Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of carbamoyl phosphate synthetase determined to 2.1 Å resolution

James B. Thoden,[a] Frank M. Raushel,[b] Matthew M. Benning,[a] Ivan Rayment[a] and Hazel M. Holden[a]*

[a]Institute for Enzyme Research, The Graduate School, and Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin–Madison, 1710 University Avenue, Madison, Wisconsin 53705, USA, and [b]Department of Chemistry, Texas A&M University, College Station, Texas 77843, USA

Correspondence e-mail: holden@enzyme.wisc.edu

Received 10 March 1998
Accepted 29 April 1998

PDB Reference: carbamoyl phosphate synthetase, 1jdb.

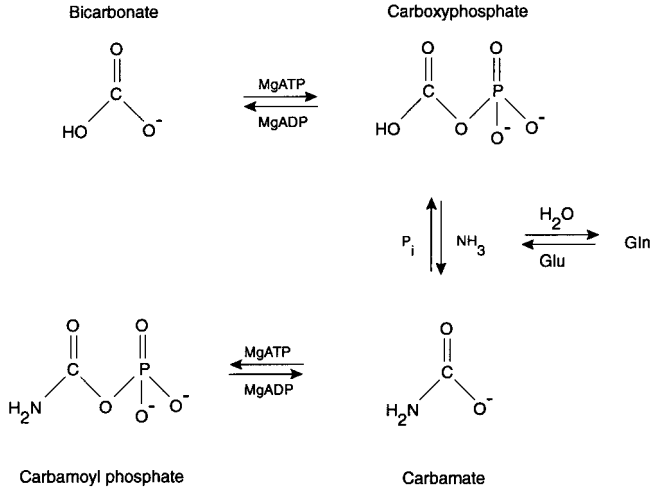
Carbamoyl phosphate synthetase catalyzes the formation of carbamoyl phosphate from one molecule of bicarbonate, two molecules of Mg^{2+}ATP and one molecule of glutamine or ammonia depending upon the particular form of the enzyme under investigation. As isolated from *Escherichia coli*, the enzyme is an α,β-heterodimer consisting of a small subunit that hydrolyzes glutamine and a large subunit that catalyzes the two required phosphorylation events. Here the three-dimensional structure of carbamoyl phosphate synthetase from *E. coli* refined to 2.1 Å resolution with an *R* factor of 17.9% is described. The small subunit is distinctly bilobal with a catalytic triad (Cys269, His353 and Glu355) situated between the two structural domains. As observed in those enzymes belonging to the α/β-hydrolase family, the active-site nucleophile, Cys269, is perched at the top of a tight turn. The large subunit consists of four structural units: the carboxyphosphate synthetic component, the oligomerization domain, the carbamoyl phosphate synthetic component and the allosteric domain. Both the carboxyphosphate and carbamoyl phosphate synthetic components bind Mn^{2+}ADP. In the carboxyphosphate synthetic component, the two observed Mn^{2+} ions are both octahedrally coordinated by oxygen-containing ligands and are bridged by the carboxylate side chain of Glu299. Glu215 plays a key allosteric role by coordinating to the physiologically important potassium ion and hydrogen bonding to the ribose hydroxyl groups of ADP. In the carbamoyl phosphate synthetic component, the single observed Mn^{2+} ion is also octahedrally coordinated by oxygen-containing ligands and Glu761 plays a similar role to that of Glu215. The carboxyphosphate and carbamoyl phosphate synthetic components, while topologically equivalent, are structurally different, as would be expected in light of their separate biochemical functions.

1. Introduction

Carbamoyl phosphate synthetase from *Escherichia coli*, hereafter referred to as CPS, catalyzes the first committed step in the separate biosynthetic pathways for the production of arginine and pyrimidine nucleotides. The heterodimeric protein is composed of a small amidotransferase subunit of molecular weight ~42 kDa that is complexed to a larger synthetase subunit of molecular weight ~118 kDa (Matthews & Anderson, 1972; Trotta *et al.*, 1971). Amino-acid sequence alignments (Piette *et al.*, 1984; Nyunoya & Lusty, 1983; Werner *et al.*, 1985) have demonstrated that the C-terminal half of the small subunit is homologous to the *trp*G-type glutamine amidotransferase domains of CTP synthetase (Weng *et al.*, 1986), GMP synthetase (Zalkin *et al.*, 1985) and *p*-aminobenzoate synthetase (Kaplan & Nichols, 1983; Kaplan *et al.*,

1985), among others. In addition, analysis of the primary structure of the larger subunit has revealed the unexpected presence of an internal tandem repeat such that the region of polypeptide chain defined by Met1–Glu403 is 40% identical to that delineated by Asn554–Asn936 (Nyunoya & Lusty, 1983). Site-directed mutagenesis experiments have further indicated that each of these components contains a binding site for one of the two Mg^{2+} ATP molecules required for the catalytic assembly of carbamoyl phosphate (Post *et al.*, 1990). These nucleotide-binding components are homologous in both primary and tertiary structure to the catalytic domains observed in biotin carboxylase (Waldrop *et al.*, 1994), D-alanine:D-alanine ligase (Fan *et al.*, 1994) and glutathione synthetase (Yamaguchi *et al.*, 1993). Chemical modification and site-directed mutagenesis experiments have further identified the C-terminal domain of the large subunit, defined by Ser937–Lys1073, as the primary binding site for the positive and negative allosteric effectors, ornithine and uridine 5′-monophosphate (UMP), respectively (Cervera *et al.*, 1996; Czerwinski *et al.*, 1995).

According to most biochemical data, CPS catalyzes the assembly of carbamoyl phosphate *via* a mechanism that requires at least four separate chemical reactions (Raushel *et al.*, 1978; Raushel & Villafranca, 1979). In the initial step, the first molecule of Mg^{2+}ATP phosphorylates bicarbonate to generate the reactive intermediate, carboxyphosphate. Concurrent with this reaction, glutamine is hydrolyzed *via* a thioester intermediate to glutamate and ammonia (Lusty, 1992). The final intermediate, namely carbamate, results from a nucleophilic attack of the ammonia on the carboxyphosphate intermediate. In the last chemical event, carbamate is phosphorylated by a second molecule of Mg^{2+} ATP to form carbamoyl phosphate.



The recent X-ray crystallographic investigation of CPS from *E. coli* to 2.8 Å resolution has yielded substantial progress towards understanding the manner in which the enzyme coordinates the regulated assembly of carbamoyl phosphate (Thoden *et al.*, 1997). As shown in Fig. 1, the CPS α,β-heterodimer may be envisioned as containing five major components. One of these components, the glutamine amidotransferase or small subunit, is positioned at the top of the large subunit and contains two structural domains defined by Leu1 to Leu153 and Asn154 to Lys381. The four other components are located in the large subunit and are color coded in green, yellow, blue and red for the regions defined by Met1–Glu403, Val404–Ala553, Asn554–Asn936 and Ser937–Lys1073, respectively. The regions defined by Met1–Glu403 and Asn554–Asn936 and referred to as the carboxyphosphate and carbamoyl phosphate synthetic components, respectively, are further broken down into three structural motifs, referred to as the *A*, *B* and *C* domains. Those regions delineated by Val404–Ala553 and Ser937–Lys1073 are referred to as the oligomerization and allosteric domains, respectively.

Perhaps the most intriguing result from the initial structural investigation of CPS was the location of the three active sites, which were shown to be separated by a linear distance of nearly 100 Å. This fact alone dictates that the reactive and unstable intermediates formed at one active site must be channeled through the interior of the protein for reaction at the next active site without their prior release into the bulk solvent. As such, the ammonia that is generated from the hydrolysis of glutamine within the amidotransferase subunit must diffuse directly through a molecular tunnel to the site of formation of carboxyphosphate. The carbamate intermediate synthesized at this site must subsequently diffuse to the homologous synthetic component on the opposite side of the protein where it is phosphorylated by the second molecule of ATP to form carbamoyl phosphate.

As expected, the carboxyphosphate and carbamoyl phosphate synthetic components are folded in a nearly identical fashion and the specific amino acids interacting with the bound nucleotides are similar. Moreover, in the folded conformation of the intact protein, these two homologous synthetic components, with their three structural domains, are related by a local twofold rotational axis. This observation suggests that the development of CPS activity evolved from an ancestral α_2-homodimeric precursor. In addition to describing the overall molecular architecture of CPS, our initial X-ray study (Thoden *et al.*, 1997) also defined the binding sites for both the positive allosteric effector, ornithine, and the negative allosteric effector, UMP. The binding site for ornithine is located at the interface between the allosteric domain and the carbamoyl phosphate synthetic component, while the binding site for UMP is located within the allosteric domain. The effect of these inhibitors or activators is primarily through the modulation of the Michaelis constant for Mg^{2+}ATP (Braxton *et al.*, 1992, 1996).

While the initial structural investigation of CPS defined the overall molecular architecture of the enzyme and the location of the three active sites, the protein model was based on an 'averaged' electron-density map and was not refined by least-squares analysis. Consequently, many of the details regarding the solvent structure, the coordination around the metals associated with the nucleotides, the active-site geometries and the subunit:subunit interactions between the small and large subunits of the α,β-heterodimer and between the α,β-heterodimers in the complete $(\alpha,\beta)_4$-tetramer were largely

unaddressed. In an effort to define more fully the various structural/functional relationships in this most remarkable enzyme, X-ray data have now been collected to 2.1 Å resolution at the Stanford Synchrotron Radiation Laboratory (SSRL). These data have allowed the model to be sufficiently improved such that, for the first time, many of these remaining questions can be effectively answered, as will be presented here.

2. Methods and materials

2.1. Purification and crystallization procedures

Protein employed in this investigation was purified as previously described (Mareya & Raushel, 1994). Crystals were grown by batch at 277 K (Thoden *et al.*, 1995, 1997) and subsequently stabilized with a synthetic mother liquor containing 1.0 *M* tetraethylammonium chloride, 8%(*w*/*v*) poly(ethylene glycol) 8000, 150 m*M* KCl, 2.5 m*M* ornithine, 2.5 m*M* $MnCl_2$, 5 m*M* ADP and 25 m*M* *N*-2-hydroxyethylpiperazine-*N*′-3-propanesulfonic acid (HEPPS; pH 7.4). Glutamine, at a concentration of 25 m*M*, was also added to the synthetic mother liquor in an attempt to trap the glutamyl thioester intermediate known to occur at Cys269 in the small subunit. Following overnight equilibration, the crystals were rapidly transferred to a cryoprotectant solution containing 1.4 *M* tetraethylammonium chloride, 8%(*w*/*v*) poly(ethylene glycol) 8000, 250 m*M* KCl, 2.5 m*M* ornithine, 2.5 m*M* $MnCl_2$, 5 m*M* ADP, 25 m*M* glutamine, 7.5%(*v*/*v*) ethylene glycol and 25 m*M* HEPPS (pH 7.4). They were then suspended in a thin film of the cryoprotectant solution in a loop of 20 µm surgical thread. The crystals were flash cooled to 123 K in a stream of nitrogen gas. At non-cryogenic temperatures, the unit-cell dimensions for these crystals were a = 154.1, b = 166.4 and c = 338.4 Å (space group $P2_12_12_1$). However, during the flash-cooling process, the crystals exhibited a significant and non-uniform change in the crystalline lattice dimensions. Specifically, there was a decrease in a by 10.3 to 143.8 Å (−6.7%), an increase in b by 1.3 to 167.7 Å (0.8%) and a decrease in c by 15.4 to 323.0 Å (−4.6%), all of which reduced the unit-cell volume by 10%. Quite remarkably, these crystals still maintained excellent X-ray diffraction properties with low mosaic character.

2.2. X-ray data collection and processing

To assess crystal quality, partial X-ray data sets to 2.8 Å resolution were initially collected with radiation from a rotating-anode X-ray generator at 123 K. Crystals demonstrating high-quality X-ray diffraction properties were subsequently stored under liquid-nitrogen conditions until X-ray data-collection time became available at SSRL. The data employed for the least-squares model refinement described here were collected on beamline 7-1 with the MAR300 image-plate system. A low-resolution X-ray data set, consisting of 90 1° frames with the direct beam centered on the detector and a crystal-to-detector distance of 420 mm, was collected first. These frames were collected with a constant number of photons per frame. The detector was then translated up into its offset position, again at a crystal-to-detector distance of 420 mm. A total of 200 1° frames were collected. Typically the crystals displayed diffraction to a nominal resolution of 1.8 Å, but the completeness of the X-ray data dropped off significantly beyond 2.1 Å because of a number of factors, including overlapping reflections and hardware limitations of the detector. The frames were processed with *DENZO* and scaled with *SCALEPACK* (Otwinowski, 1986). From 290 images, 1 652 264 measurements were integrated, which reduced to 404 727 unique reflections after scaling. Scaling statistics are presented in Table 1.

Because of the large changes in the unit-cell parameters upon flash cooling,

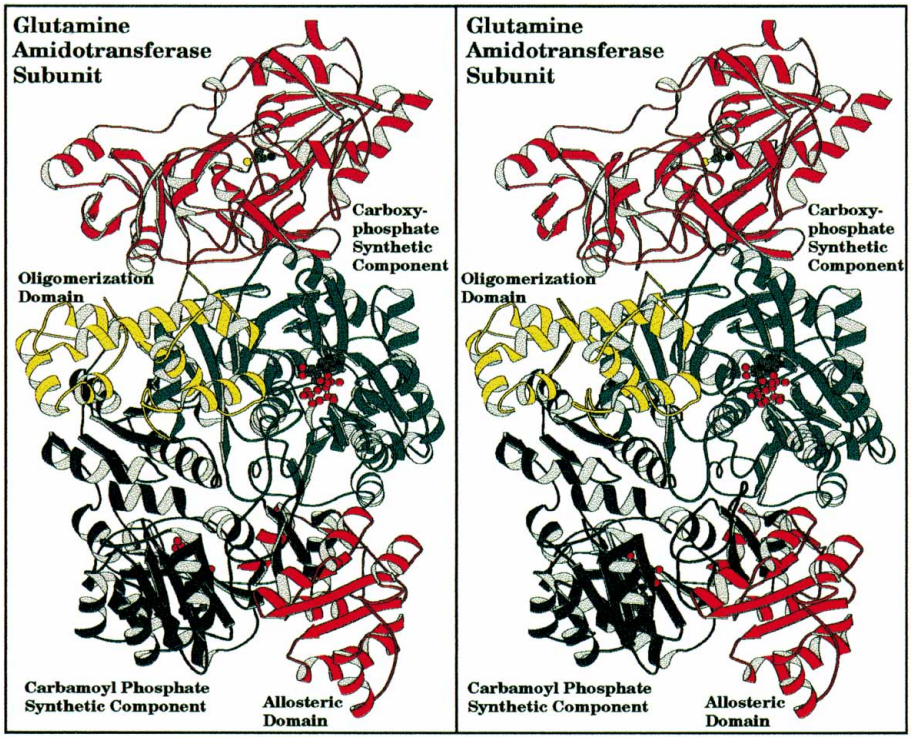


Figure 1
Ribbon representation of the CPS α,β-heterodimer. The small subunit is displayed in magenta. In the large subunit, the polypeptide-chain regions defined by Met1–Glu403, Val404–Ala553, Asn554–Asn936 and Ser937–Lys1073 are depicted in green, yellow, blue and red, respectively. These regions represent the carboxyphosphate synthetic component, the oligomerization domain, the carbamoyl phosphate synthetic component and the allosteric domain, respectively. The color scheme shown here is maintained throughout the paper. The active-site cysteine in the small subunit (Cys269) and the two ADP molecules in the large subunits are shown in ball-and-stick representations.

Table 1
Intensity statistics.

R factor = $(\sum |I - \bar{I}| / \sum I) \times 100$.

	Resolution range (Å)										
	Overall	30.0–4.52	3.59	3.14	2.85	2.65	2.49	2.37	2.26	2.18	2.10
Independent reflections	404724	45759	45349	44840	42543	41573	40508	39144	37577	35080	32559
Completeness (%)	90	99	100	99	94	92	90	87	88	78	73
Average I/average $\sigma(I)$	22.9	52.2	54.1	42.3	21.9	15.2	10.3	7.6	5.7	4.3	3.2
R factor (%)	5.9	4.3	5.2	6.7	7.7	8.7	10.3	12.3	14.6	17.9	20.3

the structure reported here was necessarily solved by molecular replacement (Rossmann, 1972) with the software package *AMoRe* (Navaza, 1994) and employing as a search model the complete $(\alpha,\beta)_4$-tetrameric form of CPS previously determined at 2.8 Å resolution (Thoden *et al.*, 1997). Following rigid-body refinement, the model was subjected to least-squares analysis at 2.1 Å resolution with the software package *TNT* (Tronrud *et al.*, 1987). After 15 cycles of refinement, the R factor decreased from 35.9 to 27.3%. With over 5800 amino-acid residues in the asymmetric unit, the goal of the model-building process was to lower the R factor as much as possible using an averaged α,β-heterodimer before finally rebuilding the entire $(\alpha,\beta)_4$-tetramer in the asymmetric unit. Consequently, to expedite the refinement and rebuilding processes, the electron densities corresponding to the four α,β-heterodimers in the asymmetric unit were averaged according to the algorithm of Bricogne (1976) and one α,β-heterodimer was rebuilt into the averaged map. Following this rebuilding process, the entire $(\alpha,\beta)_4$-tetramer was reconstructed from the averaged α,β-heterodimer and placed back into the unit cell. Additional cycles of least-squares refinement decreased the R factor to 22.3%. Again, the electron-density map was averaged and the α,β-heterodimer rebuilt on the basis of this map. In addition, approximately 1000 solvent molecules were positioned into the averaged electron-density map. Following this second rebuilding procedure of the α,β-heterodimer, the complete CPS $(\alpha,\beta)_4$-tetramer was reconstructed, placed back into the unit cell and refined. The R factor decreased to 19.1%. At this point the entire $(\alpha,\beta)_4$-tetramer model was adjusted in the asymmetric unit and additional solvent molecules were added. Following ten more cycles of refinement, the R factor was reduced to 17.9%. Note that non-crystallographic symmetry constraints were not applied during the least-squares refinement. The final model includes eight ADP molecules, 12 manganese ions, four ornithines, eight glutamines, nine inorganic phosphates, 33 potassium ions, 29 chloride ions, 4857 water molecules and four tetraethylammonium ions. Relevant refinement statistics are presented in Table 2. The following side chains (located in the large subunits) were modeled as multiple conformations: Glu59, Arg130, Glu419, Asp430, Gln519, Asp521, Asp758, Arg912 and Gln967 in α,β-heterodimer I; Ser29, Arg130, Arg145, Arg185, Arg343, Asp430, Arg509, Ser556, Met772, Glu951, Glu955, Gln967 and Arg1021 in α,β-heterodimer II; Ser29, Glu186, Met336, Glu512, Asp518 and Lys833 in α,β-heterodimer III; and Arg509, Gln811, Arg1004, Asn1007 and Glu1009 in α,β-heterodimer IV. Also, Asp368 in the small subunit of α,β-heterodimer II adopted multiple conformations. The only residues adopting dihedral angles well outside the allowed regions of the Ramachandran plot are Cys269 and Ala356, both of which are located in the small subunit. These residues will be more fully described in §3.

Table 2
Least-squares refinement statistics.

R factor = $\sum |F_o - F_c| / \sum |F_o|$ where F_o is the observed structure-factor amplitude and F_c is the calculated structure-factor amplitude. The torsional angles were not restrained during the refinement.

Resolution limits (Å)	30.0–2.1
R factor (%)	17.9
No. of reflections used	404724
No. of protein atoms	44402
No. of solvent atoms	5334
Weighted r.m.s. deviations from ideality	
Bond length (Å)	0.011
Bond angle (°)	2.61
Planarity (trigonal) (Å)	0.004
Planarity (other planes) (Å)	0.009
Torsional angle (°)	17.6

3. Results

3.1. The solvent structure

Perhaps one of the more remarkable aspects of the refined CPS model presented here is the quality of the solvent structure. Note that the X-ray data employed for the model refinement were collected at 123 K and the CPS crystals displayed a solvent content of nearly 60%. Indeed, the number of water molecules, 4857, is an under- rather than overestimate of the solvent structure in that there are more than 5800 amino-acid residues in the asymmetric unit. The average temperature factor for the water molecules is 39.9 Å^2. In addition to these water molecules, four tetraethylammonium ions, eight glutamines, nine inorganic phosphates, 29 chloride ions and 33 potassium ions were also located. Shown in Fig. 2(*a*) is an example of the electron density for one of the observed tetraethylammonium ions. The tetraethylammonium ion displayed in Fig. 2(*a*) is associated with α,β-heterodimer I; the other three are located in similar positions in heterodimers II, III and IV. As can be seen, this ion is surrounded by rings of water molecules forming either pentagonal or hexagonal arrays. These types of arrays, which are observed for each of the four tetraethylammonium ions in the CPS asymmetric

unit, have been observed previously, for example, in the high-resolution structural investigation of crambin (Teeter, 1984, 1991). In addition to these rings of water molecules, there is a chloride ion also located within 5 Å of the tetraethylammonium nitrogen. The electron density corresponding to this ion was originally modeled as a water molecule. However, the temperature factor dropped to an anomalously low value indicating that, in fact, the density corresponded to something more electron dense, such as a chloride ion. As shown in Fig. 2(*b*), the tetraethylammonium ion is wedged in a solvent cavity formed mostly by the carboxyphosphate synthetic component of the large subunit and is located within approximately 12 Å of the ADP molecule associated with this synthetic unit. Each of the α,β-heterodimers of the complete $(\alpha,\beta)_4$-tetramer contains at least two large pockets of solvent occupied by over 100 water molecules.

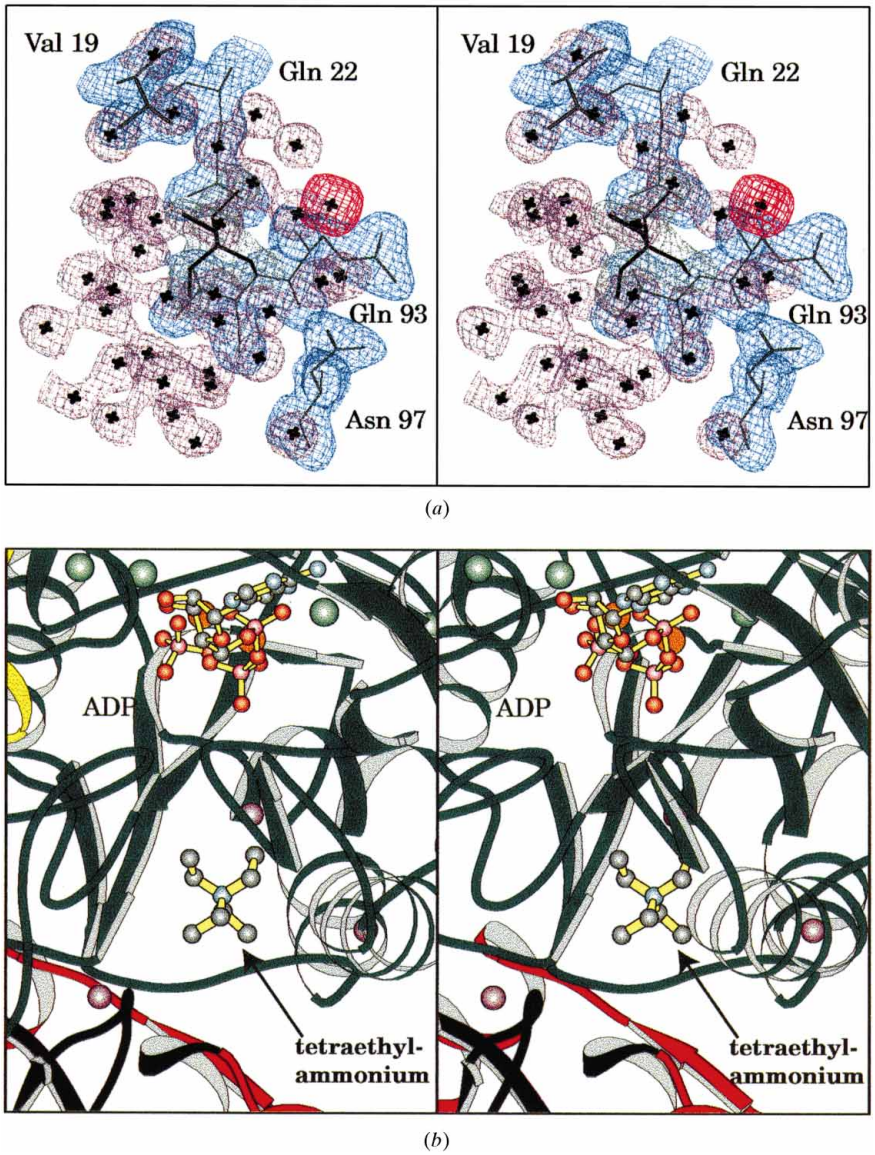


Figure 2
(*a*) Shown is a portion of the electron-density map corresponding to one of the ordered tetraethylammonium ions. The map was contoured at 1σ and calculated with coefficients of the form $(2F_o - F_c)$. Protein atoms, water molecules, the tetraethylammonium ion and the chloride ion are displayed as blue, lavender, green and red electron densities, respectively. (*b*) The tetraethylammonium ion is bound within a solvent cavity formed largely by the polypeptide chain corresponding to the carboxyphosphate synthetic component. Bound chloride and potassium ions are indicated by the purple and green spheres, respectively. The manganese ion is depicted as an orange sphere.

3.2. The small subunit

For the sake of simplicity, the remainder of the discussion will refer to α,β-heterodimer I in the asymmetric unit unless otherwise indicated. As can be seen in Fig. 3, the three-dimensional architecture of the small subunit can be described in terms of two distinct motifs defined by Leu1 to Leu153 and Asn154 to Lys382 and referred to as the N-terminal and C-terminal domains, respectively. These two domains are connected by a type-II turn formed by Leu153 to Met156. The electron densities for residues Leu1, Ala381 and Lys382 were weak for all four copies of the small subunit in the asymmetric unit and hence were not included in the present model. In addition, the small subunit contains one *cis*-proline at position 358. The polypeptide-chain backbone atoms for each small subunit in the asymmetric unit correspond with an average root-mean-square (r.m.s.) deviation of 0.26 Å.

The tertiary structure of the N-terminal domain is dominated by seven β-strands that fold into two layers of β-sheet oriented nearly perpendicular to each other. These β-strands range in length from four to ten residues. One of these layers contains four strands of anti-parallel β-sheet while the other is composed of four parallel β-strands. The four α-helices in the N-terminal domain of the CPS small subunit range in length from five to 11 amino-acid residues. Connecting these β-strand and α-helices are a number of reverse turns, including seven type-I, one type-II and one type-III. The C-terminal domain of the CPS small subunit, containing 228 amino acids, is constructed from six α-helices ranging in length from four to 15 residues, ten β-strands ranging in length from three to 17 residues, 13 type-I turns and one type-III turn.

While three-dimensional structural searches failed to reveal any significant homology between the N-terminal domain of the CPS small subunit and other proteins of known structure, it is absolutely clear that the C-terminal domain of the small subunit is highly homologous to the N-terminal domain of GMP synthetase and to other members of the *trp*G-type amido-

transferases (Tesmer *et al.*, 1996). A superposition of the C-terminal domain of the CPS small subunit onto the N-terminal domain of GMP synthetase is depicted in Fig. 4. In addition to the topological similarities in their three-dimensional structures, both of these enzymes contain a catalytic triad composed of a glutamate, a histidine and a cysteine residue. The positions of these residues are indicated in Fig. 4. It is clear from an electron-density map, calculated with coefficients of the form $(F_o - F_c)$, that Cys269 in the CPS small subunit became oxidized to sulfenic acid or a similar type of moiety during the crystallization and X-ray data-collection processes. Consequently, the different orientations observed for the cysteine side chains in the CPS small subunit and GMP synthetase are most likely due to variations in the redox states of these sulfhydryl groups. As observed in GMP synthetase and in those enzymes belonging to the α/β-hydrolase fold (Ollis *et al.*, 1992), the dihedral angles of the nucleophile, Cys269, lie in a disallowed region of the Ramachandran plot (φ = 59.1°, ψ = −96.0°). This cysteine residue resides in a so-called 'nucleophile elbow' and serves as a bridge between one β-strand and the following α-helix.

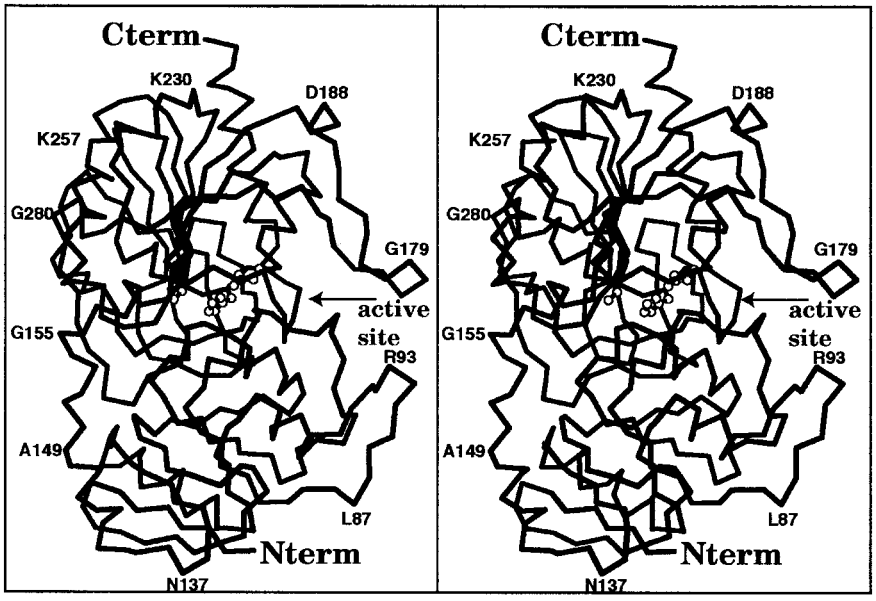


Figure 3
The small subunit of CPS. Shown here is an α-carbon trace of the small subunit. Various amino-acid residues are labeled to aid the viewer in following the course of the polypeptide chain. The catalytic triad is shown in ball-and-stick representation.

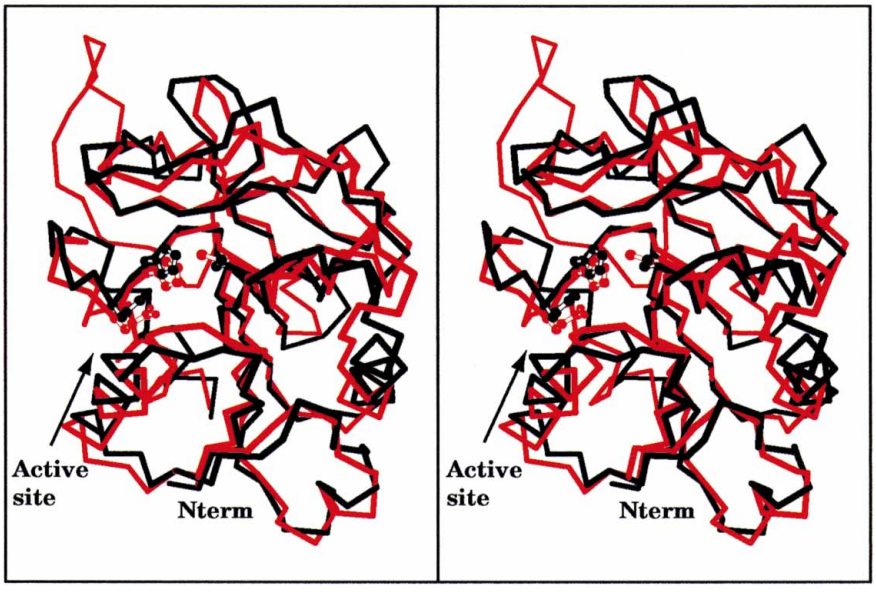


Figure 4
Superposition of the C-terminal domain of the CPS small subunit onto the N-terminal domain of GMP synthetase. α-carbon traces for GMP synthetase and the CPS small subunit are displayed in red and black, respectively. Those residues constituting the catalytic triad are shown in ball-and-stick representations. X-ray coordinates for GMP synthetase were obtained from the Brookhaven Protein Data Bank.

Those residues located within approximately 6 Å of Cys269 S^{γ} are depicted in Fig. 5. Although the carboxylate group of Glu355 is approximately 7.8 Å from Cys269, it is also included in the figure since it constitutes part of the putative catalytic triad. The region of polypeptide chain immediately following Glu355 is quite strained, with Ala356 adopting φ and ψ angles of 60.1 and −133.5°, respectively, and Pro358 oriented in the *cis* conformation. The imidazole ring of His358 is oriented such that $N^{\delta 1}$ lies within 2.6 Å of $O^{\varepsilon 1}$ of Glu355 and $N^{\varepsilon 2}$ is located at 4.0 Å from S^{γ} of Cys269. As indicated by the red spheres in Fig. 5, the active site of the small subunit is relatively hydrophilic, with five ordered water molecules located near Cys269. While the crystals employed in the X-ray data collection were soaked in a solution containing glutamine, it was clear from the electron-density maps that glutamine did not bind in the small subunit active site. However, from site-directed mutagenesis experiments, it is known that His312 is important for substrate binding (Miran *et al.*, 1991). Indeed, when this residue is replaced with an asparagine, the K_m for glutamine increases significantly (Miran *et al.*, 1991). Specifically, the K_m for glutamine in the overall reaction of CPS in the H312N mutant was found to be 20 m*M* and in the glutaminase reaction alone the K_m was determined to be 14 m*M*. The K_m values for the wild-type enzyme in the same study were 0.10 and 0.3 m*M*, respectively. Consequently, it can be speculated that the region of the active site near His312, as shown in Fig. 5, is important for proper substrate positioning. Experiments designed to explore the manner in which glutamine binds in the small subunit active

site are in progress. The subunit:subunit interactions in the CPS α,β-heterodimer are quite extensive. Specifically there are 35 direct hydrogen bonds between protein atoms that link the large and small subunits together. In addition, there are numerous water molecules that act as bridges between the two subunits. A listing of these interactions is given in Table 3.

3.3. The large subunit

The model for the large subunit described here begins at Pro2 and ends at Lys1073 with two breaks in the polypeptide chain between Arg715 and Ala724 and Ala741 and Val750. For the sake of simplicity the large subunit will be discussed below in terms of the four structural units: the carboxyphosphate synthetic component (Met1–Glu403), the oligomerization domain (Val404–Ala553), the carbamoyl phosphate synthetic component (Asn554–Asn936) and the allosteric domain (Ser937–Lys1073). There are five *cis*-prolines in the large subunit. As described in the initial structural investigation (Thoden *et al.*, 1997), the C_α atoms of the carboxyphosphate and carbamoyl phosphate synthetic components are related by a nearly exact twofold rotational axis and, indeed, the *cis*-prolines at positions 165 and 252 align with those at positions 711 and 797, respectively. The fifth *cis*-proline at position 999 residues in the allosteric domain of the large subunit. The C_α atoms for all four large subunits in the asymmetric unit superimpose with an average r.m.s. deviation of 0.28 Å, while the C_α atoms for all four α,β-heterodimers superimpose with an average r.m.s. deviation of 0.36 Å. Note that those amino-acid residues lining the subunit:subunit interface of the α,β-heterodimer are contributed by both the N- and C-terminal domains of the small subunit but only by the carboxyphosphate synthetic component and the oligomerization domain of the large subunit. There are no direct interactions between the small subunit and the carbamoyl phosphate synthetic component.

3.3.1. The carboxyphosphate synthetic component (Met1–Glu403). An α-carbon trace of this structural unit, shown in Fig. 6(*a*), can be described in terms of three domains, *A*, *B* and *C*. The *A* domain, formed by Met1 to Gly140, contains five strands of parallel β-pleated sheet flanked on one side by two α-helices and on the other by a long α-helix of 14 amino-acid residues and an additional region containing four type-I turns in tandem. These β-strands and α-helices are connected by a variety of reverse turns. Following the last β-strand in the *A* domain, the polypeptide chain folds into a rather long α-helix defined by Ala120 to Met136. Because of the dihedral angles adopted by Asp128 ($\varphi = -79.5^\circ$, $\psi = 111.2^\circ$), there is a decided kink in this region such that the helix axis defined by Ala120 to Glu 127 lies at nearly a right angle to that defined by Arg129 to Met136.

The *B* domain, delineated by Ala144 to Ser209, is composed of four strands of antiparallel β-sheet flanked on one side by two α-helices of lengths 11 and 14 residues. The second and third β-strands in the *B* domain are connected by a type-II, type-I′ and type-II series of turns starting at Pro170. The *cis*-proline at position 165 lies in a surface loop connecting the first α-helix of the *B* domain with the second strand of β-sheet. From previous studies, it is known that the *A*, *B* and *C* motifs of the carboxyphosphate synthetic unit are similar to those observed in biotin carboxylase (Waldrop *et al.*, 1994; Thoden *et al.*, 1997). Interestingly, the *cis*-proline at position 165 in CPS is conserved in biotin carboxylase as *cis*-Pro155. The *B* and *C* domains are connected by a type-II turn defined by Leu210 to Trp213.

Clearly the most complicated of the three motifs, the *C* domain, is dominated by an anti-parallel seven-stranded β-sheet flanked on one side by an α-helix (Asp258–Gly276) oriented diagonally across the sheet, and on the other side by

Table 3
List of interactions between the large and small subunits of the CPS α, β-heterodimer.

Direct protein:protein contacts

Large subunit		Small subunit		
Residue No.	Atom	Residue No.	Atom	Distances (Å)
Asn227	$O^{\delta1}$	Val305	N	3.1
Gln254	$O^{\varepsilon1}$	Asn62	$N^{\delta2}$	3.0
Gln254	$O^{\varepsilon1}$	Tyr57	O^{η}	2.7
Gln254	$N^{\varepsilon2}$	Tyr57	O^{η}	2.8
Gln254	O	Asn62	$N^{\delta2}$	3.2
Thr257	O^{γ}	Arg93	$N^{\eta1}$	2.9
Asp258	$O^{\delta2}$	Ala89	N	2.9
Lys259	N^{ζ}	Asp69	$O^{\delta2}$	3.1
Arg265	$N^{\eta1}$	Asp362	$O^{\delta2}$	2.8
Arg265	$N^{\eta2}$	Asp362	$O^{\delta1}$	3.2
Arg265	$N^{\eta2}$	Asp362	$O^{\delta2}$	3.3
Asn266	$O^{\delta1}$	His361	$N^{\delta1}$	3.4
Asp333	$O^{\delta2}$	Lys298	N^{ζ}	2.9
Glu349	O	Asn294	$N^{\delta2}$	2.8
Ser351	O^{γ}	Thr34	O	2.7
Arg389	$N^{\eta2}$	Arg116	N	3.3
Arg389	$N^{\eta2}$	Thr115	O^{γ}	3.0
Gln391	N	His59	$N^{\varepsilon2}$	2.9
Asp459	$O^{\delta2}$	Ser90	O^{γ}	2.7
Trp461	$N^{\varepsilon1}$	Asn62	$O^{\delta1}$	2.8
Arg494	$N^{\eta2}$	Arg83	$N^{\eta1}$	3.4
Arg494	$N^{\eta2}$	Arg83	$N^{\eta2}$	3.3
Lys527	N^{ζ}	Asp114	$O^{\delta1}$	2.9
Lys527	N^{ζ}	Asp114	$O^{\delta2}$	3.2
Arg528	O	Arg116	$N^{\eta2}$	3.1
Asp530	$O^{\delta2}$	Arg116	$N^{\eta1}$	3.0
Asp530	$O^{\delta2}$	Arg116	$N^{\eta2}$	3.1
Ala534	O	Arg123	N^{ε}	2.8
Ala534	O	Arg123	$N^{\eta2}$	2.9
Glu548	$O^{\varepsilon1}$	Arg83	N^{ε}	3.3
Glu549	N	Asp114	$O^{\delta1}$	3.0
Glu550	N	Asp114	$O^{\delta1}$	3.2
Glu550	$O^{\varepsilon2}$	Arg120	$N^{\eta1}$	2.8
Glu552	$O^{\varepsilon1}$	Arg116	$N^{\eta2}$	3.1
Glu552	$O^{\varepsilon2}$	Arg116	N^{ε}	2.8

Table 3 (continued)

Protein:protein contacts mediated *via* water molecules

Large subunit					Small subunit	
Residue No.	Atom	Distance (Å)	Water	Distance (Å)	Residue No.	Atom
Cys228	N	2.8	HOH	3.0	Val305	O
Cys228	O	2.6	HOH	2.9	Asp362	$O^{\delta1}$
Ala253	O	3.2	HOH	2.9	Met36	O
Lys259	N^{ζ}	2.7	HOH	2.5	Glu41	$O^{\varepsilon1}$
Lys259	N^{ζ}	2.7	HOH	2.6	Ser357	O^{γ}
Lys259	N^{ζ}	3.1	HOH	2.9	Ala68	O
Lys259	N^{ζ}	3.1	HOH	2.6	Glu71	$O^{\varepsilon2}$
Lys259	N^{ζ}	3.1	HOH	3.1	Ser357	O^{γ}
Lys259	N^{ζ}	2.7	HOH	3.1	Ala68	O
Gln262	$O^{\varepsilon1}$	2.7	HOH	2.9	His361	N
Arg265	$N^{\eta1}$	2.8	HOH	2.9	His361	N
Asn266	$N^{\delta2}$	2.8	HOH	2.9	His361	N
Pro290	O	3.0	HOH	3.0	Phe92	N
Asn292	O	3.2	HOH	3.4	Leu177	O
Asp333	$O^{\delta2}$	2.9	HOH	3.0	Asn303	O
Asp333	$O^{\delta2}$	2.7	HOH	3.3	Asn303	$N^{\delta2}$
Glu349	$O^{\varepsilon1}$	2.7	HOH	3.0	Gly293	O
Ser351	O^{γ}	3.1	HOH	2.8	Thr56	O^{γ}
Ile352	O	2.7	HOH	2.8	Thr56	O^{γ}
Arg389	$N^{\eta1}$	3.1	HOH	2.5	Thr115	O^{γ}
Arg389	$N^{\eta1}$	3.1	HOH	2.8	Thr56	O
Gln392	$O^{\varepsilon1}$	2.9	HOH	2.8	Asp84	$O^{\delta1}$
Asp459	$O^{\delta1}$	2.9	HOH	3.0	Ser90	O
Arg528	O	2.8	HOH	3.0	Arg116	$N^{\eta1}$
Asp530	$O^{\delta2}$	2.8	HOH	3.0	Arg123	$N^{\eta2}$
Cys532	O	2.6	HOH	2.8	Asn294	$N^{\delta2}$
Glu535	$O^{\varepsilon1}$	2.8	HOH	2.9	His291	N
Glu535	$O^{\varepsilon1}$	3.1	HOH	2.7	His295	$N^{\delta1}$
Glu535	$O^{\varepsilon2}$	2.4	HOH	2.4	Thr34	O^{γ}
Glu535	$O^{\varepsilon2}$	2.4	HOH	2.7	Asn294	$O^{\delta1}$
Glu535	$O^{\varepsilon2}$	3.2	HOH	2.8	Asn294	$N^{\delta2}$
Phe536	O	2.5	HOH	3.0	Arg123	$N^{\eta2}$
Tyr544	O^{η}	2.8	HOH	3.0	Arg116	$N^{\eta1}$
Glu548	$O^{\varepsilon2}$	3.0	HOH	2.4	Arg83	$N^{\eta2}$
Glu550	$O^{\varepsilon1}$	2.7	HOH	2.7	Lys117	O
Asn554	$N^{\delta2}$	3.1	HOH	3.2	Arg120	$N^{\eta2}$

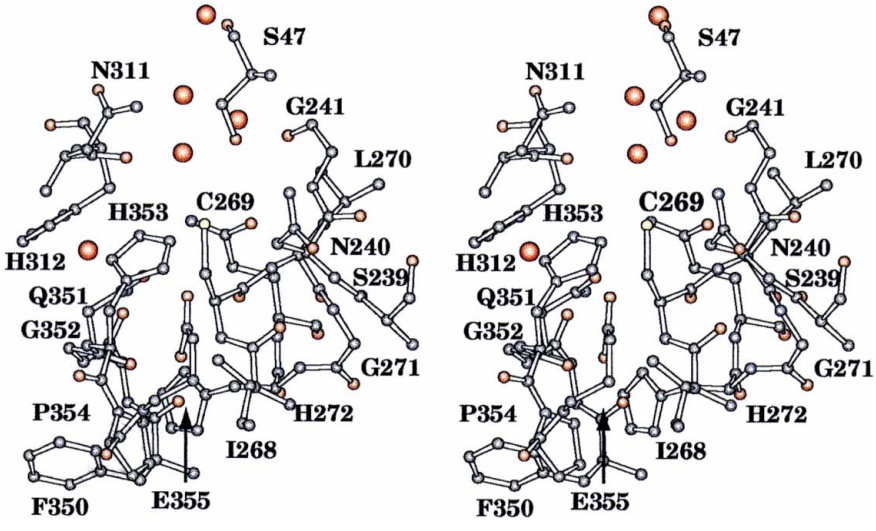


Figure 5
A close-up view of the CPS small subunit active site. Those residues located within approximately 6 Å of Cys269 are displayed. Ordered water molecules are represented as red spheres.

two smaller α-helices (Arg306–Thr315 and Ile319–Lys325). There is a decided bulge in the second β-strand of the sheet, resulting from Val231 adopting dihedral angles of $\varphi = -75.2^{\circ}$ and $\psi = -34.2^{\circ}$. In addition, there is a *cis*-proline at position 252 which resides in an extended loop connecting the third β-strand to the first α-helix of the *C* domain. This *cis*-proline residue is also conserved in biotin carboxylase at position 244.

Perhaps the most significant improvement regarding the present CPS model, as compared with the original structure solved at 2.8 Å resolution, is the definition of the active-site geometries. From this higher-resolution X-ray study it has been possible to characterize more fully the hydrogen-bonding patterns between the nucleotides and the protein, the solvent structure within the active sites, and the coordination geometries surrounding both the manganese and the potassium ions. A cartoon depicting the major interactions between the protein and the ADP/P_i is shown in Fig. 7. The adenine ring of the ADP is hydrogen bonded to the protein *via* backbone atoms, namely the carbonyl O atom of Glu208 and the amide H atom of Leu210. An additional water molecule lies at 3.0 Å from N7 of the purine ring. Glu215 anchors both the 2′- and 3′-hydroxyl groups of the nucleotide ribose and also acts as a bridge to the potassium binding site, as indicated in Fig. 8(*a*). The α- and β-phosphate moieties of the ADP interact with the guanidinium groups of Arg169 and Arg129, respectively. His243, through an imidazole nitrogen, also serves to anchor one of the β-phosphoryl oxygens. The inorganic phosphate observed in the active site of the carboxyphosphate synthetic component interacts with Arg306. Interestingly, the position of this inorganic phosphate is very similar to that observed in biotin carboxylase, another enzyme whose reaction mechanism is thought to proceed through a carboxyphosphate intermediate (Waldrop *et al.*, 1994). It can thus be speculated that this region of CPS is involved in the stabilization of the carboxyphosphate intermediate.

Both manganese ions in the active site are octahedrally coordinated by oxygen-containing ligands and are bridged by the carboxylate side chain of Glu299. Specifi-

cally, MnI is ligated by $O^{\varepsilon1}$ of Gln285, $O^{\varepsilon1}$ of Glu299, a water molecule and three phosphoryl oxygens, two from the α- and β-phosphate groups of the nucleotide and the third from the inorganic phosphate. The second metal, MnII, is coordinated by $O^{\delta1}$ of Asn301, $O^{\varepsilon1}$ and $O^{\varepsilon2}$ of Glu299, a water molecule and two phosphoryl oxygens, one contributed by the β-phosphate group of the ADP and the other by the inorganic phosphate. Bond lengths between the metals and the ligands range from 2.0 to 2.4 Å. Note that $MnCl_2$, rather than $MgCl_2$, was employed in the crystallization experiments, simply because the crystals appeared to grow somewhat better in the presence of manganese. The enzyme is fully active in the presence of manganese, however. In addition to the manganese ions, there are several potassiums located in the active site.

Since CPS is known to be activated by potassium (Anderson & Meister, 1966), the presence of these cations is not surprising. Of particular interest, however, is the binding site for the potassium ion shown in Fig. 8(*a*). In this case, the potassium ion is octahedrally coordinated by three O atoms contributed by backbone carbonyl groups (Asp238, Ala239 and Ile242) and three oxygens provided by side-chain functional groups (Asn236, Ser247 and Glu215). The bond distances between the potassium and the ligands range from 2.5 to 2.8 Å and the potassium is located at approximately 9 Å from MnI in the ADP binding site. Most likely this cation-binding pocket, which will be referred to as the 'K-loop', is the physiologically relevant site that allows direct communication from the potassium to the carboxyphosphate synthetic unit active site *via* the carboxylate group of Glu215. Clearly, the structure of CPS solved in the absence of potassium ions would serve to define more fully the structural basis of cation activation; this work is in progress.

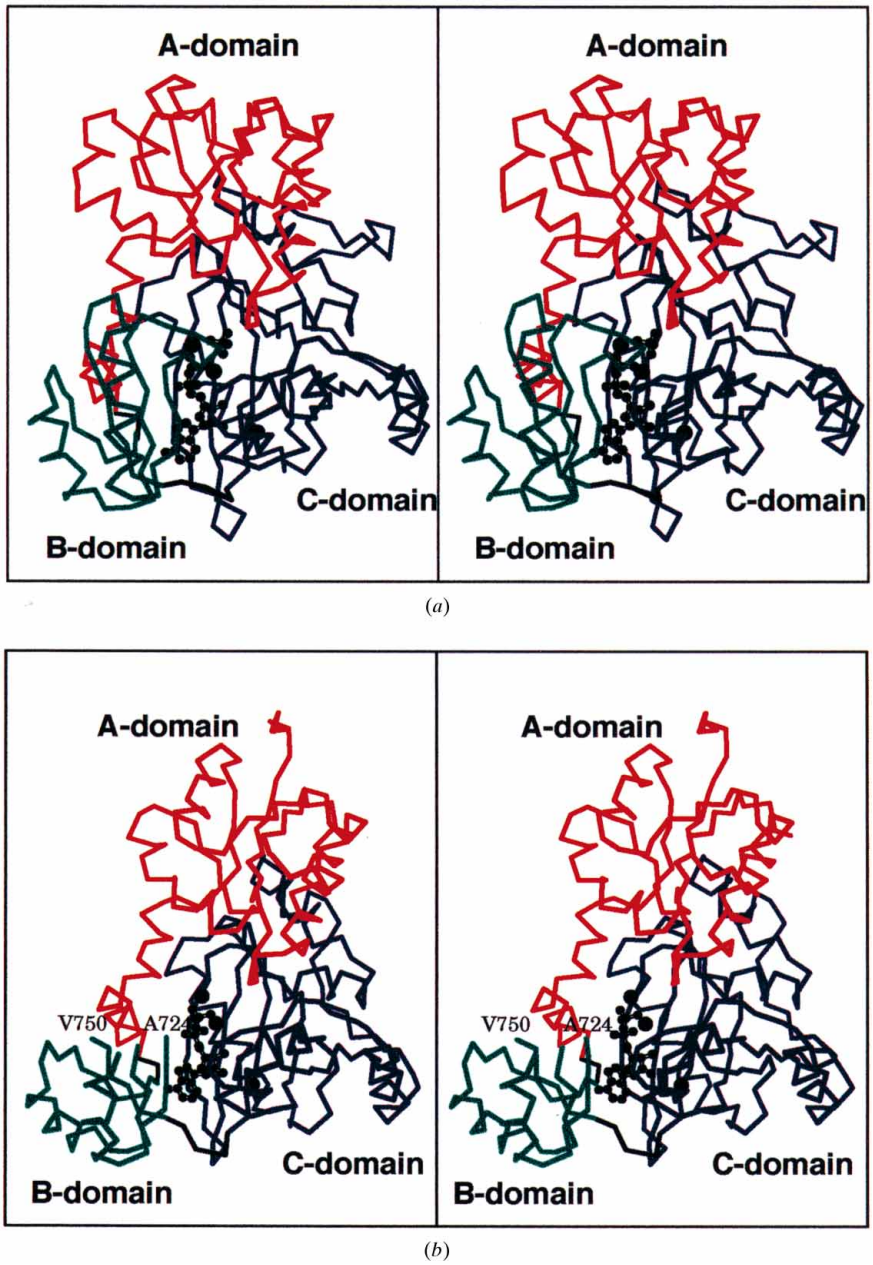


Figure 6
α-Carbon traces of the two synthetase components of the large subunit. The carboxyphosphate and carbamoyl phosphate synthetic units are shown in (*a*) and (*b*), respectively. The three structural domains, *A*, *B* and *C*, are color coded in red, green and blue, respectively. The ADP, manganese ions, inorganic phosphate and potassium ions residing in the 'K-loops' are displayed in ball-and-stick representations.

3.3.2. The oligomerization domain (Val404–Ala553). The locations of the oligomerization domains in the CPS $(\alpha,\beta)_4$-tetramer are shown in Fig. 9(*a*). Clearly the simplest of the four components of the large subunit, each oligomerization domain folds into seven α-helices and two strands of anti-parallel β-sheet. These secondary structural elements are punctuated by eight type-I turns, one type-I′ turn and one type-III turn. The antiparallel β-strand motif of the oligomerization domain provides a substantial portion of the molecular interface formed between the small and large subunits of the α,β-heterodimer. Specific electrostatic interactions are listed in Table 3.

The subunit:subunit interface between two oligomerization domains in the CPS $(\alpha,\beta)_4$-tetramer is formed by the first α-helix (Leu421–Lys429) and the loop region connecting the second and third α-helices (Gly446–Ser448). Specific electrostatic interactions between one oligomerization domain and its symmetry-related partner (indicated by *) are as follows: O of Leu415 and $N^{\eta2}$ of Arg425*; O of Pro418 and O^{γ} of Thr422*; N of Ser448 and O of Gly446*.

Indeed, the number of interactions is quite small, as would be expected for a molecular species that readily converts between an α,β-heterodimer and an $(\alpha,\beta)_4$-tetramer. On the other hand, however, there are four leucine residues that form a hydrophobic pocket, as shown in Fig. 9(*b*), and most likely these amino acids play a significant role in maintaining the proper quarternary structure of the enzyme. There are few water molecules positioned within the subunit:subunit interface.

3.3.3. The carbamoyl phosphate synthetic component (Asn554–Asn936). From amino-acid sequence analyses (Nyunoya & Lusty, 1983) it was known that the three-dimensional fold of the carbamoyl phosphate synthetic component would most likely be topologically similar to that observed for the carboxyphosphate synthetic component. Indeed, these two units are topologically but certainly not structurally equivalent, as recently concluded by Kothe *et al.* (1997). For example, while the active site of the carboxyphosphate synthetic component contains an ordered phosphate molecule, there are no such ions in the carbamoyl phosphate synthetic component. Indeed, this is in keeping with the fact that the carboxyphosphate synthetic component, but not the carbamoyl phosphate synthetic component, must provide a region for stabilization of the carboxyphosphate intermediate. An α-carbon trace of the carbamoyl phosphate synthetic component is displayed in Fig. 6(*b*). As can be seen in Fig. 6(*b*), there are two breaks in the polypeptide-chain backbone between Arg715 and Ala724 and Ala741 and Val750. These are the only breaks in the electron density for the entire α,β-heterodimer. Note that in the carboxyphosphate synthetic component show in Fig. 6(*a*), the backbone amide nitrogen of Gly175, which is part of the *B* domain, lies within hydrogen-bonding distance of the inorganic phosphate. The lack of this ion in the carbamoyl phosphate synthetic component allows the *B* domain to pull away from the main body of the molecule. This type of conformational flexibility displayed by the *B* domains in CPS has previously been observed in biotin carboxylase (Waldrop *et al.*, 1994).

The *A* domain of the carbamoyl phosphate synthetic component, delineated by Asn554–Lys686, is topologically similar to that of the carboxyphosphate synthetic component. While both the *B* and *C* domains in the carbamoyl phosphate synthetic component are topologically equivalent to those observed in the carboxyphosphate synthetic unit with the same number of β-strands, the reverse turns connecting these secondary structural elements differ. Also, the extended loop in the carboxyphosphate synthetic component, formed by residues Glu334 to Pro345, is missing in the carbamoyl phosphate synthetic unit.

A cartoon of potential electrostatic interactions between the nucleotide and the protein is displayed in Fig. 10. Two of the nitrogen atoms in the adenine ring of the nucleotide lie within hydrogen-bonding distance of O of His754 and N of Leu756. Both Gly786 and Val787, through backbone atoms, further serve to position the nucleotide in the active site. Like Glu215 in the carboxyphosphate synthetic component, Glu761 serves the same function by hydrogen bonding to both the 2′- and 3′-hydroxyl groups of the ribose. Both Arg715 and His 788, through their side-chain functional groups, interact with the α- and β-phosphate moieties of the ADP, respectively. The sole manganese ion in the active site is surrounded in an octahedral coordination sphere by two of the phosphoryl O atoms, two water molecules, $O^{\varepsilon1}$ of Gln829 and $O^{\varepsilon2}$ of Glu841. The metal–ligand bond lengths range from 1.9 to 2.4 Å with the longer lengths occurring between the manganese ion and the ordered water molecules. In addition to the manganese, there is a potassium ion that is coordinated by $O^{\varepsilon1}$ of Glu841, two β-phosphoryl O atoms from the ADP and a water molecule. These bond lengths range from 2.6 to 2.9 Å. This potassium ion is located in nearly the same position as the second manganese in the carboxyphosphate synthetic component and was originally modeled as such in the early stages of the least-squares refinement. The bond lengths between the metal and the ligands, however, refined to values longer than would be expected for a manganese ion and as such this metal was subsequently included in the model as a potassium ion.

Again, as observed in the carboxyphosphate synthetic component, there is a 'K-loop' nearby the active site, as shown in Fig. 8(*b*). In this case, the potassium ion is octahedrally coordinated by five oxygen ligands (the carbonyl O atoms of Glu783, Gln784 and Val787, O^{γ} of Ser792 and $O^{\varepsilon2}$ of Glu761) and one nitrogen ligand ($N^{\delta1}$ of His781). In the carboxyphosphate synthetic unit, Asn236 serves the same function as His781. Through the interactions with Glu761,

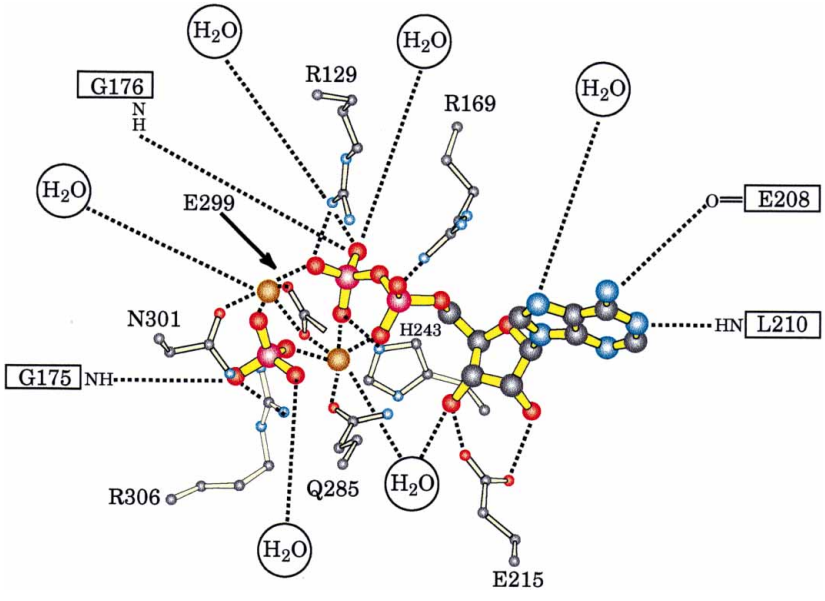


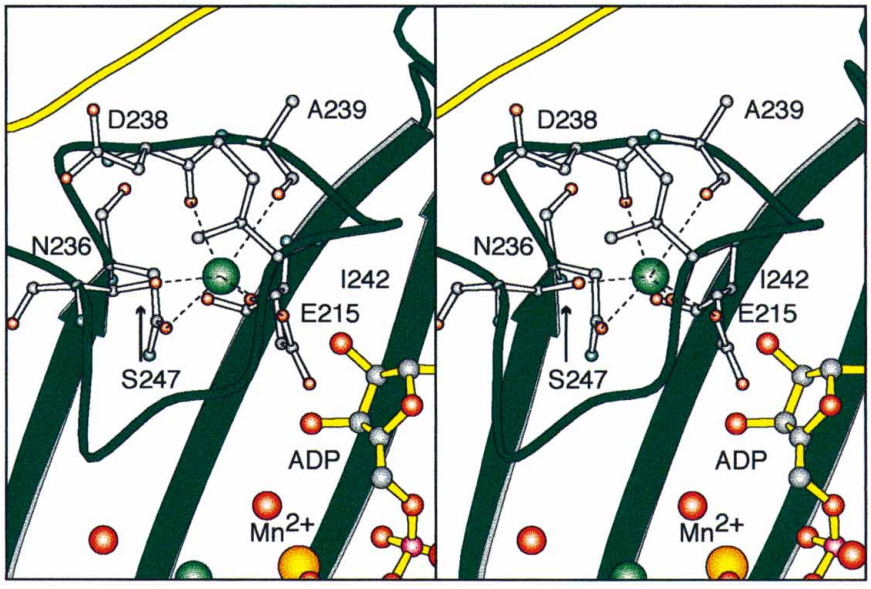
Figure 7
The active site for the carboxyphosphate synthetic component. A cartoon of potential hydrogen-bonding interactions between the ADP/P_i moiety and the protein is displayed. The manganese ions are indicated as brown spheres.

the potassium ion is linked to the carbamoyl phosphate synthetic unit active site and lies at 9.5 Å from the manganese ion.

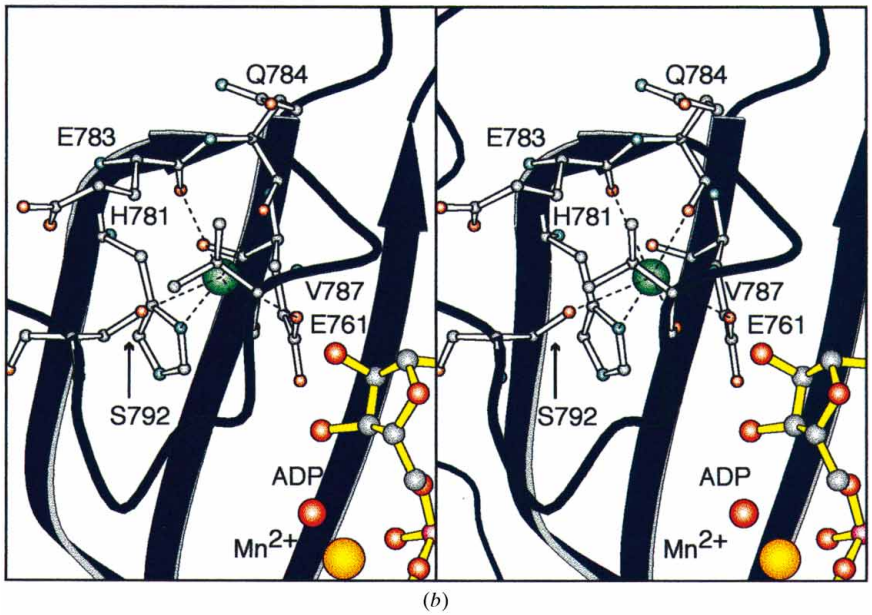
3.3.4. The allosteric domain (Ser937–Lys1073). An α-carbon trace for the allosteric domain is given in Fig. 11(*a*). This domain plays two distinct biochemical roles by providing both a molecular surface for interacting with a second α,β-heterodimer in the CPS $(\alpha,\beta)_4$-tetramer and binding pockets for effector molecules such as UMP and ornithine. The main core of the allosteric domain, a modified 'Rossmann' fold, is characterized by a five-stranded parallel β-sheet flanked on either side by two and three α-helices, respectively. Each of these α-helices ends in a type-I turn. There are two additional type-I turns, delineated by Lys993–Glu996 and Asp1057–Glu1060, and a final α-helix, formed by Val1065–Ala1070.

While previous biochemical studies have suggested that inosine 5′-monophosphate (IMP) functions as an activator of CPS activity (Boettcher & Meister, 1981), recent kinetic data suggest that under some conditions IMP exerts an inhibitory effect (Braxton *et al.*, 1992). Regardless of the actual biochemical role of IMP in modulating CPS activity, it is known, from a low-resolution structural analysis of CPS crystals soaked in IMP, that the binding site for this nucleotide is at the C-terminal end of the five-stranded parallel β-sheet (Thoden *et al.*, unpublished data). In the crystals employed in this structural investigation, glutamine and inorganic phosphate are positioned in the same binding cleft as shown in Fig. 11(*b*). There are three specific interactions between glutamine and the protein such that $N^{\varepsilon 2}$ of the side-chain carboxamide group lies at 3.1 Å from the carbonyl O atoms of Thr1061 and $O^{\varepsilon 1}$ of the carboxamide group is positioned at 3.1 Å from both $N^{\delta 2}$ of Asn1015 and O^{γ} of Thr1017. The inorganic phosphate is tightly bound to the protein through interactions of its phosphoryl O atoms with backbone amide H atoms (Gly976 and Thr977) and side-chain functional groups (O^{γ} of Thr974, O^{γ} of Thr977, N^{ζ} of Lys954 and N^{ζ} of Lys993).

The subunit:subunit interface between allosteric domains in the CPS $(\alpha,\beta)_4$-tetramer is formed primarily by the second α-helix (His975–Leu981) and the third β-strand (Arg989–Val991) of the Rossmann fold. A close-up view of the subunit:subunit interactions between neighboring allosteric domains is depicted in Fig. 12. The molecular interface is closed off on one side by a hydrogen-bonding network formed between His95 and its symmetry-related partner, His 975*. Specifically $N^{\delta 1}$ of His975 is 2.6 Å from $N^{\delta 1}$ of His975*. Additional hydrogen bonds are formed between $N^{\varepsilon 2}$ of His975 and $O^{\delta 1}$ of Asn992* (2.8 Å), and $O^{\delta 1}$ of Asn992 and $N^{\varepsilon 2}$ of His975* (2.7 Å). As in the oligomerization domain interface, there are a number of hydrophobic residues pointing towards the middle of the channel between the two subunits and these include Ile979, Ile979*, Leu990 and Leu990*. The other side of the channel is closed off by a hydrogen-bonding network formed between $N^{\delta 2}$ of Asn987 and O of Asn987* (3.1 Å), and $N^{\delta 2}$ of Asn987* and O of



(*a*)



(*b*)

Figure 8
The 'K-loops' associated with the two synthetase units of the CPS large subunit. The potassium ion located in the carboxyphosphate synthetic component is shown in (*a*). As indicated by the large green sphere, this cation is surrounded by oxygen ligands in an octahedral coordination sphere. Note the close proximity of the 2′- and 3′-hydroxyl groups of the ADP ribose to the potassium ion. The potassium ion located in the carbamoyl phosphate synthetic component is shown in (*b*). Ordered water molecules are depicted as red spheres.

Asn987 (2.9 Å). This region is not shown in Fig. 12, simply for the sake of clarity. Like the interface formed by the oligomerization domains, the subunit:subunit contact region between allosteric domains in the CPS $(\alpha,\beta)_4$-tetramer is hydrophobic with only five ordered water molecules located nearby.

3.3.5. The ornithine binding pocket. The ornithine binding pocket is formed by the fifth β-strand and the fifth α-helix of the allosteric domain and three β-strands from the carbamoyl phosphate synthetic component. A close-up view of this region is displayed in Fig. 13. One of the carboxylate O atoms of ornithine lies at 2.8 Å from the backbone amide N atoms of Thr1042 and a water molecule, and the second is positioned at 2.8 Å from O^{γ} of Thr1042 and another solvent. The α-amino group of ornithine forms a hydrogen bond with O of Tyr1040 (2.5 Å), while the δ-amino group of the effector molecule interacts with $O^{\varepsilon 1}$ of Glu783 (3.0 Å), $O^{\varepsilon 1}$ of Glu892 (2.5 Å) and O of Asp791 (2.9 Å). Ornithine acts as a bridge between the allosteric domain and the active site of the carbamoyl phosphate synthetic component such that its δ-amino group is 8.7 Å from the potassium ion residing in the 'K-loop'.

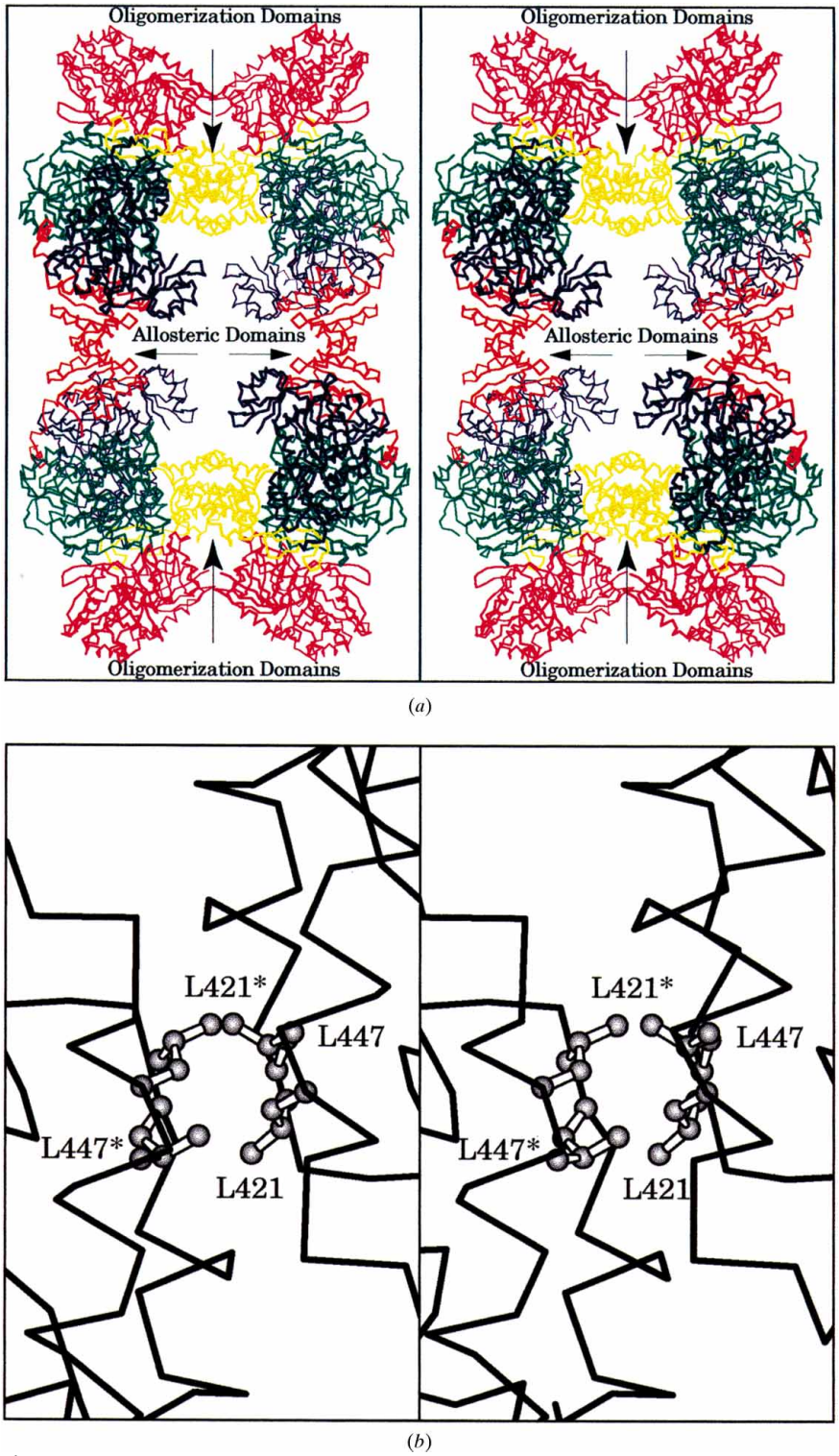


Figure 9
α-Carbon trace of the entire CPS $(\alpha,\beta)_4$-tetramer. (*a*) The oligomerization domains are as indicated. A close-up view of the subunit:subunit interface formed by these domains is given in (*b*). As indicated, the primary hydrophobic interactions occur between Leu421 and Leu447 and their symmetry-related partners, Leu421* and Leu447*.

4. Discussion

There are two undeniable facts with regard to the model of CPS presented here: (i) the carboxyphosphate and carbamoyl phosphate synthetic components, though topologically similar, are not structurally equivalent, and (ii) the three active sites of the enzyme are separated by a linear distance of nearly 100 Å. Indeed, the α,β-heterodimeric enzyme contains three distinct active sites connected to one another by two putative molecular tunnels. In addition, there are two allosteric binding sites (one for ornithine and the other for UMP or IMP) that are utilized separately for the activation or inhibition of catalytic activity through modulation of the Michaelis constant for Mg^{2+}ATP. All of these binding sites must obviously work in concert to coordinate the synthesis of carbamoyl phosphate *via* a chemical mechanism that is known to require five substrates, four distinct chemical steps and at least three reaction intermediates.

Perhaps one of the most intriguing aspects of the molecular architecture of CPS is the pair of topologically similar synthetase components within the large subunit that contain the active sites for the two phosphorylation events. The nearly exact twofold rotational axis that relates these two synthetase components to each

other strongly suggests that CPS evolved from a much simpler α_2-homodimer. Duplication and fusion of an ancestral gene that coded for an enzyme capable of catalyzing an N-ligase type of reaction may have initiated the molecular events required for the evolution of this catalytic activity. This initial genetic event was eventually followed by the mutation and specialization of the two phosphorylation components and the recruitment of both an amidotransferase subunit and an additional domain for allosteric control.

Clearly the two homologous carboxyphosphate and carbamoyl phosphate synthetic components evolved separately for their more specialized functions. The uniqueness of these two phosphorylation components is reflected in the large number of differentially conserved amino-acid residues found within each of these units. The carboxyphosphate domain must be able both to interact with the amidotransferase subunit and also to provide a path for the migration of ammonia, whereas the carbamoyl phosphate domain must act as the recipient of the signals required for the allosteric control of enzymatic activity. Moreover, each of these synthetic components interacts with the oligomerization domain in specific and quite different manners.

Can anything be learned about these two phosphorylation components by examination of the protein structure in the vicinity of these more specialized functions? For example, how similar or distinct is the protein surface of the carboxyphosphate synthetic component, as compared with the carbamoyl phosphate synthetic component, that interacts with the amidotransferase subunit? There are 19 amino acids within the carboxyphosphate synthetic unit that interact specifically with the small subunit, either through direct hydrogen bonds or interactions mediated *via* bound water molecules, as listed in Table 3. These residues of the carboxyphosphate synthetic component, namely Asn227, Cys228, Ala253, Gln254, Thr257, Asp258, Lys259, Gln262, Arg265, Asn266, Pro290, Asn292, Asp333, Glu349, Ser351, Ile352, Arg389, Gln391 and Gln392, correspond topologically to Met772, Val773, Ala798, Tyr799, Ser802, Gln803, Glu804, Asp807, Arg810, Gln811, Asn834, Glu836, Ala875, Glu882, Ile884, Pro885, Arg922, Phe924 and Ala925 of the carbamoyl phosphate synthetic component. As can be seen, only four of these amino acids are identical between the two synthetase components. Quite strikingly, Lys259 in the carboxyphosphate synthetic component corresponds to Glu804 in the carbamoyl phosphate synthetic component. These observations make it highly unlikely that the small subunit ever binds productively to the carbamoyl phosphate synthetic component. Moreover, the small subunit forms additional hydrogen bonds with 16 other residues contributed by the oligomerization domain.

Another unique structural feature of the carboxyphosphate synthetic component is a putative molecular tunnel for ammonia that leads directly from the site of the interaction with the small subunit to the binding pocket for Mg^{2+}ATP. Examination of the carbamoyl phosphate synthetic component reveals that an analogous molecular tunnel is not present. Furthermore, the amino-acid residues that form the tunnel in the carboxyphosphate synthetic component are not well conserved at the homologous sites within the carbamoyl phosphate synthetic unit. Of the 23 residues that have been identified to be within 3.2 Å of the ammonia channel, only nine of these amino acids, Ile20, Gly21, Gln22, His243, Asn283, Asn301, Arg306, Ala314 and Met378, are conserved as Ile572, Gly573, Gln574, His788, Asn827, Asn843, Arg848, Ala856 and Met911 in the carbamoyl phosphate synthetic component. It is thus highly unlikely that the carbamoyl phosphate domain can perform the functions required of the carboxyphosphate domain and *vice versa*.

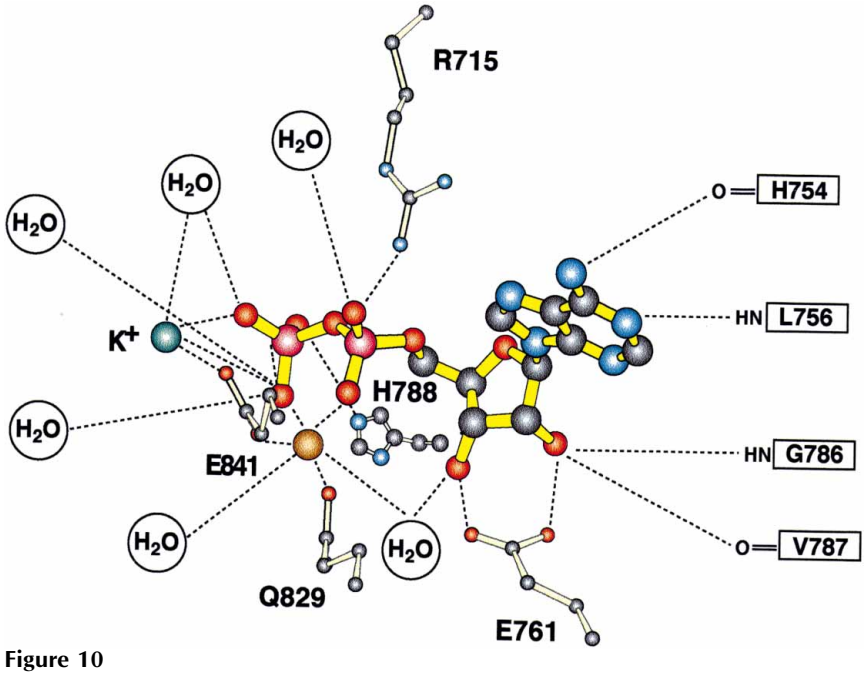


Figure 10
The active site for the carbamoyl phosphate synthetic component. A cartoon of potential hydrogen-bonding interactions between the ADP moiety and the protein is shown. The manganese ion is indicated as a brown sphere.

Recent research efforts by a number of laboratories have been made to ascertain the inherent properties of the individual domains of CPS through truncation of the large subunit. Experiments in our own laboratories have failed to demonstrate any catalytic activies for either the N-terminal or C-terminal halves of the large subunit (Raushel *et al.*, 1992). Indeed, we have shown that when as few as 14 amino acids are deleted from the C-terminal end of CPS, the rate of synthesis of carbamoyl phosphate is reduced by greater than 95% (Czerwinski *et al.*, 1995). Furthermore, there are a number of reports on the constructions and partial characterizations of fused and truncated variants of CPS that are difficult to rationalize in light of the current X-ray crystal structure of the enzyme. For example, the small and large subunits of CPS from *E. coli* have been reported to be

fused into a single polypeptide chain by joining the C-terminal end of the small subunit with Leu13 of the large subunit (Guy, Bouvier & Evans, 1997; Guy, Rotgeri & Evans, 1997). In the high-resolution structure reported here, however, the C-terminal carboxylate group of the small subunit is located more than 60 Å from the amino group of Leu13 in the large subunit. Also, Leu13 resides in a β-strand with its side chain buried in a hydrophobic pocket and its α-carbon positioned at $\sim$24 Å from the α-carbon of Met1. As a consequence, a large conformational change must occur in the fused protein and it is difficult to rationalize how such a change is possible without leading to a serious disruption of the molecular interactions between the large and small subunits of CPS.

Similar issues can be raised with regard to some of the truncated versions of the large subunit that have recently been reported as catalytically active (Guy, Rotgeri & Evans, 1997; Guy, Bouvier & Evans, 1997). Two different truncated proteins composed of either residues Leu96–Asn363 (referred to as A2) or Gly658–Lys899 (referred to as B2) from the large subunit have been reported to catalyze separately the synthesis of carbamoyl phosphate at rates that are approximately 16 times faster than that of the native enzyme. These 'mini-proteins' are assumed to form homodimers such that each of the monomeric units is able to catalyze one of the two phosphorylation events with only ammonia as the nitrogen source. From the structure presented here, however, it is difficult to rationalize the manner in which these monomeric units might form homodimers since those residues known to form the intermolecular contacts between the two synthetase units in the native intact enzyme have been deleted in these mini-proteins. If, indeed, homodimeric units are formed, then most likely the intermolecular contacts are non-native.

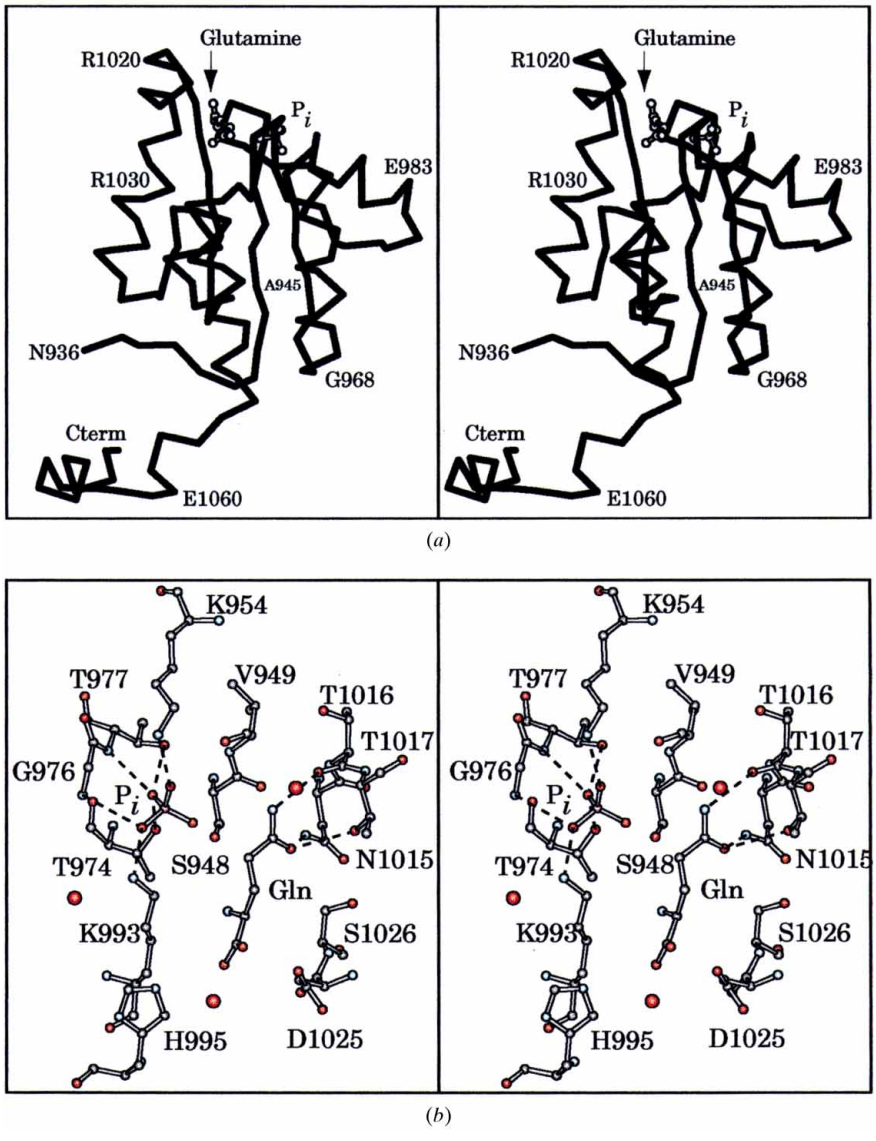


Figure 11
Shown in (*a*) is an α-carbon trace of the allosteric domain. The positions of the glutamine and phosphate molecules observed binding in this domain are indicated by ball-and-stick representations. Various amino-acid residues are labeled to aid the viewer in following the course of the polypeptide chain. A close-up view of the region near the bound glutamine and inorganic phosphate is given in (*b*). Potential hydrogen bonds are indicated by the dashed lines.

Additional problems can also be cited with the slightly larger constructions from the N-terminal and C-terminal halves of the large subunit of CPS (Guy, Bouvier & Evans, 1997; Guy, Rotgeri & Evans, 1997). It has been reported that truncated proteins containing only amino-acid residues Met1–Asn363 (referred to as A1–A2) or Gly507–Lys899 (referred to as B1–B2) are each able to form homodimers and associate with the small subunit. These proteins apparently catalyze the formation of carbamoyl phosphate with either glutamine or ammonia as the nitrogen source. Since these longer constructions have retained many of the amino acids required for the association of the N-terminal and C-terminal domains of the large subunit, the formation of homodimers is certainly feasible. However, it is not clear how the association with the small subunit is maintained. In the A1–A2 construct, 19 of the 35 amino-acid residues now known to make direct or indirect hydrogen-bond interactions with the small subunit have been deleted. Moreover, it is difficult to rationalize the manner in which the A1–A2 truncated protein can associate with the small subunit whereas the A2 construct cannot, since there is not a single known molecular contact between the small subunit and those residues of A1–A2 that have been eliminated to form A2 (residues Met1–Ala95). Similar problems occur in attempting to understand, based on the current X-ray model of CPS, the manner in which the B1–B2 construction is able to

catalyze the formation of carbamoyl phosphate with glutamine as a nitrogen source (Guy, Rotgeri & Evans, 1997; Guy, Bouvier & Evans, 1997). From the three-dimensional model of CPS, it is known that the sites of interaction between the small subunit and the N-terminal half of the large subunit are not well conserved in the C-terminal half of CPS. Even if a B1–B2 homodimer were to form and associate with the small subunit in a manner similar to that observed for the wild-type enzyme, it is still not obvious how ammonia would be directed to either of the two phosphorylation domains since there is not an analogous molecular tunnel located in the C-terminal half of the large subunit of CPS. In order for these many uncertainties to be clarified, the truncated and fused proteins must be purified to homogeneity and fully characterized for their ability to catalyze all of the partial and full biosynthetic reactions.

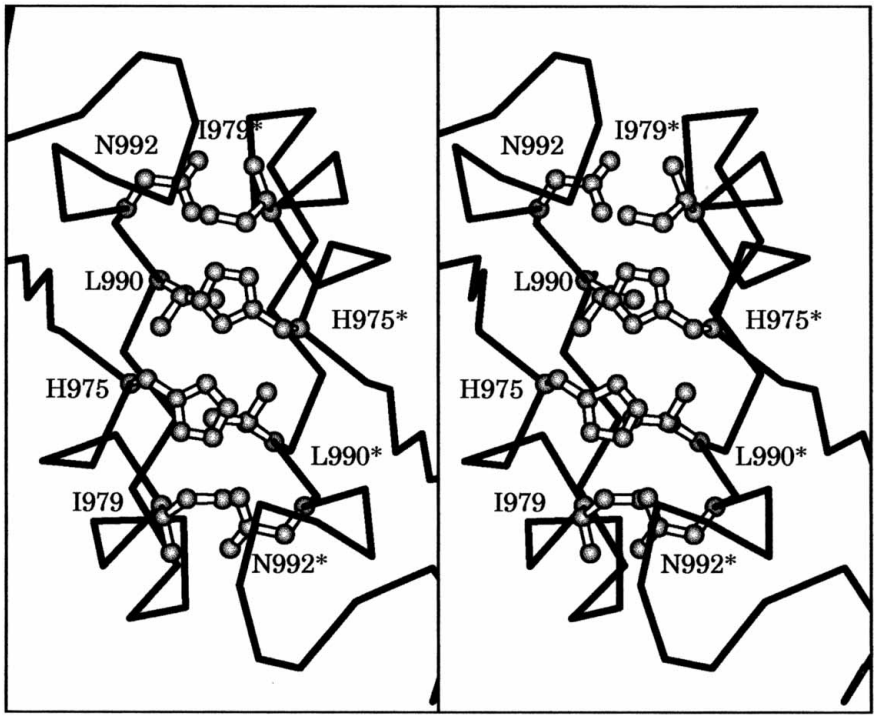


Figure 12
The subunit:subunit interface formed between two allosteric domains. Primary hydrophobic interactions occur between Ile979 and Leu990 and their symmetry-related partners. There is a hydrogen-bonding network, formed by His975 and Asn992 and their symmetry-related partners, His975* and Asn992*, that serves to shield Ile979 and Leu990 from the solvent. An additional hydrogen-bonding network, formed by Asn987 and Asn987* but not shown here for clarity, closes off the other side of the subunit:subunit interface.

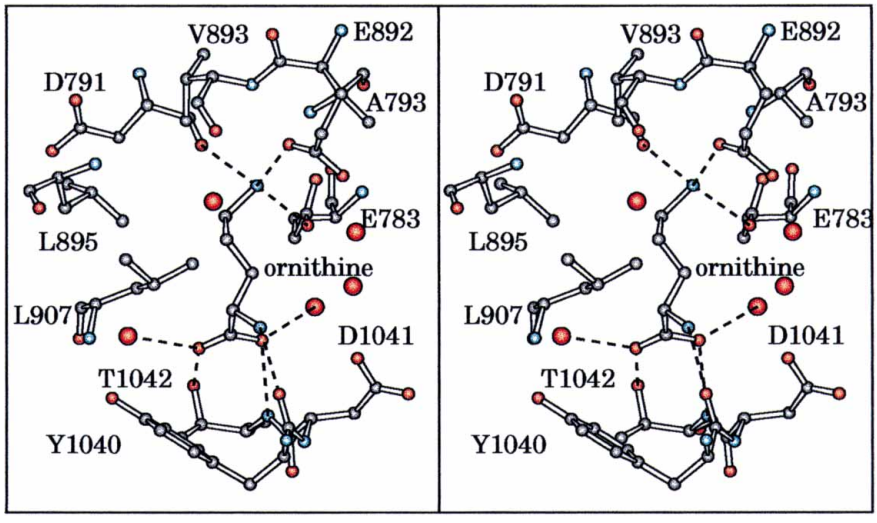


Figure 13
A close-up view of the ornithine binding pocket. Ordered water molecules are indicated by the red spheres. The carboxylate group of ornithine interacts with the allosteric domain while the δ-amino group of the side chain forms hydrogen bonds with amino-acid residues from the carbamoyl phosphate synthetic component.

Prior to the structural investigation of CPS, it was unknown as to whether the enzyme contained three specific and spatially distinct active sites or rather a single active site whereby the bicarbonate was positioned in the center and then sequentially phosphorylated, amidated and finally rephosphorylated without moving. While there are still discrepancies between some of the structural and biochemical data as discussed above, one fact is that the active sites are spatially distinct and that the ammonia generated by the hydrolysis of glutamine in the small subunit must, in some manner, be transported to the ATP-binding site formed by the carboxyphosphate synthetic component of the large subunit. The possibility of a molecular tunnel allowing for such transport between the small and large subunits is especially intriguing. Both visual inspection of the CPS model and a computational search with the software package *GRASP* (Nicholls *et al.*, 1991) have indicated the presence of such a tunnel, as depicted in Fig. 14. This molecular pathway was approximated visually as a series of base points, separated by several ångströms, along its path and refined by interpolating between these base points in steps of approximately 0.25 Å while also searching for the location that was farthest from any neighbouring protein atom, as described previously (Thoden *et al.*, 1997).

The molecular tunnel begins at the base of the small subunit active site where a crown of charged amino-acid side chains (Asp45, Lys202 and His353) face, towards its interior. Note that His353 constitutes part of the 'catalytic triad' of the small subunit. Other than these residues, however, the rest of the tunnel from the small subunit to the interface of the large

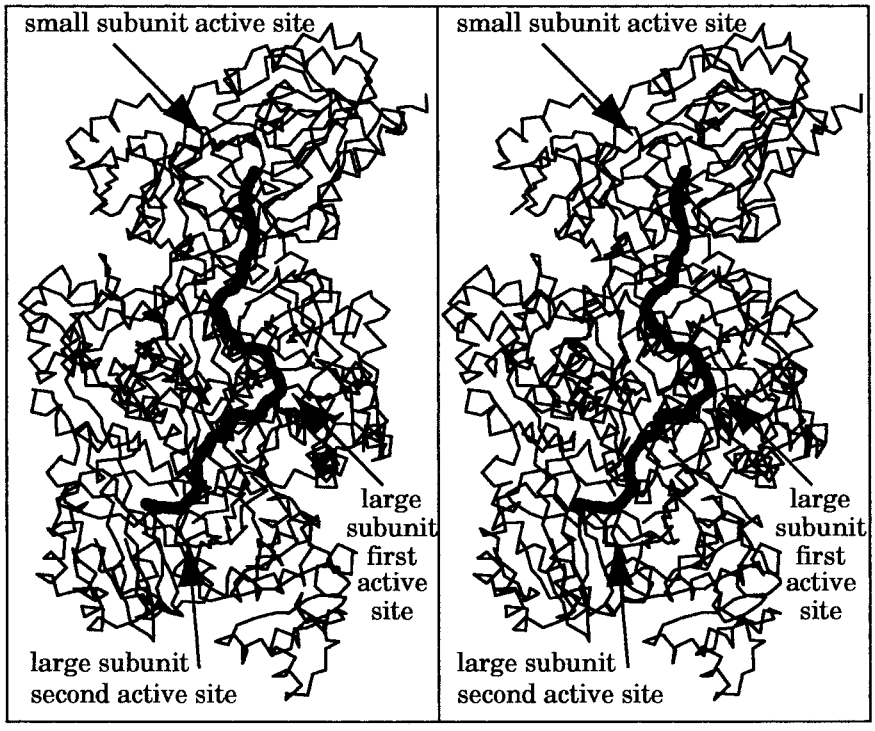


Figure 14
A putative molecular tunnel characterizing the CPS α,β-heterodimer. A possible pathway for the transport of the two reactive intermediates, ammonia and carbamate, is indicated by the black spheres. This tunnel, which first transports ammonia, leads from the active site of the small subunit to the active site of the carboxyphosphate synthetic component. The carbamate formed by the reaction of ammonia with the highly reactive carboxyphosphate intermediate is subsequently transported through the tunnel leading from the first active site of the large subunit to the second as indicated in the figure.

subunit is lined with non-reactive side chains and backbone atoms. Those residues within 3.5 Å of the center of the putative pathway include Ser35, Met36, Gly293, Ala309, Asn311, *cis*-Pro358 and Gly359. The lack of reactive amino-acid residues lining the tunnel is in keeping with the biochemical need to transport ammonia in the unprotonated state for its subsequent reaction with the carboxyphosphate generated at the first active site of the large subunit. As the tunnel extends from the interface between the small and large subunits to the active site of the carboxyphosphate synthetic component, the interior is once again lined, for the most part, with non-reactive residues, including Ser233, Ile234, Ala251, Tyr261, Asn283, Asn301, Ser307, Leu310, Ala311, Ala314, Thr315 and Ile352. The only reactive side chains within this portion of the tunnel are contributed by Glu217 and Cys232. This tentative molecular pathway from the small subunit to the first active site of the large subunit leads directly to the inorganic phosphate observed binding in the active site of the carboxyphosphate synthetic component. Indeed, as suggested above, this region of CPS may be responsible for stabilizing the carboxyphosphate intermediate and perhaps providing the molecular surface for its subsequent reaction with ammonia. The carbamate thus formed must be effectively transported to the active site of the carbamoyl phosphate synthetic component. The portion of the tunnel lying between the two active sites of the large subunit is somewhat less hydrophobic, with the side-chain carboxylate group of Glu604 pointing towards the pathway and a cluster of charges, Glu577, Arg848, Lys891 and Glu916, located near the opening to the second active site of the large subunit. Those non-reactive residues lining the molecular tunnel extending between the two active sites of the large subunit include Ile18, Val19, Ile20, Gly21, Gln22, Ala23, Met174, Gly175, Met378, Val381, Gly575, Gln829, Asn843, Thr849, Met911, Ser913 and Thr914. Approximately 25 water molecules lying within 2 Å of the pathway have been located. Whether these water molecules remain in the tunnel during the reaction is unknown. It is informative to note that many of the residues lining the putative tunnel are absolutely conserved among 22 of 24 primary structural alignments of CPS molecules including Ile20, Gln22, Ala23, Gly175, Glu217, Ala251, Asn283, Asn301, Leu310, Ala311, Thr315, Met378, Glu577, Glu604, Gln829, Asn843, Arg848, Met911, Ser913, Thr914 and Glu916 of the large subunit and Asp45, Lys202, Asn311, His353 and Gly359 of the small subunit. Furthermore, many of the residues that are not strictly conserved are replaced with amino-acid residues of comparable chemical reactivities.

While both visual inspection of the current three-dimensional model of CPS and computational analyses have indicated the presence of molecular cavities, the question still remains as to whether the molecular pathway, as outlined above, is correct. Clearly many experiments, including structural investigations of various site-directed mutant proteins, need to be conducted in order to confirm or refute the existence of such a molecular pathway. What is obvious at this time, however, is that there must be some type of structural pathway connecting the three active sites of the CPS α,β-heterodimer.

We thank Dr W. W. Cleland for critically reading this manuscript and Jessica E. Raushel for amino-acid sequence alignments. The computational expertise of Dr Gary Wesenberg is gratefully acknowledged. We also thank the three referees who took the time to review this manuscript and who provided insightful comments and suggestions. This research was supported in part by grants from the NIH (GM55513 to HMH and DK30343 to FMR) and the NSF (BIR-9317398 shared instrumentation grant to IR).

References

Anderson, P. M. & Meister, A. (1966). *Biochemistry*, **5**, 3164–3169.
Boettcher, B. & Meister, A. (1981). *J. Biol. Chem.* **256**, 5977–5980.

Braxton, B. L., Mullins, L. S., Raushel, F. M. & Reinhart, G. D. (1992). *Biochemistry*, **31**, 2309–2316.
Braxton, B. L., Mullins, L. S., Raushel, F. M. & Reinhart. G. D. (1996). *Biochemistry*, **35**, 11918–11924.
Bricogne, G. (1976). *Acta Cryst.* A**32**, 832–837.
Cervera, J., Bendala, E., Britton, H. G., Bueso, J., Nassif, Z., Lusty, C. J. & Rubio, V. (1996). *Biochemistry*, **35**, 7247–7255.
Czerwinski, R. M., Mareya, S. M. & Raushel, F. M. (1995). *Biochemistry*, **34**, 13920–13927.
Fan, C., Moews, P. C., Walsh, C. T. & Knox, J. R. (1994). *Science*, **266**, 439–443.
Guy, H. I., Bouvier, A. & Evans, D. R. (1997). *J. Biol. Chem.* **272**, 29255–29262.
Guy, H. I., Rotgeri, A. & Evans, D. R. (1997). *J. Biol. Chem.* **272**, 19913–19918.
Kaplan, J. B., Merkel, W. K. & Nichols, B. P. (1985). *J. Mol. Biol.* **183**, 327–340.
Kaplan J. B. & Nichols, B. P. (1983). *J. Mol. Biol.* **168**, 451–468.
Kothe, M., Eroglu, B., Mazza, H., Samudera, H. & Powers-Lee, S. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 12348–12353.
Lusty, C. J. (1992). *FEBS Lett.* **314**, 135–138.
Mareya, S. M. & Raushel, F. M. (1994). *Biochemistry*, **33**, 2945–2950.
Matthews, S. L. & Anderson, P. M. (1972). *Biochemistry*, **11**, 1176–1183.
Miran, S. G., Chang, S. H. & Raushel, F. M. (1991). *Biochemistry*, **30**, 7901–7907.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Nicholls, A., Sharp, K. A. & Honig, B. (1991). *Proteins Struct. Funct. Genet.* **11**, 281–296.
Nyunoya, H. & Lusty, C. J. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4629–4633.
Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G. & Goldman, A. (1992). *Protein Eng.* **5**, 197–211.
Otwinowski, Z. (1986). PhD thesis, Yale University, New Haven, CT, USA.
Piette, J., Nyunoya, H., Lusty, C. J., Cunin, R., Weyens, G., Crabeel, M., Charlier, D., Glansdorff, N. & Pierard, A. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 4134–4138.
Post, L. E., Post, D. J. & Raushel, F. M. (1990). *J. Biol. Chem.* **265**, 7742–7747.
Raushel, F. M., Anderson, P. M. & Villafranca, J. J. (1978). *Biochemistry*, **17**, 5587–5591.
Raushel, F. M., Miles, B. W., Post, L. W. & Post, D. J. (1992). *Bioorg. Med. Chem. Lett.* **2**, 319–322.
Raushel, F. M. & Villafranca, J. J. (1979). *Biochemistry*, **18**, 3424–3429.
Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon and Breach.
Teeter, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.
Teeter, M. M. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20**, 577–600.
Tesmer, J. J. G., Klem, T. J., Deras, M. L., Davisson, V. J. & Smith, J. L. (1996). *Nature Struct. Biol.* **3**, 74–86.
Thoden, J. B., Holden, H. M., Wesenberg, G., Raushel, F. M. & Rayment, I. (1997). *Biochemistry*, **36**, 6305–6316.
Thoden, J. B., Raushel, F. M., Mareya, S., Tomchick, D. & Rayment, I. (1995). *Acta Cryst.* D**51**, 827–829.
Tronrud, D. E., Ten Eyck, L. F. & Mathews, B. W. (1987). *Acta Cryst.* A**43**, 489–501.
Trotta, P. P., Burt, M. E., Haschemeyer, R. H. & Meister, A. (1971). *Proc. Natl Acad. Sci. USA*, **68**, 2599–2603.
Waldrop, G. L., Rayment, I. & Holden, H. M. (1994). *Biochemistry*, **33**, 10249–10256.
Weng, M., Makaroff, C. A. & Zalkin, H. (1986). *J. Biol. Chem.* **261**, 5568–5574.
Werner, M., Feller, A. & Pierard, A. (1985). *Eur. J. Biochem.* **146**, 371–381.
Yamaguchi, H., Kato, H., Hata, Y., Nishioka, T., Kimura, A., Oda, J. & Katsube, Y. (1993). *J. Mol. Biol.* **229**, 1083–1100.
Zalkin, H., Argos, P., Narayama, S. V. L., Tiedeman, A. A. & Smith, J. M. (1985). *J. Biol. Chem.* **260**, 3350–3354.