Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Impact of natural variation in bacterial F17G adhesins on crystallization behaviour

Lieven Buts,[a] Adinda Wellens,[a] Inge Van Molle,[a] Lode Wyns,[a] Remy Loris,[a] Martina Lahmann,[b] Stefan Oscarson,[c] Henri De Greve[a] and Julie Bouckaert[a]*

[a] Laboratorium voor Ultrastructuur, Vlaams Interuniversitair Instituut voor Biotechnologie and Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium, [b]Institutionen för Kemi, pl. 8, Kemivägen 10, S-412 96 Göteborg, Sweden, and [c]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, Sweden

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

Received 27 January 2005
Accepted 27 May 2005

PDB References: F17a-G–GlcNAc, 2bsc, r2bscsf; F17a-G–GlcNAc(β1-)OMe, 1zpl, r1zplsf; F17b-G–chitobiose, 2bs7, r2bs7sf; F17b-G–GlcNAc, 2bs8, r2bs8sf; F17e-G–GlcNAc, 2bsb, r2bsbsf; F17f-G–GlcNAc, 1zk5, r1zk5sf.

Since the introduction of structural genomics, the protein has been recognized as the most important variable in crystallization. Recent strategies to modify a protein to improve crystal quality have included rationally engineered point mutations, truncations, deletions and fusions. Five naturally occurring variants, differing in 1–18 amino acids, of the 177-residue lectin domain of the F17G fimbrial adhesin were expressed and purified in identical ways. For four out of the five variants crystals were obtained, mostly in non-isomorphous space groups, with diffraction limits ranging between 2.4 and 1.1 Å resolution. A comparative analysis of the crystal-packing contacts revealed that the variable amino acids are often involved in lattice contacts and a single amino-acid substitution can suffice to radically change crystal packing. A statistical approach proved reliable to estimate the compatibilities of the variant sequences with the observed crystal forms. In conclusion, natural variation, universally present within prokaryotic species, is a valuable genetic resource that can be favourably employed to enhance the crystallization success rate with considerably less effort than other strategies.

1. Introduction

Homologous proteins from other species are often considered if suitable crystals cannot be obtained for the protein under study. This idea was pioneered more than 50 years ago (Kendrew *et al.*, 1954; Campbell *et al.*, 1972) and has been revitalized in the era of structural genomics. Potential drawbacks of the structural genomics approach are that the homologous proteins may be rather poor structural models for each other and that the cloning, expression and purification of the homologues may require significant changes to the established protocols. An alternative strategy is therefore to stick with the target protein and to employ site-directed mutations to alter its surface properties and crystal-packing possibilities. Other approaches, such as truncations, deletions, fusions, limited proteolysis and deglycosylation of proteins, have become very popular and aim to produce better diffracting crystals through improving protein solubility, decreasing inherent molecular flexibility and increasing molecular packing possibilities. These methods still require the successful engineering of a soluble and stable protein that maintains a native structure. Intraspecies variation in the genes, best known in prokaryotes, represents a naturally available alternative that has already been selected for success. We set out to utilize the natural variation of the *Escherichia coli* F17G fimbrial adhesin to enhance the chance of obtaining well diffracting crystals without the need to change cloning, expression and purification strategies.

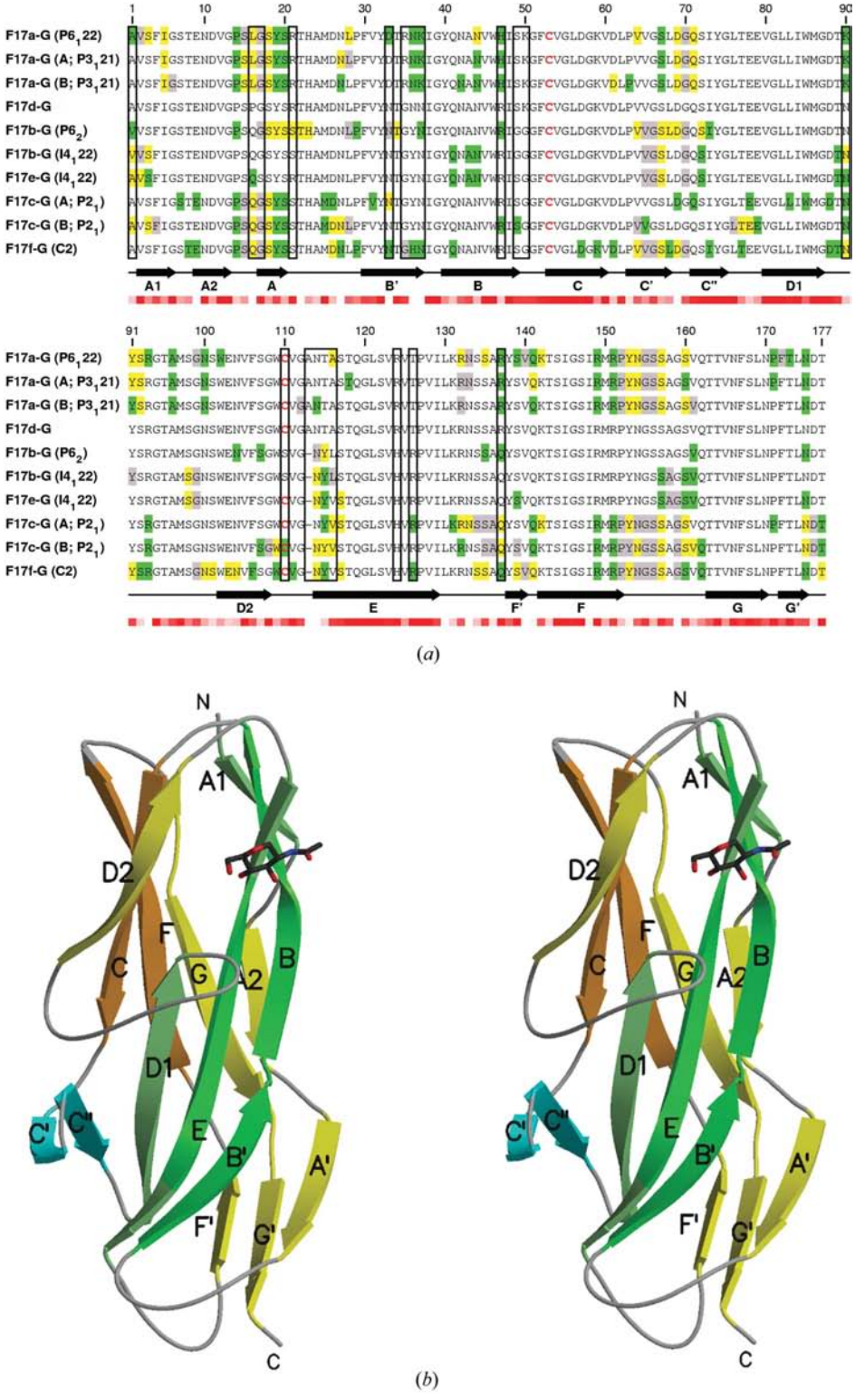

Figure 1
(*a*) Alignment of the amino-acid sequences of variants a–f of the F17G lectin domain and (*b*) stereoview of the crystal structure of the F17f-G lectin domain with labelling of the β-strands. Amino-acid differences are marked with boxes. Residues involved in crystal-packing contacts are marked in grey when only main-chain atoms are involved in the packing, in green when only side-chain atoms are involved and in yellow when both main-chain and side-chain atoms contribute to the contact. Secondary-structure elements are labelled as in the structure shown and the solvent accessibility of each residue is shown by means of a colour gradient ranging from dark red for completely buried residues to white for maximally exposed residues. Only sequence information is available for variant d.

Pathogenic bacteria use hair-like adhesive organelles called fimbriae or pili to establish an initial foothold in host colonization. F17 fimbriae are one of the virulence factors found on enterotoxigenic and septicaemic *E. coli* strains affecting ruminants. They are flexible 3 nm wide wire-like structures with an overall morphology and carbohydrate specificity that is clearly distinct from the type 1 pili (Hung *et al.*, 2002) and P pili (Dodson *et al.*, 2001) of uropathogenic *E. coli* strains. F17 pili, like type 1 and P pili, are assembled *via* the chaperone/usher pathway (Soto & Hultgren, 1999). The adhesins of the chaperone/usher pathway are two-domain proteins consisting of an amino-terminal lectin domain linked to a carboxy-terminal pilin domain. The binding of F17-fimbriated bacterial cells to intestinal villi depends on the sugar-binding activity of the F17G adhesin, located at the tip of the pilus. This interaction can be inhibited by *N*-acetylglucosamine (GlcNAc; Girardeau, 1980; Bertin *et al.*, 1996).

The F17 fimbriae family comprises various members (Fig. 1): the F17a and F17d fimbriae, produced by bovine enterotoxigenic *E. coli* (ETEC) strains (Lintermans *et al.*, 1988; Bertels *et al.*, 1989), the F17b fimbriae, expressed by *E. coli* strains isolated from septicaemic calves and lambs (El Mazouari *et al.*, 1994), the F17c fimbriae, associated with bovine diarrhoea or septicaemia and with lambs showing nephrosis (Bertin *et al.*, 1996), and F17e (*E. coli* strain CK210) and F17f (*E. coli* CK377 strain) fimbriae expressed by non-enterotoxigenic *E. coli* isolated from lambs and goat kids (Cid *et al.*, 1999). G fimbriae from human uropathogenic *E. coli* strains (Saarela *et al.*, 1995) were found to be identical in amino-acid sequence to F17c fimbriae (Martin *et al.*, 1997). All amino-acid substitutions are located in the amino-terminal lectin domain, where they may fine-tune the specificity for the host's carbohydrate receptors. On the other hand, the carboxy-terminal pilin domain of the

F17G family, which joins the adhesin non-covalently to the rest of the fimbria, is identical in all six variants. The F17a-G and F17d-G adhesins vary in six amino-acid positions, while F17b-G and F17e-G adhesins only vary in five positions. The F17e-G and F17f-G adhesins differ in only one amino acid from the F17c-G adhesin, Arg21Ser and His36Tyr, respectively. The largest difference (18 amino-acid substitutions) is observed between the F17a-G and the F17b-G adhesins.

The amino-acid variations occur predominantly on the surface of F17G surrounding a conserved binding pocket for *N*-acetylglucosamine (Buts, Bouckaert *et al.*, 2003). A more detailed characterization of the structural and biochemical properties of the different natural variants of the F17G lectin domain is of interest for a better understanding of the diversity in host specificity of different *E. coli* isolates. The crystal structures of variants a (F17a-G; Buts, Bouckaert *et al.*, 2003) and c (F17c-G or GafD; Merckel *et al.*, 2003) of the F17G lectin domain in complex with *N*-acetylglucosamine (GlcNAc) have been determined previously. We set out to crystallize the four other identified variants of F17G adhesin (b, d, e and f) and studied the effect of the natural sequence variation on the crystallization behaviour of the F17G lectin domains.

2. Materials and methods

2.1. Products

GlcNAc, GlcNAc(β1–4)GlcNAc and GlcNAc(β1–2)Man were purchased from Dextra Laboratories. GlcNAc(β1–3)Gal (Dahmen *et al.*, 1985) and GlcNAc(β1-)OMe were synthesized.

2.2. Expression and purification

The lectin domains of the variants F17a-G, F17b-G, F17c-G, F17d-G, F17e-G and F17f-G (Lintermans *et al.*, 1991; El Mazouari *et al.*, 1994; Cid *et al.*, 1999) were amplified using the primers F17-35 (5′-ACAAATTTTTATAAGGTCTTTCTG-GCTGTATTC-3′) and F17-36 (5′-CCCGAATTCTATGTGT-CATTCA GCGTAAATGGATT-3′) and cloned under the control of the T7 promoter. The resulting constructs were transformed into *E. coli* strain BL21-AI (Invitrogen). The BL21-AI cells were grown in Luria–Bertani (LB) medium to OD_{600} = 0.5 and induced with 0.2% L-arabinose. Cells were subjected to osmotic shock and the periplasmic extract was loaded onto a GlcNAc–agarose affinity column (Sigma) in buffer *A* (20 m*M* TES buffer pH 7). Bound protein was eluted with buffer *A* containing 200 m*M* GlcNAc. The protein was concentrated to 8–12 mg ml^{-1}, further purified by gel filtration (Superdex 75 HR, Amersham Biosciences) and again concentrated to about 10 mg ml^{-1} in 20 m*M* Tris pH 8 with 150 m*M* NaCl prior to crystallization.

In the case of F17a-G, two peaks eluted from the gel-filtration column. Analysis on a 12.5% SDS–PAGE revealed that both peaks contained pure F17a-G protein. Only the second peak was observed for the other F17G variants b, d, e and f. The ratio of the two peaks varied between different purification rounds and depending on the age of the protein

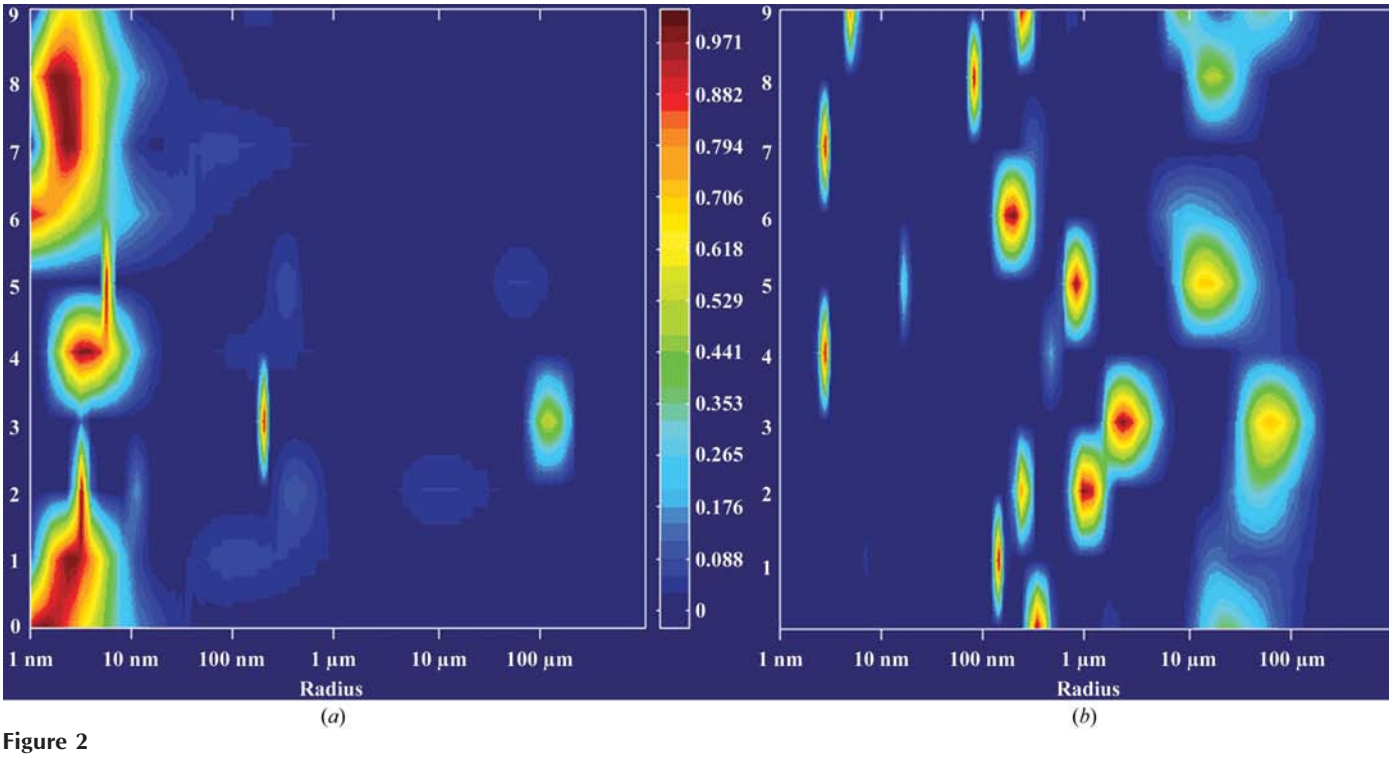

Figure 2
DLS size-distribution diagrams of F17a-G showing (*a*) a monodisperse pattern for the second eluted peak from gel filtration and (*b*) large aggregates in the first eluted peak. A similar pattern as in (*b*) is obtained upon aging of the F17a-G solution used in (*a*). The hydrodynamic radius is displayed on the *x* axis and each of the ten measurements along the *y* axis. The intensities for the distributions are indicated by the colours blue (absent) to dark red (most prominent).

Table 1
Crystallization conditions.

Variant	Compatible carbohydrate ligands†	Crystal form	Theoretical pI‡	Protein concentration (mg ml^{-1})	Precipitant
F17a-G	GlcNAc, GlcNAc(β1–2)Man, GlcNAc(β1–3)Gal, GlcNAc(β1–4)GlcNAc, GlcNAc(β1-)OMe	$P6_122$	8.6	5.5	10%(*v*/*v*) 2-propanol, 20%(*w*/v) PEG 4000, 100 m*M* HEPES pH 7.5
F17a-G	Sugar-free, GlcNAc, GlcNAc(β1-)SeMe	$P6_122$	8.6	10	30%(*w*/*v*) PEG 4000, 100 m*M* NaAc pH 4.6, 100 m*M* NH_4Ac (Buts, Bouckaert *et al.*, 2003)
F17a-G	GlcNAc(β1-)OMe	$P3_121$	8.6	5.5	10%(*v*/*v*) 2-propanol, 20%(*w*/*v*) PEG 4000, 100 m*M* HEPES pH 7.5
F17b-G	GlcNAc, GlcNAc(β1–2)Man, GlcNAc(β1–3)Gal, GlcNAc(β1–4)GlcNAc	$P6_2$	5.5	16	1.2 *M* Li_2SO_4, 10 m*M* $NiCl_2$, 100 m*M* Tris pH 8.5
F17b-G	GlcNAc	$I4_122$	5.5	16	10% ethanol, 1.5 *M* NaCl
F17c-G	GlcNAc	$P2_1$	5.5	10–30	10%(*w*/*v*) PEG 6000, 100 m*M* HEPES pH 7.5, 5%(*v*/*v*) MPD (Merckel *et al.*, 2003)
F17e-G	GlcNAc	$I4_122$	6.1	11	30% PEG 400, 0.1 *M* $CdCl_2$, 100 m*M* NaAc pH 4.6
F17f-G	GlcNAc	$C2$	5.8	12	20%(*v*/*v*) 2-propanol, 20% PEG 4000, 100 m*M* Na citrate pH 5.6

† Co-crystals with specific carbohydrate in the presence of 5–10 m*M* sugar; 226 m*M* GlcNAc was used for the preparation of F17c-G crystals (Merckel *et al.*, 2003). ‡ The calculated pI of F17d-G is 8.0.

solution and possibly also on the concentration step prior to gel filtration. Both peaks of the F17a-G protein preparation were analysed using dynamic light scattering (DLS; Fig. 2). Following centrifugation of the samples at 18000*g* for 60 min at 277 K, 30 µl of sample was injected into a flow cell of a Laser-Spectroscatter 210 instrument (RiNA Netzwerk RNA-Technologien, Berlin, Germany) at room temperature and illuminated by a 30 mW, 660 nm wavelength solid-state laser. Data were collected during ten measurements of 20 s each.

2.3. Crystallization and data collection

Crystallization conditions were initially screened using the Crystal Screen, Crystal Screen II, Crystal Screen Cryo (Hampton Research) and Wizard Screen I and II (Emerald BioSystems) kits. When necessary, initial conditions were further optimized by variations in pH, precipitant concentration and protein concentration in order to obtain crystals suitable for data collection (Table 1). Competition soaks with F17b-G crystals were performed by first washing an F17b-G crystal in its mother liquor in the absence of any sugar and subsequently transferring it to a drop of precipitant solution enriched with 50–100 m*M* of the desired sugar. Soak times were 1 h or longer.

Data were collected at 100 K on EMBL beamlines X11, X13, BW7A and BW7B of the DESY synchrotron (Hamburg, Germany) and on beamline ID14-1 of the ESRF synchrotron (Grenoble, France) (Fig. 3). Crystals of variants a and b were transferred to mother liquor enriched to a concentration of 30% 2-propanol for a few minutes and subsequently frozen directly in the cold stream. Variants e and f were frozen in concentrated precipitant solution. Data for variants b, e and f were collected using a single resolution pass. Typically, multiple passes with different exposure times were needed for

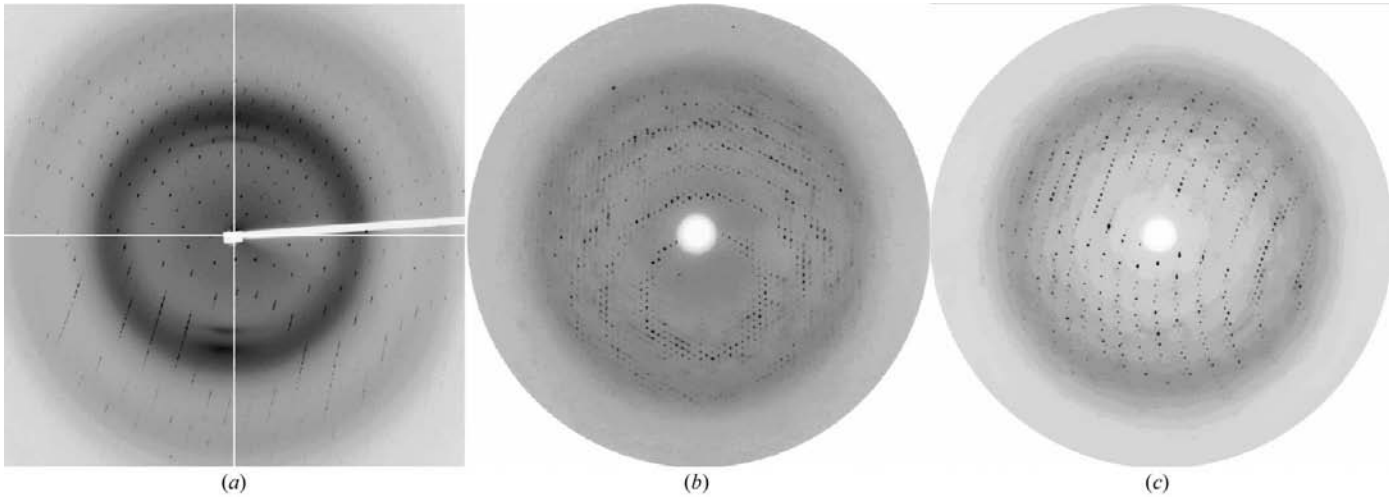

Figure 3
Diffraction patterns of (*a*) F17a-G in complex with GlcNAc(β1-)OMe extending to 1.7 Å resolution, (*b*) F17b-G in complex with GlcNAc extending to 2.3 Å resolution and (*c*) F17e-G in complex with GlcNAc extending to 2.4 Å resolution.

Table 2
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Variant	Space group	Unit-cell parameters (Å, °)	Resolution range (Å)	R_{merge}	Completeness (%)	$I/\sigma(I)$	Redundancy	Beamline	PDB code
F17a-G	$P6_122$	$a = b = 42.5$, $c = 272.9$	36–1.10 (1.14–1.10)	0.069 (0.781)	99.5 (98.3)	16 (3)	10.0 (5.8)	BW7B	2bsc
F17a-G	$P3_121$	$a = b = 42.8$, $c = 288.2$	20–1.70 (1.76–1.70)	0.058 (0.335)	90.2 (70.0)	18 (2.5)	5.0 (1.8)	ID14-1	1zpl
F17b-G	$P6_2$	$a = b = 86.7$, $c = 57.0$	35–2.10 (2.18–2.10)	0.040 (0.407)	99.6 (99.8)	29 (5.5)	9.3 (6.4)	X13	2bs7
F17b-G	$I4_122$	$a = b = 96.5$, $c = 97.7$	23–2.24 (2.32–2.24)	0.120 (0.209)	96.8 (91.1)	10 (5)	5.1 (4.5)	X11	2bs8
F17c-G	$P2_1$	$a = 43.0$, $b = 70.5$, $c = 56.1$, $\beta = 104.5$	20–1.70	See Merckel *et al.* (2003)					1oio
F17e-G	$I4_122$	$a = b = 96.7$, $c = 96.2$	15–2.40 (2.49–2.40)	0.071 (0.79)	99.7 (100.0)	23 (4.5)	14.5 (14.7)	X11	2bsb
F17f-G	C2	$a = 73.7$, $b = 42.0$, $c = 53.4$, $\beta = 97.5$	21–1.38 (1.43–1.38)	0.048 (0.242)	99.7 (99.0)	16 (5)	3.9 (3.6)	BW7A	1zk5

F17a-G crystals to avoid overloads while still collecting the data to the highest possible resolution. All data were integrated using *DENZO* and scaled and merged using *SCALEPACK* (Otwinowski & Minor, 1997). Molecular-replacement calculations were performed with *AMoRe* (Navaza, 2001) and *PHASER* (Read, 2001; Storoni *et al.*, 2004) using the refined structure of GlcNAc-bound F17a-G (1.65 Å resolution structure; PDB code 1o9w) as the search model (Buts, Bouckaert *et al.*, 2003).

2.4. Crystal-packing analysis

Prior to comparative crystal-packing analysis, a representative with the best data quality was chosen for each F17G crystal form (Table 2). These include the 1.1 Å resolution structure of variant a bound to GlcNAc (PDB code 2bsc), the 1.7 Å resolution co-crystal structure of variant a bound to GlcNAc(β1-)OMe (PDB code 1zpl), F17b-G bound to GlcNAc(β1–4)GlcNAc in the 2.1 Å resolution structure in space group $P6_2$ (PDB entry 2bs7), the F17c-G–GlcNAc complex (PDB code 1ioi), the structures of variants b and e in space group $I4_122$ (PDB codes 2bs8 and 2bsb, respectively) and the 1.4 Å resolution structure of variant f (PDB code 1zk5). The atomic positions and temperature factors of these structures were fully refined using *CNS* (Brünger *et al.*, 1998). For each of the F17G–carbohydrate complexes reported in Table 1, the carbohydrate moiety was clearly defined by the electron density. Surface areas buried through crystal packing were calculated using *NACCESS* (Hubbard & Thornton, 1993; Table 3). The packing contacts in the series of F17G crystals were analyzed using *PyMOL* (DeLano, 2002) and the *CCP*4 program *CONTACT* (Collaborative Computational Project, Number 4, 1994), using an intermolecular cutoff distance of 4 Å. Subsequently, these interactions were carefully inspected both visually with the graphical program *TURBO* (Roussel & Cambillau, 1989) and analytically by means of the *CCP*4 program *SFCHECK*. The total number of atoms involved in packing contacts was calculated and the number of atoms in packing contacts affected by the sequence differences with the other F17G variants is given as an indication of the relative compatibilities of the crystal forms in Table 3. In those crystal forms containing more than one molecule per asymmetric unit, as is the case in the trigonal crystal form of F17a-G and in the F17c-G crystals, the packing of the separate F17G molecules was analyzed to allow a direct comparison between crystal packings.

3. Results

3.1. Crystallization and packing analysis of F17a-G

The crystallization conditions of variants a (residues 1–177; Buts, Bouckaert *et al.*, 2003) and c (GafD residues 1–178; Merckel *et al.*, 2003) of the F17G lectin domain have been described previously (Table 1). Attempts to crystallize the F17G lectin domains of variants b, d, e and f using the crystallization conditions described for F17a-G were unsuccessful. *Ab initio* screening for crystallization conditions resulted in crystals of variants b, e and f, whereas no diffraction-quality crystals could be obtained for variant d. Despite the long *c* axis (Table 2), F17a-G crystals have the potential to diffract to atomic resolution. Diffraction has been observed to 1.1 Å resolution for the GlcNAc complex, although most crystals diffract to about 1.7 Å resolution (Fig. 3*a*). This is consistent with the rather dense packing of the crystals (V_M values of 1.8–2.0 Å^3 Da^{-1}, corresponding to a solvent content of around 35%).

Interestingly, crystallization of F17a-G was speeded up dramatically by the addition of the first-eluted peak from the gel filtration to the second peak that eluted more in accordance with the molecular weight of the F17a-G lectin domain. Crystal formation could be seen immediately upon setting up hanging drops through the glittering of the forming crystals and in less than a day the crystals were fully formed. In contrast, it took several days to several weeks to obtain crystals when only F17a-G from the second peak was used for crystallization. The DLS experiments show that the first peak consisted of F17a-G aggregates, while the second peak contains monodisperse and probably monomeric F17a-G (Fig. 2*a*). Protein solution kept at 277 K aggregates over time to convert to a similar size-distribution pattern as for the first peak (Fig. 2*b*). Crystallization of an aged protein solution from the second peak does not result in sufficiently large single crystals but only gives large quantities of microcrystals.

Table 3
Statistical analysis of crystal packing and compatibilities of the variant F17G sequences with the observed F17G crystal forms.

			No. of atoms involved in packing contacts in an observed crystal form for F17G (rows) that is affected by introducing a variant F17G sequence a–f (columns) into this crystal form					
Crystal form	Surface area buried in crystal packing (Å^2)	No. of atoms involved in packing contacts	F17a-G	F17b-G	F17c-G	F17d-G	F17e-G	F17f-G
F17a-G ($P6_122$)	2575	145	—	20	19	14	21	19
F17a-G ($P3_121$, mol. *A*)	2355	140	—	15	15	12	16	15
F17a-G ($P3_121$, mol. *B*)	2358	123	—	16	16	14	18	16
F17b-G ($P6_2$)	1570	92	19	—	3	9	9	3
F17b-G ($I4_122$)	1015	48	10	—	3	7	3	3
F17c-G ($P2_1$, mol. *A*)	2135	160	20	3	—	13	2	0
F17c-G ($P2_1$, mol. *B*)	2360	155	16	6	—	9	2	0
F17e-G ($I4_122$)	1180	60	9	10	0	7	—	0
F17f-G (*C*2)	2915	190	25	6	3	13	5	—

Aggregation thus correlates very well with the protein's nucleation and crystallization behaviour (Kadima *et al.*, 1990).

Crystals of the complexes of F17a-G with specific disaccharides and monosaccharides, GlcNAc(β1–2)Man, GlcNAc(β1–3)Gal, GlcNAc(β1–4)GlcNAc, GlcNAc and GlcNAc(β1-)OMe, and sugar-free F17a-G were all obtained using the same precipitant (Table 1). A carbohydrate larger than a disaccharide would almost certainly interfere with the crystal packing of F17a-G. Although the crystallization conditions differ significantly from those previously published, both in composition and pH (Table 1), the F17a-G crystals have the same space-group symmetry ($P6_122$) and unit-cell parameters. An exception is the complex of F17a-G with GlcNAc(β1-)OMe, which also occurs in crystals with space group $P3_121$ despite having identical crystallization conditions as for the $P6_122$ crystals (Table 1). The reduction of the sixfold to threefold symmetry results in the presence of two independent molecules in the asymmetric unit. Their packing is mediated in essence by the same residues as in space group $P6_122$ (Fig. 1), but a slightly lower number of atomic contacts and a smaller surface area buried in crystal packing can be noted (Table 3).

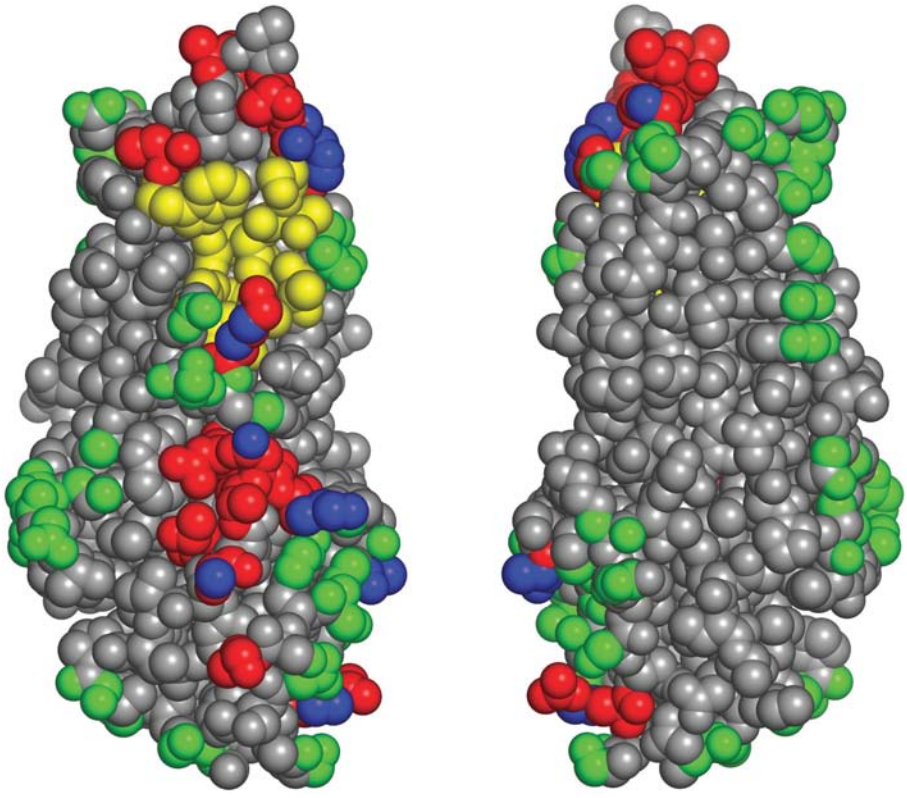

Figure 4
Incompatibility of the F17b-G amino-acid sequence with the crystal lattice interactions of the F17a-G variant in space group $P6_122$. Shown are two orientations (180° apart) of a CPK surface of the lectin domain of F17a-G. Atoms involved in crystal lattice interactions in F17a-G are coloured red. Side-chain atoms from residues that differ between F17a-G and F17b-G are indicated in green. Atoms from side chains both involved in crystal packing and differing between F17a-G and F17b-G are blue. Residues involved in monosaccharide binding are indicated in yellow.

The analysis of crystal packing and sequence compatibilities indicates that the $P6_122$ crystal lattice of F17a-G is least compatible with other F17G lectin-domain sequences (Table 3). This is further exemplified by overlaying the sequence of variant b, which is most different from variant a, onto the F17a-G structure (Fig. 4). Between variant a and all other variants, Lys37 and Lys90 change to asparagines, His47 to an arginine, Asp33 to an asparagine and Arg35 to glycine (Fig. 1). All these residues except Arg35 have side chains involved in crystal packing in the F17a-G crystals. We demonstrate that the replacement of His47 by an arginine (Fig. 5*a*) and of Lys37 (Fig. 5*b*) and Lys90 by asparagines in all variants but F17a-G are incompatible with the crystal packing for F17a-G. The His47Arg substitution alone would suffice to disallow the packing of the $P6_122$ crystals for any of the other variants (Fig. 5*a*). The Arg47 side chain cannot be modelled in a low-energy

conformation that does not clash either internally or with a symmetry-related molecule in the F17a-G crystal form, apart from the obvious electrostatic repulsion. Since Lys37 and Lys90 contribute significantly to the crystal lattice contact in F17a-G crystals (Fig. 5*b*), their substitution results in the disruption of the packing interactions with Asn153 and Asp69, respectively. In contrast, the buried lysine residue at position 50 is not involved in the crystal packing of F17a-G.

3.2. Crystallization and crystal packing of F17b-G

Crystals of the lectin domain of the F17b-G variant were obtained in complex with GlcNAc in a hexagonal and a tetragonal crystal form (Table 1). Co-crystals of F17b-G with GlcNAc(β1–2)Man were obtained with the $P6_2$ space group. In both crystal forms there is one molecule in the asymmetric unit and the V_M values are around 3.4 and 3.0 Å^3 Da^{-1}, respectively (65 and 60% solvent content). The crystals have

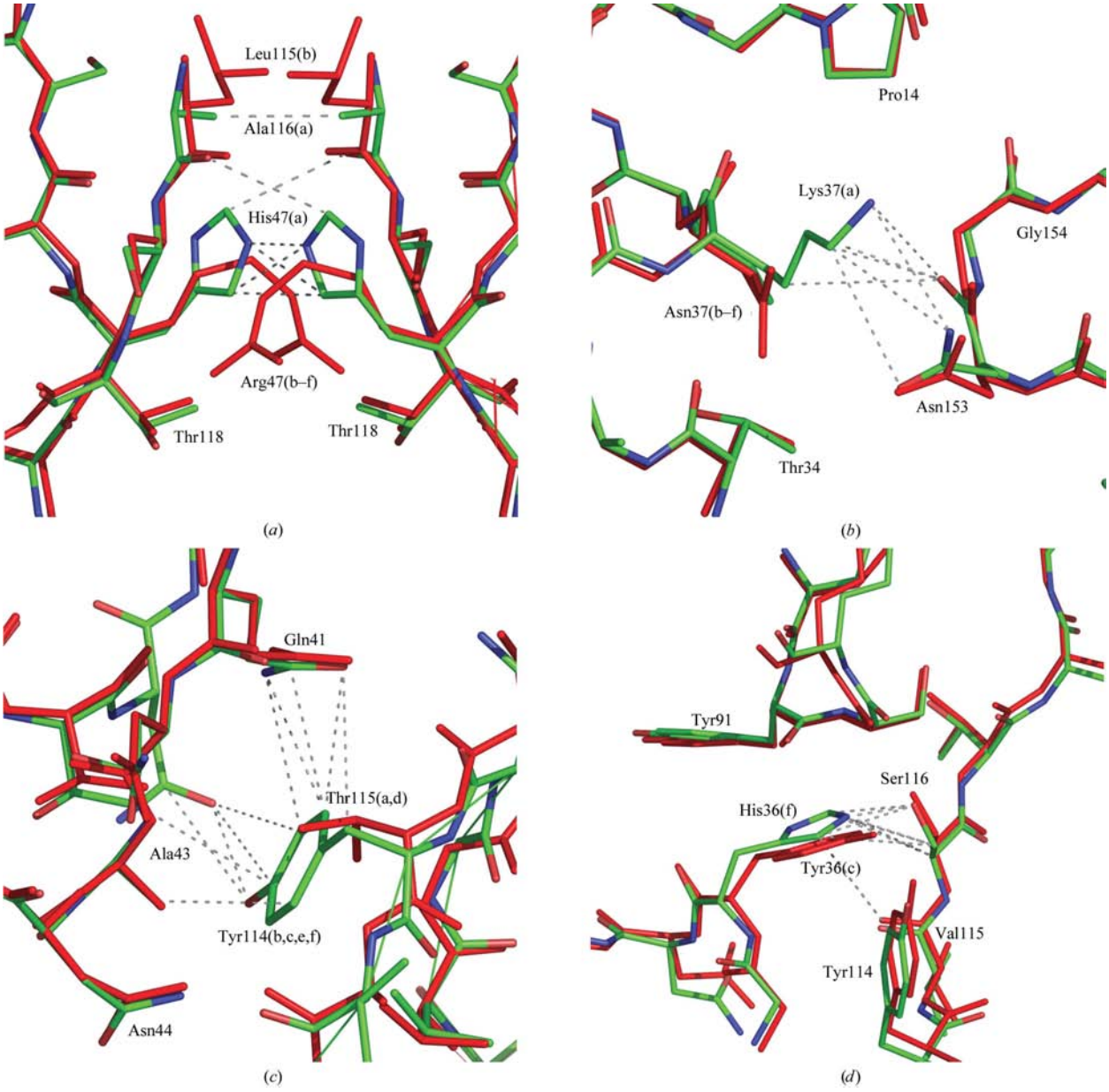

Figure 5
In-depth examples of crystal contacts affected by the sequence variation in the F17G family. In each panel, the residues involved in crystal-packing contacts are shown in green for C, blue for N and red for O atoms. Grey dotted lines present the atomic contacts (defined with a cutoff of 4 Å). The structure of a variant sequence was superimposed on the lattice contact of the genuine crystal packing and is shown in red. (*a*) Crystal contact in the F17a-G crystals involving His47 around a crystallographic twofold axis and superimposition of the corresponding F17b-G residues (red). His47 is replaced by Arg47 in all variants apart from F17a-G, leading to steric hindrance in the crystal packing of variant a. (*b*) Lys37, replaced by asparagine in all other variants, contributes significantly to a crystal lattice contact in F17a-G crystals. The replacement Asn37 is unable to make contacts. (*c*) The D2-E loop (residues 113–116; Fig. 1) is involved in crystal-packing contacts in all the observed crystal forms. In variants b, c, e and f, Tyr114 is systematically involved in packing interactions (variant c and its packing contacts shown in green), while the corresponding residue Thr115 in F17a-G (red) is not involved in packing. (*d*) Crystal contact observed in the F17f-G crystals, involving the His36 side chain. The single amino-acid change in variant c (His36 substituted by Tyr) is sufficient to render the F17f-G crystal packing unavailable to F17c-G.

relatively high overall B factors (40–50 Å^2) and their diffraction limit ranges between 2.1 and 2.6 Å (Fig. 3*b*). The tetragonal crystals of the F17b-G protein are isomorphous to the crystal form of variant e. From the statistical analysis (Table 3), it indeed appears that the $I4_122$ crystal form of F17b-G is more compatible with the other F17G sequences than the $P6_2$ crystal form. However, a possible disadvantage of the tetragonal crystal form is the involvement of the carbohydrate-binding site in extensive packing contacts. There appeared to be plenty of room to bind larger oligosaccharides in the $P6_2$ crystals of F17b-G, prompting us to try competition soaks of the F17b-G–GlcNAc(β1–2)Man co-crystals with GlcNAc(β1–3)Gal and GlcNAc(β1–4)GlcNAc. The crystals withstood this procedure and data were collected (Table 1).

3.3. Crystallization of F17d-G

Variant d grew as spherulites made of very fine needles in several conditions that are a combination of 2-propanol and PEG 4000, resembling the crystallization conditions of F17a-G, to which it is most similar in sequence (Fig. 1). The calculated pI of F17d-G is also closest to that of F17a-G (Table 1). One condition [25% PEG MME 2000, 0.1 *M* MES pH 6.5, 0.2 *M* $(NH_4)_2SO_4$] resulted in single microneedles that remained too small to allow diffraction. Packing analysis reveals that overall the F17d-G sequence has a high number of residues that are incompatible with each of the observed F17G crystal forms (F17d-G column in Table 3), including that of F17a-G. The F17a-G and F17d-G adhesins differ in six amino acids (Leu16Pro, Asp33Asn, Arg35Gly, Lys37Asn, His47Arg and Lys90Asn). All variants but F17a-G and F17d-G have Ser49Gly, Lys50Gly, Thr115Tyr, Arg124His, Thr126Arg and Arg137Gln substitutions, as well as a deletion of Ala113 (Fig. 1). Arg137 or Glu136 are involved in packing in all F17G crystal forms except for the $I4_122$ space group. The conservation of Ser49, Lys50, Ala113, Arg124 and Thr126 between variants a and d does not render F17d-G more amenable for crystallization using F17a-G crystallization conditions, since these residues are not involved in packing contacts in the F17a-G crystals. The presence of Thr115 in F17d-G instead of the equivalent and essential Tyr114 in variants b, c, e and f (Fig. 5*c*) rules out all packings other than that of F17a-G. The His47Arg substitution in turn appears to make the $P6_122$ packing of variant a inaccessible to F17d-G, as it does for the other variants (Fig. 5*a*).

3.4. Crystallization and crystal packing of F17e-G

Only a single crystal of the lectin domain of variant e (F17e-G) in complex with GlcNAc was obtained (Table 1). The GlcNAc monosaccharide is extensively involved in packing contacts, facing a GlcNAc molecule from a symmetry-related F17e-G molecule. The crystal is isomorphous with those of the tetragonal F17b-G variant bound to GlcNAc (Table 2) and diffracted to 2.4 Å resolution (Fig. 3*c*). The Matthews coefficient is 3.0 Å^3 Da^{-1}, corresponding to 50% solvent. As such, b and e are the only two variants that are observed to form isomorphous crystals. Indeed, these two variants are more similar to each other, differing by mostly conservative substitutions at five amino-acid positions (Val1Ala, Gly17Ser, Gly49Ser, Ser110Cys and Leu115Val) (Fig. 1). Positions 1 and 115 are involved in packing contacts in the $I4_122$ crystal form. The predominant contributions to these packing contacts are made by the main chain of Val1 in variant b and Ala1 in variant e (Fig. 1). Limited variation of the side chain is permitted. Contributions to packing contacts of the Val115 main-chain atoms in variant b are taken over by the side chain of Leu115 in the crystals of variant e. This observation explains why the two variant F17G sequences b and e are compatible within the same crystal lattice, despite the fact that quite a few atoms seem to be affected by the sequence variation (Table 3). Variant c differs from F17e-G only in positions 17 and 21. Of these only Gly17 is involved in the $P2_1$ crystals of F17c-G for both crystallographically independent molecules (Fig. 1).

3.5. Crystallization and crystal packing of F17f-G

A crystal of the lectin domain of F17f-G could be obtained in the presence of 10 m*M* GlcNAc (Table 1), but not in its absence or in the presence of the disaccharide GlcNAc(β1–2)Man. Again a different unit cell and different space group were found (Table 2). The crystal diffracted beyond 1.4 Å resolution and had a mosaicity of 0.8°. The asymmetric unit contains a single molecule and the Matthews coefficient of 2.2 Å^3 Da^{-1} indicates a solvent content of 43%. Analysis of the crystal packing confirms that the carbohydrate-binding site with the bound GlcNAc residue is involved in a tight lattice contact, excluding the possibility of binding of a second saccharide ring and probably disallowing the removal of the saccharide without disrupting the crystal lattice. This crystal form contains the highest number of lattice contacts of all variants (Table 3), which could be correlated with the capacity of the tiny needle-like crystals to diffract to high resolution. One crystal contact in the F17f-G crystals involves the His36 side chain (Fig. 1). Variant c differs from F17f-G only at position 36, containing a Tyr that is not involved in a packing contact. Replacement of Tyr36 of F17c-G by His as in F17f-G would be sufficient to render the F17f-G crystal packing unavailable to F17c-G (Fig. 5*d*). Three substitutions (Ser17Gly, Arg21Ser and His36Tyr) occur between variants e and f. Most likely the same His36Tyr mutation would again suffice to render the crystal form of F17f-G unavailable to F17e-G, as was shown for F17c-G (Fig. 5*d*).

4. Discussion

In the last couple of years, much attention has been focused on surface engineering of proteins to improve crystallization efficiency and crystal quality by altering packing interactions (Lawson *et al.*, 1991; McElroy *et al.*, 1992; Jenkins *et al.*, 1995; Schwede *et al.*, 1999; D'Arcy *et al.*, 1999; Dale *et al.*, 2003; Derewenda, 2004). Significant improvement in crystallizability could be observed by the removal of long charged side chains

with high conformational entropy such as lysine and glutamate (Longenecker *et al.*, 2001; Mateja *et al.*, 2002; Czepas *et al.*, 2004). On the other hand, both F17a-G and F17f-G crystallize with a molecular packing involving multiple solvent-exposed lysine and glutamate residues (Fig. 1) while displaying the capacity to diffract to atomic resolution (Fig. 3*a*, Table 2). Similar examples (Trp to Glu; Zhang *et al.*, 1997; Phe to Lys and Trp to Glu; Goldgur *et al.*, 1998) make it apparent that a positive or negative outcome of amino-acid substitutions is still very difficult to predict, even when it is known which residues are involved in packing contacts in the crystal. The six variants of the F17G lectin domain are surface-engineered by nature itself and illustrate the promise of this strategy to make proteins more amenable to crystallization.

Our results are indeed compatible with the notion that substitutions of a small number of residues scattered over the surface of the protein results in proteins that differ widely in crystallization propensity as well as in the resolution at which data can be obtained. Despite their differences, all six F17G variants have been purified by affinity chromatography on GlcNAc–agarose beads followed by gel filtration, without significantly changing the conditions. Of interest in our series of variants is that one or a few amino-acid substitutions are sufficient to move from straightforward crystal formation (F17a-G) to a protein that is practically impossible to crystallize (F17d-G). Because different variants crystallize in different crystal forms, we analysed the lattice contacts in each crystal form to evaluate whether they are compatible with the sequences of other variants (Fig. 1, Table 3). In general, the different variants are compatible only with their own crystal form. Notable exceptions are F17b-G, which is compatible with the crystal packing of F17e-G as also observed experimentally, and F17f-G, which in theory would also be compatible with the crystal packing of F17c-G and F17e-G (Table 3).

No clear link is apparent between the chemical nature of the interacting groups in packing contacts, the pI of the protein and the composition and pH of the precipitant solution. The isomorphous $I4_122$ crystals of variants F17b-G and F17e-G grew in very different solutions (10% ethanol with 1.5 *M* NaCl for F17bG; 30% PEG 400 with 100 m*M* $CdCl_2$ and 100 m*M* sodium acetate pH 4.6 for F17e-G), but nevertheless exhibit essentially identical packing interactions. The lattice contacts are almost entirely identical, with the exception of two relatively small changes (Val1Ala and Leu115Val). Strikingly, the same crystal form is obtained by either involving the side chain of Leu115 or the main chain of Val115. On the other hand, different crystal packings resulting from the same crystallization condition or even from the same drop (possibly generated during a different stage in the evolution of the drop) are quite common. An example provided by this study is the two different space groups observed for F17a-G bound to GlcNAc(β1-)OMe.

After detailed inspection of the affected packing contacts, no consistent patterns or general rules could be observed correlating the sequence changes with crystallization behaviour. This might have been expected from conflicting results available in the literature and the current inability to find a generally applicable logic behind crystallization. Therefore, we applied a statistical approach, summarized in Fig. 1 and Table 3, which provides a powerful insight into the relative compatibilities of the F17G sequences, one which is supported by our observations. Obvious from this aproach is the incompatibility of the $P6_122$ lattice of F17a-G with the other sequences. The His47Arg substitution present in the five variants b, c, d, e and f appears to be sufficient to explain this phenomenon (Fig. 5*a*). For example, the F17d-G protein most resembles F17a-G both in sequence and pI, but not in its facile crystallization behaviour, and this can be essentially attributed to the His47Arg substitution (Fig. 5*a*) combined with the loss of the Lys37 and Lys90 side chains (Fig. 5*b*). The relatively large incompatibility of the F17d-G sequence with the different crystal forms obtained could also be deduced from the packing analysis (Table 3). A limitation of the statistical approach is that compensatory effects that alleviate certain sequence/crystal packing incompatibilities, as observed for the tetragonal crystal forms of variant b and e, cannot be predicted. Visual inspection of the residues involved thus remains a necessity as one is seeking to rationalize the chemistry of a certain lattice contact.

In view of the enormous wealth of prokaryotic diversity (Fernández, 2005), which has been strongly affected by lateral gene transfer (Dutta & Pan, 2002), there is a large and underexplored source of natural variation present in prokaryotic organisms, offering a ready-to-use crystallographic tool to enhance the chances for obtaining crystals. Because selection has already taken place, natural variants can be expected to be stable proteins exhibiting only subtle differences in their function, detailed binding properties or immunological profiles. Since these differences depend largely on interactions of the protein with its environment, one might also expect a tendency for the variations to be localized on the surface. For the F17G adhesin, this holds true as all highly to moderately solvent-accessible areas (such as the small β-strand C') are involved in lattice contacts in one crystal form or another (Fig. 1). Residues from extended buried areas, such as β-strands *C* and *E*, are clearly less involved. Exceptions are arginine, glutamate and aspartate residues that still allow packing contacts *via* their long side chains, despite most of the residue being inaccessible to solvent.

Moreover, variants of a protein generating crystals displaying different space-group symmetries can improve the knowledge available on these proteins by showing different conformations of surface-exposed regions or revealing serendipitous ligand-binding or metal-ion-binding sites. They also enhance the probability of finding crystals with a molecular packing that does not directly involve binding sites and as such allows the binding of small molecules in an isomorphous crystal form. In this work, the hexagonal crystal form of variant b of the F17G adhesin is the only one that lends itself to a structural study of the interactions with oligosaccharides larger than disaccharides, thus providing a tool for further interaction studies. The extra information gained from structural studies on several natural variants will support the understanding of the functional implications of

the variation and its role in the evolution of bacterial genomes.

5. Conclusion

In their recent review, Dale *et al.* (2003) argue that the protein rather than the precipitating agent should be considered as the most important variable in screening for crystallization conditions. Our observations support this proposal. We would like to emphasize that the ways to influence the protein variable can be straightforward, without inherently involving engineering work such as site-directed mutagenesis of surface-located residues, the search for ligands to make crystallizable complexes, enzymatic deglycosylation or limited proteolysis, but often can be found directly at the source within the same species. Such a situation occurs frequently in prokaryotes, where several sequences of a given gene are known that only differ to a limited extent. By natural selection, variant proteins will be generated that probably do not differ significantly in biochemical and physical features such as the general structure and the function of the protein. Usually, the sequence differences in a given gene between strains allows sufficient variation at the protein level to avoid the need for artificial mutagenesis. On the other hand, the sequence variation may be sufficiently limited to avoid elaborate reoptimization of expression and purification protocols, which is a potential drawback of the structural genomics approach using homologous proteins from other organisms. In addition, since the variations will tend to affect surface-exposed residues for the fine-tuning of functions and systemic interactions of the protein, they may lend themselves very well to generating alternative molecular-packing interactions. Our example, the fimbrial adhesin F17G, displays natural variation to broaden the host specificity of enterotoxigenic *E. coli*. A supplementary advantage of this strategy is the extra functional information that can be gained from the observation of the different crystal forms as a starting point in the search for the functional implications of the structural diversity.

JB is a postdoctoral fellow of the Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO-Vlaanderen). Financial support for this project was provided by the Research Council of the Vrije Universiteit Brussel (OZR846, OZR727 and OZR728) and the FWO-Vlaanderen (FWOAL241). The authors gratefully acknowledge the use of synchrotron beamtime at the EMBL beamlines X11, X13, BW7A and BW7B at DESY (Hamburg, Germany) and the ID14-1 beamline at ESRF (Grenoble, France).

References

Bertels, A., Pohl, P., Schlicker, C., Vandriessche, E., Charlier, G., De Greve, H. & Lintermans, P. (1989). *Vlaams Diergeneeskd. Tijdschr.* **58**, 118–122.

Bertin, Y., Girardeau, J. P., Darfeuille-Michaud, A. & Contrepois, M. (1996). *Infect. Immun.* **64**, 332–342.

Buts, L., Bouckaert, J., De Genst, E., Loris, R., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L. & De Greve, H. (2003). *Mol. Microbiol.* **49**, 705–715.

Buts, L., Loris, R., De Genst, E., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L., De Greve, H. & Bouckaert, J. (2003). *Acta Cryst.* D**59**, 1012–1015.

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. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Campbell, J. W., Duee, E., Hodgson, G., Mercer, W. D., Stammers, D. K., Wendell, P. L., Muirhead, H. & Watson, H. C. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 165–170.

Cid, D., Sanz, R., Marín, I., De Greve, H., Ruis-Santa-Quiteria, J. A., Amils, R. & de la Fuente, R. (1999). *J. Clin. Microbiol.* **37**, 1370–1375.

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

Czepas, J., Devedjiev, Y., Krowarsch, D., Derewenda, U., Otlewski, J. & Derewenda, Z. S. (2004). *Acta Cryst.* D**60**, 275–280.

D'Arcy, A., Stihle, M., Kostrewa, D. & Dale, G. (1999). *Acta Cryst.* D**55**, 1623–1625.

Dahmen, J., Gnosspelius, G., Larsson, A. C., Lave, T., Noori, G., Palsson, K., Frejd, T. & Magnusson, G. (1985). *Carbohydr. Res.* **138**, 17–28.

Dale, G. E., Oefner, C. & D'Arcy, A. (2003). *J. Struct. Biol.* **142**, 88–97.

DeLano, W. L. (2002). *The PyMOL User's Manual.* DeLano Scientific, San Carlos, CA, USA.

Derewenda, Z. S. (2004). *Structure*, **12**, 529–535.

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

Dutta, C. & Pan, A. (2002). *J. Biosci.* **27**, 27–33.

El Mazouari, K., Oswald, E., Hernalsteens, J. P., Lintermans, P. & De Greve, H. (1994). *Infect. Immun.* **62**, 2633–2638.

Fernández, L. A. (2005). *Mol. Microbiol.* **55**, 5–15.

Girardeau, J.-P. (1980). *Ann. Microbiol. (Paris)*, **131**B, 31–37.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9150–9154.

Hubbard, S. J. & Thornton, J. M. (1993). *NACCESS.* Department of Biochemistry and Molecular Biology, University College London.

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

Jenkins, T. M., Hickman, A. B., Dyda, F., Ghirlando, R., Davies, D. R. & Craigie, R. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 6057–6061.

Kadima, W., McPherson, A., Dunn, M. F. & Jurnak, F. A. (1990). *Biophys. J.* **57**, 125–132.

Kendrew, J. C., Parrish, R. G., Marrack, J. R. & Orlans, E. S. (1954). *Nature (London)*, **174**, 946–949.

Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thomas, C. D., Shaw, W. V. & Harrison, P. M. (1991). *Nature (London)*, **349**, 541–544.

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.

Lintermans, P. F., Pohl, P., Bertels, A., Charlier, G., Vandekerckhove, J., Van Damme, J., Shoup, J., Schlicker, C., Korhonen, T., De Greve, H. & Van Montagu, M. (1988). *Am. J. Vet. Res.* **49**, 1794–1799.

Longenecker, K. L., Garrard, S. M., Sheffield, P. J. & Derewenda, Z. S. (2001). *Acta Cryst.* D**57**, 679–688.

McElroy, H. H., Sisson, G. W., Schottlin, W. E., Aust, R. M. & Villafranca, J. E. (1992). *J. Cryst. Growth*, **122**, 265–272.

Martin, C., Rousset, E. & DeGreve, H. (1997). *Res. Microbiol.* **148**, 55–64.

Mateja, A., Devedjiev, Y., Krowarsch, D., Longenecker, K., Dauter, Z., Otlewski, J. & Derewenda, Z. S. (2002). *Acta Cryst.* D**58**, 1983–1991.

Merckel, M. C., Tanskanen, J., Edelman, S., Westerlund-Wikstrom, B., Korhonen, T. K. & Goldman, A. (2003). *J. Mol. Biol.* **331**, 897–905.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Roussel, A. & Cambillau, C. (1989). *TURBO-FRODO. Silicon Graphics Geometry Partners Directory*, pp. 71–78. Silicon Graphics, Mountain View, CA, USA.
Saarela, S., Taira, S., Nurmiaholassila, E. L., Makkonen, A. & Rhen, M. (1995). *J. Bacteriol.* **177**, 1477–1484.
Schwede, T. F., Badeker, M., Langer, M., Retey, J. & Schulz, G. E. (1999). *Protein Eng.* **12**, 151–153.
Soto, G. E. & Hultgren, S. J. (1999). *J. Bacteriol.* **181**, 1059–1071.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Zhang, F., Basinski, M. B., Beals, J. M., Briggs, S. L., Churgay, L. M., Clawson, D. K., DiMarchi, R. D., Furman, T. C., Hale, J. E., Hsiung, H. M., Schoner, B. E., Smith, D. P., Zhang, X. Y., Wery, J. P. & Schevitz, R. W. (1997). *Nature (London)*, **387**, 206–209.