2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile *Sulfolobus solfataricus*: possible determinants of protein stability

Michael Hennig[1], Beatrice Darimont[2†], Reinhard Sterner[2], Kasper Kirschner[2] and Johan N Jansonius[1*]

[1]Department of Structural Biology and [2]Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstr. 70, CH-4056 Basel, Switzerland

Background: Recent efforts to understand the basis of protein stability have focussed attention on comparative studies of proteins from hyperthermophilic and mesophilic organisms. Most work to date has been on either oligomeric enzymes or monomers comprising more than one domain. Such studies are hampered by the need to distinguish between stabilizing interactions acting between subunits or domains from those acting within domains. In order to simplify the search for determinants of protein stability we have chosen to study the monomeric enzyme indole-3-glycerol phosphate synthase from the hyperthermophilic archaeon *Sulfolobus solfataricus* (sIGPS), which grows optimally at 90°C.
Results: The 2.0 Å crystal structure of sIGPS was determined and compared with the known 2.0 Å structure of the IGPS domain of the bifunctional enzyme from the mesophilic bacterium *Escherichia coli* (eIGPS). sIGPS and eIGPS have only 30% sequence identity, but share high structural similarity. Both are single-domain $(\beta/\alpha)_8$ barrel proteins, with one (eIGPS) or two (sIGPS) additional helices inserted before the first β strand. The thermostable sIGPS has many more salt bridges than eIGPS. Several salt bridges crosslink adjacent α helices or participate in triple or quadruple salt-bridge clusters. The number of helix capping, dipole stabilizing and hydrophobic interactions is also increased in sIGPS.
Conclusions: The higher stability of sIGPS compared with eIGPS seems to be the result of several improved interactions. These include a larger number of salt bridges, stabilization of α helices and strengthening of both polypeptide chain termini and solvent-exposed loops.

Structure 15 December 1995, **3**:1295–1306
Key words: archaea, $(\beta/\alpha)_8$ barrel, helix stabilization, salt bridges

Introduction

Three steps in tryptophan biosynthesis are catalyzed by enzymes of the $(\beta/\alpha)_8$ barrel topology, also designated the triosephosphate isomerase (TIM) barrel fold [1,2]. In *Escherichia coli*, two of these enzymes, indole-3-glycerol phosphate synthase (IGPS) and *N*-(5′-phosphoribosyl)-anthranilate isomerase (PRAI), are joined covalently to give a bifunctional enzyme (eIGPS:PRAI). Its X-ray structure was determined several years ago [3], and subsequently refined at 2.0 Å resolution [4]. The artificially separated recombinant eIGPS and ePRAI domains have practically the same enzymatic properties as in the bifunctional context [5].

In order to understand the stability of proteins from hyperthermophilic organisms, we decided to investigate the indole-3-glycerol phosphate synthase from the archaeon *Sulfolobus solfataricus* (sIGPS). This microorganism grows in hot, acidic and sulfurous mud pools in the Solfatara volcanic crater, near Naples, Italy and is therefore adapted to life at temperatures of about 90°C [6]. Accordingly, sIGPS is very thermostable [7]. In contrast to eIGPS, which is naturally fused to ePRAI, sIGPS is monofunctional and monomeric [8]. A structure-based sequence alignment of eIGPS and sIGPS is given in Figure 1, which also defines the numbering convention used in the structural comparisons. eIGPS and sIGPS share about 30% sequence identity.

Inspection of the three-dimensional structures of thermostable proteins has highlighted features that might account for their higher stability [9–20]. However, several reports either focus on the structure of a single hyperthermostable form [14] or compare two forms, the melting temperatures of which differ by only 20°C [9,11,12]. One of the few structural comparisons between a hyperthermostable and a thermolabile form of a protein was reported for glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The comparison of GAPDH from the hyperthermophilic bacterium, *Thermotoga maritima*, with that of the psychrophile lobster, *Homarus americanus*, revealed additional intramolecular salt bridges in the thermostable enzyme as potentially stabilizing interactions [13]. However, GAPDH is a homotetramer, the protomers of which consist of two clearly defined domains, and the structural analysis was limited to 2.5 Å resolution. Reduced lengths of loops and improved subunit interactions were observed in a comparison of thermostable and thermolabile citrate synthases [15]. Further structural comparisons at high resolution of

*Corresponding author. †Present address: Department of Cellular and Molecular Pharmacology, University of California, San Francisco, CA 94143-0450, USA.

© Current Biology Ltd ISSN 0969-2126

```
                    α0                                  α00          β1                  α1          β2
                HHHHHHHHHHH                         HHHHHHHHHH     SSSSS                HHHHHHHH     SSSSS
          1        10        20                  30        40        50        60        70        80        90
sIGPS     MPRYLKGWLKDVVQLSLRRPSFRAS----------RQRPIISLNERILEFNKRNITAIIAEYKRKSPS-GLDFER-DPIEYSKFMERYAVGLSILTEEKYFNGS
                |   |         |                |                || | § §  §§§ §      ||          ||   § §| |||§ §|
eIGPS        MQTVLAKIVADKAIWVEARKQQQPLASFQNEVQPSTRHFYDALQ----GARTAFILECKKASPSKGVIRDDFDPARIAAIYKHYASAISVLTDEKYFQGS
             1        10        20        30        40            50        60        70        80        90
                HHHHHHHHHHHHHHHHHHH                                SSSSSS                HHHHHHHH     SSSSS
                         α0                                         β1                    α1          β2

              α2         β3         α3         β4           α4             β5      α5          β6             α6
          HHHHHHHHHH    SSSSS    HHHHHHHHHH   SSSSS     HHHHHHHHHHHHH     SSSS  HHHHHHHHH    SSSSS          HHHHHH
                100       110       120       130       140       150       160       170       180       190
sIGPS     YETLRKIASSVSIPILMKDFIVKESQIDDAYNLGADTVLLIVKILTERELESLLEYARSYGMEPLIEINDENDLDIALRIGARFIGINSRDLETLEINKENQRK
             |         ||| §|§|    ||  |    ||  §|    |        |   | |  |  | |   |      |   ||   §|§ §|§  | |      |
eIGPS     FNFLPIVSQIAPQPILCKDFIIDPYQIYLARYYQADACLLMLSVLDDDQYRQLAAVAHSLEMGVLTEVSNEEEQERAIALGAKVVGINNRDLRDLSIDLNRTRE
             100       110       120       130       140       150       160       170       180       190       200
          HHHHHHHHHHH  SSSSS    HHHHHHHHH    SSSSS     HHHHHHHHHHHH    SSSSSS  HHHHHHHHH    SSSSSS          HHHHH
              α2         β3         α3         β4           α4             β5      α5          β6             α6

                     β7         α7         β8   α8′      α8
          HH       SSSSS     HHHHHHHHH    SSSS HHHHH   HHHHHH
             200       210       220       230       240
sIGPS     LISMIPSNVVKVAESGISERNEIEELRKLGVNAFLIGSSLMRNPEKIKEFIL----
          |      ||    |§§|       ||      | |§|§| ||
eIGPS     LAPKLGHNVTVISESGINTYAQVRELSHFA-NGFLIGSALMAHDDLHAAVRRVLLG
                     210       220       230       240       250
          H        SSSSS     HHHHHHH      SSSS HHHH    HHHHHHHHH
                     β7         α7         β8    α8′       α8
```

Fig. 1. Sequence alignment and position of secondary structural elements in sIGPS and eIGPS. The alignment of the amino acid sequences is based on the superposition of the three-dimensional structures of IGPS from *E. coli* and *S. solfataricus*. The single-letter code for amino acid residues is used. The amino acid sequence of sIGPS was deduced from the sequence of the *trp*C coding region [8]. The numbering scheme and the assignment and position of the secondary structural elements is shown above and below the sequence for sIGPS and eIGPS, respectively. Identical residues between sIGPS and eIGPS are indicated by (|) and invariant residues for IGPSs from various sources are marked by (§).

thermostable and thermolabile forms of monomeric proteins possessing a single domain are needed for more detailed comparative deductions.

Here we report the heterologous expression, purification, crystallization and crystal structural analysis of sIGPS at 2.0 Å resolution. The three-dimensional superposition of sIGPS and eIGPS reveals great structural similarity for the main-chain fold in the core of the molecules. The greater thermostability of sIGPS relative to eIGPS appears to be attributable to a combination of improved stabilization of the N-terminal extension of the polypeptide chain, a significant increase in the number of salt bridges, better electrostatic compensation of helix macrodipole charges, and an increased number of residues with positive dihedral angles at the C termini of several helices.

Results and discussion

Preparation of sIGPS

*Trp*C, the gene for sIGPS, was amplified from genomic DNA of *S. solfataricus* by the polymerase chain reaction (PCR) using complementary primers derived from the published DNA sequence [8]. The amplified fragment was cloned into the expression vector pDS56/RBS II [21], to yield the clone pDS SS-1. After expression in a strain of *E. coli* that lacks the entire *trp* operon, sIGPS was found exclusively in the soluble cell extract. During the purification, advantage was taken of the thermostability of sIGPS to remove most of the host proteins by heating the soluble cell extract at 75°C for 10 min. Further purification steps included ion-exchange chromatography on DEAE Sepharose and gel filtration on Sephacryl S-200. The final yield was about 5 mg of pure sIGPS per liter of bacterial culture. Gel electrophoresis in the presence of SDS demonstrated a band of the expected molecular mass (M_r 28 600) and >95% purity. The identity of the recombinant sIGPS was confirmed by N-terminal protein sequencing (Fig. 1).

Biochemical properties of sIGPS and comparison with eIGPS:PRAI

It is of interest to compare the enzymatic activity of the thermostable form (sIGPS) to that of the relatively thermolabile form (eIGPS:PRAI). Because eIGPS:PRAI denatures rapidly at temperatures above 40°C only sIGPS was investigated at both 25°C and 60°C. Measurements at temperatures above 60°C are inaccurate due to the chemical instability of the substrate 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5-phosphate (CdRP).

At pH 7.5 and 25°C, the k_{cat} value is 36-fold lower for sIGPS (0.033 s^{-1}) than for eIGPS:PRAI (1.2 s^{-1}), and the K_M value for the substrate CdRP is 2-fold lower (0.045 μM *versus* 0.090 μM), so that the efficiency parameter k_{cat}/K_M^{CdRP} is about 18-fold lower for the thermostable enzyme. Moreover, sIGPS has a lower dissociation constant than eIGPS:PRAI for the product

indole-3-glycerol phosphate (IGP) (K_p^{IGP}= 0.015 μM for sIGPS and 0.6 μM for eIGPS:PRAI). At 60°C the k_{cat} value of sIGPS is ~30-fold higher than at 25°C, whereas K_M^{CdRP} increases only 1.5-fold. Thus, the catalytic efficiency is increased by a factor of 20 at 60°C, confirming the observation of Andreotti *et al.* [7] that the specific activity of sIGPS increases with increasing temperature.

sIGPS unfolds slowly and aggregates upon addition of the denaturant guanidinium chloride (data not shown). Therefore, we used the irreversible loss of enzymatic activity as a biologically significant, albeit operational, criterion of thermostability. Figure 2 shows that the inactivation is an exponential process for both eIGPS in the bifunctional enzyme eIGPS:PRAI and sIGPS. These homologous proteins decay with similar half-lives at very different temperatures: 1.8 min at 44°C for eIGPS:PRAI and 4.4 min at 89°C for sIGPS. These data confirm the slow thermal inactivation of sIGPS observed by Andreotti *et al.* [7] at 80°C in another buffer at pH 8.5.

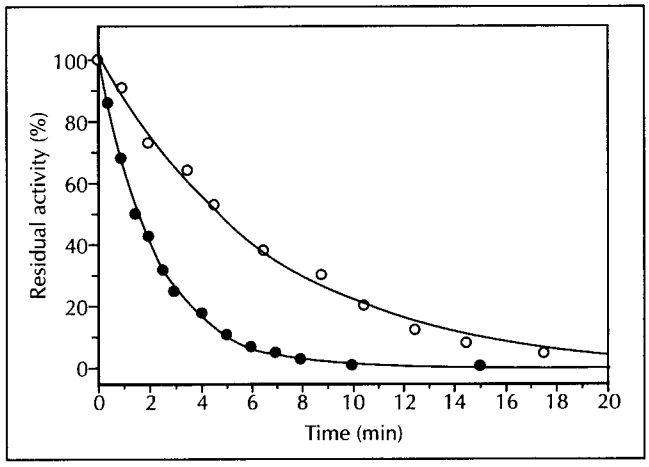

Fig. 2. The difference in thermal stabilities of sIGPS and eIGPS:PRAI. sIGPS (○) was incubated at 89°C (pH 7.0) and eIGPS:PRAI (●) was incubated at 44°C (pH 7.2). (See the Materials and methods section for reaction conditions.) After the indicated incubation times, aliquots were chilled on ice, and the residual IGPS activity determined as described in the Materials and methods section. Fits of the data points (—) to a single exponential decay gave half-lives of 4.4 min for sIGPS at 89°C and 1.8 min for eIGPS:PRAI at 44°C.

Crystallization of sIGPS

Crystals appeared within three days and grew to a maximum size of 0.3×0.3×2.0 mm^3 in about one week. They belong to the orthorhombic space group $P2_12_12_1$ with unit cell dimensions a=58.0 Å, b=73.8 Å and c=104.3 Å, α=β=γ=90°. On the basis of the space-group symmetry, unit cell volume (4.5×10^5 $Å^3$) and molecular weight of 28 600 Da, a packing density (V_m) of 3.90 $Å^3$ Da^{-1} (or 1.95 $Å^3$ Da^{-1}) can be calculated for one (or two) molecules in the asymmetric unit, respectively. Self-rotation calculations did not indicate non-crystallographic symmetry. Crystal density measurements supported a value for V_m of 3.90 $Å^3$ Da^{-1}, corresponding to one molecule in the asymmetric unit and a solvent volume fraction of 68.5%. This relatively low density for the crystal of a globular protein was later confirmed by the crystal structure, which reveals large solvent channels. The limited contact areas between individual molecules do not prevent formation of a well-ordered and stable crystal lattice. The active site faces a wide solvent channel, a favourable situation for further investigation of the catalytic mechanism by diffusing ligands into the crystal (MH, BD, KK and JNJ, unpublished data).

Description of the overall structure

sIGPS consists of an eight-stranded parallel β barrel core, surrounded by an outer layer of eight parallel α helices (Fig. 3). The canonical $(\beta/\alpha)_8$ barrel of sIGPS begins with Ile48 (Fig. 1), and is preceded by an N-terminal extension of 46 residues. The N-terminal residue of the polypeptide chain, Pro2 (labelled as -N- in Fig. 3), lies over helix α4 and is hydrogen bonded to the core of the molecule. Residues 7–17 and 33–42 form α helices (α0 and α00). The loop that connects the two helices runs along the contact area of the helices α2 and α3, while α00 covers the N-terminal end of the β barrel (bottom of Fig. 3). The only additional deviation from the canonical $(\beta/\alpha)_8$ barrel is the extra helix, helix α8′, that is located in the loop β8–α8. It is also present in eIGPS, ePRAI and the α subunit of tryptophan synthase [22] and its N terminus is the binding site of the phosphate-ester moiety of substrate and product (see below). In several other TIM-barrel proteins, such as narbonin, concanavalin B and hevamine, this additional helix is also present but is apparently not involved in substrate binding [23]. The distribution of observed dihedral angles (Fig. 4) shows that most residues are clustered in the energetically favoured regions, attesting to the reliability of the structure.

Figure 3 also shows the superposition of sIGPS onto eIGPS [3,4]. sIGPS has the same $(\beta/\alpha)_8$ barrel fold as eIGPS with a root mean square deviation (rmsd) of 1.73 Å for 239 optimally superimposed Cα atoms (as calculated using 'O' [24]). The structure-based sequence alignment of the first 45 residues of sIGPS and eIGPS (Fig. 1) differs from that presented previously [7]. For example, only Leu5, Val9 and Arg19 of eIGPS are conserved in sIGPS, and one non-overlapping gap needs to be introduced into each of the two sequences. Although the major portion of the earlier alignment [7] is confirmed, maximizing the number of identical residues turned out to be an unreliable criterion for optimal alignment. The partitioning of residues into α helices and β strands is very similar: 40% and 15% for sIGPS, compared with 36% and 16% for eIGPS. The secondary structural elements are identical in length or differ, at most, by two residues (Figs 1,3). A net increase in secondary structure is not the reason for the improved thermostability of sIGPS, even though sIGPS possesses an additional α helix, helix α00, compared with eIGPS.

The first five residues of sIGPS form a loop that is connected through several hydrogen bonds, including two salt bridges, to the core of the molecule. In contrast, the first

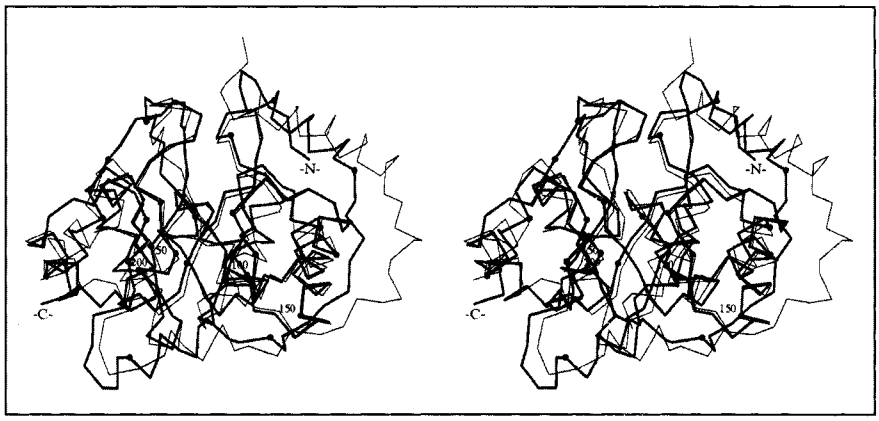

Fig. 3. Stereo drawing of superimposed Cα traces of sIGPS (thick lines) and eIGPS (thin lines) viewed with the β barrel axis vertical in the plane of the figure. For the sIGPS structure, the N- and C-terminal residues (2 and 248, respectively) are labelled, every tenth residue is marked by a closed circle, and every fiftieth residue is numbered. (Drawn using MOLSCRIPT [52].)

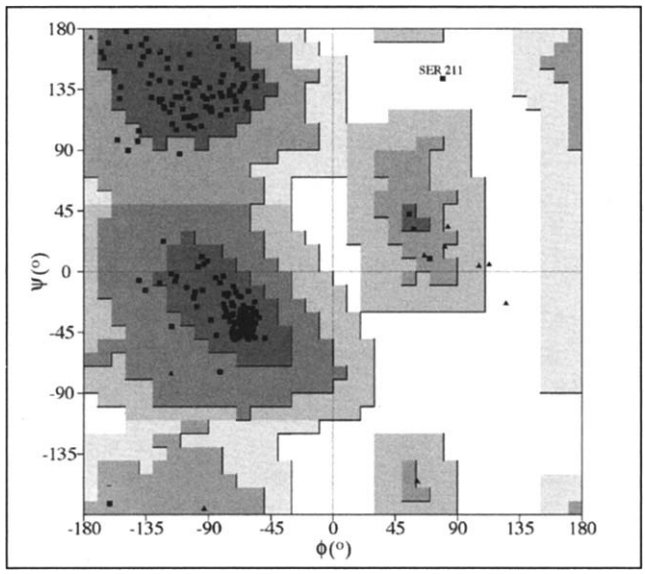

Fig. 4. Ramachandran diagram of the distribution of dihedral angles φ and ψ for the final structure of sIGPS. Glycine residues are represented by triangles, all other residues by squares. Dark areas correspond to the energetically most favourable φ,ψ angles (see text). The single residue (Ser211) with disallowed φ,ψ angles is labelled. (Drawn with PROCHECK [51].)

four residues of the polypeptide chain in eIGPS are fully solvent-exposed and rather flexible. Helix α0 of sIGPS is shorter than its equivalent in eIGPS (see Fig. 1). In addition, the part of the polypeptide chain that connects helix α0 to the first residue of strand β1 is less exposed to the solvent in sIGPS than that of eIGPS. Moreover, the part that runs across the N-terminal end of the β barrel forms a new helix designated α00 in sIGPS, where the corresponding segment of eIGPS has no secondary structure. In summary, the N-terminal extension of sIGPS is itself more structured and fits more tightly to the $(\beta/\alpha)_8$ barrel than that of eIGPS. This difference may contribute significantly to the higher thermostability of sIGPS.

Phosphate ion binding site

As in the case of eIGPS [4], a phosphate ion was detected in sIGPS during crystallographic refinement (Fig. 5). Its position, which is close to the N-terminal turn of helix α8′, corresponds to the potential binding site of the phosphate-ester moiety of the substrate and is highlighted in Figure 6. Both sIGPS and eIGPS were crystallized from a solution containing ~1.3 M ammonium sulphate and 0.05 M phosphate. Because the stereochemistry and chemical properties of phosphate and sulphate ions are rather similar we cannot exclude the alternative possibility that sulphate ions bind at this site (T Knöchel, MH, BD, A Merz, KK and JNJ, unpublished data). The ion has well-defined electron density in sIGPS and its coordination and hydrogen-bonding network to the protein and bound water molecules are presented in Figure 5. The phosphate ion connects two neighbouring loops (β7–α7 and β8–α8) via hydrogen bonds to backbone amide groups. One of the peptide amide groups hydrogen bonding to the ion is provided by Gly212 from the loop β7–α7. The adjacent residue, Ser211, which is invariant in IGPSs (see Fig. 1), is in a disallowed region of the Ramachandran plot (Fig. 4), but was unambiguously traced in the electron density. Ser215, the equivalent residue in eIGPS, has similar dihedral angles [4]. The side chain of Lys53 (located in loop β1–α1) forms a salt bridge to the ion, and several water molecules complete the coordination (Fig. 5).

Comparison of the phosphate-binding site of sIGPS with that of eIGPS reveals both similarities and differences. The key differences are: firstly, the greater distance between the phosphate ion and loop β7–α7 in eIGPS such that only water-mediated hydrogen bonds are possible; and secondly, that the loop β1–α1 in eIGPS contributes an additional positively charged side chain (Lys61) to the coordination of the phosphate ion [4], the corresponding residue being deleted in sIGPS (Fig. 1). The similarities include Lys53 (Lys55 in eIGPS) and the N-terminal turn of α8′ which have almost identical positions in the two structures and display identical hydrogen-bonding patterns. In summary, the hydrogen-bond network around the bound phosphate ion restricts the mobilities of the loops β1–α1, β7–α7 and β8–α8 in sIGPS and eIGPS, both directly and via water molecules.

Surface accessibility

Recently, an extremely thermostable form of aldehyde ferredoxin oxidoreductase was found to have a significantly lowered solvent-accessible surface to volume ratio when compared with several non-homologous proteins

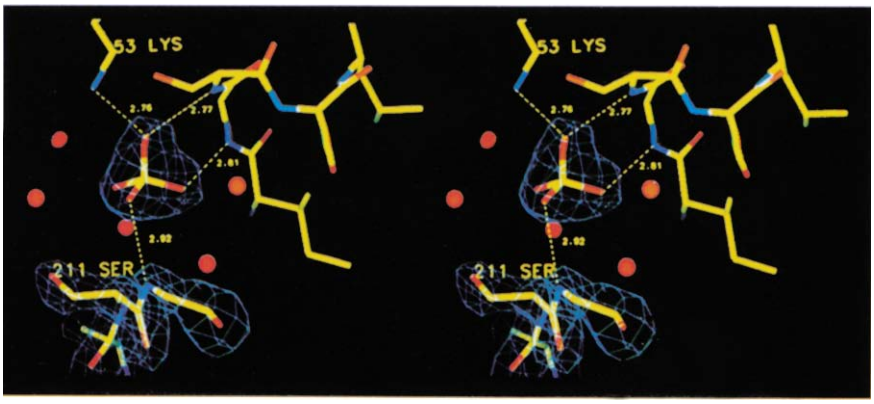

Fig. 5. Stereoview of the phosphate ion binding site and residue Ser211 in sIGPS. Ser211 is the only residue with ϕ,ψ angles in a 'forbidden' region of the Ramachandran plot (Fig. 4). The electron density in blue (contoured at 4σ) was calculated with coefficients (F_o–F_c) and α_c phases from the refined structure with the phosphate ion and residues 210–212 omitted. Atoms are shown in stick representation coloured yellow, red and blue for carbons, oxygens and nitrogens, respectively. Water molecules are shown as red spheres. (Drawn with 'O' [24].)

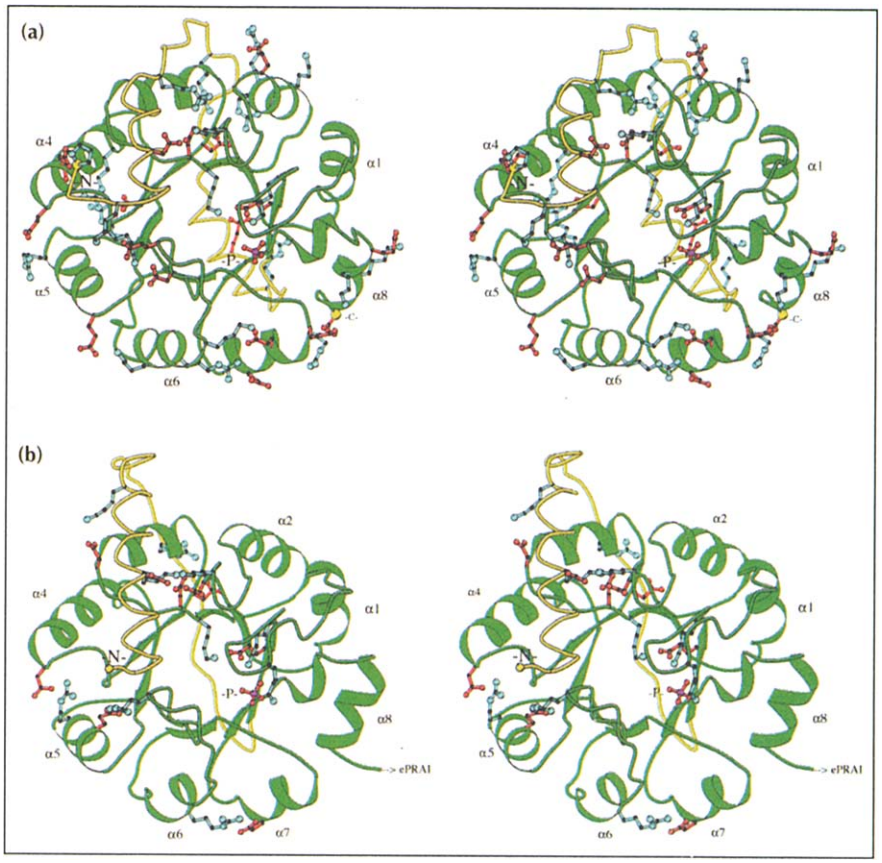

Fig. 6. Comparison of salt bridges in sIGPS and eIGPS. Negatively and positively charged side chains are coloured red and blue, respectively. The TIM-barrel fold is drawn in green, with the N-terminal extension preceding strand β1 in yellow. The N and C termini of the polypeptide chains are highlighted by yellow spheres. **(a)** sIGPS has a network of surface salt bridges, which mainly connect adjacent α helices. **(b)** eIGPS has far fewer salt bridges. The salt bridges around the phosphate ion (active site) are, however, well conserved. (Drawn with MOLSCRIPT [52].)

selected from the Protein Data Bank [14]. It is plausible that a reduced surface area and optimized packing of the atoms in the core of a structure increases protein stability. In particular, the influences of internal cavities and packing of internal side chains on protein stability have been discussed in the literature [16–18,25]. However, it is difficult to assess the degree of packing merely from the surface to volume ratio because the latter is strongly dependent on the size of the molecule. Proteins of high molecular weight would, in general, be more stable by this criterion. Likewise, multidomain proteins with their smaller solvent-accessible surface area (due to interdomain contacts) would be expected to be more stable than single-domain proteins [26].

The accessible surface area for the monomeric sIGPS molecule, comprising 2004 protein atoms, is 10 900 Å^2 (as calculated by DSSP [27]). The eIGPS domain, which is naturally fused to the ePRAI domain, comprises 2001 protein atoms. Within the bifunctional enzyme, the eIGPS domain has an accessible surface area of 10 530 Å^2 (the surface area of eIGPS which is covered by the ePRAI domain comprises 870 Å^2). Therefore, sIGPS as a single-domain monofunctional protein has a larger accessible surface area than eIGPS fused to ePRAI.

The ratio A_o/A_c, where A_o and A_c are the observed and calculated surface areas, respectively, and $A_c = 15.0 \times N^{0.866}$ (N is the total number of protein atoms in the molecule as described in [14]), has a value of 1.00 for sIGPS. This value is in keeping with those observed for other proteins with the $(\beta/\alpha)_8$ fold (e.g. 0.90 for narbonin from *Vicia narbonensis* and 1.05 for concanavalin B from *Canavalia ensiformis*; MH, unpublished data). Thus,

for sIGPS at least, thermostability is not correlated with a reduced accessible surface area.

Salt bridges

The contribution of salt bridges in making a protein thermostable has been discussed for more than 20 years [28]. Salt bridges have been shown to stabilize the α-helical conformation of short model peptides [29]. If the residues contributing to the bridge reside on adjacent secondary structural elements, salt bridges appear to stabilize the tertiary structures of certain proteins [13,18,30–33]. But there are also reports which negate a contribution of salt bridges to thermostability [34]. IGPS provides a good model system for making comparative deductions from thermostable and relatively thermolabile forms, because crystal structures at 2.0 Å resolution are available for both eIGPS [4] and sIGPS. Accurate X-ray structures are essential for salt bridge analysis, because side chains of charged residues frequently show only weak electron density in crystal structures determined at medium resolution. Table 1 presents the salt bridges of sIGPS and eIGPS, defined as hydrogen-bonded ion pairs. The average temperature factors of the side chains and the distances between hydrogen-bonded oxygen and nitrogen atoms are listed, together with the solvent accessibilities of the residues. This last parameter allows a more quantitative description of the salt bridges between the two extreme states — 'exposed' and 'buried'.

Table 1. Comparison of intramolecular salt bridges within sIGPS and eIGPS.

Positively charged side chain		Side chain B-factor* (Å^2)	Accessibility† (Å^2)	Position‡	Negatively charged side chain	Side chain B-factor* (Å^2)	Accessibility† (Å^2)	Position‡	Distance§ (Å)
sIGPS									
Pro2	NH#	24.1	266	N terminus	Glu141 Oε1	24.0	52	α4	3.31
Pro2	NH#	24.1	266	N terminus	Glu141 Oε2	24.0	52	α4	2.74
Lys6	Nζ	42.5	104	N–α0	Glu185 Oε2	55.2	82	β6–α6	3.53
Arg19	Nη1	21.2	20	α0–α00	Asp111 Oδ1	15.2	2	β3	3.05
Arg26	Nη2	23.7	39	α0–α00	Asp128 Oδ2	17.5	2	α3–β4	3.20
Arg28	Nη1	56.0	37	α0–α00	Asp128 Oδ1	17.5	2	α3–β4	2.81
Arg43	Nη1	63.5	132	α00–β1	Glu39 Oε2	58.3	88	α00	3.09
Lys53	Nζ	21.7	9	β1–α1	Glu51 Oε1	17.4	3	β1	3.17
Lys53	Nζ	21.7	9	β1–α1	PO_4^{2-}	19.0	–	–	2.88
Lys87	Nζ	34.4	110	β2–α2	Asp11 Oδ1	47.0	44	α0	3.11
Arg97	Nη1	53.1	98	α2	Glu94 Oε2	58.0	79	α2	3.15
Lys98	Nζ	47.2	118	α2	Glu94 Oε1	58.0	79	α2	2.85
Lys110	Nζ	19.6	24	β3	Glu51 Oε2	17.4	3	β1	2.73
Lys135	Nζ	29.9	30	β4–α4	Asp183 Oδ1	29.4	40	β6–α6	3.03
Arg150	Nε	19.1	50	α4	Glu155 Oε1	19.5	16	α4–β5	2.96
Arg150	Nη1	19.1	50	α4	Glu147 Oε1	48.0	106	α4	2.91
Arg150	Nη2	19.1	50	α4	Glu155 Oε2	19.5	16	α4–β5	2.84
Arg171	Nη1	39.4	141	α5	Glu139 Oε2	34.3	67	α4	2.63
Arg175	Nη1	48.8	146	α5–β6	Glu155 Oε1	19.5	16	α4–β5	3.19
Lys191	Nζ	39.9	68	α6	Glu218 Oε2	48.0	21	α7	2.79
Arg195	Nη1	59.1	50	α6	Glu218 Oε1	48.0	21	α7	3.20
Arg195	Nη1	59.1	50	α6	Glu221 Oε2	60.0	72	α7	3.17
Arg195	Nη2	59.1	50	α6	Glu221 Oε2	60.0	72	α7	3.25
Lys196	Nζ	39.7	71	α6	Glu163 Oε2	31.5	90	α5	2.93
Arg216	Nε	63.5	94	α7	Leu248 CO1**	55.8††	–	C terminus	3.03
Arg216	Nη2	63.5	94	α7	Leu248 CO2**	55.8††	–	C terminus	2.75
Lys242	Nζ	37.8	33	α8	Glu215 Oε2	58.4	86	β7–α7	3.60
Lys244	Nζ	27.7	94	α8	Glu241 Oε1	36.8	99	α8′–α8	3.14
eIGPS									
Lys12	Nζ	9.5	35	α0	Asp89 Oδ1	7.6	4	β2–α2	2.76
Lys12	Nζ	9.5	35	α0	Asp115 Oδ1	11.3	1	β3–α3	2.75
Lys20	Nζ	19.6	59	α0	Asp119 Oδ1	13.1	31	β3–α3	3.04
Lys55	Nζ	23.8	22	β1–α1	PO_4^{2-}	24.2	–	–	2.78
Lys55	Nζ	23.8	22	β1–α1	Glu53 Oε1	8.0	2	β1	3.62
Lys91	Nζ	11.4	132	β2–α2	Asp11 Oδ1	32.4	35	α0	2.56
Lys114	Nζ	10.1	25	β3	Glu53 Oε2	8.0	2	β1	2.77
Arg127	Nη1	14.1	17	α3	Asp132 Oδ1	10.1	3	α3–β4	3.47
Arg172	Nη2	8.9	46	α5	Asp143 Oδ2	28.4	64	α4	2.91
Arg189	Nη1	24.1	71	β6–α6	Glu169 Oε2	4.4	8	α5	3.09
Arg199	Nε	12.8	63	α6	Glu225 Oε1	14.4	75	α7	3.00

*The mean values are given for individual side chains. The mean B-factors for all side-chain atoms in sIGPS and eIGPS are 29.3 Å^2 and 23.1 Å^2, respectively. †Solvent-accessible surface of the residue. ‡The location of residues in secondary structural elements is as assigned in Figure 1. §All salt bridges with distances <3.7 Å between the nitrogen and the nearest oxygen are included. #α-amino group of N-terminal residue. **α-carboxylate group of C-terminal residue. ††Mean B-factor for the backbone atoms of the residue.

sIGPS has a total of 24 salt bridges, of which four (Pro2–Glu141, Arg150–Glu155, Arg195–Glu221 and Arg216–Leu248) are double hydrogen bonded. This compares with only 11 salt bridges in eIGPS. It is interesting that such a large number of salt bridges is maintained in the structure of sIGPS, in spite of electrostatic screening by the high concentration of ammonium sulphate (1.3 M) in the mother liquor. Very recently, the structure of sIGPS was re-determined in two new crystal forms, in each of which the protein molecules are packed differently. These crystal forms were grown with polyethylene glycol as the precipitating agent. In one case neither phosphate nor sulphate ions were present; in the second case only sulphate was present at 0.2 M. Almost no change in the number and pairing of salt bridges was observed (T Knöchel, MH, BD, A Merz, KK and JNJ, unpublished data).

The salt bridges of sIGPS are unevenly distributed over the molecule (see Fig. 6a, Table 1). Notably, both the N-terminal α-amino group and the C-terminal α-carboxylate group are stabilized by a salt bridge in sIGPS. Perhaps this measure prevents 'fraying' at the two termini. More importantly, sIGPS contains several unusual triple and quadruple clusters of salt bridges.

Five salt bridges fix the N-terminal extension (residues 1–47) containing the helices α0 and α00 as well as the intervening loop α0–α00 to the $(\beta/\alpha)_8$ core. The N terminus of helix α4 is clamped to the C terminus of helix α5 by the ion pair Glu139–Arg171. A similar clamp (Arg172–Asp143) occurs between helices α4 and α5 in eIGPS (Fig. 6b, Table 1). Another cluster of salt bridges is located in the active site of sIGPS and involves Glu51, Lys53 and Lys110 (Fig. 6a, Table 1), but this cluster is preserved in eIGPS (Fig. 6b, Table 1). The following salt bridges of sIGPS have no equivalent in eIGPS: the N terminus of helix α5 is clamped to the C terminus of helix α6 by Glu163–Lys196. Helices α6 and α7 are tied together by the two clamps Lys191–Glu218 and Arg195–Glu221. In eIGPS only a single salt bridge straddles helices α6 and α7 (Fig. 6b, Table 1). The loop β7–α7 in sIGPS is tied to the N terminus of helix α8 by Glu215–Lys242, and Arg216 at the N terminus of helix α7 is clamped to the α-carboxylate group of the C-terminal residue Leu248. In contrast, no salt bridges are observed between helices α7 and α8 in eIGPS (Fig. 6b, Table 1).

Two triple and two quadruple salt bridges are found only in the thermostable IGPS. Arg26 and Arg28 on loop α0–α00 interact with Asp128 on loop α3–β4; the interaction of Glu94 with both Arg97 and Lys98 probably stabilizes helix α2. In a quadruple salt bridge, Arg150 on helix α4 is hydrogen bonded to both Glu147 (α4) and Glu155 (loop α4–β5), and the latter is also hydrogen bonded to Arg175 (loop α5–β6). This cluster appears to both stabilize helix α4 and to link two adjacent loops. The salt bridges between Lys191 and Arg195 on helix α6 and Glu218 and Glu221 on helix α7 act as a double clamp between these two helices.

Not only is the number of salt bridges important for evaluating the contribution of these interactions to protein stability, but their structural context is of primary importance [17,25]. The rather different values for their contribution to stability estimated by mutagenesis studies [32,34] might be a direct consequence of the structural context of the interactions. It is apparent that stable salt bridges need to be buttressed by additional and sterically optimized side-chain or main-chain interactions. Such support, an example of which is presented in Figure 7, is required to compensate for the entropically unfavourable immobilization of the two side chains. Figure 7 shows that the loops α0–α00 (Arg26 and Arg28) and α3–β4 (Asp128) are connected by a triple salt bridge. In addition, the main-chain atoms of the C-terminal turns of α3 and α2 are linked to the charged side chains by hydrogen bonds. These interactions enlarge the hydrogen-bonding network in the examined region considerably and indicate that this cluster, which is absent in eIGPS, may stabilize the structure of sIGPS significantly.

Although salt bridges are much fewer in number in eIGPS than in sIGPS, the triple salt bridge in the active site is a feature common to both proteins. In contrast, most of the salt bridges stabilizing α helices and the α0–α00 loop in sIGPS are absent from eIGPS. The reduced number of salt bridges between adjacent helices of eIGPS is only partially compensated by the interdomain

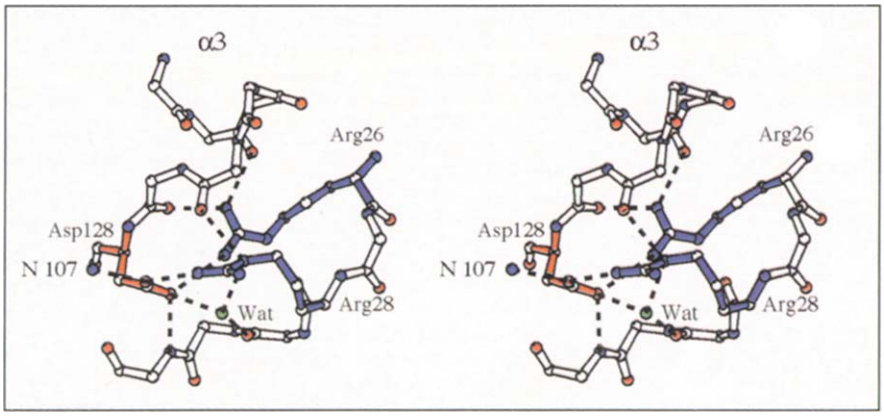

Fig. 7. Stereoview of the triple salt bridge between Arg26, Arg28 (bonds in blue) and Asp128 (red) and its support by hydrogen bonding. The following stretches of the polypeptide chain are interconnected by hydrogen bonds: loop α0–α00 to Arg26/Arg28; loop α3–β4 to Asp128 and to the main-chain carbonyl group of Gly126; helix α3 to the main-chain carbonyl group of Tyr123; strand β3 to the main-chain amino group of Ile107; helix α2 to the main-chain carbonyl group of Ala100 and a water molecule.

contact to ePRAI in the naturally bifunctional form of eIGPS. Only helices α7 and α8 of eIGPS contact ePRAI (see [4]). In summary, it appears that the larger number of salt bridges of sIGPS as well as the architectural role of several of these clamps between adjacent α helices might contribute markedly to the higher thermostability of sIGPS relative to eIGPS.

Helix capping and the compensation of dipoles

An α helix can be stabilized through charge compensation of the positive pole of its macrodipole (at its N terminus) by an adjacent carboxylate group, even if the latter does not hydrogen bond to the N-terminal residues of the α helix [17,29]. All eight helices of the TIM-barrel fold of sIGPS have at least one aspartate or glutamate residue in their first turn or in the N-cap position preceding the helix (Fig. 1). Helices α2 and α7 in eIGPS have no such stabilizing residues.

The loops connecting the secondary structural elements are of special interest in TIM-barrel enzymes. The loops αi–βi+1 are generally shorter and more regular than the loops βi–αi and seem to stabilize the tertiary conformation of the molecule [35]. In TIM-barrel proteins, several αi–βi+1 loops include a residue with a positive ϕ angle at a position designated the C-cap of the α helix. The positive dihedral angle allows the main-chain carbonyl oxygen to hydrogen bond with solvent molecules [36]. It is therefore assumed to stabilize α helices. This conformation, however, is unfavourable for all amino acids except glycine. In sIGPS, four loops have glycine and one has asparagine at this position (Table 2). The asparagine, Asn44, is located in the region of left-handed α helices in the Ramachandran plot (Fig. 4). eIGPS has two non-glycine residues with positive dihedral angles at equivalent positions.

Hydrophobic interactions

It is generally accepted that hydrophobic interactions contribute to protein thermostability. Therefore, changes in the size or packing of hydrophobic clusters influence the stability of proteins. The fraction of carbon atoms that are buried (zero solvent accessibility) gives a measure of overall hydrophobic interaction within a protein [25]. Of the 1273 carbon atoms in sIGPS, 773 (60.7%) are buried. Likewise, 779 of the 1280 carbon atoms in eIGPS (60.8%) are buried. Thus, in this respect there is no difference between sIGPS and eIGPS.

Table 2. C-cap residues with positive dihedral angles in sIGPS.

Loop position	Residue in sIGPS	Dihedral ϕ,ψ (°)	Residue in eIGPS	Dihedral ϕ,ψ (°)
α00–β1	Asn44	+52,+46	–*	–
α3–β4	Gly126	+115,+7	Gln130	+86,+6
α4–β5	Gly153	+80,+20	Glu157	+60,+57
α5–β6	Gly173	+84,+29	Gly177	+75,+59
α7–β8	Gly226	+106,+4	Ala230	–108,+30

The positions and backbone ϕ,ψ angles of the corresponding residues in eIGPS are also shown. *Helix α00 is not present in eIGPS.

A well-packed hydrophobic cluster covers the N-terminal ends of the β barrels of both sIGPS and eIGPS. The side chains of the first layer inside the β barrel, from residues located in strands β2, β4, β5 and β6 [4], are mostly hydrophobic, small and poorly conserved (Gly79, Thr129, Phe176 and Ala229 in sIGPS; Ala83, Ala133, Val180 and Gly232 in eIGPS), but the replacements are compensatory in terms of volume. The cavity is filled by hydrophobic and bulky side chains from the segment preceding strand β1. Interestingly, Leu33 and Ile37 in sIGPS occupy almost the same positions as Phe39 and Leu43 in eIGPS, although the local conformations are very different (helix α00 *versus* loop segment). The next layer of side chains of the β barrel is formed by invariant residues (Ser81, Leu131, Gly178 and Leu231 in sIGPS; Ser85, Leu135, Gly182 and Leu234 in eIGPS). In summary, the hydrophobic cluster in the N-terminal half of the β barrel core is quite conserved and cannot be responsible for differential stabilization of sIGPS over eIGPS.

A further hydrophobic cluster is located at the C-terminal end of the barrel (Fig. 8). It is formed in sIGPS by side chains from the beginning of helix α0, strand β4 (Ile133) and the following loops: β2–α2 (Tyr88), β3–α3 (Phe112) and β4–α4 (Ile136). The only tryptophan residue in sIGPS (Trp8) is located in the centre of the cluster and is surrounded by at least eight hydrophobic side chains. The participation of Leu5, Trp8 and Leu9 in the hydrophobic interaction clearly stabilizes the conformation of helix α0 and fixes it tightly to the C-terminal face of the β barrel. In eIGPS, Val4 takes the position of Trp8 which may result in somewhat weaker interactions and a smaller contribution to the stabilization of helix α0.

Biological implications

Investigations of the thermostability of proteins have revealed a variety of potentially stabilizing properties. Because these forces contribute to stability in all proteins, their relative importance in stabilizing hyperthermostable proteins is of considerable theoretical and practical interest. Therefore, a comparison of functionally identical and structurally similar proteins adapted to life at very different temperatures is instructive. Indole-3-glycerol phosphate synthase (IGPS) is a relatively simple model protein for comparative deductions, because it is monomeric and has only a single domain that belongs to the widespread TIM-barrel family. Comparison of the high-resolution X-ray structures of IGPS from the hyperthermophile *Sulfolobus solfataricus* (sIGPS) and the mesophile *Escherichia coli* (eIGPS) indicates features which can explain the greatly enhanced thermal stability of sIGPS.

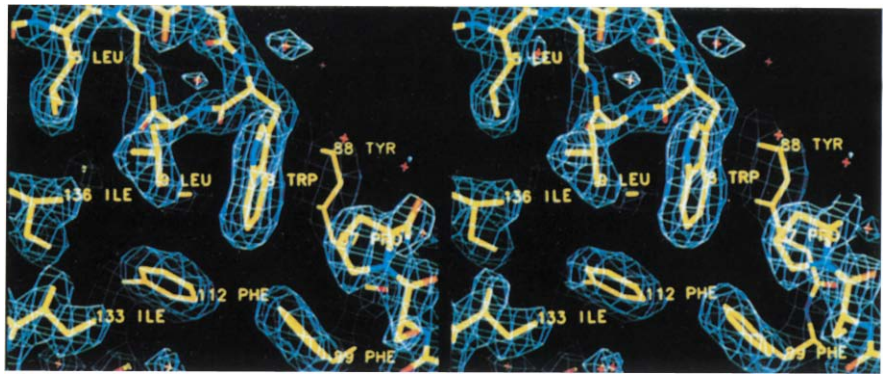

Fig. 8. Stabilizing contacts of the N-terminal part of the polypeptide chain in sIGPS with the core of the molecule. The stereoview shows the hydrophobic cluster surrounding Trp8. The electron density in blue (contoured at 1.3σ) was calculated with coefficients ($2F_o$–F_c) and α_c phases from the refined structure. Atoms are shown in stick representation coloured yellow, red and blue for carbons, oxygens and nitrogens, respectively. (Drawn with 'O' [24].)

sIGPS has 24 salt bridges and eIGPS only 11. A closer inspection reveals the importance of structural context in evaluating their potentially stabilizing role in sIGPS. Many salt bridges connect either adjacent helices or neighbouring loops, setting up a network of structural clamps. Some salt bridges form clusters which include three or four charged residues and maintain hydrogen bonds to main-chain atoms or to uncharged side chains.

However, the stability of sIGPS is also increased by the improved stability of the ends of the polypeptide chain in sIGPS. The N-terminal extension to the canonical TIM-barrel is attached more tightly to the core of the protein in sIGPS than in eIGPS. Helices are stabilized in sIGPS by negatively charged residues in the N-terminal turn and by the exchange to glycine of residues at the C-cap position. Other properties, such as an increased proportion of secondary structure or a reduced accessible surface area, have been proposed to play important roles in stabilizing other thermostable proteins. Because these factors are similar in sIGPS and eIGPS, it is clear that there are distinct ways for proteins to achieve high thermostability.

Materials and methods

DNA constructions and protein purification

*Trp*C of *S. solfataricus* [8] was amplified from the genomic DNA of strain M4 (a kind gift of Dr C Schleper) by PCR using the following two synthetic oligonucleotides as primers: SS*trp*C-1 (5′-AACGCATGCCACGTTATCTTAAAGG-3′) containing an *Sph*I site (underlined) and SS*trp*C-2 (3′-TTCTAATTTCTTAAATATGATATCGACGTCAAA-5′) containing a *Pst*I site. The amplified fragment was cloned into the corresponding sites of the expression vector pDS 56/RBS II [21] to yield pDS SS-1. Sequencing of the insert of pDS SS-1 confirmed the expected DNA sequence, with the exception of a silent T→A transversion in codon 213 (Ile ATT→ATA). The *trp*C gene in pDS SS-1 is transcribed from the T5 promotor that is regulated by the *lac* operator [21]. *E. coli* strain W3110 *trp*R *tna*2 Δ*trp*EA2 [37], which lacks the entire *trp* operon, was cotransformed with both pDS SS-1 and with the plasmid pDM. The latter encodes the *lac* repressor constitutively [21]. Expression of *trp*C is induced by isopropyl thiogalactoside (IPTG), which binds to (and inactivates) the *lac* repressor. Transformed cells were grown at 37°C in Luria broth medium supplemented with 0.1 mg ml^{-1} ampicillin (for pDS SS-1) and 0.025 mg ml^{-1} kanamycin (for pDM). Expression was induced with 1 mM IPTG after the cell culture reached a density of OD_{600}=0.6 and cells were harvested by centrifugation after a further 5 h incubation.

For protein purification the cells were suspended at 0°C in buffer (0.1 M potassium phosphate buffer at pH 7.8, containing 2 mM EDTA, 1 mM 1,4-dithioerythritol (DTE) and 0.3 mM phenylmethylsulfonylfluoride) and sonicated. Protamine sulphate solution was added to the soluble fraction to precipitate nucleic acids [38], which were removed by centrifugation. Heating the supernatant for 10 min at 75°C precipitated the major portion of contaminating proteins. The suspension was centrifuged, and the supernatant was dialyzed and subsequently resolved with a linear gradient (0.005–0.1 M) of the above buffer (with the same supplements) on a column of DEAE Sepharose fast flow. sIGPS, which eluted around 0.018 M buffer concentration, was concentrated by ultrafiltration (Amicon PM-10 membranes) and purified to better than 95% purity by gel filtration on Sephacryl-S200 in the above buffer. The bifunctional enzyme from *E. coli* (eIGPS:PRAI) was purified as described previously [38].

Analytical methods

The purification achieved by the individual steps of the procedure was monitored by SDS-PAGE, and by analytical gel filtration on a calibrated Superose 12 column as described in [39] sIGPS gave a single peak of UV-absorbing material (data not shown). The elution of the nearly spherical proteins was retarded, for unknown reasons, leading to spuriously low values of apparent molecular mass (19 000 instead of 28 600 for sIGPS and 21 000 instead of 28 900 for eIGPS). The extinction coefficient ($E_{280}^{0.1\%}$=0.606 cm^2 mg^{-1}) was determined by second derivative UV spectroscopy [40].

Steady-state enzyme kinetics made use of the fluorescence of the reaction product IGP by monitoring the fluorescence excited at 280 nm and emitted at 350 nm [41]. The buffer used previously (Tris-HCl [5]) was replaced by *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propane sulfonic acid) (HEPPS) buffer because, at pH 7.5 and 25°C, sIGPS was more active in 0.05 M HEPPS than in 0.05 M Tris-HCl (data not shown).

To avoid product inhibition by IGP, k_{cat} and K_M^{CdRP} were determined by analysis of the dependence of initial reaction velocity on the substrate concentration [42]. The product inhibition constant, K_P^{IGP}, was determined separately by fitting the entire progress curve of the practically irreversible reaction catalyzed by IGPS to the integrated form of the rate equation [5,43].

For measuring the kinetics of thermal inactivation, samples of sIGPS and eIGPS:PRAI were incubated at 89°C and 44°C, respectively. Both proteins had a concentration of 4 μM in 0.05 M HEPPS, 4 mM KMgEDTA, 2 mM dithiothreitol (DTT), pH=7.0 at 89°C for sIGPS and pH=7.2 at 44°C for eIGPS:PRAI [44]. Aliquots were taken after different times and chilled on ice. Subsequently, the residual IGPS activities were measured in the above buffer at 50°C (sIGPS) and 25°C (eIGPS:PRAI).

In the case of sIGPS, due to the heat-lability of the substrate, mainly initial velocities were measured. For normalization, entire progress curves were recorded at enzyme concentrations that allowed the reaction to be completed within 3 min. Entire progress curves were recorded for eIGPS:PRAI and analyzed according to [5,43]. This analysis showed that K_M^{CdRP} and K_p^{IGP} of the remaining active enzyme were unchanged after the heat treatment, which confirms that the surviving fraction of enzyme is identical to the native enzyme.

Crystallization

The purified enzyme was dialyzed against 50 mM potassium phosphate buffer at pH 5.0 containing 5 mM EDTA and 2 mM DTE. Crystallization was carried out with the vapour diffusion method in hanging drops. At 4°C, the dialyzed enzyme solution containing 20 mg ml^{-1} protein was mixed with an equal volume of the reservoir solution prepared in the dialysis buffer and 26–30% saturated ammonium sulphate (1.0–1.2 M [45]) as precipitant. Crystal density measurements were performed using a gradient of xylene and carbon tetrachloride [46].

Data collection

The needle-shaped crystals were mounted in glass capillaries and X-ray diffraction data were measured on a MAR-Research image plate area detector system mounted on a modified Elliott GX-20 rotating anode generator (CuKα, 40 kV and 50 mA). A rotation angle of 1° was used for each exposure. The crystals were oriented with the b* axis about parallel to the rotation axis and rotated over at least 100° during data collection. Data evaluation was performed with the MOSFLM program package [47]. The integrated intensities were scaled with the program ROTAVATA and merged using AGROVATA. The reflection intensities were finally converted to structure-factor amplitudes using TRUNCATE [47]. Details of data collection parameters and statistics of the measured data are summarized in Table 3.

Molecular replacement

Truncated models based on the coordinates of the eIGPS domain of the bifunctional enzyme (R-factor 17.3%, rmsds from bond length 0.010 Å, Brookhaven PDB code 1PII) were used in molecular replacement computations with X-PLOR [48]. The N-terminal residues (1–48) and C-terminal residues (238–248; numbered according to sIGPS unless otherwise stated) were deleted from the search model because of lack of sequence homology for these residues. In addition, non-homologous side chains were trimmed to alanine or glycine.

A Patterson integration radius of 5 Å and a resolution range 10–3.5 Å did not lead to a clear solution. Therefore, the 30 highest cross-rotation peaks were tried in Patterson correlation refinement, consisting of 20 steps of rigid-body refinement of the molecule. A maximum was found for the third highest peak of the cross-rotation function. Translation-function calculations using all data between 10.0 Å and 3.5 Å gave a solution 1.6σ higher than the second peak. The initial R-factor at this resolution was 51.3 %.

Table 3. X-ray data and refinement statistics.

No. of crystals	3
Measured reflections	205 248
Unique reflections	29 607
Resolution range (Å, all data)	15.0–2.0
Redundancy	6.9
$\langle I/\sigma \rangle$	9.2
Reflections >3σ (%)	81.3
Completeness (%)	96.0
R_{sym}*	0.073
No. of protein atoms	2003
No. of water molecules	239
No. of phosphate ions	1
Rms distances (1–2) (Å)	0.009
Rms bond angles (°)	1.6
Residues with most favoured φ,ψ angles (%)	94.2
R-factor† (10.0–2.0 Å, all data)	0.177
R-factor† (10.0–2.0 Å, data >3σ)	0.169
Average B-factor for protein atoms (Å^2)	25.5
Average B-factor for water molecules (Å^2)	51.2

$^*R_{sym}=\sum_{hkl}\sum_i |I(hkl)_i-\langle I(hkl)\rangle| / \sum_{hkl}\sum_i \langle I(hkl)\rangle$

†R-factor$=\sum_{hkl} |F(hkl)_o-F(hkl)_c| / \sum_{hkl} F(hkl)_o$

Refinement and quality of the structure

The correctly positioned molecule was refined for 10 cycles as a single rigid entity using reflections in the resolution range 14.0–4.0 Å, which reduced the R-factor to 47.4%. Further refinement of 15 rigid segments (each segment contained one β strand or α helix) lowered the R-factor to 43.5%. After energy-minimization of the model ('preparation step') and calculation of the weight, a simulated annealing protocol was applied. The initial temperature of 3000 K was decreased in time steps of 0.5 fsec to 300 K resulting in an R-factor of 32.5% for all data between 14 Å and 2.5 Å. $2F_o-F_c$ and F_o-F_c electron-density maps were calculated and a first round of manual corrections of the model was performed using the molecular graphics program FRODO [49]. After positional and B-factor refinement a new electron-density map was calculated and the model corrected and extended. A phosphate ion could be identified and was included in subsequent refinement cycles. After 12 rounds of manual rebuilding, positional and B-factor refinement, and inclusion of 239 water molecules, a final R-factor of 17.7% for all data between 10.0 Å and 2.0 Å was obtained.

Quality parameters of the final model are summarized in Table 3. The full polypeptide chain could be traced unambiguously. The model comprises all 247 residues of the protein, as well as 239 water molecules and a bound phosphate ion. Isolated, residual electron density was observed in a solvent region bordering the loop between strand β1 and helix α1 (designated loop β1–α1) which could not be explained by water molecules. The density even appeared again in an electron-density map based on a data set collected from a crystal grown after extensive dialysis of the native protein followed by denaturation (with guanidinium hydrochloride), renaturation (by dialysis) and subsequent crystallization. Further experiments are under way to identify its origin.

A Ramachandran diagram (Fig. 4) [50,51] for the molecular structure shows that the φ,ψ values for all non-glycine residues

except Ser211 fall within energetically most favoured (94.2%) or additionally allowed (5.3%) regions. Ser211 is located close to the phosphate ion binding site and its conformation is similar in eIGPS. Figure 5 shows this part of the structure superimposed on the corresponding electron density.

The coordinates for sIGPS have been deposited with the Brookhaven Protein Data Bank (entry code 1IGS).

Acknowledgements: The authors thank C Schleper for a gift of *S. solfataricus* cells, and HM Szadkowski and R Müller for excellent technical assistance in purification and crystallization, respectively, T Schirmer for discussions and G Kleemann and T Knöchel for help in preparing figures. Thanks are due to R Jaenicke for making available a manuscript prior to publication. This work was supported in part by the Swiss National Science Foundation grants 31-36432.92 to JNJ and 31-32369.91 to KK.

References

1. Banner, D.W., *et al.*, & Waley, S.G. (1975). Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* **255**, 609–614.
2. Reardon, D. & Farber, G.K. (1995). The structure and evolution of α/β barrel proteins. *FASEB J.* **9**, 497–503.
3. Priestle, J.P., *et al.*, & Jansonius, J.N. (1987). Three-dimensional structure of the bifunctional enzyme *N*-(5′-phosphoribosyl)anthranilate isomerase–indole-3-glycerol-phosphate synthase from *Escherichia coli. Proc. Natl. Acad. Sci. USA* **84**, 5690–5694.
4. Wilmanns, M., Priestle, J.P., Niermann, T. & Jansonius, J.N. (1992). Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerol-phosphate synthase from *Escherichia coli* refined at 2.0 Å resolution. *J. Mol. Biol.* **223**, 477–507.
5. Eberhard, M., Tsai-Pflugfelder, M., Bolewska, K., Hommel, U. & Kirschner, K. (1995). Indoleglycerol phosphate synthase-phosphoribosyl anthranilate isomerase: comparison of the bifunctional enzyme from *Escherichia coli* with engineered monofunctional domains. *Biochemistry* **34**, 5419–5428.
6. de Rosa, M., Gambacorta, A. & Bu'lock, J.D. (1975). Extremely thermophilic acidophilic bacteria convergent with *Sulfolobus acidocaldarius. J. Gen. Microbiol.* **86**, 156–164.
7. Andreotti, G., Tutino, M.L., Sannia, G., Marino, G. & Cubellis, M.V. (1994). Indole-3-glycerol-phosphate synthase from *Sulfolobus solfataricus* as a model for studying thermostable TIM-barrel enzymes. *Biochim. Biophys. Acta* **1208**, 310–315.
8. Tutino, M.L., Scarano, G., Marino, G., Sannia, G. & Cubellis, M.V. (1993). Tryptophan biosynthesis genes *trp*EGC in the thermoacidophilic archaebacterium *Sulfolobus solfataricus. J. Bacteriol.* **175**, 299–302.
9. Skarzynski, T. & Wonacott, A. (1988). Coenzyme induced conformational changes in D-glyceraldehyde-3-phosphate dehydrogenase from *Bacillus stearothermophilus. J. Mol. Biol.* **203**, 1097–1118.
10. Day, M.W., *et al.*, & Rees, D.C. (1992). X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus. Protein Sci.* **1**, 1494–1507.
11. Davies, G.J., Gamblin, S.J., Littlechild, J.A. & Watson, H.C. (1993). The structure of a thermally stable 3-phosphoglycerate kinase and a comparison with its mesophilic equivalent. *Proteins* **15**, 283–289.
12. Stark, W., Pauptit, R.A., Wilson, K.S. & Jansonius, J.N. (1992). The structure of neutral protease from *Bacillus cereus* at 0.2-nm resolution. *Eur. J. Biochem.* **207**, 781–791.
13. Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of holo-glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 511–521.
14. Chan, M.K., Mukund, S., Kletzin, A., Adams, M.W.W. & Rees, D.C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463–1469.
15. Russell, R.J.M., Hough, D.W., Danson, M.J. & Taylor, G.L. (1994). The crystal structure of citrate synthase from the Archaeon, *Thermoplasma acidophilum. Structure* **2**, 1157–1167.
16. Jaenicke, R. (1991). Protein stability and molecular adaptation to extreme conditions. *Eur. J. Biochem.* **202**, 715–728.
17. Matthews, B.W. (1993). Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **62**, 139–160.
18. Jaenicke, R., Schurig, H., Beaucamp, N. & Ostendorp, R. (1996). Structure and stability of "hyperstable proteins": glycolytic enzymes from the hyperthermophilic bacterium *Thermotoga maritima. Adv. Protein Chem.*, in press.
19. Teplyakov, A.V. & Wilson, K.S. (1990). Crystal structure of thermitase at 1.4 Å resolution. *J. Mol. Biol.* **214**, 261–279.
20. Ishikawa, K., *et al.*, & Morikawa, K. (1993). Crystal structure of ribonuclease H from *Thermus thermophilus* HB8 refined at 2.8 Å resolution. *J. Mol. Biol.* **230**, 529–542.
21. Certa, U., *et al.*, & Mous, J. (1986). Subregions of a conserved part of the HIV gp41 transmembrane protein are differently recognized by antibodies of infected individuals. *EMBO J.* **5**, 3051–3056.
22. Wilmanns, M., Hyde, C.C., Davies, D.R., Kirschner, K. & Jansonius, J.N. (1991). Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* **30**, 9161–9169.
23. Hennig, M., Jansonius, J.N., Terwisscha van Scheltinga, A., Dijkstra, B.W. & Schlesier, B. (1995). Crystal structure of concanavalin B at 1.65 Å resolution. An "inactivated" chitinase from seeds of *Canavalia ensiformis. J. Mol. Biol.* **254**, 237–246.
24. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
25. Honig, B. & Yang, A.-S. (1995). Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27–55.
26. Miller, S., Lesk, A.M., Janin, J. & Chothia, C. (1987). The accessible surface area and stability of oligomeric proteins. *Nature* **328**, 834–836.
27. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
28. Perutz, M. & Raidt, H. (1975). Stereochemical basis of heat stability in bacterial ferredoxins and in hemoglobin A2. *Nature* **255**, 256–259.
29. Chakrabartty, A. & Baldwin, R.L. (1995). Stability of α-helices. *Adv. Protein Chem.* **46**, 141–173.
30. Walker, J.E., Wonacott, A.J. & Harris, J.I. (1980). Heat stability of a tetrameric enzyme, D-glyceraldehyde-3-phosphate dehydrogenase. *Eur. J. Biochem.* **108**, 581–586.
31. Kelly, C.A., Nishiyama, M., Ohnishi, Y., Beppu, T. & Birktoft, J.J. (1993). Determinants of protein thermostability observed in the 1.9 Å crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus. Biochemistry* **32**, 3913–3922.
32. Horovitz, A., Serrano, L., Avron, B., Bycroft, M. & Fersht, A.R. (1990). Strength and cooperativity of contributions of surface salt bridges to protein stability. *J. Mol. Biol.* **216**, 1031–1044.
33. Shih, D.T., Luisi, B.F., Miyazaki, G., Perutz, M.F. & Nagai, K. (1993). A mutagenic study of the allosteric linkage of His(HC3)146b in haemoglobin. *J. Mol. Biol.* **230**, 1291–1296.
34. Sali, D., Bycroft, M. & Fersht, A.R. (1991). Surface electrostatic interactions contribute little to stability of barnase. *J. Mol. Biol.* **220**, 779–788.
35. Urfer, R. & Kirschner, K. (1992). The importance of surface loops for stabilizing an eightfold βα barrel protein. *Protein Sci.* **1**, 31–45.
36. Serrano, L., Sancho, J., Hirshberg, M. & Fersht, A.R. (1992). α-Helix stability in proteins. *J. Mol. Biol.* **227**, 544–559.
37. Schneider, W.P., Nichols, B.P. & Yanofsky, C. (1981). Procedure for production of hybrid genes and proteins and its use in assessing significance of amino acid differences in homologous tryptophan synthetase α polypeptides. *Proc. Natl. Acad. Sci. USA* **78**, 2169–2173.
38. Kirschner, K., Szadkowski, H., Jardetsky, T.S. & Hager, V. (1987). Phosphoribosylanthranilate isomerase-indoleglycerol-phosphate synthase from *Escherichia coli. Methods Enzymol.* **142**, 386–397.
39. Eder, J. & Kirschner, K. (1992). Stable substructures of eightfold βα-barrel proteins: fragment complementation of phosphoribosyl anthranilate isomerase. *Biochemistry* **31**, 3617–3625.
40. Levine, R.L. & Federici, M.M. (1982). Quantitation of aromatic residues in proteins: model compounds for second-derivative spectroscopy. *Biochemistry* **21**, 2600–2606.
41. Hankins, C.N., Largeu, M. & Mills, S.E. (1975). A rapid spectrofluorimetric assay for indoleglycerol phosphate synthase. *Anal. Biochem.* **69**, 510–517.
42. Cornish-Bowden, A. & Wharton, C.W. (1988). *Enzyme Kinetics.* IRL Press Ltd., London.
43. Hommel, U., Eberhard, M. & Kirschner, K. (1995). Phosphoribosyl anthranilate isomerase catalyzes a reversible Amadori reaction. *Biochemistry* **34**, 5429–5439.
44. Good, N.E. & Izawa, S. (1972). Hydrogen ion buffers. *Methods Enzymol.* **24**, 53–68.

45. Di Jeso, F. (1968). Ammonium sulphate conversion monogram for 0°C. *J. Biol. Chem.* **243**, 2022–2023.
46. Mikol, V. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins: Measuring Crystal Density.* Oxford University Press, New York.
47. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
48. Brünger, A.T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* Yale University Press, New Haven and London.
49. Jones, T.A. (1985). Interactive computer graphics: FRODO. *Methods Enzymol.* **115**, 157–171.
50. Ramachandran, G.N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Adv. Protein Chem.* **28**, 283–437.
51. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 946–950.
52. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.

Received: **15 Aug 1995**; revisions requested: **11 Sep 1995**; revisions received: **11 Oct 1995**. Accepted: **19 Oct 1995**.