.A., and Wormald, M.R. 2004. Statistical analysis of the protein environment of *N*-glycosylation sites: Implications for occupancy, structure, and folding. *Glycobiology* **14**: 103–114.
- Petrescu, A.-J., Wormald, M.R., and Dwek, R.A. 2006. Structural aspects of glycomes with a focus on *N*-glycosylation and glycoprotein folding. *Curr. Opin. Struct. Biol.* **16**: 600–607.
- Schneider, T.R. and Sheldrick, G.M. 2002. Substructure solution with ShelxD. *Acta Crystallogr. D Biol. Crystallogr.* **58**: 1771–1779.
- Szymanski, C.M. and Wren, B.W. 2005. Protein glycosylation in bacterial mucosal pathogens. *Nat. Rev. Microbiol.* **3**: 225–237.
- Szymanski, C.M., Burr, D.H., and Guerry, P. 2002. Campylobacter protein glycosylation affects host cell interactions. *Infect. Immun.* **70**: 2242–2244.
- Terwilliger, T.C. 2000. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**: 965–972.
- Wacker, M., Linton, D., Hitchen, P.G., Nita-Lazar, M., Haslam, S.M., North, S.J., Panico, M., Morris, H.R., Dell, A., Wren, B.W., et al. 2002. *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**: 1790–1793.
- Wacker, M., Feldman, M.F., Callewart, N., Kowarik, M., Clarke, B.R., Pohl, N.L., Hernandez, M., Vines, E.D., Valvano, M.A., Whitfield, C., et al. 2006. Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci.* **103**: 7088–7093.
- Young, N.M., Brisson, J.-R., Kelly, J., Watson, D.C., Tessier, L., Lanthier, P., Jarrell, H.C., Cadotte, N., St. Michael, F., Aberg, E., et al. 2002. Structure of the *N*-linked glycan present on multiple glycoproteins in the gram-negative bacterium, *Campylobacter jejuni*. *J. Biol. Chem.* **277**: 42530–42539.