

High-resolution structure of NodZ fucosyltransferase involved in the biosynthesis of the nodulation factor^{***}

Krzysztof Brzezinski¹, Tomasz Stepkowski², Santosh Panjekar³, Grzegorz Bujacz^{2,4} and Mariusz Jaskolski^{1,2}✉

¹Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland; ²Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; ³EMBL Hamburg Outstation, c/o DESY, Hamburg, Germany; ⁴Institute of Technical Biochemistry, Technical University of Łódź, Poland

Received: 28 June, 2007; revised: 22 August, 2007; accepted: 24 August, 2007
available on-line: 30 August, 2007

The fucosyltransferase NodZ is involved in the biosynthesis of the nodulation factor in nitrogen-fixing symbiotic bacteria. It catalyzes α 1,6 transfer of L-fucose from GDP-fucose to the reducing residue of the synthesized Nod oligosaccharide. We present the structure of the NodZ protein from *Bradyrhizobium* expressed in *Escherichia coli* and crystallized in the presence of phosphate ions in two crystal forms. The enzyme is arranged into two domains of nearly equal size. Although NodZ falls in one broad class (GT-B) with other two-domain glycosyltransferases, the topology of its domains deviates from the canonical Rossmann fold, with particularly high distortions in the N-terminal domain. Mutational data combined with structural and sequence alignments indicate residues of potential importance in GDP-fucose binding or in the catalytic mechanism. They are all clustered in three conserved sequence motifs located in the C-terminal domain.

Keywords: glycosyltransferase, fucosyltransferase, NodZ, nodulation, nitrogen fixation

INTRODUCTION

The rhizobium genes *nod*, *nol* and *noe*, all involved in the biosynthesis of the nodulation (Nod) factor, belong to the most specific component of the rhizobium genome (Stepkowski & Legocki, 2001).

The Nod factor (Fig. 1a) is a signaling molecule necessary for the recognition of the rhizobium symbiont by the legume host (DeBelle *et al.*, 2001), yellow lupine in the case of *Bradyrhizobium* sp. WM9 (Stepkowski *et al.*, 2003), and for the induction of the root nodule meristem (Denarie *et al.*, 1996) during the ini-

✉ **Corresponding author:** Mariusz Jaskolski, Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznań, Poland; tel. (48 61) 829 1274; fax: (48 61) 829 1505; e-mail: mariuszj@amu.edu.pl

★ Some of the calculations were carried out in the Poznan Metropolitan Supercomputing and Networking Center.

* Due to unfortunate PDB renaming conventions, the phosphate ions in the PDB entry 2HHC (form I) are numbered differently than in the entry 2HLH (form II). To simplify discussion, this paper uses a common numbering scheme, as in the entry 2HLH. In the PDB entry 2HHC, the numbers of the phosphate groups 2 and 3 are swapped.

† **Protein Data Bank accession codes:** The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank with the accession codes 2HHC (native form I), 2HLH (native form II) and 2OCX (Se-Met derivative).

Abbreviations: AGT, phage T4 α -glucosyltransferase; BGT, phage T4 β -glucosyltransferase; DTT, 1,4-dithiothreitol; FucT, fucosyltransferase; Fut8, human α 1,6-fucosyltransferase; GlcNAc, N-acetyl-D-glucosamine; GlgA, glycogen synthase; GP, glycogen phosphorylase; GPGTF, glycogen phosphorylase/glycosyltransferase; GtfB, TDP/UDP-glucose: aglycosyl-vancomycin glucosyltransferase; GtfD, UDP- β -L-4-epi-vancosamine: vancomycin-pseudoaglycone vancosaminyltransferase; GT, glycosyltransferase; GT-A, glycosyltransferase type A; GT-B, glycosyltransferase type B; MurG, UDP-N-acetylglucosamine:N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase; NDP, nucleoside diphosphate; ORF, open reading frame; OtsA, α,α -trehalose-phosphate synthase; PMSF, phenylmethylsulfonyl fluoride; r.m.s.d., root-mean-square deviation; ST1, α 2,3-sialyltransferase; SpsA, spore coat polysaccharide biosynthesis glycosyltransferase from *Bacillus subtilis*; WaaG, UDP-glucose:L-glycero-D-manno-heptose II α 1,3-glycosyltransferase I; UGT71G, UDP-glucose:flavonoid glycosyltransferase; VvGT1, UDP-glucose:anthocyanidin 3-O-glycosyltransferase.

tiation of the symbiotic process in which the bacteria infect the root nodules and fix atmospheric nitrogen. The Nod factor is a lipo-chitooligosaccharide formed by three-to-six β 1,4-linked *N*-acetyl-D-glucosamine (GlcNAc) residues and a fatty acid acyl group attached to the nitrogen atom at the non-reducing end (Denarie *et al.*, 1996). The number of GlcNAc units of the chitin backbone, the nature of the fatty acid, and possible presence of additional substituents, for instance, saccharide residues such as D-arabinose or L-fucose (Stacey *et al.*, 1994; Mergaert *et al.*, 1996), are crucial for host specificity and nodulation efficiency (Olsthoorn *et al.*, 1998). With a few exceptions, most of the nodulation genes are unique to the symbiotic bacteria and do not have counterparts in other organisms. For this reason, the origin of the nodulation loci, i.e., their emergence and evolution, are currently unknown. This issue is of special interest considering that symbiotic nitrogen fixation is responsible for the assimilation of 170 million tons of N_2 annually and that it involves both very ancient proteins, such as the nitrogenase complex that evolved over three billion years ago (Raymond *et al.*, 2004), as well as relatively new nodulation functions that appeared not earlier than 60 million years ago, or presumably later, following the explosive radiation of Leguminosae plants in early Eocene (Lavin *et al.*,

2005). Determination of the structure of the nodulation proteins should facilitate the elucidation of their origin and our understanding of how this remarkable function has evolved. Apart from the evolutionary aspects, isolation of the nodulation proteins and determination of their structure will illuminate the biochemistry of Nod factor synthesis.

In this paper, we present the crystal structure of bacterial α 1,6-fucosyltransferase NodZ from *Bradyrhizobium* sp. WM9 (EC 2.4.1.68), which according to the CAZy database (carbohydrate-active enzyme database at <http://www.cazy.org/>) (Coutinho & Heenrissat, 1999; Coutinho *et al.*, 2003) belongs to family GT-23 of glycosyltransferases. The enzyme is involved in the biosynthesis of the Nod factor by catalyzing the transfer of L-fucose from the GDP- β -L-fucose donor to the reducing residue of the chitin oligosaccharide backbone, preferably formed by five or six GlcNAc units. The reaction has to take place before the attachment of a fatty acid group (Quesada-Vinces *et al.*, 1997; Quinto *et al.*, 1997) (Fig. 1b).

Carbohydrates and glycosylated proteins play a number of cellular functions in Bacteria, Archaea and Eucarya. The biosynthesis of complex carbohydrates and protein glycosylation are carried out by glycosyltransferases (GTs), which account for about 1–2% of all ORFs in the genomes (Coutinho *et al.*,

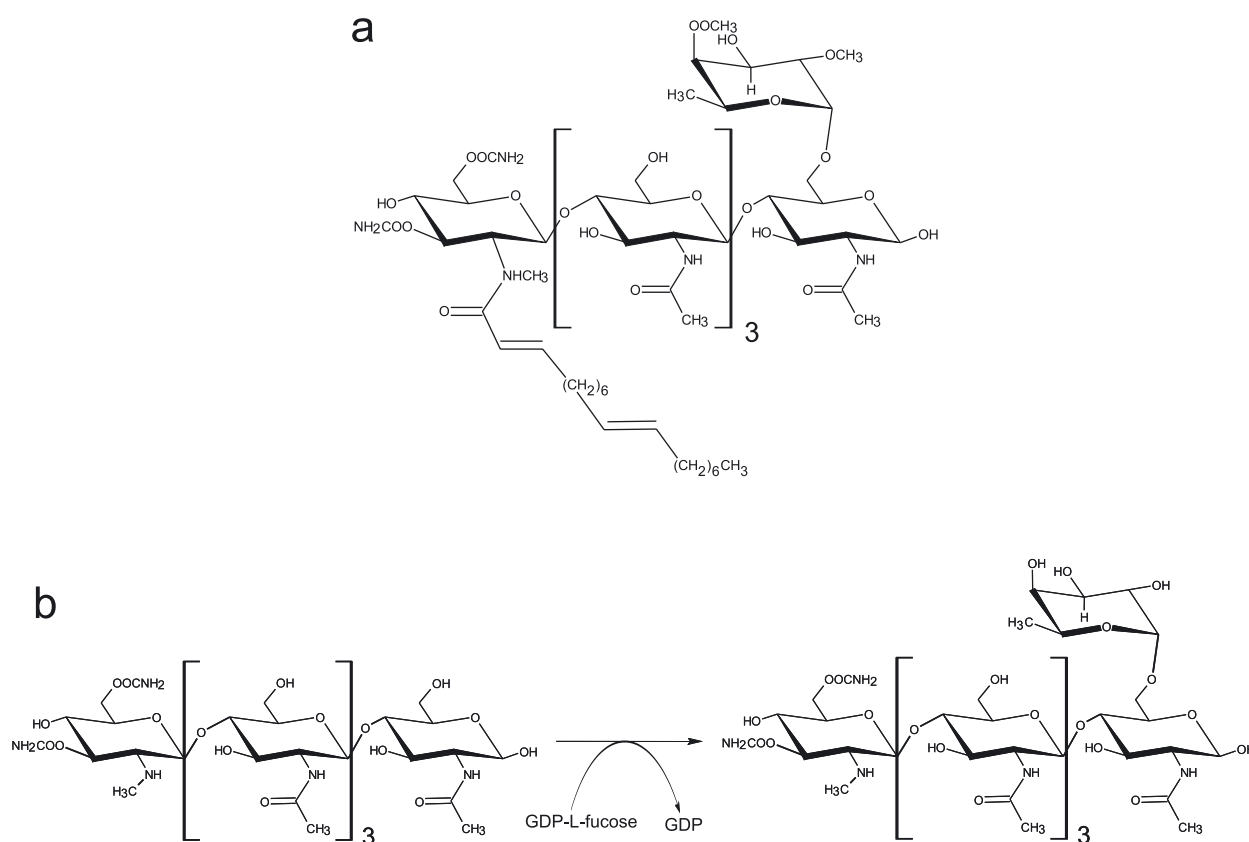


Figure 1. The structure and fucosylation reaction of the nodulation factor.

(a) Structure of the Nod factor from *Bradyrhizobium* sp. WM9 and (b) the reaction catalyzed by NodZ fucosyltransferase.

2003; Davies *et al.*, 2005). They exhibit a remarkable diversity of their donor and acceptor molecules, being able to generate a practically unlimited range of products. On the level of amino-acid sequences, GTs are a family of highly dissimilar proteins, currently classified into 90 groups in the CAZy database. Despite low sequence similarity, there are only two distinct folds adopted by these proteins. One of them, GT-A, is typified by the SpsA protein from *Bacillus subtilis* (Charnock & Davies, 1999), whereas the other one, GT-B, is defined by phage T4 β -glucosyltransferase (Vrielink *et al.*, 1994). The GT-A fold consists of two closely associated $\beta/\alpha/\beta$ domains that tend to form a central β -sheet of eight or more β -strands, usually flanked on both sides with α -helices. The GT-B fold consists of two domains, each containing a Rossmann-like motif. The GT-B enzymes share a similar structural type with glycogen phosphorylases and UDP-*N*-acetylglucosamine 2-epimerases, and all together form the GPGTF superfamily (Campbell *et al.*, 2000; Wrabl & Grishin, 2001).

Fucosyltransferases are found in eukaryotic and prokaryotic organisms. According to the CAZy database, they catalyze the transfer of L-fucose with an inversion of the anomeric configuration to various acceptors, such as oligosaccharides or proteins, using the same GDP- β -L-fucose donor. By analogy to other fucosyltransferases, it can be assumed that NodZ is also an inverting enzyme but, so far, this speculation has not been confirmed experimentally. The transfer of L-fucose creates an α 1,2-linkage at a galactose residue, α 1,3/1,4 or α 1,6-linkage at an *N*-acetyl-D-glucosamine residue, or takes place directly at a serine or threonine side chains on protein glycosylation (Ma *et al.*, 2006). Despite a lack of overall sequence conservation, three peptide motifs have been noted in prokaryotic and eukaryotic α 1,2- and α 1,6-fucosyltransferases and protein-O- α -fucosyltransferases (Breton *et al.*, 1998; Oriol *et al.*, 1999; Chazalet *et al.*, 2001; Martinez-Duncker *et al.*, 2003). α 1,3/1,4-Fucosyltransferases form a distinct group which share two common peptide motifs absent in other fucosyltransferases (Breton *et al.*, 1998). The reaction catalyzed by α 1,3/1,4-fucosyltransferases requires the presence of Mn²⁺ or Mg²⁺ (Murray *et al.*, 1996) while for the other fucosyltransferases divalent metal ions are not mandatory. So far two crystal structures of fucosyltransferases have been determined, α 1,3-FucT from *Helicobacter pylori* (GT-10) (Sun *et al.*, 2007) and human α 1,6-FucT (GT-23) (Ihara *et al.*, 2007).

MATERIALS AND METHODS

Protein expression and purification. Preliminary data on the cloning, expression, purification and crystallization of recombinant NodZ fucosyl-

transferase have been reported previously (Brzezinski *et al.*, 2004) but the limited diffraction of the crystals obtained from that material, traced to protein nonhomogeneity (detected using native PAGE), has led to a modification of the purification procedure. Briefly, after expression in BL21 CodonPlus[®] (DE3)-RIPL cells, the pellet was resuspended in the lysis buffer (5 mM imidazole, 500 mM NaCl, 20 mM Tris/HCl, pH 7.9) with the addition of 1 mM PMSF, 5 mM β -mercaptoethanol and 20 mg/L lysozyme. After 1 h incubation on ice, cells were disrupted with sonication on ice and centrifuged for removal of cell debris. The supernatant was loaded onto a His-Bind[®] column (Novagen) equilibrated with 50 mM NiSO₄. The protein was eluted in a gradient (0.1–1 M) of imidazole in 500 mM NaCl, 20 mM Tris/HCl, pH 7.9, 5 mM β -mercaptoethanol, and dialyzed against a buffer containing 20 mM Tris/HCl, pH 8.0, 50 mM NaCl, 5 mM DTT, 1 mM EDTA and 5% (w/v) glycerol. The C-terminal His₆ tag following immediately the NodZ sequence was not removed. The protein solution was concentrated to 10 mg/mL and stored at 193 K.

Preparation of selenomethionyl NodZ fucosyltransferase. Se-Met-labeled NodZ fucosyltransferase was produced using metabolic inhibition of the methionine biosynthetic pathway (Van Duyne *et al.*, 1993) in the BL21 CodonPlus[®] (DE3)-RIPL *E. coli* strain. Recombinant *E. coli* cells were grown at 303 K in LB medium containing 100 μ g/mL ampicillin and 34 μ g/mL chloramphenicol till turbidity. Two milliliter of culture were transferred to 1 L of M9 minimal medium containing additionally 2 mM MgSO₄, 0.1 mM CaCl₂, 50 mg/L EDTA, 4 g/L glucose, vitamins (1 mg/L of riboflavin, thiamine, pyridoxine monohydrochloride, niacinamide), a mixture of amino acids (100 mg/L of L-arginine, L-aspartate, L-cysteine, L-glutamate, glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-phenylalanine, 150 mg/L of L-valine, 400 mg/L of L-serine, L-threonine) and trace amounts of FeCl₃, CuCl₂, CoCl₂, MnCl₂, ZnCl₂, and H₃BO₃. Bacteria were grown at 303 K until the OD_{600nm} reached 1.0. After overnight incubation on ice, the culture was diluted with fresh minimal medium to the final OD_{600nm} of 0.2. The culture was agitated at 303 K until the OD_{600nm} reached 0.8, at which point a mixture of amino acids for the metabolic inhibition of the methionine biosynthetic pathway (50 mg/L of L-isoleucine, L-leucine, L-valine, 100 mg/L of L-lysine, L-phenylalanine, L-threonine) and 100 mg/L of solid DL-selenomethionine were added. The culture was cooled to 291 K and after 30 min incubation, 0.3 mM IPTG was added to induce the production of Se-Met-labeled NodZ for 16 h. The purification procedure was as described for the native enzyme except that the concentration of β -mercaptoethanol in the lysis and elute buffers was

increased to 7.5 mM. The substitution of the six methionine sites with selenomethionine was confirmed by mass spectrometry (not shown).

Crystallization. The native enzyme was crystallized using the hanging-drop vapor-diffusion technique with KH_2PO_4 replacing the previously reported $\text{NH}_4\text{H}_2\text{PO}_4$ precipitant (Brzezinski *et al.*, 2004). The protein solution was dialyzed against buffer containing 20 mM Tris/HCl, pH 8.0, 50 mM NaCl, 5 mM DTT, 1 mM EDTA, and concentrated to 10 mg/mL. Because of fast precipitation, the protein solution was preincubated (ratio 2:1) with a buffer containing 100 mM $\text{NH}_4\text{H}_2\text{PO}_4$ and 100 mM Tris/HCl, pH 7.5, for 12 h and then centrifuged, and the supernatant was used to set up the crystallizations. The best crystals were obtained by mixing 4 μl of protein solution with 2 μl of reservoir solution con-

taining 250 mM KH_2PO_4 at 292 K (form I) or at 277 K (form II). The Se-Met derivative was crystallized as the native form I, but the precipitant concentration was increased to 300 mM.

Data collection and processing. The crystals were transferred to a cryo-protectant solution containing the mother liquor and 15% (v/v) (2R,3R)-(-)-2,3-butanediol, and flash-frozen to 100 K in a gas nitrogen stream. X-Ray diffraction data for the native form I were measured at the X11 EMBL beamline and for form II at the BW7A EMBL beamline of the DESY synchrotron to the resolution of, respectively, 1.54 and 1.95 Å. In both cases, the crystals are hexagonal and belong to space group $P6_522$ with one protein molecule in the asymmetric unit, but they differ in the lattice parameters and unit cell volume (Table 1). A crystal of the

Table 1. Crystallographic data and refinement statistics for NodZ.

Data collection and processing statistics					
	Se-MAD data				
Data set	peak	edge	remote	native form I	native form II
Beamline		BW7A		X11	BW7A
Wavelength (Å)	0.9787	0.9790	0.9537	0.8128	1.0200
Temperature (K)			100		
Space group			$P6_522$		
Cell dimensions (Å)					
<i>a</i>		124.2		124.2	130.1
<i>c</i>		96.2		96.1	83.3
Mosaicity (°)		0.56		0.41	0.34
Resolution range (Å)	20.00–2.40 (2.49–2.40) ^a	20.00–2.40 (2.49–2.40)	20.00–2.20 (2.28–2.20)	15.00–1.54 (1.57–1.54)	15.00–1.95 (2.01–1.95)
Total reflections	712549	699422	964096	1499323	364665
Unique reflections	32442	31449	22641	64545	30716
Completeness (%)	100.0 (100.0)	99.6 (96.1)	100.0 (100.0)	100.0 (100.0)	99.9 (100.0)
Redundancy	22.0 (17.5)	22.2 (13.7)	42.6 (43.1)	24.4 (17.8)	11.7 (11.7)
$\langle I/\sigma(I) \rangle$	33.3 (4.5)	33.7 (4.5)	56.0 (13.4)	72.5 (3.8)	58.2 (11.9)
R_{merge}^b	0.114 (0.679) ^c	0.108 (0.619) ^c	0.085 (0.346)	0.050 (0.802)	0.045 (0.270)
	0.110 (0.670) ^d	0.107 (0.611) ^d			
Refinement statistics					
Resolution (Å)			20.00–2.20	15.00–1.54	15.00–1.95
No. of reflections in					
working set			21436	61199	29626
test set			1062	1270	1085
R/R_{free}^e			0.184/0.217	0.182/0.197	0.161/0.206
No. of atoms (protein/water)			2410/180	2437/310	2344/274
No. of ions (phosphate/Tris)			3/1	3/1	4/0
R.m.s.d. from ideality					
bond lengths (Å)			0.017	0.017	0.017
bond angles (°)			1.532	1.705	1.599
Average B factor (Å ²)			32.7	29.0	27.0
Ramachandran statistics (%)					
most favored regions			93.4	92.9	92.0
allowed regions			6.6	7.1	8.0
PDB code			2OCX	2HHHC	2HLH

^aThe values in parentheses correspond to the last resolution shell; ^b $R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$, where $\langle I(\text{hkl}) \rangle$ is the average intensity of reflection hkl; ^cFor the MAD data, treating Bijvoet reflections as equivalent; ^dFor the MAD data, treating Bijvoet reflections as non-equivalent; ^e $R = \sum_{\text{hkl}} | |F_o(\text{hkl})| - |F_c(\text{hkl})| | / \sum_{\text{hkl}} |F_o(\text{hkl})|$, where F_o and F_c are the observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement.

Se-Met derivative, isomorphous with form I, was used for a multiwavelength anomalous diffraction (MAD) experiment at the Se *K*-absorption edge at the EMBL beamline BW7A of the DESY synchrotron. All diffraction images were processed and scaled with DENZO and SCALEPACK from the HKL suite (Otwinowski & Minor, 1997). Data collection and processing statistics for all the experiments are shown in Table 1.

Structure determination and refinement. The structure was solved on-the-fly using the peak and inflection-point Se-Met MAD data and the software pipeline AutoRickshaw (Panjikar *et al.*, 2005). The structure solution was completed (using SHELXD (Schneider & Sheldrick, 2002) and SHELXE (Sheldrick, 2002)) while the inflection-point experiment was still under way. The remote- λ data set was used for automatic model building in the ARP/wARP (Morris *et al.*, 2004) procedure, which traced 275 out of the 330 residues, in this number 266 with side chains, and subsequently for maximum-likelihood refinement in Refmac5 (Murshudov *et al.*, 1997), which included two TLS (Winn *et al.*, 2001) groups, one for each domain. Difference electron density maps calculated in model rebuilding cycles which alternated with the Refmac5 calculations revealed only some of the missing residues, indicating that the C-terminus as well as parts of some surface loops might be disordered. At the conclusion of the refinement, residues 179–190, 245–255 and 318–330 were not modeled due to lack of electron density. The final Se-Met NodZ model includes 180 water molecules plus three phosphate ions and one Tris ion (Table 1).

Native form I — The Se-Met NodZ model was transferred to the nearly identical unit cell of the native crystal form I, for which diffraction data were collected to 1.54 Å resolution. The refinement was carried out as for the Se-Met derivative crystal structure. In the final rounds of refinement, 310 water molecules were included in the model together with three phosphate anions and one Tris cation. As in the Se-Met NodZ structure, electron density was not defined for some parts of the model and residues 179–190, 245–255, 305–306 and 318–330 were not modeled.

Native form II — The structure of native form II was solved by molecular replacement using the program MolRep (Vagin & Teplyakov, 1997) and the structure of form I as the search model. The correct solution was characterized by an *R* factor of 0.40 and a correlation coefficient of 0.58. The refinement was carried out as above. Two hundred seventy four water molecules and three phosphate anions were included in the model. The three surface loops, residues 179–192, 245–256, 305–308, and the C-terminus (318–330) could not be traced in the electron density.

Model building and validation. The Xfit program from the XtalView package (McRae, 1999) was used for manual modeling in the electron density maps. Stereochemical quality of the models was assessed with PROCHECK (Laskowski *et al.*, 1993). The refinement statistics for the native and Se-Met structures are reported in Table 1.

RESULTS AND DISCUSSION

Overall structure of NodZ

NodZ is a monomeric protein consisting of 330 amino-acid residues. The structure is arranged into nearly equal but differently folded N- and C-terminal domains, separated by a deep central cleft (Figs. 2a and 3a). The two domains are connected by a short linker, residues 147–150. The N-terminal domain (residues 2–146) consists of a central β -sheet (β N3- β N4- β N2- β N1- β N6- β N5) with only one strand (β N3) in antiparallel orientation. The β -sheet is flanked by three 3_{10} helices (α N3, α N4, α N5) on one side and four α -helices (α N1, α N2, α N6, α N7) on the other. The C-terminal domain (residues 151–318) has a Rossmann-like fold, with a central six-stranded β -sheet (β C3- β C2- β C1- β C4- β C5- β C6). Its last (β C5- β C6) segment is a hairpin rather than a β - α - β crossover, disrupting the Rossmann geometry and the parallel strand organization (Fig. 2b). The C-terminal β -sheet is flanked by two α -helices (α C2, α C3) on one side and by three α -helices (α C1, α C4, α C5) on the other side.

Superposition of 277 C α atoms from both native crystal forms is characterized by an r.m.s.d. of 0.41 Å. Both models have poor electron density definition of three fragments within the C-terminal domain and the C-terminus itself, indicating conformational flexibility and disorder. Fragment 1 (Fig. 2a, 3a) corresponds to a surface loop located near a potential nucleotide-sugar binding site and includes residues responsible for catalysis (Takahashi *et al.*, 2000; Chazalet *et al.*, 2001). Disorder in this region is also observed in other substrate-free GT-B structures, GtfB (Mulichak *et al.*, 2001) and Fut8 (Ihara *et al.*, 2007). The other two loops (fragments 2 and 3) are also located on the surface but their role in substrate binding is unknown.

Ions associated with the NodZ molecule

The high resolution and quality of the maps allowed us to exclude with certainty the possibility of metal coordination in the reported structures. In both native crystal forms, phosphate anions could be identified. In addition, form I cocrystallized with a Tris cation. All three P_i-1-3(I)^{*} phosphate ions in

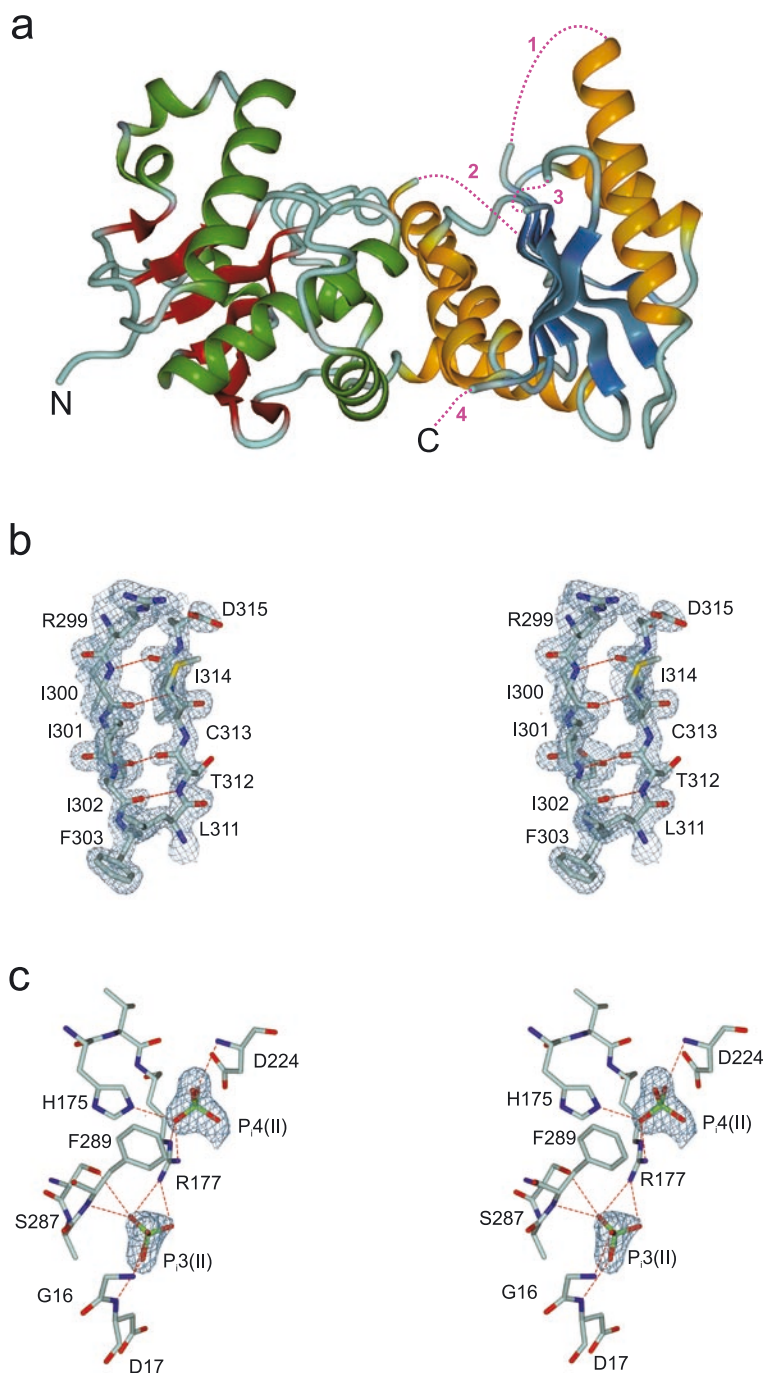


Figure 2. The structure and topology of NodZ fucosyltransferase.

(a) The tertiary structure of NodZ. The N-terminal domain is colored with green α - and 3_{10} -helices and red β -strands, and the C-terminal domain is colored with gold α -helices and blue β -strands. The numbers 1–4 and the dashed lines correspond to gaps in the model. (b) Stereoview of the antiparallel β C5- β C6 element in the structure of form I, shown with $2F_o$ - F_c electron density map contoured at the 1.0σ level. The dashed lines represent antiparallel β -sheet main-chain hydrogen bonds. (c) Stereoview of phosphates P_i 3-4(II) bound in the structure of form II, shown in $2F_o$ - F_c electron density map contoured at the 0.9σ level. The structural figures were prepared with DINO (www.dino3d.org).

form I are located at the protein surface. Phosphate P_i 3(I) forms crystallographic dimers which interact with the Tris ions. Ions P_i 1-2(I) are not engaged in any intermolecular interactions. Ions P_i 1-2 interact with the NodZ molecule in the same way in both forms. However, due to different crystal packing, P_i 1(II) has a different environment and participates in an interaction between symmetry related protein molecules. Ions P_i 3-4(II) are found only in form II and are located within the interdomain cleft. P_i 3(II) is coordinated by amino-acid residues from both domains whereas P_i 4(II) interacts only with the C-terminal domain (Fig. 2c). The interactions of the

phosphate ions are listed in Supplemental Table 2 (on-line edition only at www.actabp.pl).

NodZ and other GPGTF members

While sequence homology within the group of NodZ fucosyltransferases from different rhizobia is very high, there is almost no sequence similarity (identity below 20%) between NodZ and the GPGTF superfamily. Nevertheless, NodZ shares some structural features with such GPGTF members as GT-Bs and 2-epimerases (Campbell *et al.*, 2000; Wrabl & Grishin, 2001) (Fig. 4). All these enzymes are folded

as two distinct N- and C-terminal domains, separated by a deep cleft. Each domain contains a β -sheet core associated with helices. A DALI (Holm & Sander, 1998) search with the NodZ model as a probe has revealed similarity to a number of Rossmann-fold enzymes with a nucleotide-binding function, such as the GPGTF superfamily members or oxidoreductases. The closest matches are human α 1,6-fucosyltransferase Fut8 (Figs. 3b and 4) (Ihara *et al.*, 2007) (PDB code 2DE0) with a DALI Z-score of 8.9 and multifunctional sialyltransferase ST1 from *Pasteurella mul-*

ticida (Ni *et al.*, 2006) (PDB code 2EX0) with a DALI Z-score of 7.0. The r.m.s.d. based on structural alignment of about 120 common C α atoms (out of 460) is 3.0 Å for both Fut8 and ST1. Nevertheless, sequence identity is very low (17 and 9% for Fut8 and ST1, respectively). The next DALI Z-scores (6.7 and 6.6) are for UDP-*N*-acetylglucosamine 2-epimerase (Campbell *et al.*, 2000) (PDB code 1F6D) and for ADP-heptose lipopolysaccharide heptosyltransferase (WaaF) (unpublished, PDB code 1PSW) with the respective r.m.s. deviations of 4.5 and 4.0 Å (for 204/376 and 167/319 C α atoms) and about 9% sequence identity over the superposed positions. There are also other GT-B proteins with a DALI Z-score about 6–5 but with sequence identity below 20%. On the other hand, the DALI search showed a very low structural similarity to an α 1,3-fucosyltransferase (Sun *et al.*, 2007) (PDB code 2NZW), a member of the GPGTF superfamily (Z-score 2.2, r.m.s.d. 4.9 Å for 138/349 C α atoms, 6% sequence identity).

There are several significant differences between the NodZ fucosyltransferase and the other GPGTF superfamily members. The first discrepancy is in the topology of the N-terminal domain. In NodZ, it is not arranged with the Rossmann topology found in the other enzymes. In addition to a completely different pattern of connectivity, the central β -sheet includes one antiparallel segment (within a β -hairpin) and is six-stranded while in most GPGTF proteins the β -sheet consists of seven parallel strands, sometimes with a few extra ones. The differences in the N-terminal domain topology influenced the DALI search, with the hits based mainly on the similarity of the C-terminal domain of NodZ to a Rossmann-type domain with a six-stranded β -sheet core. However, the C-terminal domain of NodZ differs from the all-parallel topology found in other GPGTFs, as its last segment (β C5- β C6) is an antiparallel hairpin. Another point of difference is the absence in NodZ of a C-terminal helix which typically extends from the C-terminal domain and crosses to the N-terminal domain

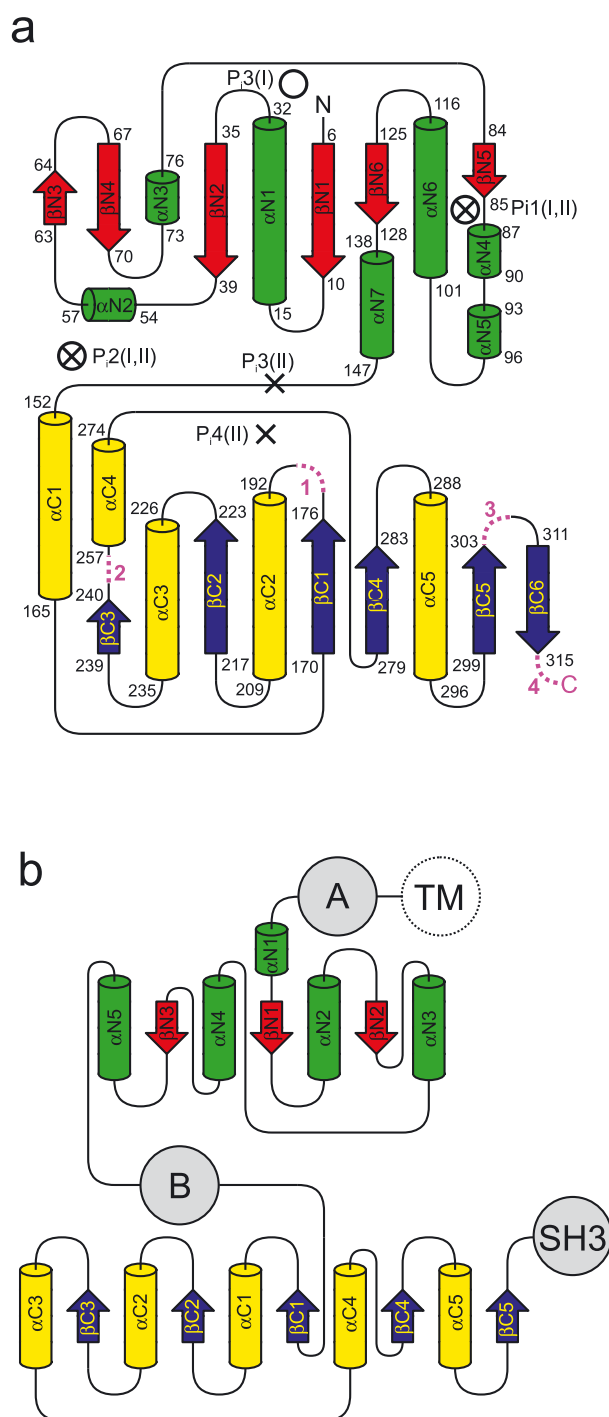


Figure 3. Topology diagrams

(a) A topology diagram of NodZ. The β -strands and helices are numbered consecutively in the N- and C-terminal domains. The inclusive residue numbers correspond to DSSP (Kabsch & Sander, 1983) secondary structure assignments. The α N3, α N4 and α N5 elements are 3_{10} -helices. The numbered dashed lines correspond to gaps in the model, as in Fig. 2a. The positions of the phosphate ions are indicated by circles (crystal form I) and by crosses (form II). (b) A topology diagram of human α 1,6-fucosyltransferase Fut8 (Ihara *et al.*, 2007). The β -strands and helices are numbered consecutively in the N- and C-terminal parts of the catalytic domain. TM, transmembrane domain absent in the crystal structure of the enzyme; SH3, C-terminal SH3 domain; A and B correspond to fragments of the sequence not included in the catalytic domain of Fut8. Topology diagrams prepared in TopDraw (Bond, 2003).

(Campbell *et al.*, 2000). Absence of this helix is also observed in the structures of sialyltransferase ST1, in Fut8, and in WaaF.

In common with other GPGTF members, the C-terminal domain of NodZ possesses the key β - α - β structure called the GPGTF motif (Wrabl & Grishin, 2001). In most structures, the N-terminal end of the α -helix interacts with the pyrophosphate group of the donor substrate to neutralize the charge using the positive pole of its dipole. This interaction occurs generally between the main-chain N-H groups of two consecutive residues of the α -helix and the α -phosphate group of the donor. In the NodZ structure, this pattern corresponds to residues Ala226 and Gln227.

The first signature of the GPGTF motif, present in several retaining and inverting GT-B enzymes, is an acidic residue located in the middle of the α -helix, which is responsible for the binding of the ribose fragment of the donor molecule in OtsA, BGT, UGT71G, WaaG, AGT, VvGT1, and MurG (Morera *et al.*, 1999; Ha *et al.*, 2000; Gibson *et al.*, 2004; Shao *et al.*, 2005; Lariviere *et al.*, 2005; Martinez-Fleitez *et al.*, 2006; Offen *et al.*, 2006). On the basis of structural alignments, this position is occupied by Asp230 in NodZ. The presence of a basic (GP) or hydrophobic (GlgA) residue in this particular position can indicate a different catalytic mechanism (Henrissat & Davies, 2000), or a differ-

ent mode of ribose binding, as predicted for GlgA, where a tyrosine is proposed to bind the ribose moiety (Cid *et al.*, 2000).

The second signature of the β - α - β motif, present in all retaining GT-Bs and in 2-epimerase but only in some inverting enzymes, is an acidic residue located between the first β -strand and the α -helix. In NodZ, this position is occupied by Asp224. Most known glycosyltransferase complex structures include an NDP ligand and the residue located in the second signature is not involved in any interactions. In the complexes of OtsA, WaaG and Fut8 with NDP-sugar ligands, the acidic residue forms an interaction with the sugar moiety of the substrate.

Two other characteristic positions of the GPGTF motif have a structural function. The third position, preceding the first β -strand, is mostly occupied by a charged residue which is always in a right-handed helical conformation and corresponds to Lys215 in NodZ. The fourth position, located in a left-handed turn of the α - β loop, is mostly occupied by glycine or aspartate. In NodZ, this position corresponds to Asp238.

Potential GDP- β -L-fucose binding site

Our attempts to crystallize a binary complex between NodZ and GDP- β -L-fucose have been un-

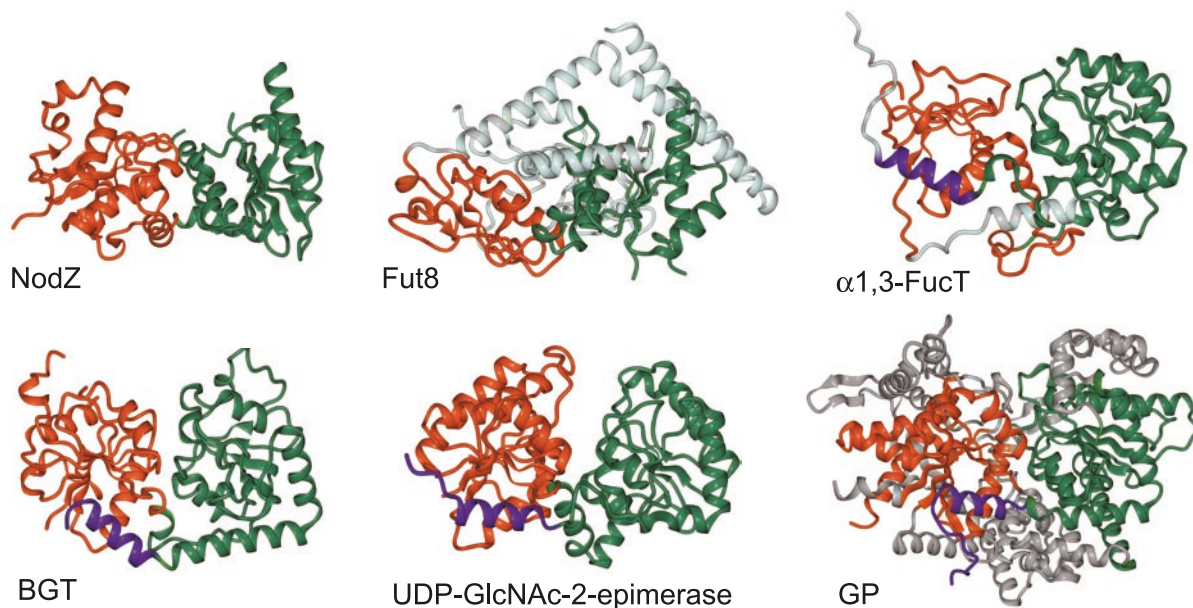


Figure 4. NodZ fucosyltransferase and selected members of the GPGTF superfamily sharing a similar, two-domain topology.

Fut8, human α 1,6-fucosyltransferase (PDB code 2DE0); α 1,3-FucT, *Helicobacter pylori* α 1,3-fucosyltransferase (2NZW); BGT, phage T4 β -glucosyltransferase (1C3J); UDP-GlcNAc-2-epimerase (1F6D); GP, glycogen phosphorylase (1A8I). The equivalent N- and C-terminal domains are colored red and green, respectively. The C-terminal α -helix reentering the N-terminal domain, present in α 1,3-FucT, BGT, 2-epimerase and in GP, but not in NodZ and Fut8, is colored blue. The additional insertions in the Fut8, α 1,3-FucT and GP structures are colored gray.

successful. By analogy to other GT-B enzymes, the binding site should be located in the C-terminal domain. Its exact location based on sequence alignment is difficult to predict due to the low identity level. Some knowledge about the donor substrate binding in other GPGTFs is helpful in some way, but the main problem is the presence of pyrimidine bases in the known complex structures of the inverting enzymes. However, some information about the GDP- β -L-fucose binding site can be deduced from phylogeny and biochemical studies of fucosyltransferases as well as from some structural features of other enzymes classified in the GPGTF superfamily.

Phylogenetic studies have revealed a relationship between the prokaryotic and eukaryotic α 1,2- and α 1,6-fucosyltransferases and also protein-O-fucosyltransferases (Breton *et al.*, 1998; Oriol *et al.*, 1999; Chazalet *et al.*, 2001; Martinez-Duncker *et al.*, 2003), with three conserved sequence motifs (Fig. 5a). Mapped onto the NodZ structure, these three motifs are located in the C-terminal domain (Fig. 5b). Mutational analysis in motif I, carried out for *Azorhizobium caulinodans* NodZ (Chazalet *et al.*, 2001) and human α 1,6-fucosyltransferase Fut8 (Takahashi *et al.*, 2000; Ihara *et al.*, 2007) showed the role of the individual amino-acid residues. In both cases, mutation of the residue corresponding to Arg177 in *Bradyrhizobium* NodZ to Ala resulted in complete loss of activity. The positively charged side chain of the Arg residue is frequently in contact with the phosphate groups of the substrate nucleotide. In the above studies, even the conservative Arg \rightarrow Lys mutation has led to a complete loss of substrate binding ability (Takahashi *et al.*, 2000; Chazalet *et al.*, 2001). The geometry of the Arg side chain, which can be a donor in five hydrogen bonds, and its different distribution of positive charge, seem to be crucial for substrate binding (Takahashi *et al.*, 2000). In several structures, the pyrimidine ring of the substrate interacts with the Arg side chain (GtfD (Mulichak *et al.*, 2004), MurG, AGT, VvGT1, UGT71G1, 2-epimerase). The role of Arg177 and its environment is also supported by substrate specificity analysis of human α 1,6-fucosyltransferase. These results show that binding of the substrate depends mostly on the guanine ring and the pyrophosphate group (Ihara *et al.*, 2006). In contrast to α 1,3-fucosyltransferases which bind different purines (Murray *et al.*, 1996), α 1,6-fucosyltransferases show high affinity only for the guanine base (Ihara *et al.*, 2006).

Residues His175, His178, Asp224, Asp270, Phe283 and Ser287 located near Arg177, are highly conserved among α 1,2-, α 1,6- and O-fucosyltransferases. However, mutagenesis of the first His residue carried out using the human enzyme had no

effect on enzymatic activity (Takahashi *et al.*, 2000) despite the high conservation level.

Mutations of the residue corresponding to His178 to Ala or Lys in the human enzyme (Takahashi *et al.*, 2000) and to Ala or Arg in *Azorhizobium* NodZ (Chazalet *et al.*, 2001) have resulted in significantly decreased activity, but the conclusions are different. In the case of human fucosyltransferase, the data suggest an interaction of this residue with the β -phosphoryl group of GDP (Takahashi *et al.*, 2000), while kinetic analysis of the *Azorhizobium* enzyme (Chazalet *et al.*, 2001) indicates an indirect role of this His residue in substrate binding and possible participation in the enzymatic mechanism or in the maintenance of active enzyme conformation.

Substitution of *Azorhizobium* NodZ Asn185 (Asn180 in *Bradyrhizobium*) by Ala dramatically decreased the enzyme activity, but substitution by Asp, present in related fucosyltransferases, completely abolished the activity. His178 and Asn180 are located in a conformationally flexible loop and could not be fully modeled in the present structures.

Mutation of the conserved residue in human Fut8 located at the second signature position of the GPGTF motif (Asp224 in *Bradyrhizobium* NodZ) and also of the Asp residue located in motif III (Asp270 in *Bradyrhizobium* NodZ) either to Ala or Asn in the human and *Azorhizobium* enzymes have resulted in complete loss of activity. In the previously described complexes of OtsA, WaaG and Fut8 with a substrate, the acidic residue located at the second position of the GPGTF motif interacts with the transferred sugar moiety of the substrate. By analogy, we postulate that Asp224 in *Bradyrhizobium* NodZ is likely to interact with the L-fucose moiety. The function of Asp270, on the other hand, presents a puzzle. Mutagenesis of this position in the human and *Azorhizobium* proteins did not explain its role in the enzymatic reaction. In the *Bradyrhizobium* NodZ structure, Asp270 is located in a C-terminal pocket of highly negative electrostatic potential (Fig. 5c).

Mutations of Fut8 residues corresponding to Phe283 and Ser287 in *Bradyrhizobium* NodZ to Ala also inactivate the enzyme. Tyr382 in Fut8 and Phe283 in *Bradyrhizobium* NodZ are located in completely different parts of the sequence. In the human enzyme, the Tyr residue precedes motif II, while in the *Bradyrhizobium* protein the Phe283 residue is located in motif III. However, superposition of the predicted substrate binding sites shows close proximity of these residues (Fig. 5d).

CONCLUSIONS

The crystal structure of the first member of the NodZ fucosyltransferase family of enzymes involved

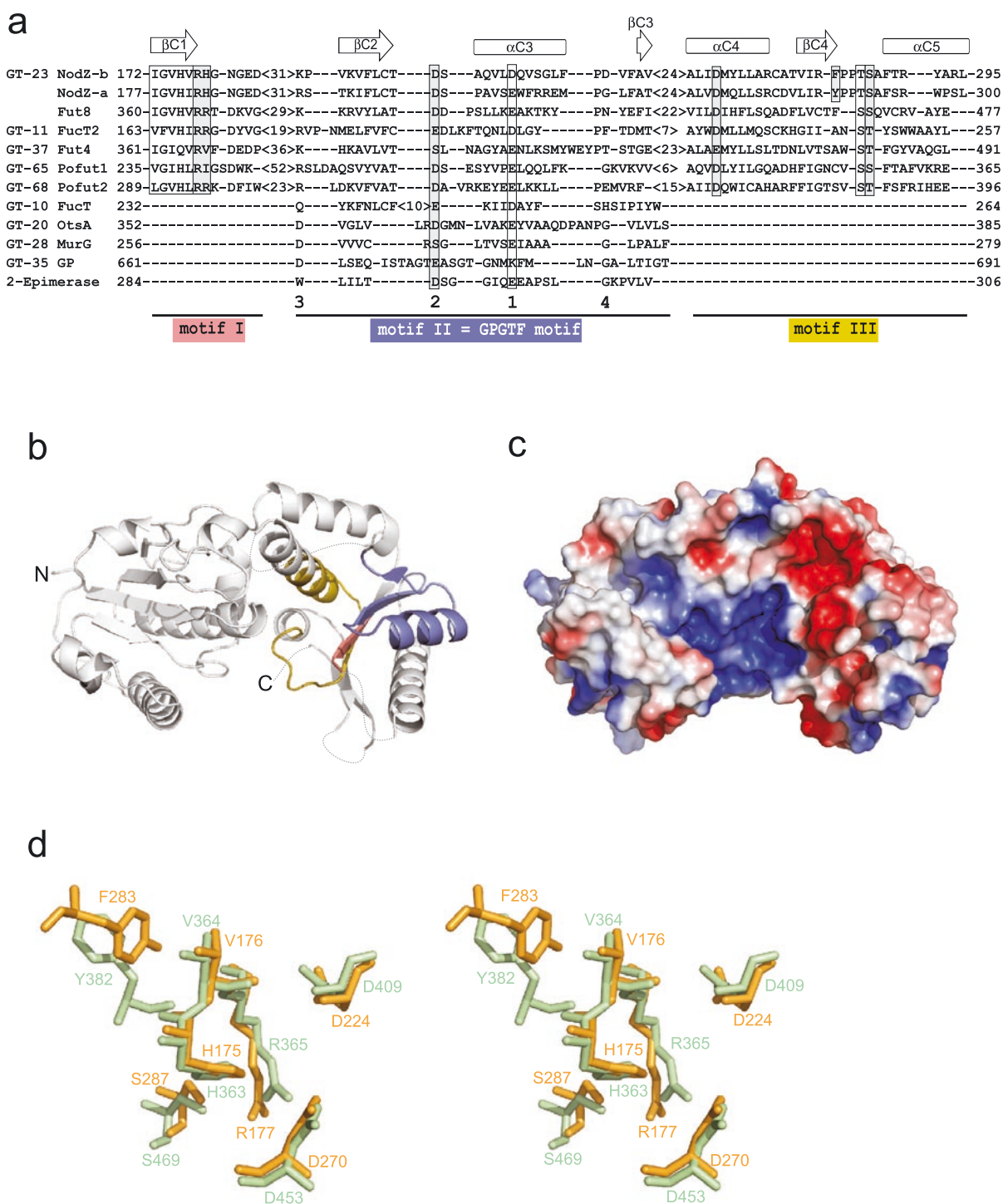


Figure 5. Conservation of sequence motifs in NodZ and other GPGTF enzymes.

(a) Multiple sequence alignment (Higgins *et al.*, 1994) and structure-based sequence alignment for the three conserved sequence motifs (I, II, III) found in α 1,2-, α 1,6-, and protein-O-fucosyltransferases, based on the structure of *Bradyrhizobium* NodZ (NodZ-b) and the following sequences retrieved from GenBank (www.ncbi.nlm.nih.gov): NodZ-a, α 1,6-FucT from *Azorhizobium caulinodans* (L18897); Fut8, α 1,6-FucT from *Homo sapiens* (D89289); FucT2, α 1,2-FucT from *Helicobacter pylori* (AF076779); Fut4, xyloglucan α 1,2-FucT from *Arabidopsis thaliana* (AF417474); Pofut1, protein-O-FucT from *Homo sapiens* (AF375884); Pofut2, protein-O-FucT from *Homo sapiens* (AJ575591). At the bottom of the table, the characteristic motif (corresponding to motif II) of the GPGTF enzymes with known structures has been added, arranged according to structural alignment with NodZ-b. The signatures 1-4 of the GPGTF motif are numbered beneath the alignment table. The following PDB coordinates have been used: NodZ-b, this work, 2HHC; Fut8, 2DE0; FucT, α 1,3-FucT from *Helicobacter pylori*, 2NZW; OtsA, 1GZ2; MurG, 1NLM; GP, 1A8I; 2-epimerase, 1F6D. GT families (left column) were ascribed according to the CAZy database (Coutinho & Henrissat, 1999; Coutinho *et al.*, 2003). The secondary structure elements (top) correspond to NodZ-b (this work). Highly conserved residues are boxed. Residues implicated in interactions with

in the biosynthesis of the Nod factor shows that, in agreement with other members of the GPGTF superfamily, the protein is folded into two similarly sized domains. However, there are significant topological differences between the NodZ fucosyltransferase and the other GPGTF enzymes as the N-terminal domain does not follow the common pattern and in fact is not folded into a Rossmann-like structure. The C-terminal domain represents a variant of the Rossmann fold with the main deviation seen in the topology of the central core, where the last (sixth) strand has an antiparallel orientation in contrast to the canonical all-parallel topology. The absence of the C-terminal α -helix, which typically extends from the C- to the N-terminal domain, is another idiosyncratic feature of the NodZ enzyme.

Phylogenetic analyses (Breton *et al.*, 1998; Oriol *et al.*, 1999; Chazalet *et al.*, 2001; Martinez-Duncker *et al.*, 2003) combined with the results of mutational studies of a related NodZ enzyme (Chazalet *et al.*, 2001) as well as comparison with the structure of human α 1,6-fucosyltransferase (Ihara *et al.*, 2007) have allowed us to postulate the residues that could be potentially involved in NodZ substrate binding. All of them are located in the C-terminal domain, within three sequence motifs that are conserved among prokaryotic and eukaryotic α 1,2- and α 1,6-fucosyltransferases. In particular, Arg177 can be implicated in binding the guanine ring of the GDP- β -L-fucose sugar donor, while Asp224 is probably involved in interactions with the β -L-fucose moiety.

Acknowledgements

This work was supported in part by a subsidy from the Foundation for Polish Science to M.J.

REFERENCES

- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**: 10037–10041.
- Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**: 311–312.
- Breton C, Oriol R, Imbert A (1998) Conserved structural features in eukaryotic and prokaryotic fucosyltransferases. *Glycobiology* **8**: 87–94.
- Brzezinski K, Rogozinski B, Stepkowski T, Bujacz G, Jaskolski M (2004) Cloning, purification, crystallization and preliminary crystallographic studies of *Bradyrhizobium* fucosyltransferase NodZ. *Acta Cryst D* **60**: 344–346.
- Campbell RE, Mosimann SC, Tanner ME, Strynadka NCJ (2000) The structure of UDP-N-acetylglucosamine 2-epimerase reveals homology to phosphoglycosyl transferase. *Biochemistry* **39**: 14993–15001.
- Charnock SJ, Davies GJ (1999) Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**: 6380–6385.
- Chazalet V, Uehara K, Geremia RA, Breton C (2001) Identification of essential amino acids in the *Azorhizobium caulinodans* fucosyltransferase NodZ. *J Bacteriol* **183**: 7067–7075.
- Cid E, Gomis RR, Geremia RA, Guinovart JJ, Ferrer JC (2000) Identification of two essential glutamic acid residues in glycogen synthase. *J Biol Chem* **275**: 33614–33621.
- Cohen GE (1997) ALIGN: a program to superimpose protein coordinates, accounting for insertions and deletions. *J Appl Cryst* **30**: 1160–1161.
- Coutinho PM, Henrissat B (1999) Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert HJ, Davies G, Henrissat B, Svensson B, eds) pp 3–12. The Royal Society of Chemistry, Cambridge.
- Coutinho PM, Deleury E, Davies GJ, Henrissat B (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol* **328**: 307–317.
- Davies GJ, Gloster TM, Henrissat B (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr Opin Struct Biol* **15**: 637–645.
- Debelle F, Moulin L, Mangin B, Denarie J, Boivin C (2001) *nod* Genes and Nod signals and the evolution of the rhizobium legume symbiosis. *Acta Biochim Polon* **48**: 359–365.
- Denarie J, Debelle F, Prome JC (1996) *Rhizobium* lipo-chitooligosaccharide nodulation factors: signaling molecules mediating recognition and morphogenesis. *Annu Rev Biochem* **65**: 503–535.
- Gibson RP, Tarling CA, Roberts S, Withers SG, Davies GJ (2004) The donor subsite of trehalose-6-phosphate synthase: binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. *J Biol Chem* **279**: 1950–1955.
- Ha S, Walker D, Shi Y, Walker S (2000) The 1.9 Å crystal structure of *Escherichia coli* MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis. *Protein Sci* **9**: 1045–1052.
- Henrissat B, Davies G (2000) Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. *Plant Physiol* **124**: 1515–1519.
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Figure 5. continued

GDP- β -L-fucose are shown on gray background. (b) *Bradyrhizobium* NodZ structure with the three motifs shown in pink (motif I), blue (II) and yellow (III), as in (a). (c) The molecular surface of *Bradyrhizobium* NodZ colored by electrostatic potential from red (negative) to blue (positive), using Adaptive Poisson-Boltzmann Solver program (Baker *et al.*, 2001) and visualized in PyMol (DeLano Scientific LLC, <http://pymol.sourceforge.net>). (d) Stereoview of the superposition of residues from the potential active sites of NodZ (orange) and Fut8 (green). The superposition, based on Ca atoms only, was calculated in ALIGN (Cohen, 1997).

- Holm L, Sander C (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**: 316–319.
- Ihara H, Ikeda Y, Taniguchi N (2006) Reaction mechanism and substrate specificity for nucleotide sugar of mammalian α 1,6-fucosyltransferase — a large-scale preparation and characterization of recombinant human FUT8. *Glycobiology* **16**: 333–342.
- Ihara H, Ikeda Y, Toma S, Wang X, Suzuki T, Gu J, Miyoshi E, Tsukihara T, Honke K, Matsumoto A, Nakagawa A, Taniguchi N (2007) Crystal structure of mammalian α 1,6-fucosyltransferase, FUT8. *Glycobiology* **17**: 455–466.
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**: 2577–2637.
- Lariviere L, Sommer N, Morera S (2005) Structural evidence of a passive base-flipping mechanism for AGT, an unusual GT-B glycosyltransferase. *J Mol Biol* **352**: 139–150.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**: 283–291.
- Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst Biol* **54**: 575–594.
- Ma B, Simala-Grant JL, Taylor DE (2006) Fucosylation in prokaryotes and eukaryotes. *Glycobiology* **16**: 158R–184R.
- Martinez-Duncker I, Michalski JC, Bauvy C, Candelier JJ, Mennesson B, Codogn P, Oriol R, Mollicone R (2003) A new superfamily of protein-O-fucosyltransferases, α 2-fucosyltransferases, and α 6-fucosyltransferases: phylogeny and identification of conserved peptide motifs. *Glycobiology* **13**: 1C–5C.
- Martinez-Fleites C, Proctor M, Roberts S, Bolam DN, Gilbert HJ, Davies GJ (2006) Insights into the synthesis of lipopolysaccharide and antibiotics through the structures of two retaining glycosyltransferases from family GT4. *Chem Biol* **13**: 1143–1152.
- McRee DE (1999) XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**: 156–165.
- Mergaert P, D'Haese W, Fernandez-Lopez M, Geelen D, Goethals K, Prome JC, van Montagu M, Holsters M (1996) Fucosylation and arabinosylation of Nod factors in *Azorhizobium caulinodans*: involvement of nolK, nodZ as well as noeC and/or downstream genes. *Mol Microbiol* **21**: 409–419.
- Morera S, Imberty A, Aschke-Sonnenborn U, Ruger W, Freemont PS (1999) T4 phage β -glucosyltransferase: substrate binding and proposed catalytic mechanism. *J Mol Biol* **292**: 717–730.
- Morris RJ, Zwart PH, Cohen S, Fernandez FJ, Kakaris M, Kirillova O, Vonrhein C, Perrakis A, Lamzin VS (2004) Breaking good resolutions with ARP/wARP. *J Synchrotron Rad* **11**: 56–59.
- Mulichak AM, Losey HC, Walsh CT, Garavito RM (2001) Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure* **9**: 547–557.
- Mulichak AM, Lu W, Losey HC, Walsh CT, Garavito RM (2004) Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry* **43**: 5170–5180.
- Murray BW, Takayama S, Schultz J, Wong CH (1996) Mechanism and specificity of human α -1,3-fucosyltransferase V. *Biochemistry* **35**: 11183–11195.
- Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst D* **53**: 240–255.
- Ni L, Sun M, Yu H, Chokhawala H, Chen X, Fisher AJ (2006) Cytidine 5'-monophosphate (CMP)-induced structural changes in a multifunctional sialyltransferase from *Pasteurella multocida*. *Biochemistry* **45**: 2139–2148.
- Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J* **25**: 1396–1405.
- Olsthoorn MM, Lopez-Lara IM, Petersen BO, Bock K, Haverkamp J, Spaink HP, Thomas-Oates JE (1998) Novel branched nod factor structure results from α -(1 \rightarrow 3) fucosyl transferase activity: the major lipo-chitin oligosaccharides from *Mesorhizobium loti* strain NZP2213 bear an α -(1 \rightarrow 3) fucosyl substituent on a nonterminal backbone residue. *Biochemistry* **37**: 9024–9032.
- Oriol R, Mollicone R, Cailleau A, Balanzino L, Breton C (1999) Divergent evolution of fucosyltransferases genes from vertebrates, invertebrates and bacteria. *Glycobiology* **9**: 323–334.
- Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**: 307–326.
- Panjikar S, Parthasarathy V, Lamzin VS, Weiss MS, Tucker PA (2005) Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Cryst D* **61**: 449–457.
- Quesada-Vincens D, Fellay R, Nasim T, Viprey V, Burger U, Prome JC, Broughton WJ, Jabbouri S (1997) *Rhizobium* sp. strain NGR234 NodZ protein is a fucosyltransferase. *J Bacteriol* **179**: 5087–5093.
- Quinto C, Wijffes AH, Bloemberg GV, Blok-Tip L, Lopez-Lara IM, Lugtenberg BJ, Thomas-Oates JE, Spaink HP (1997) Bacterial nodulation protein NodZ is a chitin oligosaccharide fucosyltransferase which can also recognize related substrates of animal origin. *Proc Natl Acad Sci USA* **94**: 4336–4341.
- Raymond J, Siefert JL, Staples CR, Blankenship RE (2004) The natural history of nitrogen fixation. *Mol Biol Evol* **21**: 541–554.
- Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Cryst D* **58**: 1772–1779.
- Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell* **17**: 3141–3154.
- Sheldrick GM (2002) Macromolecular phasing with SHELXE. *Kristallogr* **217**: 644–650.
- Stacey G, Luka S, Sanjuan J, Banfalvi Z, Nieuwkoop AJ, Chun JY, Forsberg LS, Carlson R (1994) NodZ, a unique host-specific nodulation gene, is involved in the fucosylation of the lipooligosaccharide nodulation signal of *Bradyrhizobium japonicum*. *J Bacteriol* **176**: 620–633.
- Stepkowski T, Legocki AB (2001) Reduction of bacterial genome size and expansion resulting from obligate intracellular lifestyle and adaptation to soil habitat. *Acta Biochim Polon* **48**: 367–384.
- Stepkowski T, Swiderska A, Miedzinska K, Czaplinska M, Swiderski M, Biesiadka J, Legocki AB (2003) Low sequence similarity and gene content of symbiotic clusters of *Bradyrhizobium* sp. WM9 (*Lupinus*) indicate early

- divergence of "lupin" lineage in the genus *Bradyrhizobium*. *Antonie Van Leeuwenhoek* **84**: 115–124.
- Sun HY, Lin SW, Ko TP, Pan JF, Liu CL, Lin CN, Wang AH, Lin CH (2007) Structure and mechanism of *Helicobacter pylori* fucosyltransferase. A basis for lipopolysaccharide variation and inhibitor design. *J Biol Chem* **282**: 9973–9982.
- Takahashi T, Ikeda Y, Tateishi A, Yamaguchi Y, Ishikawa M, Taniguchi N (2000) A sequence motif involved in the donor substrate binding by α 1,6-fucosyltransferase: the role of the conserved arginine residues. *Glycobiology* **10**: 503–510.
- Vagin A, Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Cryst* **30**: 1022–1025.
- Van Duyne GD, Standaert RF, Karplus PA, Schreiber SL, Clardy J (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229**: 105–124.
- Vrielink A, Ruger W, Driessen HP, Freemont PS (1994) Crystal structure of the DNA modifying enzyme β -glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J* **13**: 3413–3422.
- Winn MD, Isupov MN, Murshudov GN (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst D* **57**: 122–133.
- Wrabl JO, Grishin NV (2001) Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J Mol Biol* **314**: 365–374.