Acta Cryst. (1996). D**52**, 93–104

X-Ray Structure of the Signal Transduction Protein P_{II} from *Escherichia coli* at 1.9 Å

P. D. Carr,* E. Cheah, P. M. Suffolk, S. G. Vasudevan,† N. E. Dixon and D. L. Ollis

Centre for Molecular Structure and Function, Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia

(Received 9 February 1995; accepted 2 June 1995)

Abstract

The structure of the bacterial signal transduction protein P_{II} has been refined to an R factor of 13.2% using 3σ data between 10 and 1.9 Å. The crystals exhibited twinning by merohedry and X-ray intensities were corrected using the method of Fisher & Sweet [Fisher & Sweet (1980). *Acta Cryst.* A**36**, 755–760] prior to refinement. Our earlier 2.7 Å structure [Cheah, Carr, Suffolk, Vasudevan, Dixon & Ollis (1994). *Structure*, **2**, 981–990] served as a starting model. P_{II} is a trimeric molecule, each subunit has a mass of 12.4 kDa and contains 112 amino-acid residues. The refined model includes all 1065 protein atoms per subunit plus 312 water molecules. The high-resolution refinement confirms the correctness of our 2.7 Å model, although it leads to a redefinition of the extent of various secondary-structural elements. The monomeric structure of P_{II} exhibits an interlocking double $\beta\alpha\beta$ fold. This is a stable fold found in a number of proteins with diverse functions. The association of the protein into a trimer leads to a new structure which we describe in detail. The effects of crystal packing forces are discussed and potential interaction sites with other proteins and effector molecules are identified.

1. Introduction

P_{II} is a signal transduction protein that plays a critical role in the regulation of nitrogen metabolism in bacteria. It transmits a signal between the protein that monitors the levels of available cellular nitrogen, uridylyl transferase/uridylyl removing enzyme (UTase/UR), and enzymes, NR_I and NR_{II}, that regulate the activity and transcription of glutamine synthetase (GS), a key enzyme in nitrogen assimilation. The key structural change associated with signalling is the covalent attachment of an uridylyl monophosphate (UMP) moiety to residue Tyr51 in P_{II}, which occurs under conditions of low cellular nitrogen levels. Descriptions of this process can be found in Cheah *et al.* (1994); Magasanik (1993); Stock, Ninfa & Stock (1989); Kamberov, Atkinson, Feng, Chandran & Ninfa (1994).

In addition to the protein–protein interactions, which occur with P_{II} and at least three enzymes, recent reports suggest that binding of the small-molecule ligands adenosine triphosphate (ATP) and 2-ketoglutarate are important for making P_{II} a suitable substrate for uridylyl transferase (UTase), the enzyme responsible for the uridylylation of Tyr51 under low nitrogen conditions (de Mel *et al.*, 1994).

In this paper we present the refined structure of unmodified P_{II} from *Escherichia coli* at 1.9 Å resolution, describe the detailed structure of the molecule, and identify potential sites for interaction with other proteins and effector molecules.

2. Experimental

The crystals of P_{II} used in this study were grown using conditions described earlier (Vasudevan *et al.*, 1994). The crystals are hexagonal, space group $P6_3$, with unit-cell parameters $a = 61.6$, $c = 56.3$ Å and have one subunit in the asymmetric unit. Two native data sets were used during the refinement. Both data sets were collected using single-axis rotation geometry on an imaging-plate system [R-AXIS II (Sato *et al.*, 1992)] mounted on a Rigaku RU200 rotating-anode generator producing Cu $K\alpha$ radiation at a power of 5 kW (50 kV, 100 mA). The first data set which was collected at room temperature contained observations for 6774 unique reflections with an R_{merge} of 2.5%. Of these reflections, 5733 with $I > 5\sigma(I)$ were used in our refinements and they comprised 93.3% of the unique data between 10.0 and 2.2 Å. A higher resolution data set was obtained at 277 K and was used in the latter stages of refinement. This second data set contained 45 387 measurements of 11 079 reflections with an R_{merge} of 7.1% for data with $I > 3\sigma(I)$. 7976 reflections with $I > 3\sigma(I)$ were used in our refinement and they comprised 83.1% of the unique data between 10.0 and 1.9 Å. Table 1 shows the data completeness as a function of resolution shell.

† Current address: Division of Biochemistry and Physiology, Department of Molecular Sciences, James Cook University, Townsville, Queensland 4811, Australia.

Table 1. *Completeness of the 1.9 Å native data set*

d_{min} (Å)	d_{max} (Å)	Number	% Complete	R_{shell}	R_{merge}
4.01	10.00	775	78.9	0.082	0.082
3.22	4.01	793	81.7	0.085	0.083
2.82	3.22	796	83.6	0.122	0.093
2.57	2.82	784	81.6	0.146	0.102
2.39	2.57	804	83.8	0.164	0.119
2.25	2.39	798	84.1	0.166	0.166
2.14	2.25	787	82.0	0.182	0.121
2.05	2.14	746	78.4	0.183	0.126
1.97	2.05	744	76.6	0.187	0.129
1.90	1.97	709	75.7	0.189	0.132

Both of the crystals exhibited twinning by merohedry. The twin element relating the twins is a twofold rotation parallel to **a**. Reflections with indices *hkl* superimpose reflections on $kh\bar{l}$. In crystals exhibiting a twin fraction near 50% such an operation appears to raise the symmetry of the point group from $P6_3$ to $P6_322$. This additional approximate symmetry was observed in precession photographs obtained from some crystals of P_{II} (data not shown).

The determination of the degree of twinning and subsequent correction of intensities were carried out using the method of Fisher & Sweet (1980) using local space-group specific programs. Following the methodology of Fisher & Sweet plots were constructed of the number of intensities calculated to be negative after correction assuming different values for the twin fraction. The plot obtained for our second data set is shown in Fig. 1. The plots we obtained did not show a straight line intercepting the base line as predicted, but rather had a pronounced curvature in the region of interest. To determine which was the most appropriate value for the twin fraction we used a partially refined model and refined it against data detwinned at various twin fractions (Carr, Cheah & Ollis, 1996). All refinements consisted of 400 cycles of postitional refinement using *X-PLOR* (Brünger, 1992) and detwinned X-ray data with $I > 3\sigma(I)$. On the basis of Figs. 1(*a*) and 1(*b*) we chose 0.17 as the best value. The first native had a twin fraction of 0.061.

2.1. *Refinement*

The starting model for the refinement was our 2.7 Å structure with all 1065 protein atoms and no solvent molecules. This model had undergone initial refinement using *X-PLOR* (Brünger, 1992) with data from our first native data set. This initial refinement consisted of 300 cycles of positional refinement followed by a simulated-annealing step where the temperature of the system was increased to 4000 K then allowed to cool to 300 K in 25 K steps using time steps of 0.5 fs between energy calculations. The model was subjected to a further 60 cycles of positional refinement and 15 cycles of unrestrained *B*-factor refinement. The resulting model (Cheah *et al.*, 1994) had an *R* factor of 24.5% with r.m.s. deviations from ideality of 0.01 Å for bond lengths and 1.49° for bond angles.

Subsequent rounds of refinement were also performed using *X-PLOR*. Our refinement stratagem was to run positional refinement using the Powell conjugate-gradient method (Powell, 1977) until convergence, to follow this with cycles of individual restrained isotropic *B*-factor refinement, then to refine water occupancies (setting any occupancy values greater than 1.0 to 1.0), and finally to repeat the *B*-factor refinement. An extra round of positional refinement was sometimes added at this point.

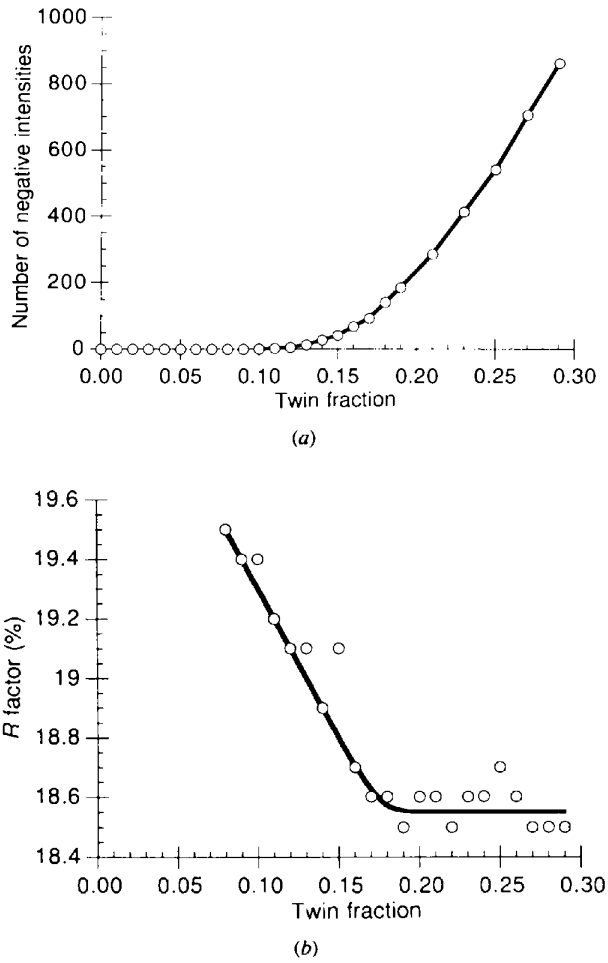


Fig. 1. (*a*) Graph showing the number of intensities calculated as negative as a function of assumed twin fraction for our second native data set. This was used to estimate the twin fraction according to the method of Fisher & Sweet (1980). (*b*) The variation of *R* factor with twin fraction. Each data point is the result of 400 cycles of positional refinement using *X-PLOR* (Brünger, 1992) on the same partially refined model using detwinned X-ray data with $I > 3\sigma(I)$.

Table 2. *Progress of refinement*

Resolution 10–2.7 Å

Starting model, no water molecules, $R = 24.5\%$
Rebuild model, 22 water molecules
589 cycles positional refinement, $R = 25.9\%$
20 cycles B refinement,* $R = 23.3\%$
Rebuild model, 110 water molecules

Resolution 10–2.2 Å

700 cycles positional refinement, $R = 27.8\%$
15 cycles B refinement, $R = 23.6\%$
200 cycles positional refinement, $R = 22.5\%$
Change weights to increase X-ray term
200 cycles positional refinement, $R = 19.6\%$
Rebuild model, 136 water molecules
400 cycles positional refinement, $R = 19.8\%$
15 cycles B refinement, $R = 18.1\%$
160 cycles positional refinement, $R = 17.8\%$

Resolution 10–1.9 Å†

258 cycles positional refinement, $R = 20.2\%$
BQrefinement,‡ $R = 18.9\%$
Rebuild model, 261 water molecules
123 cycles positional refinement, $R = 20.3\%$
BQrefinement $R = 17.1\%$
260 cycles positional refinement, $R = 15.6\%$
BQrefinement, $R = 14.8\%$
Rebuild model, 312 water molecules
600 cycles positional refinement, $R = 16.8\%$
BQrefinement, $R = 14.8\%$
300 cycles positional refinement, $R = 14.4\%$
BQrefinement, $R = 13.8\%$
Changed twin fraction to 17.0% (from 15.0%)§
153 cycles positional refinement, $R = 13.7\%$
BQrefinement, $R = 13.4\%$
72 cycles positional refinement, $R = 13.4\%$
BQrefinement, $R = 13.2\%$

* B refinement = individual restrained isotropic B-factor refinement. † Data from the second native crystal used from this point onwards. ‡ BQrefinement = 20 cycles B refinement, followed by 15 cycles water occupancy (setting any > 1.0 to 1.0), followed by 20 cycles B refinement. § A detailed study of twin fraction determination suggested this to be a more appropriate value (Carr *et al.*, 1996).

$(2|F_{obs}| - |F_{calc}|, \alpha_{calc})$ and $(|F_{obs}| - |F_{calc}|, \alpha_{calc})$ Fourier maps were calculated using the refined coordinates. Manual checking and rebuilding of the model on a Silicon Graphics workstation was then performed using the program *FRODO/TOM* (Jones, 1985) before the next round of refinement. Table 2 summarizes the progress of refinement.

3. Results and discussion

3.1. *Quality of the electron-density map*

Fig. 2 shows typical electron density from our 1.9 Å $(2|F_{obs}| - |F_{calc}|, \alpha_{calc})$ electron-density map superimposed upon the final model. Our final maps were of sufficient quality to unambiguously determine the positions of all of the protein atoms on a $(2|F_{obs}| - |F_{calc}|, \alpha_{calc})$ Fourier map; 312 ordered water molecules were also identified.

3.2. *Accuracy of the model*

The final model had an R factor of 13.2% when compared to 3σ data between 10.0 and 1.9 Å from our second native data set, and an R factor of 15.2% when compared to all data in the same resolution shell. It had an r.m.s. deviation from ideal stereochemistry of 0.017 Å for bond lengths and 1.98° for bond angles. A Luzzati plot (Luzzati, 1952) suggested a mean coordinate error of 0.15 Å. The model was checked for stereochemical correctness using the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). For all criteria used by *PROCHECK*, our model was either flagged as being better, or within acceptable limits of, typical values for 118 protein structures that had been refined

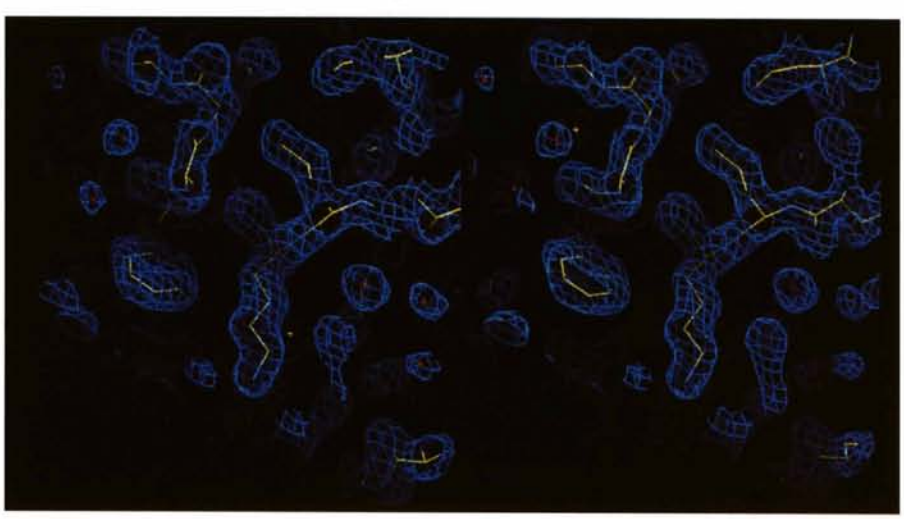


Fig. 2. A section from a $(2|F_{obs}| - |F_{calc}|, \alpha_{calc})$ electron-density map contoured at 1σ superimposed upon the final model.

at a resolution of 2.0 Å or better. The overall G factor (Laskowski, MacArthur & Thornton, 1994) was 0.2.

The main-chain torsion angles (φ, ψ) are represented on a Ramachandran plot (Ramakrishnan & Ramachandran, 1965) in Fig. 3. Of the 97 non-end residues which were neither glycine nor proline, 93 fell into the most favoured 'core' regions, three residues were in additional 'allowed' regions, and one (Tyr46) was outside these two in a 'generously allowed' region. The different regions quoted are those described by Morris, MacArthur, Hutchinson & Thornton (1992), which were taken from the observed φ–ψ distribution for 121 870 residues from 463 protein structures. Tyr46 is part of a loop (T loop, described later) which contacts neighbouring trimers. Although not hydrogen bonded to neighbouring molecules the close proximity of these neighbours limits the conformational space available to it and this may explain why the torsion angles deviate a little from the most favoured conformations.

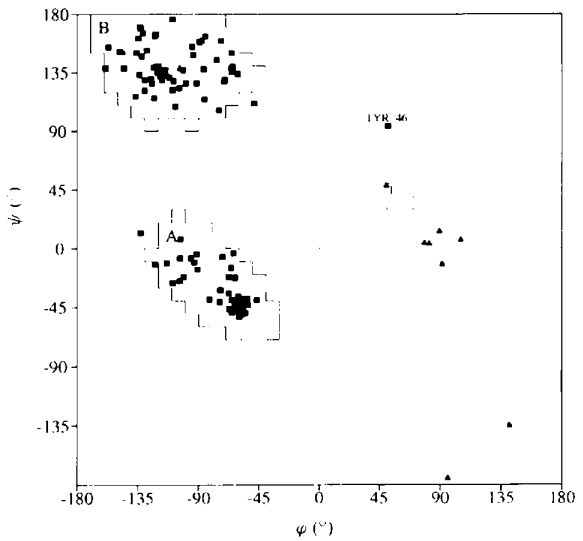


Fig. 3. Ramachandran plot of main-chain conformational angles. Glycine residues are shown as triangles, all other residues as squares.

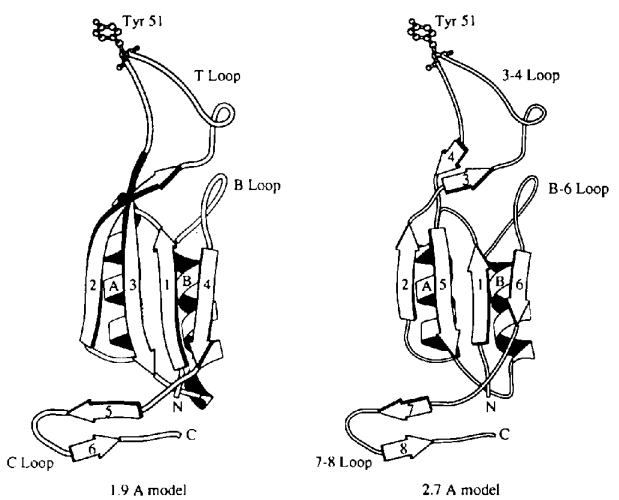


Fig. 4. *MOLSCRIPT* (Kraulis, 1991) schematic diagrams of the P_{II} monomer. (*a*) Present 1.9 Å model. (*b*) Former 2.7 Å model.

3.3. *Description of the structure of the monomer*

Although the refined high-resolution structure confirms the general correctness of our earlier model, it has led to some redefinition of the extent of secondary-structure elements within the protein. In particular strands 2 and 3 of the 2.7 Å model have now been shown to be part of a single curved β-strand. Similarly, strands 4 and 5 of the earlier model are now considered to be part of a single strand. Fig. 4 shows *MOLSCRIPT* (Kraulis, 1991) schematic diagrams for the 2.7 Å structure and the current model. Fig. 5 shows the secondary-structure assignments for P_{II}. These were made using the algorithm of Kabsch & Sander (1983) as implemented in the programs *DSSP* and *PROCHECK*. In the present assignment Gly27 is included in strand 2, since it is clearly part of the extended sheet with (φ, ψ) angles of (–146, 150°). A water-mediated hydrogen bond is formed from the carbonyl O atom of Gly27 to the backbone N atom of Val64. Strands 2 and 3 are diverging at this point, and although the backbone N atom of Gly27 and the carbonyl O atom of Val64 are not close enough to form a hydrogen bond, they are still in an extended β-conformation. Residue Glu106 is now included in strand 6. This is in an extended conformation with (φ, ψ) angles of (–86, 137°), and a hydrogen bond is formed between the carbonyl O atom of Glu106 and the backbone N atom of Arg101. The N atom of Glu106, although still running antiparallel to strand 5 forms hydrogen bonds with residues Thr104 and Gly105 that presumably helps to stabilize the hydrogen-bonded turn formed by these two residues. Fig. 6 shows the

```
1           11          21          31          41          51
MKKIDAIIKP  FKLDDVREAL  AEVGITGMTV  TEVKGFGRQK  KHTELYRGAE  YMVDFLPKVK
.EEEEEEE.G  GGHHHHHHHH  HHHT..aEEE  EEEEEES..T  T...S....S  S....EEEEE
 <- β1-> <------helix A ----------> <---- β2  ---->                <--β3

61          71          81          91          101         111
IEIVVPDSIV  DTGVDTIIRT  AQTGKIGVGK  IFVFDVARVI  RIRTGEEDCA  AI
EEEEE.GGGH  HHHHHHHHHH  H..SSTT..E  EEEEE.SEEE  ETTT.EEE.T  T.
----> <-------- helix B --------'        <-- β4 ---> <-- β5 -->   <β6>
```

Fig. 5. Secondary-structure assignments. E = extended sheet, H = α-helix, G = 3_{10} helix, S = bend, T = hydrogen-bonded turn (Kabsch & Sander, 1983). a,b extended sheet not flagged by *DSSP* (see text).

hydrogen bonding and secondary-structure assignments for P_{II} produced by the program *HERA* (Hutchinson & Thornton, 1990).

The bends in strands 2 and 3 are not the result of a systematic change in the φ and ψ angles as a function of residue number. However, the (φ, ψ) angles for Glu32 (–91, 127°), whilst being clearly in the most favoured region of the Ramachandran plot for β-sheet, are lower than the rest of this strand and a noticeable bend occurs in the Cα trace at this point. The angle subtended at residue Glu32 between the axis of the β-strand from residues 27 to 32, and the axis of the β-strand between residues 32 to 36 is 58.2°. This was calculated using the method of Ploegman, Drenth, Kalk & Hol (1978). Daffner, Chelvanayagam & Argos (1994) have studied the bending of β-strands from 946 strands of lengths greater than three amino-acid residues from 247 non-homologous protein structures available from the Protein Data Bank at Brookhaven (Bernstein *et al.*, 1977). They show that 85% of strands exhibit single bends of the type observed in P_{II}. Fig. 7(*a*) shows the hydrogen bonding for strand 2 with its symmetry mates. As will be explained later, the bends in strands 2 and 3 are crucial to the formation of the oligomeric structure of the protein as they allow these strands to participate in hydrogen bonding between two symmetry-related sheets within the trimer.

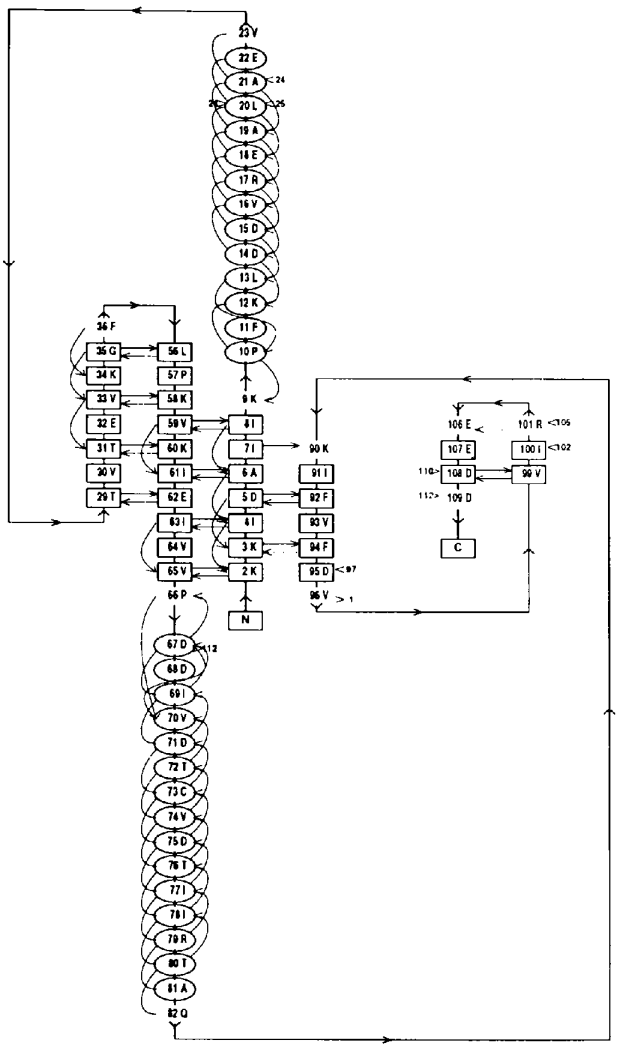


Fig. 6. Secondary structure and hydrogen bonding within the P_{II} monomer. This diagram was prepared using the program *HERA* (Hutchinson & Thornton, 1990).

The monomer consists of a central four-stranded antiparallel β-sheet with two helices on one side, forming an interlocking double $\beta\alpha\beta$ motif. There are two major loops emanating from the sheet. The first and largest between strands 2 and 3 (the *T* loop, formerly reported as the 3–4 loop) contains the functionally important Tyr51 residue at its apex. The second loop is found between strand 4 and the carboxyl terminus and contains strands 5 and 6 (the *C* loop, formerly reported as the 7–8 loop). In addition, there is a smaller loop between helix *B* and strand 4, the *B* loop (formally reported as the *B*–6 loop). A cleft formed between the *B* loop, the first four residues of the *T* loop, and the *C* loop of a neighbouring monomer, is likely to be of functional significance. The interlocking double $\beta\alpha\beta$ core of the monomer exhibits approximate twofold symmetry, relating strands 1 and 4 to strands 2 and 3. Residues 4–22 and residues 29–32 overlap with their twofold-related coordinates with an r.m.s. deviation of 0.7 Å. This additional twofold axis is approximately parallel to **b***, which corresponds to a twofold axis found in space group $P6_322$ but not $P6_3$. This may contribute to the tendency of the crystals to twin.

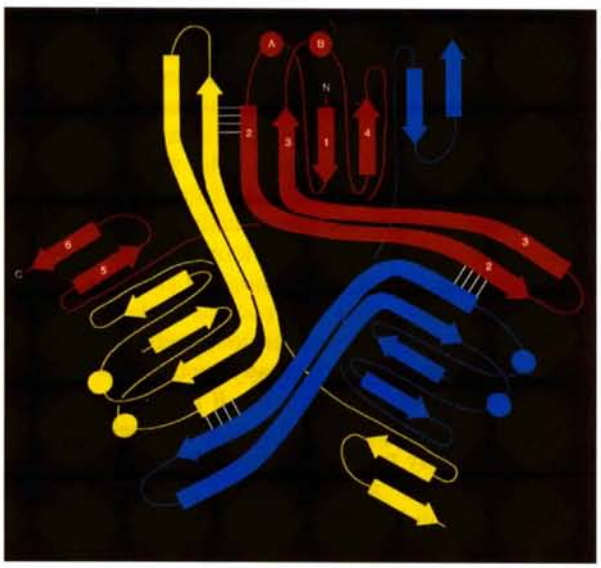
There are two helices per monomer, helix *A* extends from residue Pro10 to Val23. The first three residues are formally 3_{10} rather than α-helix. Helix *B* extends from residue Asp67 to Ala81 and again the first three residues are formally 3_{10} rather than α-helix. The helices are roughly antiparallel to each other, the axes of the helices being inclined at an angle of 149° to each other. The helices pack against the antiparallel β-sheets forming hydrophobic interactions. The β-sheets curve around the helices to maximize such interactions.

3.4. *Description of the structure of the trimer*

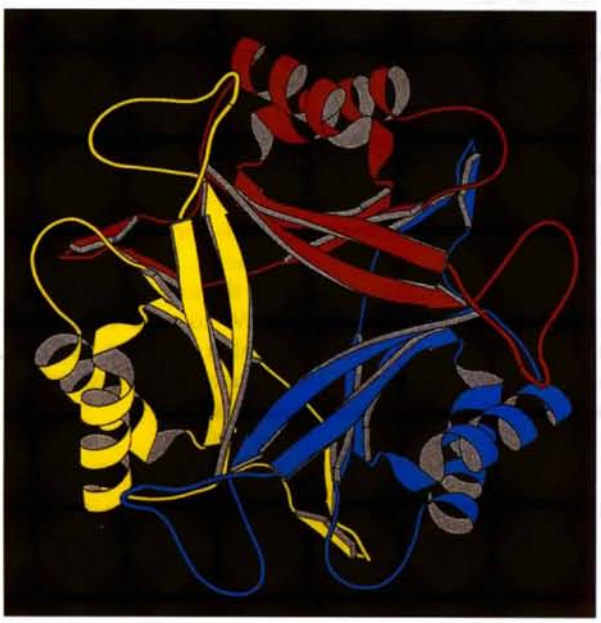
The structure of the trimer remains as described in our earlier paper but strands 2 and 3 are now continuous and span between neighbouring subunits (Fig. 7). A central cavity is formed around the threefold axis of the trimer. This cavity varies in diameter from *ca* 24 Å at the ends to 5.5 Å approximately 7.0 Å from either end. The diameter at the centre of the cavity is *ca* 12 Å. There are numerous ordered water molecules and hydrogen-bonding networks formed within the cavity.

3.5. *Loop structures form a cleft*

Except for the *T*, *B* and *C* loops, the loops connecting elements of secondary structure are short lengths of random coil. The *T* loop is a large 20-residue loop that positions the functionally important Tyr51 residue clear of the compact main body of the molecule. Residues Lys40 and Gly41 form a small hydrogen-bonded turn within this loop. The function of this turn is uncertain, although similar turns are seen in loops from other proteins that contain the interlocking double $\beta\alpha\beta$ motif. The thermal parameters for residues in the *T* loop are somewhat higher than those in the rest of the molecule, and the loop is stabilized to some extent by contacts

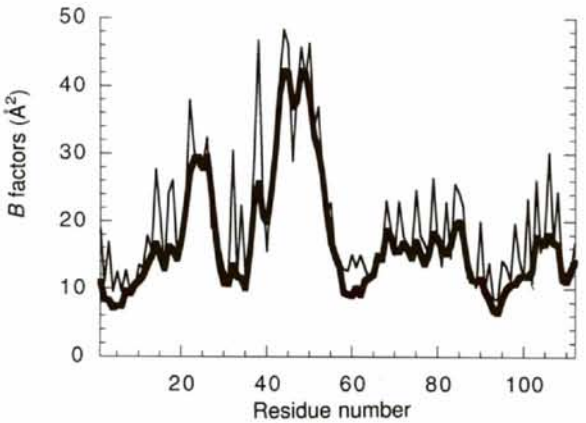


(*a*)

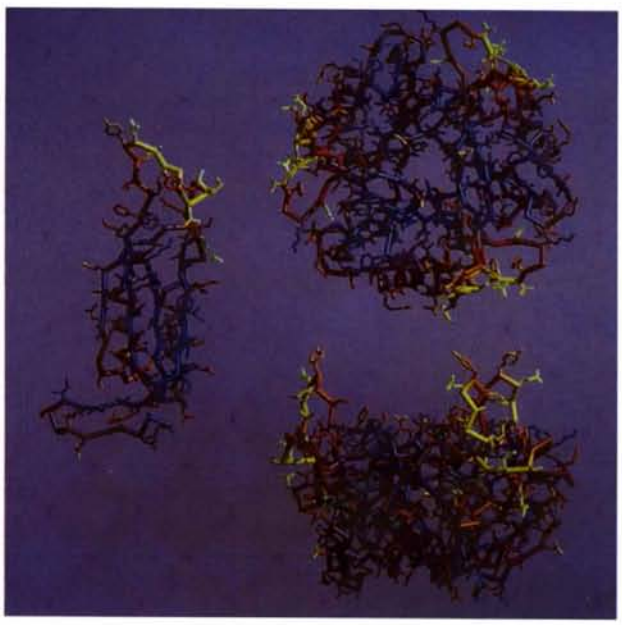


(*b*)

Fig. 7. The P_{II} trimer (*a*) schematic showing connectivity, the white lines represent hydrogen bonds between the backbone O atoms and N atoms which exist between residues 28 and 36, and residues 30 and 34. These bind the symmetry mates of strand 2 together in adjacent subunits. (*b*) *MOLSCRIPT* (Kraulis, 1991) schematic viewed down crystallographic threefold axis.



(*a*)



(*b*)

Fig. 8. *B* factors (Å^2) as a function of residue number. (*a*) Plot of main chain (thick curve) and side chain (thin curve). (*b*) *MIDAS* (Ferrin, Huang, Jarvis & Langridge, 1988) plot of structure colour coded by temperature factor in the range 5 (blue) to 50 Å^2 (yellow). The three orientations depict, on the left, the monomer; upper right, trimer viewed down the crystallographic threefold axis; lower right, trimer view perpendicular to crystallographic triad. This layout is used in subsequent figures.

with symmetry-related molecules from other trimers within the crystals. A fuller discussion of both thermal parameters and crystal contacts is given below.

The *C* loop is a 17-residue loop containing strands 5 and 6. It enables these two strands to be positioned so that they are able to participate in β-sheet formation with neighbouring subunits. There is a hydrogen-bonded turn between strands 5 and 6 which forms part of the elliptical cleft discussed below.

A nine-residue loop exists between helix *B* and strand 4, the *B* loop. A cleft is formed between this loop, the *T* loop, and the *C* loop of a neighbouring monomer (Fig. 7*b*). One side of this cleft is covered with uncharged residues while the other side is lined with basic residues that include Arg38, Lys40, Lys85 and Lys90 from one monomer, and Arg101 and Arg103 from an adjacent monomer.

3.6. *Thermal motion*

Fig. 8 shows the variation of *B* factors as a function of residue number. The main-chain atoms have *B*-factor values in the range of 7.0–43.2 Å^2 with a mean of 17.0 Å^2 and a standard deviation of 8.7 Å^2. The side-chain atoms have *B*-factor values in the range 7.8–49.8 Å^2 with a mean of 20.2 Å^2 and a standard deviation of 9.8 Å^2. Two regions show significantly higher thermal motion than the rest of the molecule. These are the *T* loop and the loop between strand 2 and helix *A*. As one would expect, main-chain *B* factors for β-sheet and helical residues are lower than the average due to their low conformational entropy. The helices exhibit higher average *B* factors than the sheets, which is not surprising given their higher degree of solvent accessibility. Fig. 9 shows the distribution of temperature factors and occupancy values for the 312 water molecules.

3.7. *Side-chain environment*

Hydrophobic residues are concentrated in the region between the curved central β-sheets and the surrounding helices and the polar residues are primarily on loops at the exterior of the molecule. There are also several polar residues in the central cavity that are involved in the formation of numerous hydrogen bonds. A detailed description of these interactions is given below.

Fig. 10 shows the distribution of charged residues within the molecule. A clustering of basic residues is seen on one side of the elliptical clefts formed at the interfaces between the monomers. The lower surface of the trimer contains mainly acidic residues while the upper surface shows an alternation of acidic and basic regions. Fig. 11 shows the electrostatic potential

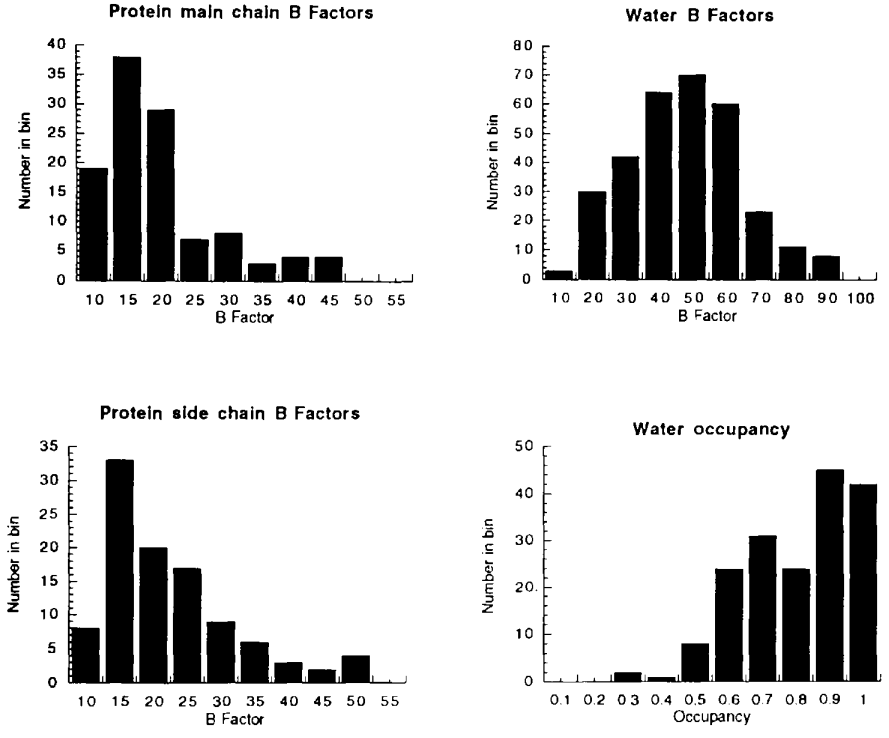


Fig. 9. Histograms of the distribution of thermal parameters for protein and water molecules and the distribution of water occupancies. The occupancies of all protein atoms were set to 1.0.

generated by such a distribution of charged residues. The large areas of opposite charge seen in the top half of the molecule suggest that electrostatics play an important role in guiding target molecules to their correct docking sites on P_{II}.

3.8. *Contents of the cavity*

The side chains pointing into the centre of the cavity are mainly polar. Specifically, residues Lys3 and Asp5 from strand 1 and residues Lys60 and Glu62 from strand 3 all point into the interior of the molecule. These residues are related by the approximate twofold of the interlocking double $\beta\alpha\beta$ core; Lys3 rotates approximately onto Lys60 and Asp5 rotates onto Glu62. The charges on these groups apparently balance one another – Lys3 approaches close to Asp5 and Glu62 from the same subunit. Lys60 makes a hydrogen bond with water Wat758 which also interacts with Asp5. The effect of the threefold symmetry is to generate two rings of alternating charged residues – each Lys3 is close to Asp5 from the same subunit and Asp5 from a symmetry-related subunits. Similarly, Lys60 and Glu62 alternate around the centre of the molecule. In addition to Wat758, other water molecules make hydrogen bonds with the polar residues located in the central cavity.

The top and bottom of the central cavity are formed by the close approach of strands from the threefold-related subunits, as shown in Fig 7. At the top of the cavity, the start of strand 2 of molecule 1 interacts with the end of strand 2 of the symmetry-related molecule 3. The end of strand 2 of molecule 1 interacts with the start of strand 2 of molecule 2. As noted above, strand 2 is bent at position 32; its backbone amide and carbonyl O atom form water-mediated hydrogen bonds with its symmetry mates. In addition, the side chain of Glu32 forms a salt link with the side chain of Lys34 of a symmetry-related subunit. Entry to the cavity from the top would be restricted by the three symmetry mates of Val33 whose side chains approach to within 5.0 Å. The situation at the bottom of the cavity is similar to the top. Here however, two strands, 4 and 5, fill the role played by strand 2 at the top of the cavity. Strands 4 and 5 are linked by two residues, Val96 and Ala97. Strand 4 of molecule 1 is hydrogen bonded to strand 5 of molecule 2 while strand 5 of molecule 1 is hydrogen bonded to strand 4 of molecule 3. Again, the backbone amides and carbonyl O atoms of Asp95 form water-mediated hydrogen bonds with their symmetry mates while the side chain of this residue forms a salt link to Lys2 of the same subunit. The side chain of Arg98 forms a salt link with the side chain of Asp71, on a symmetry-related molecule. Entry to the cavity from the bottom is most restricted by the side chains of the symmetry mates of residue Phe94 which approach to within about 6.0 Å of one another at their Cβ atoms.

Within the central cavity, electron-density features that correspond to all of the modelled water molecules appear in OMIT maps. In addition, there are two features in OMIT maps which suggest that the cavity may also contain metals. The putative metals lie on the threefold axis and have electron density in OMIT maps that is higher than expected for a water molecule. They make close approach (2.1 Å) to less electron-dense features – as expected of a water coordinated to a metal. Although suggestive of metals, these observations are of a preliminary nature and more detailed studies are necessary to confirm the existence of metals within the P_{II} trimer. A recent paper (Liu & Magasanik, 1995) indicates that Co^{2+} stimulates the interaction of P_{II} and NR_{II} to dephosphorylate NR_I phosphate.

3.9. *Crystal packing*

Crystal contacts occur between neighbouring trimers within the crystal lattice. A given monomer in one P_{II} trimer (at $z = 0$) has interactions with two other monomers within a neighbouring trimer (at $z = 1/2$). These interactions are formed with the *T* loop. This loop fits into a depression between strand 6 and helix *B* of adjacent subunits in the neighbouring trimer. Residues Lys40 and His42, found in the vicinity of the short hydrogen-bonded turn within the loop, are also involved in forming crystal contacts.

The specific interactions involve hydrogen bonds between the Nζ atom of Lys40 and the carbonyl O atoms of residues Ile110 and Val96. Furthermore, one of the Oδ atoms from Asp95 is 2.7 Å from the same N atom and a water-mediated contact exists to the O atom of Asp108 and Ala110. The Nε atom of His42 lies at distances of 2 9 and 3.3 Å from the Oε atoms of Glu107. The terminal N atoms of Arg47 are 2.7 and 3.4 Å from the Oε atoms of Glu107. The backbone N atoms of Glu50 and Tyr51 both form hydrogen bonds with one of the Oδ atoms of Asp75 (3.5 and 2.9 Å, respectively). The backbone N atom of Val53 is 3.2 Å from the carbonyl O atom of Gly105. The carbonyl O atom of Val53 is also involved in water-mediated interactions with the backbone O atom and N atom of Glu107.

3.10. *Conserved residues*

In our earlier paper we showed the alignment of the P_{II} amino-acid sequence from *Escherichia coli* with P_{II} homologues from eight other species of microrganisms and highlighted the conserved residues (Fig. 2 of Cheah *et al.*, 1994). Fig. 12 shows the location of these conserved residues in the final model. We observe them to be concentrated in three areas: (*a*) the flat top surface of the molecule; (*b*) the loop regions contributing to the formation of the elliptical cleft at the subunit interfaces;

and (*c*) some residues in the central cavity. The first two areas are potential sites of interactions with other proteins and effector molecules.

There are two areas that contain very few conserved residues, namely the exterior surfaces of the α-helices and the flat surface that forms the bottom of the molecule. The lack of conservation of specific residues in these areas would tend to suggest that these regions do not participate in specific protein–protein interactions.

4. Concluding remarks

The high-resolution structure of P_{II} reveals more clearly the details of the geometry and conformation of the main-chain and side-chain atoms compared to the 2.7 Å study. In particular, the extent of the various secondary-structural elements have been more precisely defined. The high-resolution refinement has additionally revealed the position of 312 ordered water molecules. Some of these water molecules have been shown to participate in hydrogen-bonding networks in the central cavity of the molecule which may further stabilize the trimer.

The concave arrangement of β-sheets which form around the threefold axis and are comprised of strands emanating from all three subunits is very unusual and strongly favours trimer formation. in the α/β proteins, such as triose phosphate isomerase (TIM), α-helices surround an all-β core (Banner *et al.*, 1975). Although this is also true for P_{II}, the structure of TIM is quite different to that of P_{II}. In TIM, the strands are parallel and form a barrel while in P_{II} the stands are antiparallel and do not form a barrel. In both P_{II} and TIM, the helices form exterior surfaces to the proteins, but in P_{II} the sheets bend around the helices while the opposite is true for TIM. P_{II} is also different to other trimeric $\alpha+\beta$ proteins which, like TIM, form barrel-like structures with α-helices on the outside surface. In 6-pyruvoyl tetrahydropterin synthase (PTPS) (Nar, Huber, Heizemann, Thorny & Burgisser, 1994) and human CksHs2 (Parge, Arvai, Murtari, Reed & Tainer, 1993) the individual subunits, which consist of four antiparallel strands, pack together to form a 12-stranded barrel with six α-helices on the outside surface. Although these proteins form trimers they are quite different to the type formed by P_{II}; they do not have an interlocking double $\beta\alpha\beta$ core structure. Muconolactone isomerase (MI) does have subunits with an interlocking double $\beta\alpha\beta$ structure, but this protein forms dodecamers with 52 symmetry (Katti, Katz & Wyckoff, 1989). In addition, the strands of MI form a barrel-like structure with α-helices on the external surface. The curving of the β-strands in P_{II} allows elliptical clefts to be formed at the subunit interfaces which form potential interaction sites with other molecules.

The central scaffold of the P_{II} molecule is an interlocking double $\beta\alpha\beta$ motif. This is a particularly common motif found in a variety of proteins with a range of functions. Orengo & Thornton (1993) studied 17 structures containing this topology and concluded that although they exhibit extensive structural similarities they did not all have a common function. A number of these proteins appeared to interact with RNA while others bound, or catalysed reactions involving, nucleotides. The active sites of these molecules are distributed over a large portion of the surface of the fold. Orengo & Thornton (1993) concluded that the ability of this fold to accommodate a range of functions suggests that it is a stable fold capable of adapting a number of sites to different functional requirements. The structure of P_{II} is consistent with this line of argument since its structure suggests that its functionally important areas are different to those found in other molecules with similar topologies.

Swindells & Alexandrov (1994) have noted similarity in nucleotide binding across the β-sheet of the interlocking double $\beta\alpha\beta$ domains in aspartate transcarbamylase, nucleoside diphosphate kinase and glutamine synthetase. In P_{II} the hydrophobic surfaces formed by the β-sheets are not accessible for such interactions, being buried around the threefold axis on one side and protected on the other side by α-helices. Alternatively, the two α-helices of P_{II} are at the surface of the molecule and pack together in a manner reminiscent of the helices in the human class I histocompatibility antigen (Bjorkman *et al.*, 1987). In this antigen, a groove between the two α-helices acts as an interaction site. The same role is not likely to be shared by P_{II} since there is little cross-species conservation of amino-acid residues in this region in P_{II} and its homologues. The loops of the interlocking double $\beta\alpha\beta$ proteins have been considered important for their function. Several protease inhibitors [ovomucoid (Bode, Epp, Huber, Laskowski & Ardelt, 1985), subtilisin carlesberg (Bode, Papamokos & Muskil, 1987), subtilisin bpn (Takeuchi, Satow, Nakamura & Mitsui, 1991)] use surface loops to bind to the active site. The activation domain of procarboxypeptidase may also inhibit the catalytic domain by means of a surface loop (Coll, Guasch, Aviles & Huber, 1991). In contrast, the *T* loop of P_{II} is a substrate for the UTase/UR enzyme.

Clearly, the loop structures of P_{II} are unique to it and are important to its structure and function. The *T* loop which positions the functionally important Tyr51 residue clear of the compact main body of the molecule is unlike the recognition sites found in molecules with a related function. A typical example is the Che Y protein which transmits a signal (*via* phosphorylation) in chemotaxisis (Stock, Mottonen, Stock & Schutt, 1989). The recognition site of this protein is an acid group that is in a short loop between a β-strand and an α-helix – unlike the *T* loop of P_{II}, it is part of the central

core of the protein. Combined with the fact that the position of the *T* loop is stabilized by interaction with neighbouring crystallographically related molecules, the higher than average *B* factors for atoms in this loop suggest that it may be capable of adopting more than one conformation. This may have functional significance in its recognition by UTase/UR as a substrate and subsequent uridylylation of Tyr51. One possibility is that binding of 2-ketoglutarate and ATP may stabilize one of the possible *T*-loop structures and that this structure may be important for recognition by the UTase/UR. Work is now in progress to determine the structure of P_{II} complexed with 2-ketoglutarate and ATP.

The loops of P_{II} are also important for oligomerization which is undoubtedly important for the function of the molecule. Association of the molecule results in a compact cylinder-shaped structure and creates a number of potential sites for specific interactions with other proteins and effector molecules. Evolutionary conservation suggests some of these sites are more promising than others (Fig. 12). The flat top surface of the molecule contains numerous conserved residues and is readily accessible to solvent. When this information is coupled with the electrostatic potential surface generated by these residues (Fig. 11) it suggests an important role for this surface in protein–protein interactions.

Another likely interaction site is the elliptical cleft formed at the monomer interfaces by the *B*, *T* and *C* loops. This cleft contains several conserved residues and exhibits a local positive electrostatic potential. This is again suggestive that it is a site of interaction, possibly with ATP. Other groups have noted (de Mel *et al.*, 1994; Kamberov *et al.*, 1994) that the amino-acid sequence in the *B*-loop region (Fig. 5) contains residues that form

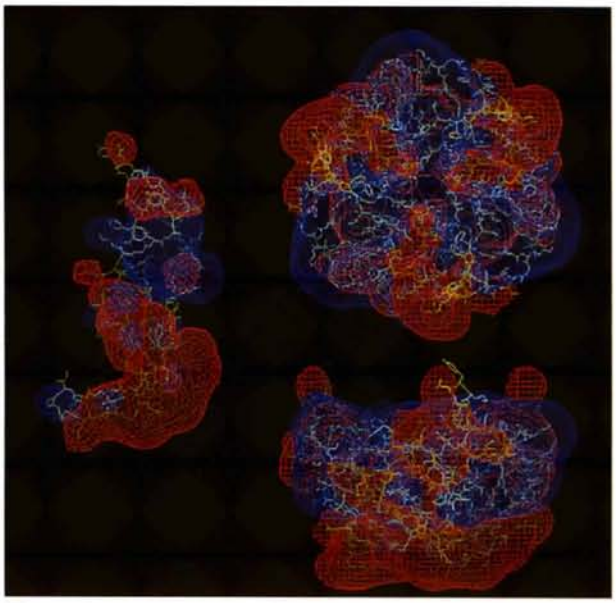


Fig. 11. Electrostatic potential generated by the P_{II} trimer. This was calculated using the program *DelPhi* (Biosym Technologies) using dielectric constants of 2.0 and 80.0, respectively, for the solute and solvent, and an ionic strength of 0.145 mol l^{-1}. The resulting potential grid was contoured at -2 kT e^{-1} (red) and $+2$ kT e^{-1} (blue).

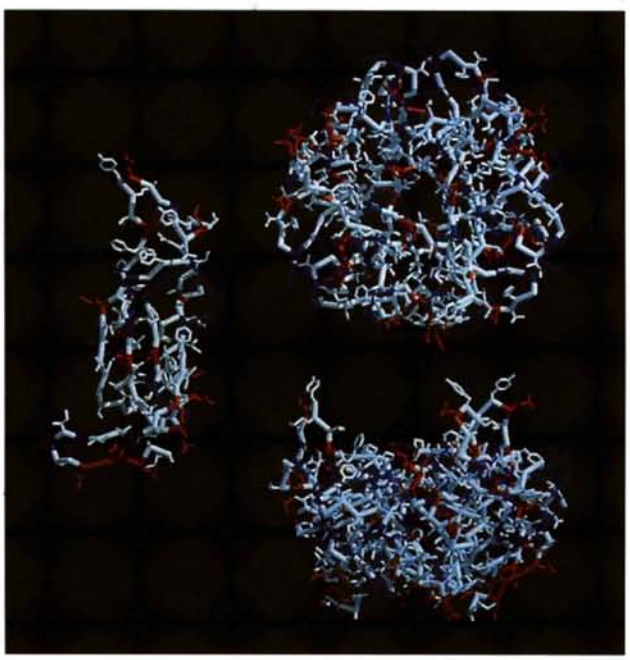


Fig. 10. Distribution of charged residues. Acidic groups (Glu, Asp) are red, basic groups (Lys, Arg) are deep blue, other residues are white. This plot was produced using the program *MIDAS* (Ferrin *et al.*, 1988).

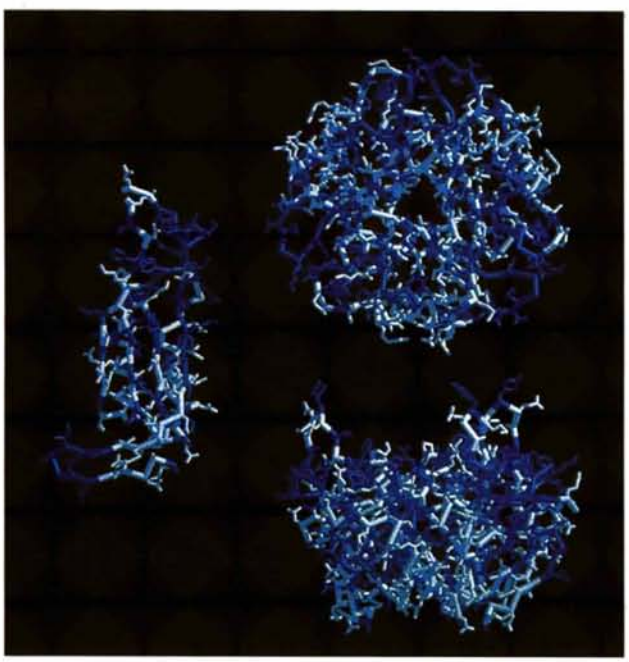


Fig. 12. *MIDAS* (Ferrin *et al.*, 1988) diagram showing conserved residues in purple, other residues are coloured white with light blue shading.

part of a 'Walker A' consensus motif that is involved in ATP binding in a number of proteins (Walker, Saraste, Runswick & Grey, 1982; Husain, Van Houten, Thomas & Sancar, 1986). The consensus motif is G-*X*-*X*-*X*-G-K-*X*-*X*-*X*-*X*-*X*-I. In P_{II} residues 84–90 consist of G-*X*-*X*-*X*-*X*-G-K. In proteins that bind ATP using this motif, it occurs at a bend following a β-strand and before an α-helix. The helix dipole stabilizing phosphate binding. In P_{II} the sequence occurs in the *B* loop where the connectivity is the reverse of this, *i.e.* α-helix, turn, β-strand. The helix dipole moment would tend to destabilize ATP binding. If the ellipical-shaped cleft in P_{II} is an ATP binding site then the mode of interaction between it and P_{II} is different to that of proteins exhibiting the Walker A motif.

The central cavity has conserved residues and is available to solvent from either side of the molecule. Although this is a large internal space filled with many water molecules, there are problems with its use as a receptor site. The central core of the molecule contains numerous groups which stabilize the trimer. Ligand binding in this region could disrupt the trimer.

Two other regions of the structure that have the potential to act as interaction sites are the flat negatively charged bottom surface of the molecule and the cleft formed by the exterior surfaces of the α-helices. Neither of these regions shows a high level of conservation in its amino-acid sequence when compared to P_{II} and related homologues from other species.

Given that P_{II} interacts with three enzymes as well as two small molecule effectors and given that its effect on the enzymes is to totally reverse their functionality upon uridylylation of P_{II}, it is somewhat surprising that P_{II} is such a small and structurally economical molecule. Nature appears to have selected a stable protein motif, the interlocking double $\beta\alpha\beta$ fold, and from this framework created a number of interaction sites which may be used to control the functionality of the larger enzymes. The formation of a stable trimeric molecule seems to have been central to this process. Increased stability of the trimer and an economical use of secondary structural elements appear to have been gained by the curving of β-strands 2 and 3 to allow their involvement in hydrogen bonding within neighbouring β-sheets. The structure of P_{II} gives us an initial glimpse into the structural basis of the very elegant system that controls bacterial nitrogen metabolism.*

The ANU supercomputing facility is thanked for making computer facilities available to us on their Fujitsu VP2200 computer by a grant of time on the machine. Mr John D. Barton is thanked for assistance in producing Fig. 7(*a*).

* Atomic coordinates have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 2PII). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: HA0139). At the request of the authors, the atomic coordinates will remain privileged until 1st May 1996.

References

Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. & Waley, S. G. (1975). *Nature (London)*, **255**, 609–614.

Bernstein, C. F., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). *Nature (London)*, **329**, 512–518.

Bode, W., Epp, O., Huber, R., Laskowski, M. L. & Ardelt, W. (1985). *Eur. J. Biochem.* **147**, 387–395.

Bode, W., Papamokos, E. & Muskil, D. (1987). *Eur J. Biochem.* **166**, 673–692.

Brünger, A. T. (1992). *X-PLOR Manual, Version* 3.1. *A System for Crystallography and NMR*. Yale University, New Haven, CT, USA.

Carr, P. D., Cheah, E. & Ollis, D. L. (1996). In preparation.

Cheah, E., Carr, P. D., Suffolk, P. M., Vasudevan, S. G., Dixon, N. E. & Ollis, D. L. (1994). *Structure*, **2**, 981–990.

Coll, M., Guasch, A., Aviles, F. & Huber, R. (1991). *EMBO J.* **10**, 1–9.

Daffner, C., Chelvanayagam, G. & Argos, P. (1994). *Protein Sci.* **3**, 876–882.

Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). *J. Mol. Graphics*, **6**, 13–27.

Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* A**36**, 755–760.

Husain, I., Van Houten, B., Thomas, D. C. & Sancar, A. (1986). *J. Biol. Chem.* **261**, 4895–4901.

Hutchinson, E. G. & Thornton, J. M. (1990). *Proteins*, **8**, 203–212.

Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.

Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.

Kamberov, E. S., Atkinson, M. R., Feng, J., Chandran, P. & Ninfa, A. J. (1994). *Cell. Mol. Biol. Res.* **40**, 175–191.

Katti, S. K., Katz, B. A. & Wyckoff, H. W. (1989). *J. Mol. Biol.* **205**, 557–571.

Kraulis, P. J. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **6**, 283–291.

Laskowski, R. A., MacArthur, M. W. & Thornton, J. M. (1994). In *From First Map to Final Model*, edited by S. Bailey. Warrington: Daresbury Laboratory.

Liu, J. & Magasanik, B. (1995). *J. Bacteriol.* **177**, 926–931.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Magasanik, B. (1993). *J. Cell. Biochem.* **51**, 34–40.

Mel, V. S. J. de, Kamberov, E. S., Martin, P. D., Zhang, J., Ninfa, A. J. & Edwards, B. F. P. (1994). *J. Mol. Biol.* **243**, 796–798.

Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364

Nar, H., Huber, R., Heizemann, C. W., Thorny, B. & Burgisser, D. (1994). *EMBO J.* **13**, 1255–1262.

Orengo, C. A. & Thornton, J. M. (1993). *Structure*, **1**, 105–120.

Parge, H. E., Arvai, A. S., Murtari, D. J., Reed, S. I. & Tainer, J. A. (1993). *Science*, **262**, 387–395.

Ploegman, J. H., Drenth, G., Kalk, K. H. & Hol, W. G. J. (1978). *J. Mol. Biol.* **123**, 577–594.

Powell, M. J. D. (1977). *Math. Program.* **12**, 241–254.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophysics J.* **5**, 909–933.
Sato, M., Yamamoto, M., Katasumi, I., Katsube, Y., Tanaka, N. & Higashi, T. (1992). *J. Appl. Cryst.* **25**, 348–357.
Stock, A. M., Mottonen, J. M., Stock, J. B. & Schutt, C. E. (1989). *Nature (London)*, **337**, 745–749.
Stock, J. B., Ninfa, A. J. & Stock, A. M. (1989). *Microbiol. Rev.* **53**, 450–490.
Swindells, M. G. & Alexandrov, N. N. (1994). *Nature Struct. Biol.* **1**, 677–678.
Takeuchi, Y., Satow, Y., Nakamura, K. T. & Mitsui, Y. (1991). *J. Mol. Biol.* **221**, 309–325.
Vasudevan, S. G., Gedye, C., Dixon, N. E., Cheah, E., Carr, P. D., Suffolk, P. M., Jeffrey, P. D. & Ollis, D. L. (1994). *FEBS Lett.* **337**, 255–258.
Walker, J. E., Saraste, M., Runswick, M. J. & Grey, N. J. (1982). *EMBO J.* **1**, 945–951.