Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of indole-3-glycerol phosphate synthase from *Thermus thermophilus* HB8: implications for thermal stability

Bagautdin Bagautdinov[a,b]* and Katsuhide Yutani[b]

[a]Japan Synchrotron Radiation Research Institute (JASRI/SPring-8), 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan, and [b]RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

Correspondence e-mail: bagautdi@spring8.or.jp

Received 12 September 2011
Accepted 28 October 2011

PDB Reference: indole-3-glycerol phosphate synthase, 1vc4.

The three-dimensional structure of indole-3-glycerol phosphate synthase (IGPS) from the thermophilic bacterium *Thermus thermophilus* HB8 (*Tt*IGPS) has been determined at 1.8 Å resolution. The structure adopts a typical $(\beta/\alpha)_8$-barrel fold with an additional N-terminal extension of 46 residues. A detailed comparison of the crystal structure of *Tt*IGPS with available structures of IGPS from the archaeon *Sulfolobus solfataricus* (*Ss*IGPS) and the bacteria *Thermotoga maritima* (*Tm*IGPS) and *Escherichia coli* (*Ec*IGPS) has been performed. Although the overall folds of the proteins are the same, there are differences in amino-acid composition, structural rigidity, ionic features and stability clusters which may account for the high thermostability of the hyperthermophilic (*Ss*IGPS and *Tm*IGPS) and thermophilic (*Tt*IGPS) proteins when compared with the mesophilic *Ec*IGPS. The thermostability of IGPS seems to be established mainly by favourable interactions of charged residues, salt bridges and the spatial distribution of relatively rigid clusters of extensively interacting residues.

1. Introduction

The biosynthesis of tryptophan from chorismate requires seven enzymatic functions. In this pathway, indole-3-glycerol phosphate synthase (IGPS) catalyzes the fifth step by converting the substrate 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5′-phosphate to the product indole-3-glycerol phosphate (Creighton & Yanofsky, 1966; Hennig *et al.*, 2002). IGPS belongs to the large and versatile family of $(\beta/\alpha)_8$-barrel enzymes but has an additional N-terminal extension of about 40 amino-acid residues. The canonical $(\beta/\alpha)_8$-barrel consists of an eight-stranded parallel barrel core surrounded by an external layer of eight parallel α-helices interfaced with the solvent. X-ray structures of IGPSs from the mesophilic bacterium *Escherichia coli* (*Ec*IGPS; Priestle *et al.*, 1987; Wilmanns *et al.*, 1992), the hyperthermophilic bacterium *Thermotoga maritima* (*Tm*IGPS; Merz *et al.*, 1999; Knöchel *et al.*, 2002) and the hyperthermophilic archaeon *Sulfolobus solfataricus* (*Ss*IGPS; Andreotti *et al.*, 1994, 1997; Hennig *et al.*, 1995; Knöchel *et al.*, 1996) have been reported. The single-domain monomers of *Ss*IGPS, *Tm*IGPS and *Ec*IGPS were used to correlate the enzyme stability with the number of salt bridges and interactions of the terminal regions of the polypeptide chains and solvent-exposed loops with the barrel core (Hennig *et al.*, 1995; Knöchel *et al.*, 1996, 2002). Generally, study of the essential features for thermostability is based on structural comparison between orthologous proteins from hyperthermophilic and mesophilic organisms. However, in order to reveal general strategies for modulating heat

tolerance in proteins, it should be of interest to incorporate intermediate moderately thermophilic proteins into these studies. With optimum growth temperatures (T_o) in the range 323–353 K, they bridge the temperature region between their mesophilic (293–323 K) and hyperthermophilic (353–393 K) counterparts. As a result, the study of proteins bridging the three classes of thermostability should improve our understanding of how proteins from mesophilic organisms are gradually stabilized to work at higher temperatures, at which they often show higher reaction rates with higher substrate solubility as well as a lower probability of microbial contamination (Vieille & Zeikus, 2001; Wiegel & Adams, 1998).

Here, an X-ray crystallographic study of native IGPS from the thermophilic bacterium *Thermus thermophilus* HB8 (*Tt*IGPS) at 1.8 Å resolution is reported. Although the asymmetric unit contained two molecules of 254 amino-acid residues each, the biologically significant unit was found to be a monomer by dynamic light scattering. To gain a more general insight into IGPS adaptation over a wide range of temperatures, the single-domain structural model of a moderately thermophilic *Tt*IGPS having T_o = 345 K was compared with the crystal structures of *Ec*IGPS (T_o = 310 K), *Tm*IGPS (T_o = 356 K) and *Ss*IGPS (T_o = 363 K). As ∼10% of all known enzymes contain $(\beta/\alpha)_8$ domains (Farber, 1993), understanding the evolution of stability in IGPS may have widespread application.

2. Materials and methods

2.1. Structure determination

The expression and purification (Laemmli, 1970) of the *Tt*IGPS protein, dynamic light-scattering study, crystallization (Chayen *et al.*, 1990; Sugahara & Miyano, 2002) and data collection (Otwinowski & Minor, 1997) are presented as Supplementary Material[1].

The protein structure was determined using the molecular-replacement method with *MOLREP* from the *CCP*4 program suite (Vagin & Teplyakov, 2010; Winn *et al.*, 2011) using the structure of *Ss*IGPS (Hennig *et al.*, 2002), with 39.7% sequence identity, as the search model. The structure was refined using *CNS* (Brünger *et al.*, 1998). *QUANTA* (Accelrys, San Diego, California, USA) was used for molecular rebuilding and visualization of the structure. The final model was produced after several rounds of model building and energy minimization followed by individual *B*-factor refinement. Water molecules were added to the model and inspected manually during refinement. Eight tetrahedrally-shaped electron-density peaks were modelled as sulfate anions. The final model has an *R* factor of 18.6% for all data in the resolution range 33.6–1.8 Å (the free *R* factor for 5% of the data was 21.8%). Details of the refinement statistics are summarized in Table 1. The residue Ser215 is found in a disallowed region in both *Tt*IGPS molecules (molecule *A*, $\varphi = 72.7^\circ$, $\psi = 144.5^\circ$; molecule *B*, $\varphi = 77.5^\circ$, $\psi = 128.1^\circ$). The corresponding residues in *Ss*IGPS (Ser211), *Tm*IGPS (Ser210) and *Ec*IGPS (Ser215) have similar disallowed φ, ψ angles. Structural analysis was performed using *PROCHECK* (Laskowski *et al.*, 1993), *VADAR* (Willard *et al.*, 2003), *APBS* (Adaptive Poisson–Boltzman Solver) (Baker *et al.*, 2001), *CAST* (Dundas *et al.*, 2006) and *STRIDE* (Frishman & Argos, 1995). Protein interactions were analyzed using *PIC* (Tina *et al.*, 2007). Stabilization centres and stabilization residues were located using *SCide* (Dosztányi *et al.*, 2003) and *SRide* (Magyar *et al.*, 2005), respectively. Figures were prepared using *PyMOL* (DeLano, 2002).

The coordinates of the *Tt*IGPS structure and the structure-factor file have been deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000) under accession code 1vc4. Coordinates for the other three IGPS structures (*Ec*IGPS, 1vc4, 1jcm; *Tm*IGPS, 1i4n; *Ss*IGPS, 1lbf) were obtained from the PDB.

Table 1
Summary of data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Crystal data	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 63.65, b = 78.19, c = 91.52
Subunits per asymmetric unit	2
Data collection and refinement	
Temperature (K)	100
Wavelength (Å)	1.5418
Resolution range (Å)	33.56–1.80 (1.88–1.80)
Total No. of reflections	43028 (5290)
No. of unique reflections	42011 (4894)
Multiplicity	6.2 (5.8)
Completeness (%)	99.5 (99.3)
Mean $I/\sigma(I)$	12.6 (2.3)
R_{merge}† (%)	6.9 (37.5)
R_{work}‡	0.186 (0.275)
R_{free}§	0.218 (0.312)
No. of protein atoms	3897
No. of water molecules	683
Mean *B* factors (Å^2)	
Protein atoms	22.6
Main chains	20.9
Side chains	24.4
Water	38.73
Ramachandran plot (%)	
Most favoured	93.7
Additionally allowed	5.6
Generously allowed	0.2
Disallowed	0.5
PDB code	1vc4

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ and $\langle I(hkl)\rangle$ are the observed intensity of measurement i and the mean intensity of the reflection with indices hkl, respectively. ‡ $R_{work} = \sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. § R_{free} is the *R* factor for a subset of 5% of the reflections that were omitted from refinement.

3. Results and discussion

3.1. Overall structure

The crystal structure of *Tt*IGPS contains two independent molecules in the asymmetric unit. Each molecule of 254 amino acids consists of an eight-stranded parallel β-barrel core surrounded by an outer layer of eight parallel α-helices (Fig. 1). The canonical $(\beta/\alpha)_8$-barrel of *Tt*IGPS, which starts at

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MV5049). Services for accessing this material are described at the back of the journal.

Ser47, is preceded by an N-terminal extension with α_0 and α_{00} helices. The α_0 helix is positioned close to the entrance of the barrel, while the α_{00} helix covers the bottom part of barrel, which is relatively compact when compared with the top half of the barrel. Another difference from the canonical $(\beta/\alpha)_8$-barrel is the extra helix α_8' in loop $\beta_8\alpha_8$. The root-mean-square deviation (r.m.s.d.) of C^α positions for molecules *A* and *B* of *Tt*IGPS is 0.48 Å (for 227 residues). The largest deviations observed are in loops $\beta_1\alpha_1$ (residues 53–66) and $\beta_6\alpha_6$ (residues 182–193), which coincide with the regions of high temperature factors or *B* values. These loops are very likely to have relatively high conformational flexibility. Because of the good agreement of the other structural elements in molecules *A* and *B*, the following description and comparisons are based on molecule *A*. The *Tt*IGPS protein has 39.7, 37.1 and 36.4% sequence identity to *Ss*IGPS, *Ec*IGPS and *Tm*IGPS, respectively (Fig. 2*a*). Superposition of the structural single-domain $(\beta/\alpha)_8$-barrel scaffolds shows that the four IGPSs are spatially homologous: *Tt*IGPS superposes on the *Ss*IGPS, *Tm*IGPS and *Ec*IGPS monomers with r.m.s.d. values for C^α positions of 1.20 Å (for 180 residues), 1.41 Å (for 192 residues) and 1.25 Å (for 187 residues), respectively (Fig. 2*b*). 14 of the 39 identical residues in all four proteins are positioned in β-barrels, two in helices and 23 in coils (Fig. 2*a*). The structural differences between the four backbones are primarily localized to the N- and C-terminal extensions: loops $\alpha_0\alpha_{00}$, $\beta_1\alpha_1$ and $\beta_6\alpha_6$.

Table 2
Amino-acid composition of the four IGPS proteins.

Amino acid	*Ss*IGPS (T_o = 363 K)	*Tm*IGPS (T_o = 356 K)	*Tt*IGPS (T_o = 345 K)	*Ec*IGPS (T_o = 310 K)
Charged				
Lys	17 (6.9%)	20 (8.0%)	8 (3.1%)	12 (4.9%)
Arg	21 (8.5%)	19 (7.6%)	25 (9.8%)	14 (5.4%)
Asp	11 (4.5%)	19 (7.6%)	9 (3.5%)	17 (6.6%)
Glu	27 (10.9%)	26 (10.4%)	29 (11.4%)	14 (5.4%)
Total	76 (30.8%)	84 (33.5%)	71 (28.0%)	57 (22.0%)
Polar				
Asn	14 (5.7%)	8 (3.2%)	3 (1.2%)	10 (3.9%)
Gln	4 (1.6%)	3 (1.2%)	1 (0.4%)	16 (6.2%)
Ser	22 (8.9%)	13 (5.2%)	13 (5.1%)	16 (6.2%)
Thr	6 (2.4%)	8 (3.2%)	6 (2.4%)	7 (2.7%)
Total	46 (18.6%)	32 (12.7%)	23 (9.1%)	49 (18.9%)
Aliphatic				
Val	12 (4.9%)	20 (8.0%)	22 (8.7%)	19 (7.3%)
Ile	29 (11.7%)	25 (10.0%)	8 (3.1%)	19 (7.3%)
Leu	29 (11.7%)	25 (10.0%)	42 (16.5%)	27 (10.4%)
Met	5 (2.0%)	4 (1.6%)	3 (1.2%)	4 (1.5%)
Total	75 (30.4%)	74 (29.5%)	75 (29.5%)	69 (26.6%)
Aromatic				
Phe	8 (3.2%)	9 (3.6%)	7 (2.8%)	10 (3.9%)
Tyr	9 (3.6%)	4 (1.6%)	4 (1.6%)	10 (3.9%)
Trp	1 (0.4%)	3 (1.2%)	—	1 (0.4%)
Total	18 (7.3%)	16 (6.4%)	11 (4.3%)	21 (8.1%)
Other				
Ala	12 (4.9%)	23 (9.2%)	33 (13.0%)	31 (12.0%)
Pro	9 (3.6%)	8 (3.2%)	16 (6.3%)	9 (3.5%)
Cys	—	1 (0.4%)	—	5 (1.9%)
Gly	11 (4.5%)	10 (4.0%)	22 (8.7%)	11 (4.2%)
His	—	3 (1.2%)	3 (1.2%)	7 (2.7%)
Total	32 (13.0%)	45 (17.9%)	74 (29.1%)	63 (24.3%)
Overall total	247 (100.0%)	251 (100.0%)	254 (100.0%)	259 (100.0%)

3.2. Active site

The catalytically active residues of $(\beta/\alpha)_8$-barrel enzymes are typically located at the C-terminal ends of the β-strands and in the loops that connect the β-strands to the subsequent α-helices ($\beta_n\alpha_n$ loops). Therefore, the top half of the barrel harbouring the active site is considered to be the catalytic face, while the opposite more compact end of the barrel, which is important for conformational stability of the fold, is considered to be the stability face (Höcker *et al.*, 2001; Wiederstein & Sippl, 2005). It is known that most of the important residues surrounding the active site in the IGPS proteins are conserved, suggesting similarities in substrate binding and the reaction mechanism (Hennig *et al.*, 2002; Darimont *et al.*, 1998). Comparison with the crystal structure of *Ss*IGPS suggests that the catalytic residues in *Tt*IGPS are Lys112, Glu160 and Lys53, which are located at the C-terminus of β_3, the C-terminus of β_5 and the N-terminus of coil $\beta_1\alpha_1$, respectively (Figs. 2*a* and S1). The catalytic residues are covered by flexible loops $\beta_1\alpha_1$ and $\beta_6\alpha_6$, which presumably assist in substrate binding and product release.

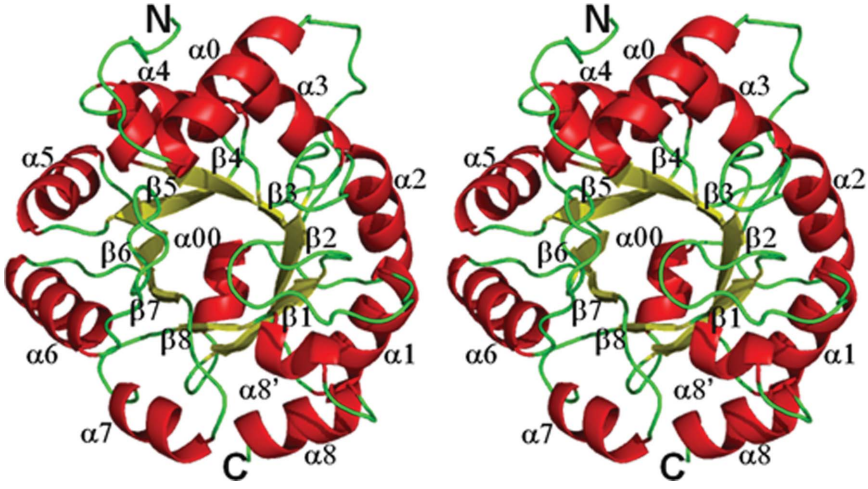

Figure 1
The overall fold structure of *Tt*IGPS. A stereoview of the *Tt*IGPS subunit with labelling of secondary-structure elements is shown. The α-helices are shown in red and drawn as spiral ribbons, while the β-strands are shown in green and drawn as arrows from the amino end to the carboxyl end of the β-strand. The view is down the respective barrel axis. The canonical $(\beta\alpha)_8$-barrel of *Tt*IGPS contains 208 amino acids and is composed of eight units, each of which consists of a β-strand and an α-helix that are connected by a $\beta\alpha$-loop.

3.3. Structural basis of thermal stability

It is well recognized that different proteins adapt to higher temperatures using different sets of structural interactions (Szilágyi & Závodszky, 2000; Vieille & Zeikus, 2001). As each class of proteins appears to have used its own mechanism for establishing protein stability under extreme conditions, comparison of functionally and structurally similar proteins from different organisms provides

one approach to identifying the factors that confer thermal stability. To this end, we compared the *Ec*IGPS, *Tt*IGPS, *Tm*IGPS and *Ss*IGPS proteins. As the protein is not necessarily thermolabile at temperatures higher than the T_o of its source, ideally the availability of an experimentally determined melting temperature (T_m) would be desirable for the analysis of thermal stability. However, in the absence of experimental T_m values for the four IGPS proteins, we assume that T_m is correlated to T_o.

3.3.1. Amino-acid composition. The compositions of the four IGPS proteins, in terms of polar charged, polar uncharged, aliphatic, aromatic and other amino acids, are shown in Table 2. The three thermostable proteins (*Ss*IGPS, *Tm*IGPS and *Tt*IGPS) contain an increased fraction of Arg and Glu and a decreased fraction of Asp, Asn, Gln, Thr, Cys, Ser and His compared with the mesophilic *Ec*IGPS. The fractions of thermolabile Gln and Cys (Russell *et al.*, 1997) are particularly low (or absent) in the thermostable IGPS proteins relative to *Ec*IGPS. A clear correlation between thermostability and the ratio of two pairs of preferred (Glu and Lys) and avoided (Gln and His) amino acids (Farias & Bonato, 2003) was observed: the ratio (E + K)/(Q + H) is high in *Ss*IGPS (11.0), *Tm*IGPS (7.0) and *Tt*IGPS (9.3), while for *Ec*IGPS it is much lower (1.1). The fraction of charged residues DEKR (Asp, Glu, Lys and Arg; Smith & Gallagher, 2008) in *Ec*IGPS (22.0%) is lower than that in *Tt*IGPS (30.0%), *Tm*IGPS (33.5%) and *Ss*IGPS (30.8%). Among charged residues, the thermostable IGPS proteins contain a large total number of Arg and Glu residues, both of which have a tendency to form multiple ion pairs and hydrogen bonds. The higher ratios for charged amino acids, which are mainly located in the helices and loops, should be important for stabilization of the exposed regions of the IGPS fold through participation in additional electrostatic interactions. Moreover, the prevalence of charged residues in the helices of thermostable IGPSs (~38; ~15%) compared with *Ec*IGPS (26; 10%) may provide charge compensation of the helix dipoles. The inner β-barrels of the four IGPSs have closely similar compositions, consisting mainly of aliphatic amino acids. Therefore, the overall differences in amino-acid composition in mesophilic and thermostable IGPS are much greater on the protein surface than in the interior, suggesting a strategy for adaptation to the environment. No correlation is observed between the thermostability of the

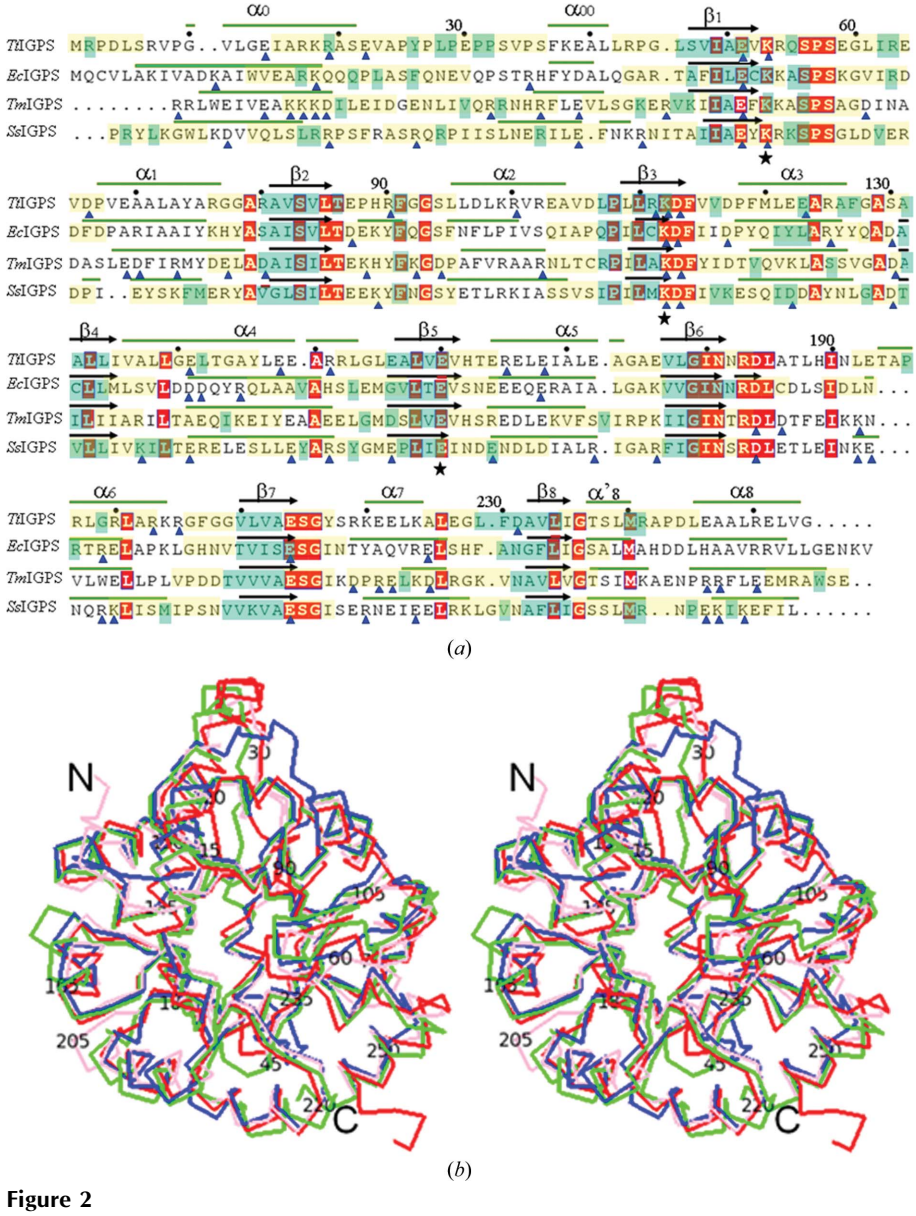

Figure 2
(*a*) Sequence alignment of the four IGPS proteins from various sources for which three-dimensional structures are available: *Tt*IGPS (from *Thermus thermophilus*; PDB entry 1vc4), *Ec*IGPS (from *Escherichia coli*; PDB entry 1jcm), *Tm*IGPS (from *Thermotoga maritima*; PDB entry 1i4n) and *Ss*IGPS (from *Sulfolobus solfataricus*; PDB entry 1lbf). The single-letter code for amino-acid residues is used; secondary structures are presented above the alignment. The symbols β_n and α_n denote β-strands and α-helices, respectively; every tenth residue of *Tt*IGPS is highlighted. Invariant residues for IGPS from various sources are shown in red. The residues involved in salt bridges are indicated by blue triangles. Yellow-shaded highlighting indicates residues forming SC clusters. Black stars indicate the amino-acid residues involved in formation of the active site. Sequence alignments were generated using *ClustalX* (Jeanmougin *et al.*, 1998). The sequence-based alignment is unambiguous and agrees with alignments based on superposition of the backbone structures. (*b*) A stereo-diagram showing the superposition of the four IGPS structures viewed with the β-barrel axis vertical in the plane of the figure. Structures are represented as C^{α}-backbone traces. The colouring is as follows: pink, *Tt*IGPS; blue, *Ss*IGPS; green, *Tm*IGPS; red, *Ec*IGPS. For the *Tt*IGPS structure, the N- and C-termini are marked and every 15th residue is numbered.

proteins and the percentage of hydrophobic residues (Ala, Phe, Gly, Ile, Leu, Met, Pro, Val and Thr) in the amino-acid composition: their total numbers are 115 (47%), 127 (51%), 153 (60%) and 131 (51%) for *Ss*IGPS, *Tm*IGPS, *Tt*IGPS and *Ec*IGPS, respectively. Notably, while the frequency of charged residues in the present IGPS proteins increases in a continuous manner from mesophile to thermophile to hyperthermophiles, the composition of some of the hydrophobic residues changes irregularly, with a significant increase in Pro, Gly, Leu and Ala observed for *Tt*IGPS relative to *Ec*IGPS but not for *Tm*IGPS and *Ss*IGPS.

3.3.2. Structural rigidity. In addition to correlation with amino-acid composition, temperature adaptation in proteins depends on structural interactions amongst amino acids and between protein and solvent as well as local rigidity. Generally, the flexible regions exemplified by high crystallographic temperature factors, or *B* factors, are relatively unstable segments in proteins (Daggett & Levitt, 1993; Vihinen, 1987; Vihinen *et al.*, 1994; Jaenicke, 1996; Parthasarathy & Murthy, 1997, 2000; Smith *et al.*, 2003). For this reason, thermostable enzymes should have relatively rigid structures to compensate for elevated thermal mobility in their natural environment. The residues in the strands of the four IGPS structures have lower *B* factors, confirming the core as the more rigid, densely packed region of the structure (Table 3, Fig. 3). The mean *B* factors of the outer atoms (located in helices, coils and loops) are higher than those in the strands by factors of 1.38, 1.48, 1.52 and 1.32 in *Ss*IGPS, *Tm*IGPS, *Tt*IGPS and *Ec*IGPS, respectively. As the structures were all determined under similar conditions (~100 K, cold nitrogen-gas stream), the observed higher values for the thermostable IGPS proteins indicate relative rigidity of their barrel core. This could be important for maintenance of the structural integrity of the active site located at the C-terminal end of the barrel at higher temperatures, allowing enzymatic function. The strands of IGPS connect to helices through $\alpha_i\beta_{i+1}$ loops at the N-terminal face of the β-barrel and $\beta_i\alpha_i$ loops at the C-terminal face of the β-barrel. Overall, $\alpha_i\beta_{i+1}$ loops are shorter and less flexible than $\beta_i\alpha_i$ loops (Table 3). However, $\beta_3\alpha_3$ and $\beta_5\alpha_5$ of the C-terminal face, which do not contain any catalytically important residues, are also short (Figs. 1 and 2*a*). Thus, the IGPS loops are classified into two types: short and less flexible, which support the barrel stability, and flexible and longer for functional activity. The amino-acid sequences of the

Table 3
B factors (Å^2) of C$^\alpha$ atoms of the four IGPS proteins.

Values in parentheses are the numbers of residues.

	*Ss*IGPS	*Tm*IGPS	*Tt*IGPS	*Ec*IGPS
All atoms	23.77 (247)	37.87 (251)	20.94 (254)	58.38 (259)
Strand atoms	18.02 (37)	26.93 (40)	14.60 (40)	45.91 (42)
Helix atoms	23.98 (99)	37.85 (97)	20.44 (104)	58.10 (95)
Loop atoms	25.32 (111)	41.85 (114)	22.15 (110)	64.89 (122)
Beside strands	24.79 (210)	39.95 (211)	22.13 (204)	60.55 (217)
N-terminal atoms	28.38 (46)	42.44 (41)	20.62 (46)	80.95 (48)
Beside N-terminal atoms	22.72 (201)	36.98 (210)	21.01 (208)	51.71 (211)
$\beta_i\alpha_i$ loops	23.47 (53)	41.54 (59)	24.57 (53)	50.36 (51)
$\alpha_i\beta_{i+1}$ loops	21.10 (31)	40.53 (34)	20.15 (29)	52.22 (32)

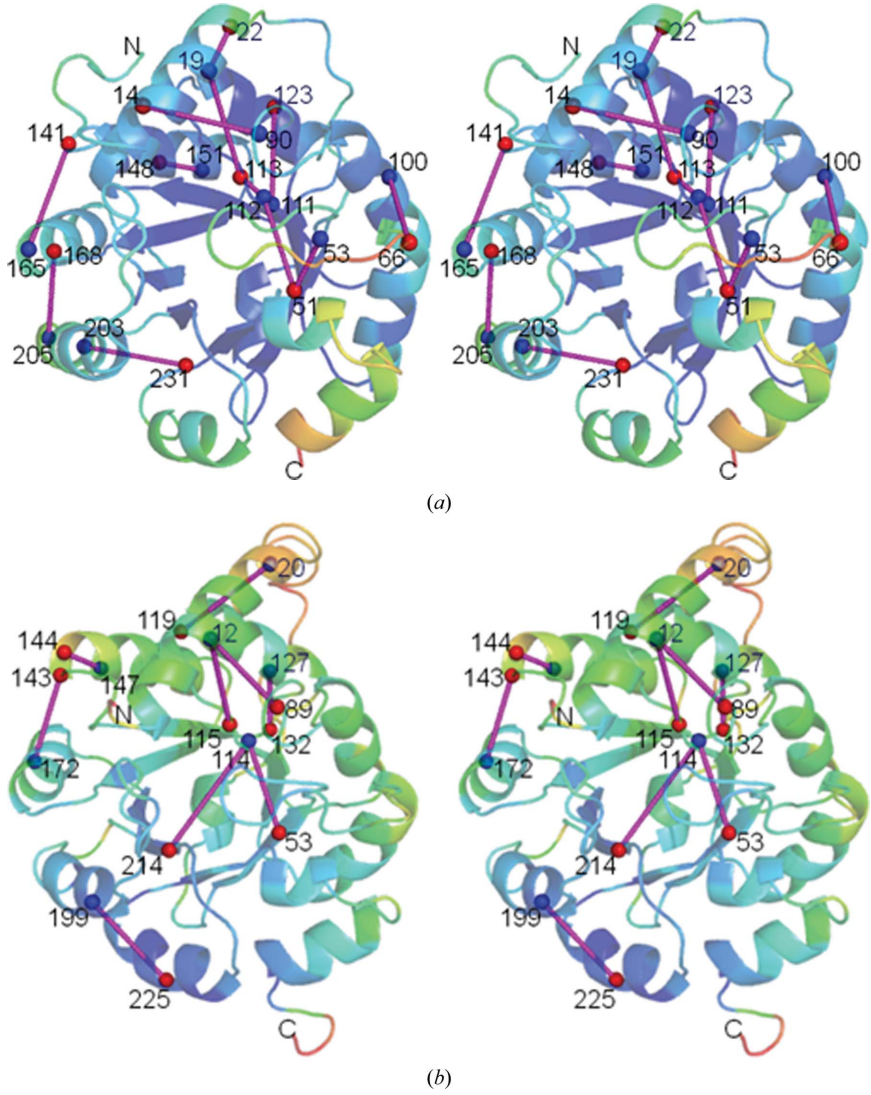

Figure 3
Stereoview of the distribution of salt bridges in the IGPS proteins (*a*) *Tt*IGPS, (*b*) *Ec*IGPS, (*c*) *Tm*IGPS and (*d*) *Ss*IGPS. The residues in salt bridges are numbered and the respective connections are shown in magenta. Negatively and positively charged residues are highlighted as red and blue spheres, respectively, at the C$^\alpha$ positions. Salt bridges are considered as hydrogen-bonded ion pairs with distances of less than 3.7 Å between the donor and the acceptor. For *Tt*IGPS and *Tm*IGPS the salt bridges common to both the *A* and *B* molecules are presented. The $(\beta/\alpha)_8$-barrel fold is colour-coded by *B* factor from dark blue for low *B* to red for high *B*. The N- and C-termini of the polypeptide chains are highlighted.

N-terminal segment show low conservation among the IGPS enzymes (Fig. 2*a*), which may be important for differences in thermostability. In fact, the N-terminally truncated variants *Ss*IGPSΔ(1–26) and *Tm*IGPSΔ(1–25) were observed to be less thermostable than the full-length proteins (Schneider *et al.*, 2005). The ratios of the *B* factors of the N-terminal atoms to those of the $(\beta/\alpha)_8$-barrel fold are 1.25, 1.15, 0.98 and 1.57 for *Ss*IGPS, *Tm*IGPS, *Tt*IGPS and *Ec*IGPS, respectively (Table 3). The highest value is for *Ec*IGPS, suggesting that its N-terminus is less associated with the $(\beta/\alpha)_8$-barrel core compared with the thermostable IGPS, in which the N-terminal and core *B* factors are more correlated.

3.3.3. Hydrophobic and hydrogen-bond interactions. Hydrophobic and hydrogen-bond interactions are important noncovalent forces that hold the polypeptide in a compact three-dimensional structure. Information about hydrophobic, hydrogen-bonding and electrostatic effects in the IGPS structures is presented in Table 4. The cavity volumes presented in Table 4 were calculated using a probe radius of 1.4 Å and with all bound solvent molecules excluded (Connolly, 1993). The accessible surface area (ASA) values of proteins and the ΔASA of nonpolar C/S and polar N/O atoms in residues were calculated from their X-ray crystal structures and those in the denatured state were calculated from their extended structures (Tanaka *et al.*, 2001). The changes in denaturation Gibbs energy owing to hydrophobic effects were calculated using the method described by Funahashi *et al.* (1999) and Takano *et al.* (1998).

We computed the optimal hydrogen bonds in the IGPS set using *WHAT IF*, a molecular-modelling and drug-design program (Vriend, 1990; Hooft *et al.*, 1996). For complete analysis, we additionally examined the nonconventional C—H···O bonds, as there is growing evidence that these may also be relevant to biopolymer structures (Desiraju, 1991; Derewenda *et al.*, 1995; Wahl & Sundaralingam, 1997). These hydrogen bonds, characterized by a C—H donor in place of the more common O—H or N—H groups, were identified using *HBAT* (Tiwari & Panigrahi, 2007). We did not include water molecules in our calculations. As the present four X-ray structures do not contain any H-atom coordinates, the program calculates the position of H atoms according to standard geometrical rules. Therefore, it presents a qualitative study of the most probable hydrogen bonds in the structures. While conventional hydrogen bonds dominate in their absolute number, 30–40% of the total number of hydrogen bonds are nonconventional C—H···O bonds (Table 4). The four IGPS monomers contain comparable numbers of conventional hydrogen bonds and an average of more than two hydrogen-bonding interactions were observed per residue. Interestingly, the mesophilic *Ec*IGPS has a larger number of possible C—H···O bonds relative to other IGPSs. In general, when polar or charged groups are buried inside proteins they make favourable electrostatic interactions as hydrogen bonds or salt bridges to compensate for the charge (Barlow & Thornton, 1983). For this reason, C—H···O bonds may be important in stabilizing folded structures, although they tend to be weaker compared with conventional hydrogen bonds owing to the lower proton-donating ability of C—H (Kollman *et al.*, 1975; Seiler *et al.*, 1987; Umeyama & Morokuma, 1977). Amongst 129 polar residues (Ser, Cys, Thr, Asn, Gln Tyr, Lys, Arg, His, Asp and Glu) in *Ec*IGPS, 27 (21%) are buried. In contrast, in *Ss*IGPS only 19 of 131 (15%), in *Tm*IGPS 17 of 124 (14%) and in *Tt*IGPS 19 of 101 (19%) polar residues are buried. It is likely that the large percentage of buried polar residues in *Ec*IGPS is related to its high content of C—H···O bonds. Although the four IGPS proteins do not appear to show a correlation between thermostability and an increase in the number of hydrogen bonds, the effect of hydrogen bonds on the total conformational

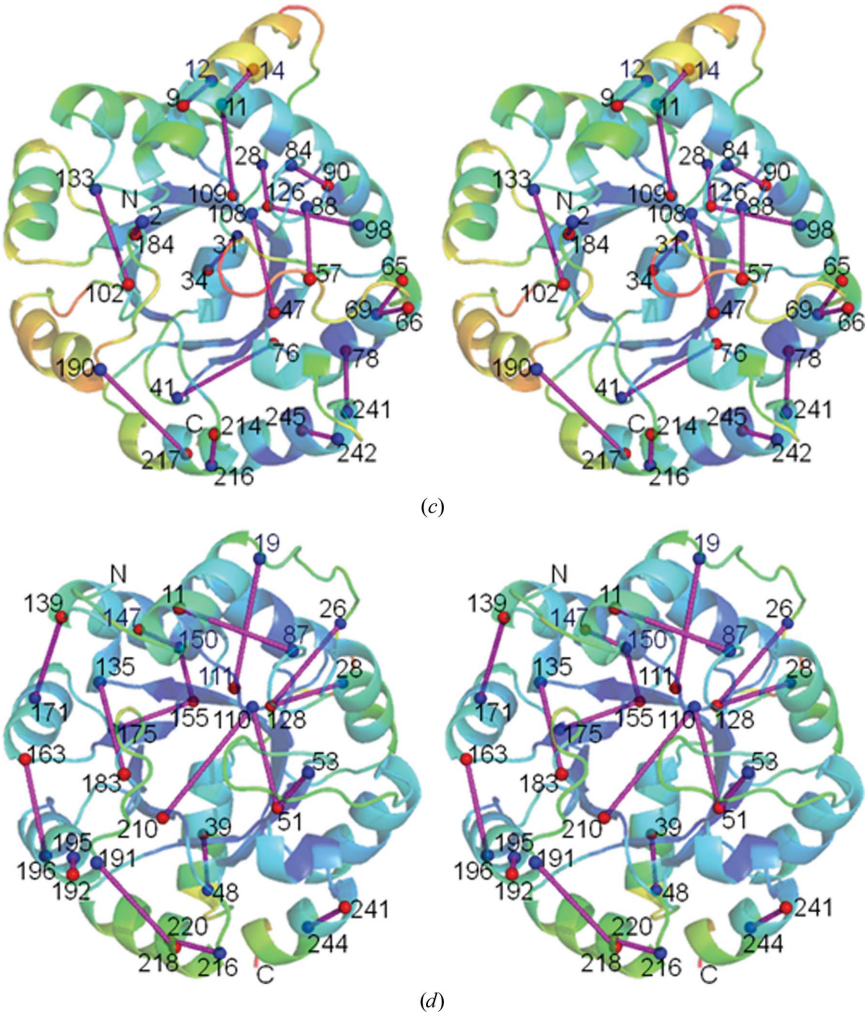

Figure 3 (continued)
Distribution of salt bridges in the IGPS proteins, (*c*) *Tm*IGPS and (*d*) *Ss*IGPS.

Table 4
Comparison of the stabilizing factors among the four IGPS proteins.

	*Ss*IGPS	*Tm*IGPS	*Tt*IGPS	*Ec*IGPS
No. of cavities† (probe 1.4 Å)	6	12	4	5
Volume of cavities† (Å^3)	219	399	130	151
ΔG_{cav}‡ (kJ mol^{-1})	−16	−29	−9	−11
Hydrophobic interactions§ (<3.5 Å)	11	7	16	25
ΔASA† (C/S) (Å^2)	16876	16865	16640	17097
ΔASA† (N/O) (Å^2)	8045	7436	7405	7632
ΔG_{hyd}‡ (kJ mol^{-1})	2633	2619	2585	2656
$\Delta\Delta G_{hyd}$¶ (kJ mol^{-1})	−23	−37	−71	—
No. of hydrogen bonds, total†††‡‡ (<3.5 Å)	496	458	514	596
No. of hydrogen bonds, nonconventional‡‡	161	155	169	250
No. of cation–π interactions§§	3	1	1	3
No. of salt bridges (<3.7 Å)	17	17	12	10
No. of salt bridges per residue	0.069	0.068	0.047	0.039
Mean d of salt bridges (Å)	3.00	3.10	2.97	3.24
Sequence range of salt-bridging residues	Δ233 (Asp11–Lys244)	Δ243 (Arg2–Glu245)	Δ217 (Glu14–Asp231)	Δ214 (Asp11–Glu225)
No. of charge interactions in salt bridges¶¶	76	81	69	36
Electrostatic contributions of the salt-bridge atoms, ΔG_{sb}¶¶ (kJ mol^{-1})	−89	−87	−73	−55
$\Delta\Delta G$ (kJ mol^{-1})	−34	−32	−18	—
No. of charge interactions¶¶	3481	3599	3474	1849
No. of repulsive interactions¶¶	1596	1706	1561	873
ΔG_{ri}¶¶ (kJ mol^{-1})	235	315	191	179
No. of attractive interactions¶¶	1885	1893	1913	976
ΔG_{ai}¶¶ (kJ mol^{-1})	−579	−603	−523	−341
Total electrostatic contributions of all atoms, ΔG_{el}¶¶ (kJ mol^{-1})	−344	−288	−332	−162
$\Delta\Delta G_{el}$ (kJ mol^{-1})	−182	−126	−170	—

† Connolly (1993). ‡ Takano *et al.* (1998) and Funahashi *et al.* (1999). § *PIC* (Tina *et al.*, 2007). ¶ $\Delta\Delta G_{hyd}$, $\Delta\Delta G_{sb}$ and $\Delta\Delta G_{el}$ represent the differences in ΔG_{hyd}, ΔG_{sb} and ΔG_{el}, respectively, of *Ss*IGPS, *Tm*IGPS and *Tt*IGPS relative to *Ec*IGPS. †† *WHAT IF* (Vriend, 1990). ‡‡ *HBAT* (Tiwari & Panigrahi, 2007). §§ *CAPTURE* (Gallivan & Dougherty, 1999). ¶¶ *FOLDX* (Schymkowitz *et al.*, 2005).

stability should be important, as shown by the high number of interactions and from the long-range character of many of them in terms of sequence separation, especially in the β-strand regions.

The number of energetically significant (above −2 kcal mol^{-1}) cation–π interactions in each protein (Table 4) was calculated using the program *CAPTURE* (Gallivan & Dougherty, 1999). As the hydrophobic, hydrogen-bond and cation–π analyses of IGPS are not correlated with T_o, it is clear that there are other mechanisms of adaptation of the IGPS protein to temperature.

3.3.4. Salt bridges. Usually, charged groups are distributed in the protein structure in such a way that the total interaction between the charges is favourable. The desolvation cost upon charge-contact formation decreases at elevated temperatures, leading to an effective stabilization at high temperatures (Vogt *et al.*, 1997; Elcock, 1998; Thomas & Elcock, 2004). The salt bridges identified by a threshold distance of 3.7 Å for *Ss*IGPS, *Tm*IGPS and *Ec*IGPS have been discussed previously (Hennig *et al.*, 1995; Knöchel *et al.*, 2002). Here, we analyze salt bridges in *Tt*IGPS and compare them with those in the other three proteins. For this, by analogy to the previous studies, salt bridges are determined when Asp or Glu side-chain carboxyl O atoms are found within 3.7 Å of the N atoms in Arg and Lys side chains, regardless of their orientation. It was found that in *Tt*IGPS 20 residues are involved in the formation of 12 salt bridges (Tables 4 and 5). The number of salt bridges in *Tt*IGPS (0.05 per residue) is higher than that for *Ec*IGPS (ten in total or 0.02 per residue), but is lower than that for *Tm*IGPS and *Ss*IGPS (17 in total or 0.07 per residue) (Table 4). The average N—O distance of the salt bridges of *Tt*IGPS is 2.97 Å and, with the exception of two long salt bridges (Asp113–Arg111 and Glu141–Arg165), they are strong salt bridges with good hydrogen-bond geometry (Table 5). The salt bridges linking portions of the protein that are far apart in the sequence are mainly localized in the N–α_4 half of the barrel. Helix α_4 is clamped to helix α_5 by the ion pair Glu141–Arg165 bridging the two halves, N–α_4 and α_5–C, of the barrel. A similar clamp between helices α_4 and α_5 is found to be a common feature of IGPSs. Another preserved cluster of salt bridges is located in the active site of *Tt*IGPS and involves Glu51–Lys53 and Glu51–Lys112, which may promote the correct conformation of the active-site residues. There are eight bridges in *Tt*IGPS linking portions of the protein that are non-adjacent in the sequence. Notably, the salt bridges in *Ss*IGPS and *Tm*IGPS involve residues from a large sequence range and that are evenly distributed over the whole structure. However, in *Tt*IGPS and *Ec*IGPS the bridge distribution is biased and the sequence range of residues involved is lower (Table 4, Figs. 2*a* and 3). Considering the number as well as the architectural role of the salt bridges, it appears that *Ec*IGPS and *Tt*IGPS are less stabilized by these interactions than *Tm*IGPS and *Ss*IGPS.

The electrostatic contributions to the free-energy change upon salt-bridge formation in IGPS were calculated using the *FOLDX* computer-program package (Guerois *et al.*, 2002; Schymkowitz *et al.*, 2005; http://foldx.crg.es/; Tables 4 and 5). To reflect the nonpoint nature of electrostatic interaction, we considered the electrostatic energy of each residue in the salt

Table 5
Intramolecular salt bridges in *Tt*IGPS.

Salt bridges that are common to both molecules *A* and *B* are listed. All salt bridges with distances of less than 3.7 Å between the N atom and the nearest O atom are included. The assignment of residues to secondary-structural elements is documented in Fig. 2.

Negatively charged side chain, position	Side-chain *B* factor† (Å^2)	ASA/residue‡ (Å^2)	Positively charged side chain, position	Side-chain *B* factor† (Å^2)	ASA/residue‡ (Å^2)	N—O distance (Å)	ΔG§ (kJ mol^{-1})
Glu14, α_0	27	67	Arg90, $\beta_2\alpha_2$	31	85	2.72	−1.55
Glu22, $\alpha_0\alpha_{00}$	25	47	Arg19, α_0	19	24	2.56	−7.57
Glu51, β_1	16	4	Lys53, $\beta_1\alpha_1$	20	23	3.00	−5.82
Glu51, β_1	16	4	Lys112, β_3	17	25	2.66	−7.82
Asp66, $\beta_1\alpha_1$	34	51	Arg100, α_2	25	80	3.14	−2.02
Asp113, β_3	16	0	Arg19, α_0	19	24	2.95	−10.71
Asp113, β_3	16	0	Arg111, β_3	14	5	3.70	−6.95
Glu123, α_3	19	14	Arg111, β_3	14	5	2.82	−9.33
Glu141, α_4	23	109	Arg165, α_5	45	148	3.67	−3.10
Glu148, α_4	21	61	Arg151, α_4	18	93	2.78	−3.26
Glu168, α_5	31	51	Arg205, $\alpha_6\beta_7$	41	85	2.89	−2.64
Asp231, $\alpha_7\beta_8$	20	31	Arg203, α_6	23	59	2.70	−7.87

† Mean *B* factor for side chain; the average *B* factor for all side-chain atoms in *Tt*IGPS is 32.4 Å. Buried salt-bridge residues belonging to the strands present low *B* factors and they should be stronger as a bridge between fixed charges. ‡ Solvent-accessible surface for individual donors and acceptors in protein structure as determined using the *WHAT IF* server (http://swift.cmbi.ru.nl/). § The electrostatic contributions of the salt-bridge residues to the changes in the free energy.

bridges. Our results showed that each salt bridge in *Tt*IGPS contributed to a reduction in the free energy ($\Delta G_{sb} < 0$) regardless of whether they were buried or exposed (Table 5). The total contribution of salt bridges to the free energy (ΔG_{sb}), as well as the total electrostatic energy (ΔG_{el}) reflecting all attractive and repulsive interactions of the charged residues, in the four IGPSs correlates with the relative thermostabilities of the proteins (Table 4).

3.3.5. Stabilization clusters. It is common knowledge that interresidue interactions in proteins are not distributed evenly within a structure and that regions of higher interaction are energetically stabilized. Therefore, the identification of clusters with dense networks of cooperative interactions is important from the viewpoint of understanding the stability of a protein. Here, we applied the stabilization-centre (SC) residues approach (Dosztányi *et al.*, 1997) to determine the spatial locations of densely interresidue-contacted clusters in the four IGPS proteins. The long-range interacting SC pairs in the IGPS structures were determined using the *SCide* program (Dosztányi *et al.*, 2003; http://www.enzim.hu/scide; Tables 6 and S1, Figs. S2 and S3). The SC residues in IGPS are relatively rigid (have lower *B* factors), which might reflect tight packing of the SC surroundings (Halle, 2002). The SC pairs in IGPS are mainly found in the core β-sheets or in the outer α-helices and loops, while a few outer-core SC contacts have been observed (Table S1; Fig. S2). Thus, the interior and outer elements of IGPS tend to interact with each other. Some of the SCs act as anchor residues forming two or more SC pairs that simultaneously connect different parts of the IGPS framework. Notably, in the thermostable IGPS proteins SCs are spread out over the structural framework and comprise more residues when compared with *Ec*IGPS (Tables 6 and S1; Figs. 2*a*, S2 and S3). The SC residues which have high surrounding hydrophobicity and conservation scores (named stabilization residues; SR) were revealed by the program *SRide* with default cutoff values (Magyar *et al.*, 2005; Gromiha *et al.*, 2004; Tables 6, S1 and S2). The SR residues are located mainly in the barrel strands and their numbers correlate with the T_o of a particular IGPS.

According to the definition of SC, there are two triplets, the central interacting SC residues plus two additional residues, one on each flanking tetrapeptide side, which form seven or more of the possible nine contacts (Dosztányi *et al.*, 1997). These cooperative noncovalent interactions of the nearest-neighbour structural environment stitch together the spatially nearby peptide segments. We considered the area composed of the SC pair as the central interacting residues and the tetrapeptide of residues neighbouring each SC residue as the SC cluster. This allowed us to graphically reveal the densely interacting structural areas of the proteins (Figs. 2*a* and 4). The densely packed SC pairs in the core β-strands yield the globular β-strand core cluster, which occupies about 16% of the total volume of the IGPS protein. Since the core SC cluster is almost identical in the four IGPSs, it was deemed to be chiefly responsible for the overall folding of the protein. The edges of the core cluster spread to the flanking $\alpha_i\beta_{i+1}$ and $\beta_i\alpha_i$ loops and in some cases further throughout the loops to the outer α-helices (Fig. 2*a*; Table S1). This is likely to be the mechanism by which the central globule SC cluster guides the arrangement of outer α-helices.

In contrast to the core SC cluster, the outer clusters form a number of discrete non-overlapping clusters which occupy structurally nonsimilar locations among the proteins (Table S1; Figs. 2*a* and 4). The external clusters have a preference for the N–α_4 part of the proteins, where the N-terminal residues intensely cluster with the α_3 and α_4 helices of the barrel. Interactions within the outer SC cluster stabilize the mutual orientation of the helices. Additionally, the SC clusters coupled to the core cluster may fix helices relative to the protein core. The outer clusters occupy a total of 60, 43, 44 and 28% of the total volume of *Ss*IGPS, *Tm*IGPS, *Tt*IGPS and *Ec*IGPS, respectively. As thermophiles present additional and/or more extended (overlapped) cluster sizes, they are possibly related to the distinct temperature-adaptation of the proteins. In fact,

the SC cluster assembled from sequentially remote parts of the polypeptide chain can only be disintegrated by simultaneously breaking several interactions. Moreover, these individually weak interactions might act cooperatively to increase the kinetic barrier to unfolding, thus keeping the protein structure intact for an extended period and/or at higher temperature.

The abundance of SC-cluster residues in the thermostable IGPSs compared with *Ec*IGPS may indicate a large effect of SCs on the thermophilicity. In the vicinity of the active site, the SC cluster should help in retaining the conformational features of the active-site residues that are required to bind the substrate and catalyze chemistry at high temperatures. At the same time, the presence of additional SC clusters close to the active site of the thermophilic IGPS proteins should restrict the conformational fluctuations that are necessary for catalytic function at low temperatures. In general, thermophilic enzymes are stable and fully active at elevated temperatures but are not functional at room temperature (Hecht *et al.*, 1989; Jaenicke & Bohm, 1998; Georis *et al.*, 2000). It was found that the catalytic efficiency of *Tm*IGPS at 298 K is 25-fold greater than that of *Ss*IGPS (Merz *et al.*, 1999). The lower number of cluster residues and different arrangement of SC clusters as well as of salt bridges (Figs. 3 and 4) possibly allow the catalytic centre of *Tm*IGPS to be more flexible and therefore more active than that of *Ss*IGPS at 298 K.

The decreased SC clustering in *Ec*IGPS indicates that some interactions are modified or lost in the mesophilic protein compared with the thermophilic proteins (Table S1; Figs. 2*a* and 4). The mutation of topologically equivalent residues in the mesophilic protein to provide the additional SC cluster residues found in the thermophilic proteins should increase protein stability. A mutation that retains the nature of the residue as well as providing additional interactions will be important for the structure of the mesophilic protein to gain the additional stability. Since the denaturation process is known to start with unfolding of the outer surface, which leads to exposure of the hydrophobic core (Caflisch & Karplus, 1994), stabilizing the protein surface with dense interactions may stabilize the entire protein. Thus, knowledge of the outer SC-cluster structure of IGPS from thermophiles may be useful for designing additional clusters in the mesophilic protein by the replacement of a few elements.

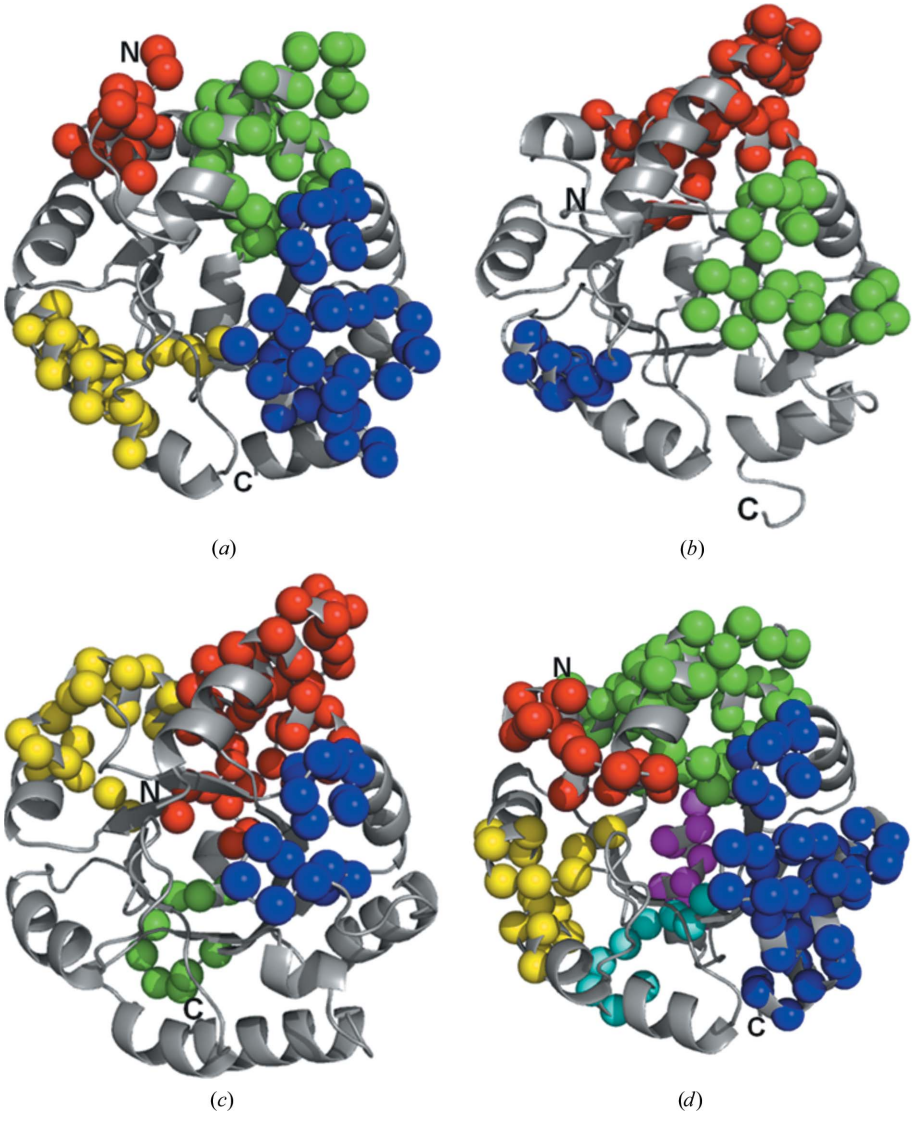

Figure 4
Distribution of the external SC clusters in the four IGPSs: (*a*) *Tt*IGPS, (*b*) *Ec*IGPS, (*c*) *Tm*IGPS, (*d*) *Ss*IGPS. The SC clusters identified are shown in a sphere representation at the C^{α} positions. Non-overlapping clusters are coloured differently. The protein molecules are shown in cartoon representation and coloured grey. The N- and C-termini of the polypeptide chains are highlighted.

Both hydrophobic and hydrophilic types of amino acids form significant fractions of the external SC clusters (Table S2). As the location of hydrophobic residues on solvent-accessible protein surfaces is unfavourable, the compositional features of SC clusters may be related to the dense interactions within the clusters. It is well known that extensive interactions of CH_2 groups of hydrophilic and hydrophobic amino acids cage hydrophobic residues by surrounding hydrophilic side chains to exclude them from solvent (Tisi & Evans, 1995; Van den Burg *et al.*, 1994; Selvaraj & Gromiha, 2003). At the same time, the charged residues reduce repulsive electrostatic interactions and favourably solvate the polar end groups thanks to these contacts.

Thus, we used the SC approach to detect densely interacting clusters in the present four IGPS structures. Essentially, this approach is based on atomic close-contact interactions and ultimately

Table 6
Stabilization centres (SC), stabilization residues (SR) and SC clusters in the four IGPSs.

	*Ss*IGPS	*Tm*IGPS	*Tt*IGPS	*Ec*IGPS
No. of SC residues	61 (25%)	44 (18%)	58 (23%)	54 (21%)
No. of hydrophobic SC residues	46	32	43	34
No. of SC residues in the strands	27	26	31	32
Sequence range of SC residues	Δ240 (Arg3–Ile243)	Δ235 (Ile15–Thr250)	Δ239 (Arg2–Arg241)	Δ215 (Arg19–Leu234)
No. of SC–SC interactions	54	36	56	51
B factor of SC residues (Å^2)	20.81	30.99	18.15	53.68
B factor beside SC residues (Å^2)	24.74	39.33	21.77	60.45
No. of SRs	20 (8.0%)	20 (8.0%)	19 (7.5%)	17 (6.6%)
No. of hydrophobic SRs	18	18	15	12
No. of SRs in the strands (%)	50	67	79	71
Sequence range of SRs	Δ181 (Ile48–Ala229)	Δ181 (Ala46–Ala227)	Δ185 (Val48–Val233)	Δ180 (Leu52–Gly232)
No. of SR–SR interactions	13	9	11	13
B factor of SRs (Å^2)	18.07	25.84	14.13	50.17
B factor beside SRs (Å^2)	24.27	41.07	21.49	59.01
No. of SC-cluster residues	204	178	187	167
No. of outer SC-cluster residues	157 (64%)	106 (42%)	115 (45%)	77 (30%)
B factor of C$^\alpha$ in outer SC clusters (Å^2)	23.39	39.82	22.54	66.46
No. of core SC cluster residues	47	70	72	90
B factor of C$^\alpha$ in core SC cluster (Å^2)	20.64	32.82	116.58	49.12
No. of residues beside SC clusters	43	75	67	92
B factor of C$^\alpha$ beside SC clusters (Å^2)	28.24	40.47	22.88	60.59

the structure is divided into common inner and several outer SC clusters. The central, conserved and globular cluster is probably a dominant driving force for protein folding, while the numerous outer differing clusters might be involved in the stabilization of the fold.

4. Conclusion

The crystal structure of thermophilic *Tt*IGPS was solved at 1.8 Å resolution and used for further analysis of the structure-stability factors of IGPS proteins. Analysis of the structural parameters of the hyperthermophilic *Ss*IGPS and *Tm*IGPS, the thermophilic *Tt*IGPS and the mesophilic *Ec*IGPS shows that, in contrast to other classes of proteins, there is little evidence for a correlation between hydrophobic packing and hydrogen bonds with the relative thermostability of the four IGPS structures. However, the composition of charged amino-acid residues and structural rigidity, as well as the numbers and distributions of salt bridges and SC clusters, do appear to correlate with thermostability. IGPS thermostability involves the formation of additional and improved salt bridges and SC-cluster interactions across a wider range of residues in the sequences. The number and the sequence range of these interactions increase in a continuous manner from the mesophilic to thermophilic to hyperthermophilic proteins. Both salt-bridge and SC-cluster interactions impose connectivity constraints on conformational motions of external and internal parts of the proteins. While an SC cluster is a locally dense area with multiple interactions, the salt bridges force backbone segments to jointly increase the local atomic density. Generally, the characteristics of hyperthermophilic IGPS proteins are the augmented numbers of salt bridges and SC clusters and their unbiased distribution over the entire structure, thereby fixing the protein parts more strongly relative to the thermophilic and mesophilic counterparts. The outer regions of the IGPS protein are stabilized mainly by the formation of salt bridges and discrete SC clusters, while the barrel strands of the inner core are stabilized mainly by the formation of globular SC clusters. The patchwork of external SC clusters in the IGPS structures is distinct for each protein. The additional SC clusters and/or enlarged SC-cluster networks found on the thermophilic protein surface indicate a special role in promoting thermo-adaptation of the protein. The present cluster description provides a potential effective strategy for converting a starting three-dimensional structure of a mesophilic protein into a thermophilic one. The approach relies on estimation of the SC clusters of the source and target proteins based on the van der Waals radii of residue atoms and replacement by a suitable conformer of a residue that enhances and/or improves local interactions to create extra SC clusters. Since mutations are made only to residues that are on or very close to the surface of the protein, it is highly unlikely that they will affect the three-dimensional architecture of the protein. Thus, proteins designed following an outer SC-cluster strategy are expected to fold and also to function properly.

Although the data set considered in this study was limited to four IGPS proteins, the reliable observation of extra outer SC-cluster residues in thermophilic proteins is significant. Since $(\beta/\alpha)_8$-barrel structures have high structural similarity, we believe that the extra outer SC clusters would work to increase thermal stability in general. The tendency of a protein to expand outer SC segments may be a natural step-by-step approach to adapt itself to higher temperatures. The identified features of SC clusters and salt bridges which make IGPS more adaptable to high temperature should be helpful for future biotechnological applications and provide a useful guide for site-directed mutagenesis aimed at improving the thermostability of $(\beta/\alpha)_8$-barrel proteins.

The authors would like to thank the staff of RIKEN Genomic Sciences Center for providing the plasmid. This work was supported by the 'National Project of Protein

Structural and Functional Analysis' funded by the MEXT of Japan. We thank Dr W. C. Stallings and Professor D. Beckett for critical reading of the manuscript.

References

Andreotti, G., Cubellis, M. V., Palo, M. D., Fessas, D., Sannia, G. & Marino, G. (1997). *Biochem. J.* **323**, 259–264.

Andreotti, G., Tutino, M. L., Sannia, G., Marino, G. & Cubellis, M. V. (1994). *Biochim. Biophys. Acta*, **1208**, 310–315.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

Barlow, D. J. & Thornton, J. M. (1983). *J. Mol. Biol.* **168**, 867–885.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

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.

Caflisch, A. & Karplus, M. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 1746–1750.

Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.

Connolly, M. L. (1993). *J. Mol. Graph.* **11**, 139–141.

Creighton, T. E. & Yanofsky, C. (1966). *J. Biol. Chem.* **241**, 4616–4624.

Daggett, V. & Levitt, M. (1993). *J. Mol. Biol.* **232**, 600–619.

Darimont, B., Stehlin, C., Szadkowski, H. & Kirschner, K. (1998). *Protein Sci.* **7**, 1221–1232.

DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.

Desiraju, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.

Dosztányi, Z., Fiser, A. & Simon, I. (1997). *J. Mol. Biol.* **272**, 597–612.

Dosztányi, Z., Magyar, C., Tusnády, G. & Simon, I. (2003). *Bioinformatics*, **19**, 899–900.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. & Liang, J. (2006). *Nucleic Acids Res.* **34**, W116–W118.

Elcock, A. H. (1998). *J. Mol. Biol.* **284**, 489–502.

Farber, G. K. (1993). *Curr. Opin. Struct. Biol.* **3**, 409–412.

Farias, S. T. & Bonato, M. C. (2003). *Genet. Mol. Res.* **2**, 383–393.

Frishman, D. & Argos, P. (1995). *Proteins*, **23**, 566–579.

Funahashi, J., Takano, K., Yamagata, Y. & Yutani, K. (1999). *Protein Eng.* **12**, 841–850.

Gallivan, J. P. & Dougherty, D. A. (1999). *Proc. Natl Acad. Sci. USA*, **96**, 9459–9464.

Georis, J., de Lemos Esteves, F., Lamotte-Brasseur, J., Bougnet, V., Devreese, B., Giannotta, F., Granier, B. & Frère, J.-M. (2000). *Protein Sci.* **9**, 466–475.

Gromiha, M. M., Pujadas, G., Magyar, C., Selvaraj, S. & Simon, I. (2004). *Proteins*, **55**, 316–329.

Guerois, R., Nielsen, J. E. & Serrano, L. (2002). *J. Mol. Biol.* **320**, 369–387.

Halle, B. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 1274–1279.

Hecht, K., Wrba, A. & Jaenicke, R. (1989). *Eur. J. Biochem.* **183**, 69–74.

Hennig, M., Darimont, B. D., Jansonius, J. N. & Kirschner, K. (2002). *J. Mol. Biol.* **319**, 757–766.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). *Structure*, **3**, 1295–1306.

Höcker, B., Jürgens, C., Wilmanns, M. & Sterner, R. (2001). *Curr. Opin. Biotechnol.* **12**, 376–381.

Hooft, R. W., Sander, C. & Vriend, G. (1996). *Proteins*, **26**, 363–376.

Jaenicke, R. (1996). *FASEB J.* **10**, 84–92.

Jaenicke, R. & Bohm, G. (1998). *Curr. Opin. Struct. Biol.* **8**, 138–748.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). *Trends Biochem. Sci.* **23**, 403–405.

Knöchel, T. R., Hennig, M., Merz, A., Darimont, B., Kirschner, K. & Jansonius, J. N. (1996). *J. Mol. Biol.* **262**, 502–515.

Knöchel, T. R., Pappenberger, A., Jansonius, J. N. & Kirschner, K. (2002). *J. Biol. Chem.* **277**, 8626–8634.

Kollman, P., McKelvey, J., Johansson, A. & Rothenberg, S. (1975). *J. Am. Chem. Soc.* **97**, 955–965.

Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Magyar, C., Gromiha, M. M., Pujadas, G., Tusnády, G. E. & Simon, I. (2005). *Nucleic Acids Res.* **33**, W303–W305.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Merz, A., Knöchel, T., Jansonius, J. N. & Kirschner, K. (1999). *J. Mol. Biol.* **288**, 753–763.

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

Parthasarathy, S. & Murthy, M. R. (1997). *Protein Sci.* **6**, 2561–2567.

Parthasarathy, S. & Murthy, M. R. (2000). *Protein Eng.* **13**, 9–13.

Priestle, J. P., Grütter, M. G., White, J. L., Vincent, M. G., Kania, M., Wilson, E., Jardetzky, T. S., Kirschner, K. & Jansonius, J. N. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 5690–5694.

Russell, R. J., Ferguson, J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). *Biochemistry*, **36**, 9983–9994.

Schneider, B., Knöchel, T., Darimont, B., Hennig, M., Dietrich, S., Babinger, K., Kirschner, K. & Sterner, R. (2005). *Biochemistry*, **44**, 16405–16412.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. (2005). *Nucleic Acids Res.* **33**, W382–W388.

Seiler, P., Weisman, G. R., Glendening, E. D., Weinhold, F., Johnson, V. B. & Dunitz, J. D. (1987). *Angew. Chem.* **99**, 1216–1218.

Selvaraj, S. & Gromiha, M. M. (2003). *Biophys. J.* **84**, 1919–1925.

Smith, N. N. & Gallagher, D. T. (2008). *Acta Cryst.* F**64**, 886–892.

Smith, D. K., Radivojac, P., Obradovic, Z., Dunker, A. K. & Zhu, G. (2003). *Protein Sci.* **12**, 1060–1072.

Sugahara, M. & Miyano, M. (2002). *Tanpakushitsu Kakusan Koso*, **47**, 1026–1032.

Szilágyi, A. & Závodszky, P. (2000). *Structure Fold Des.* **8**, 493–504.

Takano, K., Yamagata, Y. & Yutani, K. (1998). *J. Mol. Biol.* **280**, 749–761.

Tanaka, H., Chinami, M., Mizushima, T., Ogasahara, K., Ota, M., Tsukihara, T. & Yutani, K. (2001). *J. Biochem.* **130**, 107–118.

Thomas, A. S. & Elcock, A. H. (2004). *J. Am. Chem. Soc.* **126**, 2208–2214.

Tina, K. G., Bhadra, R. & Srinivasan, N. (2007). *Nucleic Acids Res.* **35**, W473–W476.

Tisi, L. C. & Evans, P. A. (1995). *J. Mol. Biol.* **249**, 251–258.

Tiwari, A. & Panigrahi, S. K. (2007). *In Silico Biol.* **7**, 651–661.

Umeyama, H. & Morokuma, K. J. (1977). *J. Am. Chem. Soc.* **99**, 1316–1332.

Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.

Van den Burg, B., Dijkstra, B. W., Vriend, G., Van der Vinne, B., Venema, G. & Eijsink, V. G. (1994). *Eur. J. Biochem.* **220**, 981–985.

Vieille, C. & Zeikus, G. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 1–43.

Vihinen, M. (1987). *Protein Eng.* **1**, 477–480.

Vihinen, M., Torkkila, E. & Riikonen, P. (1994). *Proteins*, **19**, 141–149.

Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 631–643.

Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.

Wahl, M. C. & Sundaralingam, M. (1997). *Trends Biochem. Sci.* **22**, 97–102.

Wiederstein, M. & Sippl, M. J. (2005). *J. Mol. Biol.* **345**, 1199–1212.

Wiegel, J. & Adams, M. W. W. (1998). *Thermophiles – the Keys to Molecular Evolution and The Origin of Life?* London: Taylor & Francis.

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). *Nucleic Acids Res.* **31**, 3316–3319.

Wilmanns, M., Priestle, J. P., Niermann, T. & Jansonius, J. N. (1992). *J. Mol. Biol.* **223**, 477–507.

Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.