Carbamoyl phosphate synthetase: a tunnel runs through it
Hazel M Holden*, James B Thoden† and Frank M Raushel‡

The direct transfer of metabolites from one protein to another in a biochemical pathway or between one active site and another within a single enzyme has been described as substrate channeling. The first structural visualization of such a phenomenon was provided by the X-ray crystallographic analysis of tryptophan synthase, in which a tunnel of approximately 25 Å in length was observed. The recently determined three-dimensional structure of carbamoyl phosphate synthetase sets a new long distance record in that the three active sites are separated by nearly 100 Å.

Addresses
*Department of Biochemistry, University of Wisconsin, Madison, WI 53706, USA
†e-mail: holden@enzyme.wisc.edu
‡e-mail: jim@enzyme.wisc.edu
$Department of Chemistry, and Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, USA; e-mail: raushel@tamu.edu

Current Opinion in Structural Biology 1998, 8:679–685
http://biomednet.com/elecrefl0959-440X00800679
© Current Biology Ltd ISSN 0959-440X

Abbreviations
CPS carbamoyl phosphate synthetase
GPATase glutamine phosphoribosylpyrophosphate amidotransferase

Introduction
The concept of substrate channeling, as described in [1], was originally put forth to explain the manner by which reactive intermediates are transferred from one protein to another in a metabolic pathway or shuttled from one active site to another within a single enzyme. Although the biochemical evidence for substrate channeling was substantial, the first direct structural observation of such a phenomenon was derived from the elegant X-ray crystallographic analysis of tryptophan synthase isolated from Salmonella typhimurium [2]. This investigation demonstrated that the two active sites, located on the α and β subunits of the enzyme, are separated by a distance of approximately 25 Å and are connected by a tunnel of the appropriate diameter to facilitate the diffusion of indole.

As more complicated protein structures are solved to increasingly higher resolution, it is becoming apparent that substrate channels may, indeed, be quite common. So far, the long distance record for a channel between active sites has been set by carbamoyl phosphate synthetase from Escherichia coli, the focus of this review. In this review, we present the structure of the enzyme and describe the channels that are essential for shuttling the reactive and unstable intermediates between three active site regions.

Biochemistry of carbamoyl phosphate synthetase
Carbamoyl phosphate synthetase, hereafter referred to as CPS, plays a critical role in both arginine and pyrimidine biosynthesis by providing an essential precursor, namely carbamoyl phosphate. This remarkable enzyme has been the focus of intense investigation for more than 30 years, due, in part, to both its important metabolic role and the large number of substrates, products and effector molecules that bind to it.

According to most biochemical data, CPS catalyzes the production of carbamoyl phosphate from one molecule of bicarbonate, two molecules of Mg²⁺ATP and one molecule of glutamine, as depicted in Scheme I below [3,4].

As can be seen, there are, at minimum, three reactive species—carboxyphosphate, with a half-life of approximately 70 ms [5], ammonia and carbamate, with an estimated half-life of 28 ms [6].

As isolated from E. coli, the enzyme is composed of two polypeptide chains, referred to as the large and small subunits. The small subunit catalyzes the hydrolysis of glutamine [7], while the large subunit is responsible for the two phosphorylation events [8]. In addition, the large subunit provides the regions of the polypeptide chain that are responsible for binding physiologically important monovalent cations and effector molecules, such as ornithine, an activator, and UMP, an inhibitor [9,10]. These inhibitors and activators effect the reaction primarily through the modulation of the Michaelis constant for Mg²⁺ATP [11,12].

Through the painstaking efforts of Lusty and co-workers [13,14], the genes encoding both the large and small subunits were sequenced in the early 1980s; two key features
were revealed from these studies. First, in the small subunit, the amino acid sequence of the C-terminal portion of the polypeptide chain was shown to be homologous to sequences corresponding to the N-terminal domains of trpG-type amidotransferases. Second, and more surprising, in the large subunit, a homologous repeat sequence was found, such that residues Met1 to Arg400 were 40% identical to residues Ala553 to Leu933.

Prior to the successful structural determination of the CPS (α,β)4 tetramer in 1997 [15**,16**], it was unclear how as to the enzyme was able to orchestrate the synthesis and stabilization of three separate reaction intermediates. It was tacitly assumed that the active sites were situated near one another. Quite strikingly, however, the active sites are separated by nearly 100 Å, thereby implying some type of substrate channeling, as discussed here.

The carbamoyl phosphate synthase (α,β)4 tetramer

The overall three-dimensional motif of the CPS (α,β)4 tetramer, displayed in a ribbon representation in Figure 1, exhibits nearly exact 222 symmetry, with the rotational relationships between one α,β species and the other three being 179.9°, 178.1° and 178.1° [15**]. Color coded in magenta in Figure 1, the four small subunits of the (α,β)4 tetramer are perched at the ends of the molecule. Each large subunit within the tetramer can be described in terms of four distinct components: two units referred to as the carboxyphosphate and carbamoyl phosphate synthetic components, displayed in green and blue, respectively; an oligomerization region, shown in yellow; and an allosteric motif, depicted in red. These four regions of the large subunit are delineated by Met1 to Glu403, Val404 to Ala553, Asn554 to Asn936 and Ser937 to Lys1073, respectively. The molecular interactions between the small and large subunits of the CPS α,β heterodimer are quite extensive, with 35 direct hydrogen bonds between the two polypeptide chains [16**]. Importantly, only residues in the carboxyphosphate synthetic component and oligomerization domain of the large subunit contribute to the formation of this dimeric interface. There are no direct interactions between the small subunit and the carbamoyl phosphate synthetic component of the large subunit. In contrast to the subunit–subunit interface of the α,β heterodimer, the number of interactions between one α,β heterodimer and another within the complete tetramer is minimal [16**].

In order to more fully appreciate the underlying architecture of the α,β heterodimer, the following discussion will focus on the individual parts. It should be kept in mind, however, that all of these components are intimately associated with each other and, ultimately, are dependent upon one another for the full enzymatic activity displayed by CPS.

The small subunit

As can be seen in Figure 1, the polypeptide chain of the small subunit folds into two distinct structural motifs. The N-terminal domain, formed by Leu1 to Leu153, contains two β-sheet layers oriented nearly perpendicular to one
Figure 2

Close-up stereo view of the small subunit active site, with the bound glutamyl thioester intermediate indicated by the filled black bonds.

Another. One of these sheets contains four parallel \( \beta \) strands, whereas the other consists of four antiparallel \( \beta \) strands. The C-terminal domain, delineated by Asn154 to Lys382, is dominated by a 10-stranded mixed \( \beta \) sheet and six \( \alpha \) helices. This mixed \( \beta \) sheet is the only example of this tertiary structural element in the entire CPS \( \alpha,\beta \) heterodimer, as all other \( \beta \) sheets in the enzyme run either purely parallel or antiparallel.

Although three-dimensional structural searches have thus far 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 other members of the \( \text{trpG} \)-type amidotransferase family [17]. A superposition of the C-terminal domain of the CPS small subunit onto the N-terminal domain of GMP synthetase results in a root mean square deviation between 129 structurally equivalent \( \alpha \) carbons of approximately 1.7 Å. In addition to the topological similarity of their three-dimensional structures, both of these enzymes contain a cysteine residue that is located in a ‘nucleophile elbow.’ This residue, Cys269 in CPS, serves as the active site nucleophile and, as previously observed in enzymes belonging to the \( \alpha,\beta \) hydrolase family [18], it adopts dihedral angles well outside of the allowed regions of the Ramachandran plot (\( \phi = 59.1^\circ, \psi = -96.0^\circ \)).

The reaction mechanism by which glutamine is hydrolyzed to glutamate and ammonia in the small subunit is thought to occur via nucleophilic attack on the carbonyl carbon of the sidechain carboxamide group by the thiolate anion of Cys269. Indeed, numerous biochemical studies have suggested that the reaction mechanism proceeds through the formation of a covalently bound glutamyl thioester intermediate [19]. Furthermore, the role of His353 in activating the cysteine for nucleophilic attack has been supported by site-directed mutagenesis experiments in which it was replaced with an asparagine [20]. Recent X-ray crystallographic analyses of the H353N mutant have indeed confirmed the existence of the glutamyl thioester intermediate, which was trapped in the active site [21**]. As can be seen in Figure 2, it is absolutely clear that O\( \gamma \) of Ser47 and the backbone amide hydrogen of Gly241 are in an ideal location both to position the carbonyl carbon of the substrate for nucleophilic attack by the thiolate of Cys269 and to stabilize the developing oxyanion. Besides Ser47, the only other specific sidechain group interacting with the glutamyl thioester intermediate is that of Gln273.

The investigation described in [21**] represents the first structural evidence for a glutamyl thioester intermediate in the amidotransferase family of proteins.

The large subunit

For the sake of brevity, only the two synthetase units of the large subunit, namely the carboxyphosphate (Met1–Glu403) and the carbamoyl phosphate (Asn554–Asn936) synthetic components, will be described here. Details concerning the oligomerization region and allosteric motif can be found in [15**,16**]. As expected from the extensive primary structural homology, these two synthetic components are topologically, but not structurally, equivalent. Quite strikingly, these structural units are related by a nearly exact twofold rotation axis within the large subunit, thereby suggesting that the present CPS activity arose from a primordial \( \alpha_2 \) homodimeric enzyme.
Each of the synthetase components can be broken down into three smaller motifs, referred to as the A, B and C domains, which are similar in overall structure to such domains observed in biotin carboxylase [22], d-alanine:D-alanine ligase [23], glutathione synthetase [24], succinyl-CoA synthetase [25] and purK (JB Thoden, JB Kappock, J Stubbe, HM Holden, unpublished data), among others. In all of these enzymes, including the two synthetic components of CPS, the active sites are wedged between the B and C domains, with the B domains showing conformational flexibilities that are dependent upon the nature of the molecular species occupying the active sites.

With respect to tertiary structure, the A domains of the two synthetic CPS components contain five strands of parallel β sheet, whereas the B domains are composed of four strands of antiparallel β sheet, flanked on one side by two α helices. Clearly the most complicated of the three structural motifs, the C domains are dominated by an antiparallel seven-stranded β sheet. In the model described in [16**], the active sites of both the carboxyphosphate and carbamoyl phosphate synthetic components contained bound Mn²⁺-ADP. In addition, an inorganic phosphate was observed binding in only the active site of the carboxyphosphate synthetic component. As a result of this inorganic phosphate, the B domain of
the carboxyphosphate synthetic component is closed down, relative to that of the carbamoyl phosphate synthetic component. Quite intriguing is the fact that an inorganic phosphate is located in a similar position in the active site region of biotin carboxylase, another enzyme thought to proceed through a carboxyphosphate intermediate [22]. It can thus be speculated that this region of CPS is responsible for stabilizing the carboxyphosphate intermediate.

Cartoons illustrating the interactions between the nucleotides and the protein within the two active sites of the large subunit are shown in Figure 3. The manner in which the nucleotides are linked to the protein is strikingly similar for both synthetic components. For example, Arg169 and Arg715 play similar roles, interacting with the α phosphate moieties of the ADP molecules. Likewise, both Glu215 and Glu761 serve similar functions, anchoring the 2' and 3' hydroxyl groups of the nucleotide riboses and acting as bridges to potassium-ion-binding sites.

The two nucleotide-binding sites shown in Figure 3 differ primarily by the presence of an inorganic phosphate and a second manganese ion in the carboxyphosphate synthetic unit. In this particular synthetase unit, the manganese ions are octahedrally coordinated by oxygen-containing ligands and are bridged by the carboxylate sidechain of Glu299. Specifically, one ion is ligated by O^e1 of Gln285, O^e1 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 is coordinated by O^e1 of Asn301, O^e1 and O^e2 of Glu299, a water molecule and two phosphoryl oxygens, one contributed by the β phosphate group of the ADP and the other from the inorganic phosphate. Bond lengths between the metals and ligands range from 2.0 to 2.4 Å. In the carbamoyl phosphate synthetic component, the sole manganese ion is surrounded in an octahedral coordination sphere by two phosphoryl oxygens, two water molecules, O^e1 of Gln829 and O^e2 of Glu841. Note that if a second metal were to bind to the carbamoyl phosphate synthetic component, Glu841, located in a structurally identical position to Glu299 (in the carboxyphosphate synthetic unit), would most probably serve as the bridging ligand.

CPS is known to be activated by potassium [26] and, indeed, such ions have been located approximately 9 Å from both active sites of the large subunit, in essentially identical positions. Each potassium ion is ligated to the protein via an octahedral coordination sphere comprising three carbonyl oxygens and three sidechain ligands. The potassium ion to ligand bond distances range from 2.5 to 2.8 Å. In the carboxyphosphate synthetic component, these sidechain ligands are provided by Asn236, Ser147 and Glu215, whereas in the carbamoyl phosphate synthetic component, His781, Ser792 and Glu761 serve identical roles. The replacement of Asn236 in the carboxyphosphate synthetic component with His781 in the carbamoyl phosphate synthetic component is one of
many examples in the CPS large subunit where the nonconservation of an amino acid residue does not necessarily imply the nonconservation of function.

The tunnel

The one undeniable fact to emerge from the recent structural investigations of CPS is that the three active sites contained within the αβ heterodimer are separated by a linear distance of nearly 100 Å. The carboxyphosphate, ammonia and carbamate intermediates are highly reactive, such that the reaction mechanism must be exquisitely timed. Visual inspection of the CPS model, in addition to a computational search using the software package 'GRASP' [27], indicates a possible tunnel connecting the three active sites in the αβ heterodimer. This tunnel, as depicted in Figure 4, extends from the base of the small subunit active site, towards the surface of the carboxyphosphate synthetic component and is lined, for the most part, with nonreactive sidechains and backbone atoms. Amino acid residues lying within 3.5 Å of the center of the putative pathway in the small subunit include Ser35, Met36, Gly293, Ala309, Asn311, α-Pro358 and Gly359. As the tunnel extends from the top of the carboxyphosphate synthetic component to the first active site of the large subunit, it is once again lined with mostly nonreactive residues, except for Glu217 and Cys232. The portion of the tunnel lying between the two active sites of the large subunit is somewhat less hydrophobic, with the sidechain 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. In support of the molecular tunnel, it should be noted that many of the residues lining the pathway depicted in Figure 4 are absolutely conserved among 22 out of 24 primary structural alignments of CPS molecules and those residues that are not strictly conserved are typically replaced with amino acid residues of comparable chemical reactivities [16**]. Approximately 25 water molecules have been identified as lying within 2 Å of the pathway [16**], but clearly their locations will change as the CPS reaction proceeds from substrates to the final product.

Is the tunnel between the three active sites a true pathway? It is known that there is insignificant uncoupling of the various partial reactions catalyzed by CPS when all the substrates are at saturating concentrations. Indeed, the three intermediates, carboxyphosphate, ammonia and carbamate, must be protected from the bulk solvent, thereby arguing for such a molecular motif. In support of such a tunnel is the recent structural analysis of glutamine phosphoribosylpyrophosphate amidotransferase (GPATase), which also employs glutamine as a source of reduced nitrogen [28**]. In this enzyme, the tunnel is quite short in length (20 Å) and is described as being hydrophobic in nature [28**]. The ammonia channel in CPS, however, needs to cover a greater length (over 45 Å) and necessarily is more complicated in nature. Also, in CPS, the three active sites must be tightly coordinated during the reaction cycle, whereas the situation is somewhat simpler in GPATase as there are only two active site pockets. Using site-directed mutagenesis experiments, it will be possible to 'fine tune' the position of the CPS tunnel and, indeed, this work is in progress.

Conclusions

Substrate channeling has long been thought of as an ideal method for shuttling reactive intermediates, in a coordinated manner, from one protein to another in a metabolic pathway or from one active site to another within a single enzyme species. The first compelling biophysical evidence for the existence of such a tunnel was derived from the X-ray analysis of tryptophan synthetase [2]. Within the past year, the high-resolution structural analyses of CPS [15**,16**] and GPATase [28**] have added to our understanding of molecular tunnels. Indeed, as more complicated protein structures are solved to high resolution by X-ray crystallography, these channels or tunnels may become quite commonplace.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

• of special interest
•• of outstanding interest


This paper describes the formation of a 20 Å channel, through the binding of substrate analogs, that connects the active site for glutamine hydrolysis to the phosphoribosylpyrophosphate-binding site.