Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4864–4868, August 1984
Evolution

Protein differentiation: A comparison of aspartate transcarbamoylase and ornithine transcarbamoylase from *Escherichia coli* K-12

(evolution/functional specificity/secondary structure)

JOHN E. HOUGHTON, DAVID A. BENCINI*, GERARD A. O'DONOVAN, AND JAMES R. WILD†

Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX 77843

Communicated by William N. Lipscomb, April 20, 1984

ABSTRACT **The amino acid sequence of aspartate transcarbamoylase (carbamoylphosphate:L-aspartate carbamoyltransferase, EC 2.1.3.2) has been compared with that of ornithine transcarbamoylase (carbamoylphosphate:L-ornithine carbamoyltransferase, EC 2.1.3.3). The primary sequence homology is 25–40%, depending upon the alignment of homologous residues. The homologies are incorporated into discrete clusters and are interrupted by regions of length polymorphism. The most striking homologies correspond to regions putatively involved in the binding of the common substrate, carbamoyl phosphate. Chou–Fasman predictive analysis [Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* 13, 211–222; 222–245] indicates substantial conservation of secondary structural elements within the two enzymes, even in regions whose primary sequence is quite divergent. The results reported herein demonstrate that the two enzymes, aspartate transcarbamoylase and ornithine transcarbamoylase, share a common evolutionary origin and appear to have retained similar structural conformations throughout their evolutionary development.**

The first unique step in the *de novo* pyrimidine biosynthetic pathway involves the enzyme aspartate transcarbamoylase (ATCase, carbamoylphosphate:L-aspartate carbamoyltransferase, EC 2.1.3.2), which catalyzes the nucleophilic addition of the α-amino group of aspartate to the carbonyl carbon of the energy-rich phosphoanhydride, carbamoyl phosphate. In *Escherichia coli*, as in other enteric bacteria, ATCase competes for its phosphoanhydridic substrate with ornithine transcarbamoylase (OTCase, carbamoylphosphate:L-ornithine carbamoyltransferase, EC 2.1.3.3), the sixth enzyme unique to *de novo* arginine biosynthesis. In an analogous reaction to that of ATCase, OTCase catalyzes the transfer of a carbamoyl moiety from carbamoyl phosphate to the δ-amino group of ornithine. There are two major differences between these reactions: (*i*) the amino acid acceptor of the carbamoyl moiety (i.e., aspartate vs. ornithine) and (*ii*) the actual amino group (α-amino or δ-amino) that reacts with the carbamoyl moiety. An evaluation of probable transition states for both enzymes (Fig. 1) shows that the chemical differences between these two states might be due to different charged groups in the unreactive portion of the substrate amino acid. Notwithstanding these differences, the reaction mechanisms for both enzymes are very much alike; thus, the binding of the two transition states would appear to require similar active site geometries. In addition to comparable catalytic mechanisms there are several other striking similarities that have been observed between these two enzymes. The catalytic polypeptides of each enzyme have comparable M_rs:

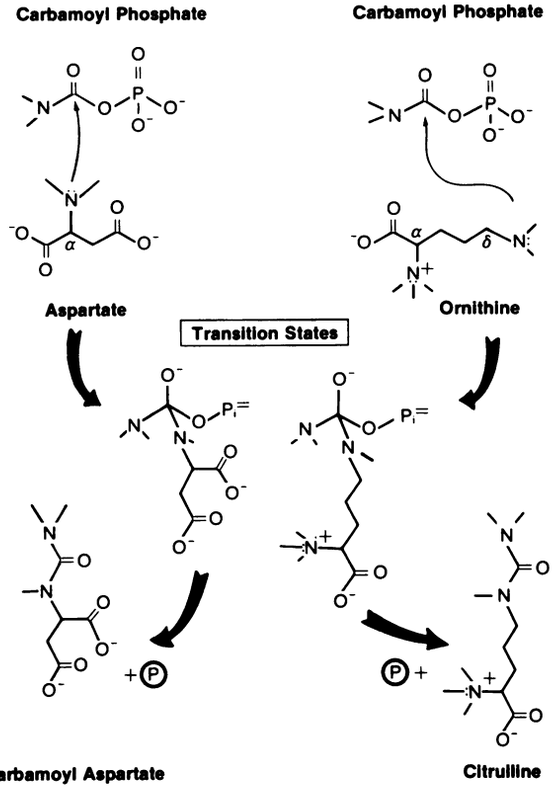

FIG. 1. Reaction of carbamoyl phosphate with aspartate and ornithine. The phosphate group of carbamoyl phosphate is displaced by the α-amino group of L-aspartate to form *N*-carbamoyl-L-aspartate in a reaction catalyzed by ATCase or by the δ-amino group of L-ornithine to form citrulline in an analogous reaction catalyzed by OTCase. The nucleophilic addition–elimination reactions incorporate similar unstable tetrahedral intermediates that decay to eliminate the phosphate groups and form the appropriate products.

ATCase, 34,000 (1, 2); and OTCase, 38,000 (3). Furthermore, the simplest functional architecture for both ATCase and OTCase is that of a catalytically active trimer, even though the native ATCase holoenzyme possesses a more complex dodecameric structure composed of two catalytic trimers linked by three regulatory dimers, $2(c_3):3(r_2)$. The regulatory subunits of ATCase can be separated from the catalytic trimers without the loss of catalytic activity by *in vitro* treatment of the purified enzyme with mercurials (4) or

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: ATCase, aspartate transcarbamoylase; OTCase, ornithine transcarbamoylase.

*Present address: Department of Chemistry, Harvard University, Cambridge, MA 02138.
†To whom reprint requests should be addressed.

heat (5). In addition, mutants lacking *pyrI*, the gene encoding the regulatory polypeptide, produce catalytically active trimers that assemble *in vivo* in the absence of regulatory subunits (6).

The genes encoding ATCase and OTCase (*pyrBI* and *argI*) are located near each other on the *E. coli* chromosome at 96.5 minutes (3, 6, 7). *E. coli* K-12 contains an additional gene, *argF*, located at 6.5 minutes, which encodes a second OTCase (OTCase-2) (7–9). The gene products of *argI* and *argF* form true isoenzymes, even though they exhibit some different physical characteristics, such as surface charge and heat stability (10). This ancillary gene (*argF*) is absent in other enteric bacteria, including *E. coli* strains B and W. Partial sequence comparisons among the products of these three genes revealed significant homology among the first 86 NH_2-terminal amino acids (*argI/argF*, 79%; *pyrB/argI*, 33%; and *pyrB/argF*, 31%) (11). Such homology, when evaluated relative to the previously observed similarities, strongly suggests a common ancestry for these enzymes and is consistent with the possible involvement of tandem duplication(s) in the early evolution of the *pyrB*-encoded ATCase and the *argI*-encoded OTCase-1 (12). The evolution of *argF*-encoded OTCase-2, on the other hand, has been proposed to have occurred either by genome duplication (13) or by transpositional events (14) that might have been intergeneric. From the observation that ATCase and OTCase are both involved in essential metabolic processes, it can be inferred that any duplicational event involving *pyrB* and *argI* must have occurred at a very early stage in biogenesis. Consequently, the divergence of the resultant enzymes represents an attractive model for the investigation of functional differentiation in proteins of such an ancestral lineage. In this paper we detail the similarities and differences in the primary structures of ATCase and OTCase-1 and evaluate their comparative secondary and supersecondary structures.

MATERIALS AND METHODS

Determination of Primary and Secondary Structure. The DNA sequences for both *pyrB* and *argI* have been determined in our laboratory (1, 3). A comparison of their attendant amino acid sequences was undertaken by using both DNA sequence homology and evident amino acid similarities. Secondary structure predictive analysis was accomplished by application of a Chou–Fasman (15, 16)-based computer algorithm, courtesy of Donald Pettigrew and David Proctor (Texas A&M University). Subsequent comparison of the predicted supersecondary structures of ATCase and OTCase with the defined crystallographic structure of ATCase was determined with reference to previously published data (17–19).

RESULTS

Primary Sequence Homology. The ATCase catalytic polypeptide is comprised of 310 amino acids (1, 2), whereas that of OTCase is 333 amino acids in length (3). No direct one-to-one correlation between the two amino acid sequences can be derived without invoking several specific regions of length polymorphism (additions or deletions) (20). As illustrated in Fig. 2, an alignment of these sequences must allow for various insertions and/or deletions (i.e., length polymorphisms) in both ATCase and OTCase, especially in regions of highly divergent sequence. An overall comparison of these sequences shows that OTCase and ATCase possess 25–40% homology; the degree of homology depends on the number of insertions or deletions invoked in maximizing that homology. This alignment revealed that there were numerous residues that shared some degree of "functional homology" in which a charged or hydrophobic amino acid in one sequence corresponded to a similar amino acid in the other (Fig. 2). If functional residue comparisons (for example, sub-

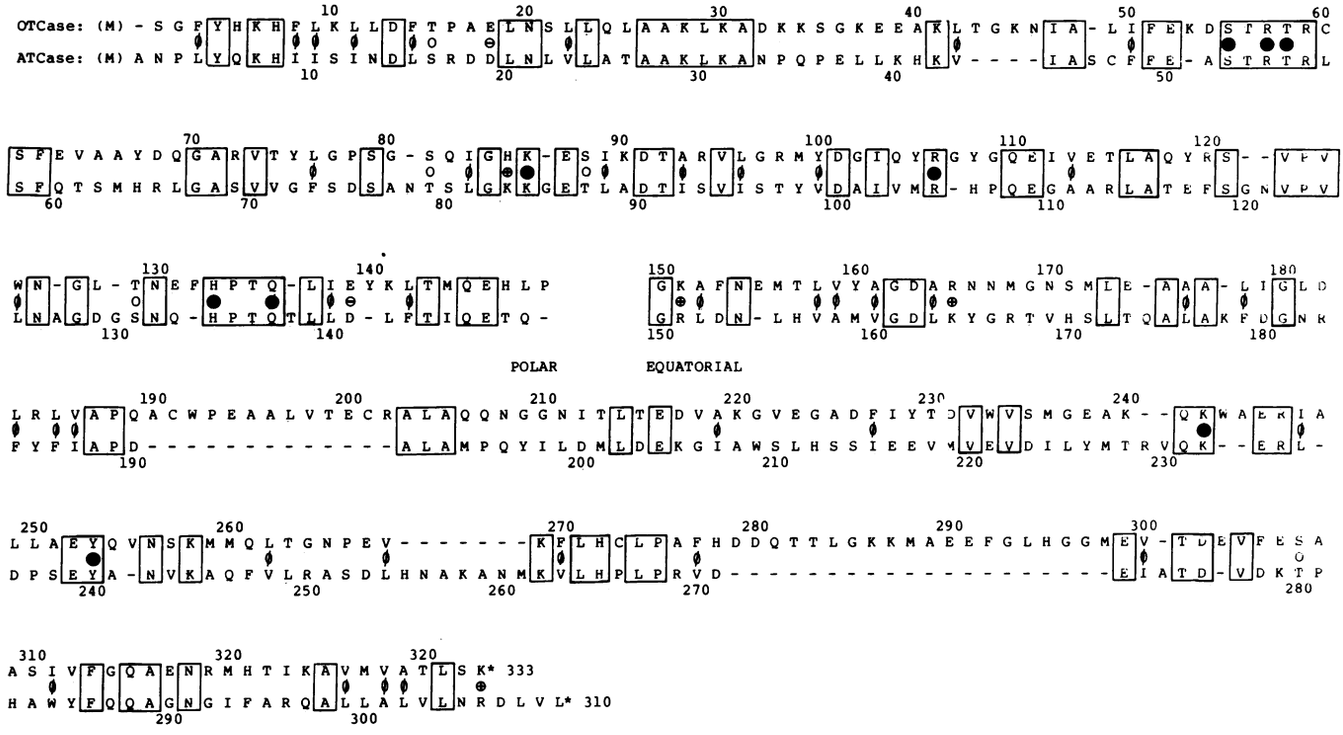

FIG. 2. Comparison of the deduced amino acid sequence of ATCase and OTCase. The amino acid sequences of ATCase (encoded by *pyrB*) and OTCase (encoded by *argI*) have been aligned to maximize primary sequence homology. The amino acid sequences were deduced from the nucleotide sequence of the respective genes (1, 3). The amino acid symbols used are A, Ala; R, Arg; N, Asn; D, Asp; C, Cys; Q, Gln; E, Glu; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; F, Phe; P, Pro; S, Ser; T, Thr; W, Trp; Y, Tyr; and V, Val. Identical amino acids are displayed in boxes, whereas homologous residues possessing similar side chain properties are denoted by various symbols; ○, hydroxyl group; ⊕, positively charged; ⊖, negatively charged; and Φ, hydrophobic. ●, Residues implicated in the catalysis of ATCase that are present in OTCase.

stitution of a threonine residue for a serine) were included, functional homology approached 50%.

X-ray crystallographic data on ATCase (17–19) have defined several specific components of secondary and supersecondary structure, which are organized into two distinct domains, designated as polar and equatorial. The primary sequence of ATCase, and comparable primary sequence of OTCase, can be essentially divided into two halves that correspond to these domains: the first half of each polypeptide (the amino terminus) forms the polar domain and the carboxyl-terminal half forms the equatorial domain. As depicted in Fig. 2, the primary sequence is more conserved in the polar domain and incorporates fewer insertions or deletions than in the equatorial domain. In addition, distinct clusters of considerable homology can be observed throughout both sequences; however, these are again more evident within the polar domain—for example, H1 (Ser-16 to Pro-34), H2 (Ser-52 to Arg-65), S5 (Pro-123 to Gly-128), and H5 (Thr-136 to Glu-149).

Conservation of Catalytically Important Residues. Recent studies with ATCase involving substrate analogues have implicated the involvement of specific residues in the binding of carbamoyl phosphate and aspartate (18, 19, 21–25). Of the 13 amino acids involved at the active site in ATCase, 9 are conserved in OTCase and there is a potential, functional substitution of histidine for lysine-83 (Table 1 and Fig. 2). The primary sequences of phosphoryl binding sites in a number of enzymes have been characterized (26, 27). These sites show some degree of conservation (Table 2) in that the hydroxy amino acid (usually serine) is separated by a single amino acid from a positively charged residue (usually lysine or arginine). This sequence possesses complete homology in ATCase and OTCase (Ser-Thr-Arg-Thr-Arg). Furthermore, these sequences are often found within Rossmann folds consisting of two consecutive $\beta\alpha\beta$-units (for a discussion, see ref. 28), as in adenylate kinase (29). The residues involved in carbamoyl phosphate binding in ATCase and the analogous residues of OTCase are found within a "Rossmann-like" fold.

Table 1. Comparison of amino acid residues implicated in catalysis

ATCase		OTCase[a]	
Cysteine[b,c]	47 (47)	Cysteine	60[d]
Serine[b,e]	52 (52)	Serine	55
Arginine[b,e]	54 (54)	Arginine	57
Threonine[b,e]	55 (55)	Threonine	58
Lysine[f,g]	83 (83)	Histidine	85
Lysine[f,g]	84 (84)	Lysine	86
Arginine[b,e]	105 (105)	Arginine	106
Histidine[b,e]	134 (134)	Histidine	133
Glutamine[b,e]	137 (137)	Glutamine	137
Tyrosine[f,h]	165 (165)	—	
Arginine[e,f]	167 (167)	—	
Lysine[f,g]	232 (224)	Lysine	242
Tyrosine[b,i]	240 (232)	Tyrosine	253

Numerals in parentheses represent the numbering system denoted by Konigsberg's unpublished provisional amino acid sequences used by Honzatko and co-workers (18, 19). Numerals outside parentheses are defined from the current sequences (1, 2).
[a]Postulated for OTCase based on homology with ATCase.
[b]Suggested from x-ray structural analyses.
[c]Ref. 18.
[d]Denotes a non-homologous residue, which may have a similar orientation and function within the active site.
[e]Ref. 19.
[f]Implicated in catalysis by direct experimentation.
[g]Refs. 21 and 25.
[h]Ref. 24.
[i]Refs. 19 and 22.

Table 2. Comparison of amino acid sequences involved in the phosphate binding sites of a number of proteins

Protein	Sequence
Myelin protein	Gly - Ser(OH) - Gly - Lys - Asp
Histone H2	- Ser(OH) - Gly - Arg - Gly
Histone H4	- Ser(OH) - Gly - Arg - Gly
Histone H1	Gly - Ser(OH) - Phe - Lys - Leu
Glycogen synthetase	Ile - Ser(OH) - Val - Arg -
Troponin I	Ile - Thr(OH) - Ala - Arg - Arg
Flavodoxin	Gly - Ser(OH) - Gly - Lys - Gly
Adenylate kinase	Gly - Ser(OH) - Gly - Lys - Gly
ATCase*	^{51}Ala - Ser(OH) - Thr - Arg - Thr - ^{56}Arg
OTCase	^{54}Asp - Ser(OH) - Thr - Arg - Thr - ^{59}Arg
Consensus sequence[†]	XXX - Ser(OH) - XXX - Arg/Lys - XXX

This table was modified from Schulz and Schirmer (26).
*See Fig. 2 for sequence orientation.
[†]XXX represents a variable amino acid residue.

Predicted Secondary Structural Conservation. Unfortunately, x-ray crystallographic data are not available for OTCase and a definitive comparison of structure cannot be undertaken. Nonetheless, there are several methods that can predict secondary structural elements such as α-helices, β-strands (sheets), and turns. One commonly utilized method is that of Chou and Fasman (15, 16), which has predicted secondary structures to an accuracy approximating 75% when compared to x-ray crystallographic data. The Chou–Fasman method has been used, with the aid of computer analysis, to predict secondary structural units within both ATCase and OTCase. As can be seen in Fig. 3, this method adequately predicts the secondary structure of ATCase, which has been defined by crystallographic analysis (17–19). The corresponding prediction for OTCase, also shown in Fig. 3, should similarly reflect the secondary structure of OTCase. Indeed, where Chou–Fasman analysis fails to predict the secondary structure for ATCase (e.g., residues 233–266), a similar secondary structure to that predicted for ATCase is predicted for OTCase (residues 235–280). It is graphically evident that the predicted secondary structure of the OTCase sequence compares favorably to that of ATCase, even in regions where little if any direct amino acid homology is apparent. Furthermore, the similarities in secondary structural architecture are more extensive within the polar domain than in the equatorial domain. The most highly conserved regions are depicted in Fig. 4. This central region is thought to be involved in catalysis (18). A notable exception to this generalization is the observation that the carboxyl-terminal "-strand-H8-S11-H9-" sequence in the equatorial domain of ATCase is strongly predicted in both ATCase' and OTCase'. This region of the peptide chain apparently folds back and reenters the polar domain in ATCase. Thus, these residues bridge the inter-domain crevice, which is thought to be of considerable importance in the catalytic activity of ATCase (18, 19).

Predicted Secondary Structural Differentiation. Even with the proposed structural similarities apparent between these two enzymes, there are distinct differences that correspond to sequences of amino acids found within one sequence and not the other (i.e., length polymorphisms). The alignment shown in Fig. 2 suggests that there are five regions that exhibit marked structural differences. All of these regions correspond to surface components of ATCase, as depicted in Fig. 4. Similar surface variants have been determined for families of differentiated proteins (20, 30). Of particular im-

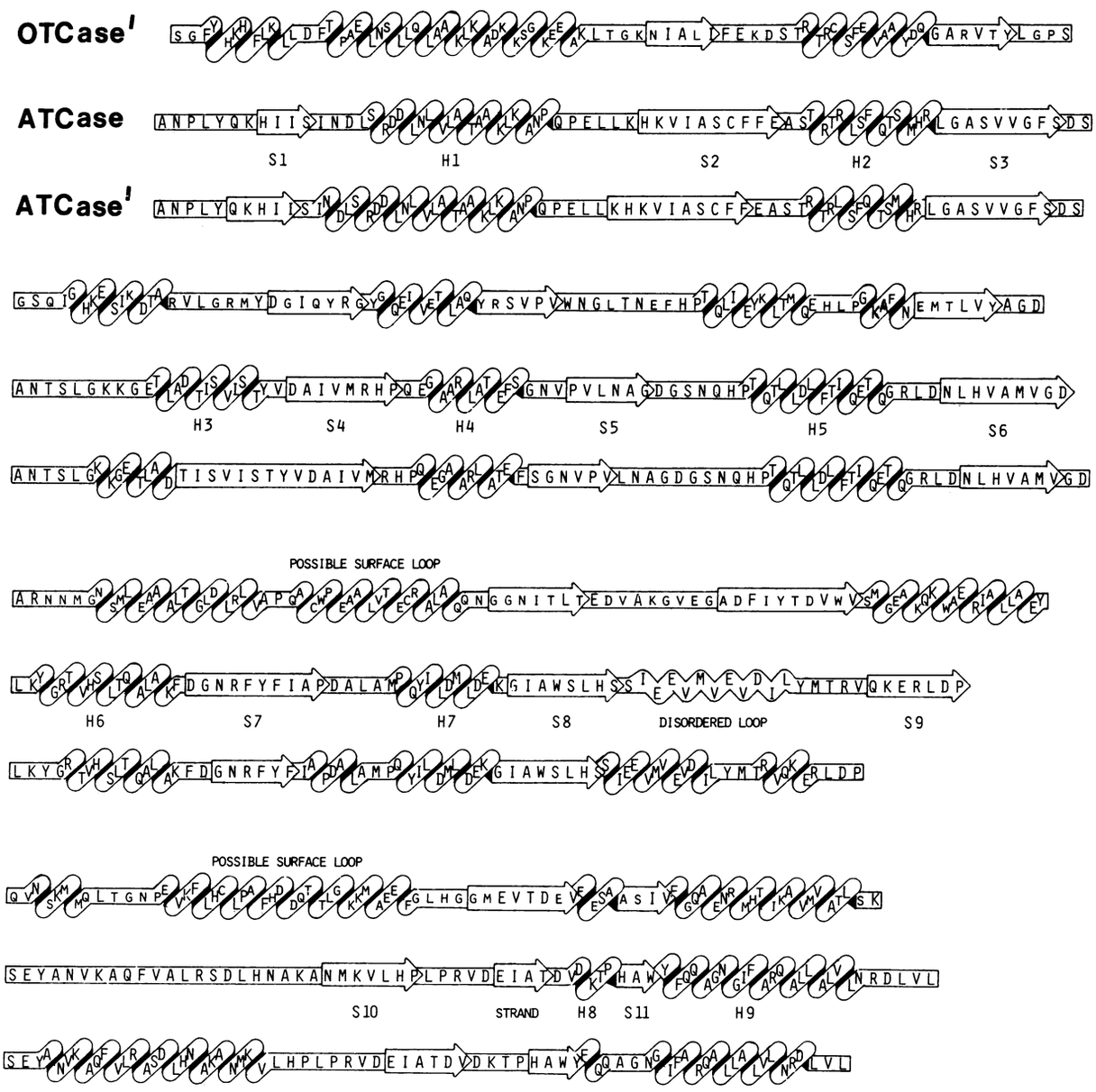

FIG. 3. Diagramatic representation of the secondary structure of ATCase and OTCase from *E. coli* as predicted by Chou–Fasman analysis of the primary sequences (15, 16). The predicted structures are indicated by ATCase' and OTCase'. The structure of ATCase as estimated by x-ray crystallographic analyses by Honzatko and Lipscomb (19) and Monaco *et al.* (17) is presented between the predicted structures and is designated ATCase. α-Helices are represented graphically and β-sheets are represented as linear arrows. Uppercase letters in the OTCase' sequence signify direct or functional amino acid homologies between ATCase and OTCase. The amino acid symbols are the same as used in Fig. 2.

portance are two extensive insertions present in OTCase but absent in ATCase. These regions contain many residues tending to form α-helices but include internal proline residues that completely exclude this possibility. These insertions may be additionally restricted from helical formation by steric constraints imposed by the neighboring sheet lattice of the equatorial domain. Thus, these regions appear to be *surface loops* located at opposite ends of the equatorial domain (S7-H7 vs. S10-H8 in Fig. 4). These surface loops are composed of 13 and 19 amino acids and are bounded by residues that correspond to amino acids at positions 190 and 191 and positions 271 and 272 of ATCase (Fig. 2). Three additional regions show varying length polymorphisms that are apparently involved in the surface regions of the enzymes. The first of these involves a turn between H1 and S2 in which OTCase contains four more residues than ATCase. Predictive analysis indicates that a similar turn function is undertaken by different residues in the two enzymes, in that the -Pro-Gln-Pro-Glu- turn in ATCase could be accommodated by a larger -Asp-Lys-Lys-Gly-Leu-Thr-Gly- sequence in OTCase. The other two regions of moderate structural dimorphism involve the residues corresponding to S8/S9 and S9/S10 of ATCase (Fig. 4). In OTCase there appear to be fewer residues in these regions.

DISCUSSION

The enzymes ATCase and OTCase possess some remarkable similarities both in their primary sequences and in the architecture of their predicted secondary structures. A comparison of these similarities supports the hypothesis that these two enzymes share a common ancestry and have diverged subsequent to tandem gene duplication. Conservation of the secondary structural elements in ATCase and

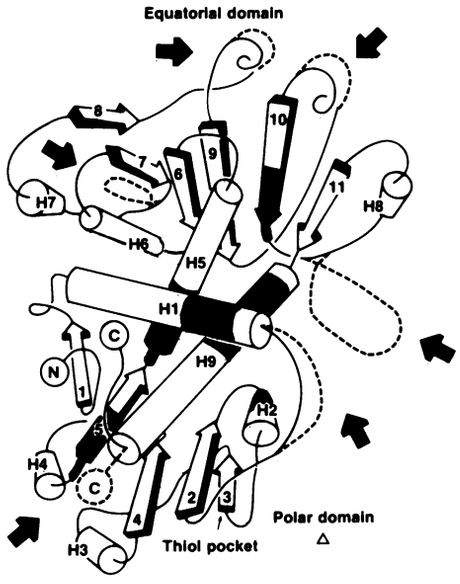

FIG. 4. Schematic representation of a single catalytic polypeptide of ATCase [adapted from Honzatko and Lipscomb (19) with permission]. The regions of length polymorphism between ATCase and OTCase are indicated by directing arrows. Dashed lines indicate the predicted structural configurations of OTCase that differ from ATCase. The α-helices of the catalytic chain are represented by cylinders designated H1, H2, etc., and the β-sheet regions are represented as broad internal arrows designated 1, 2, etc. Shaded regions denote defined clusters of rigorous amino acid sequence homology between OTCase and ATCase.

OTCase is predominantly associated with the polar domain, which is thought to be involved in the binding of the common substrate, carbamoyl phosphate. On the other hand, th equatorial domain exhibits less primary sequence homology, possesses less obvious conservation of secondary structure, and contains regions of length polymorphism. This observation is consistent with the proposal that the equatorial domain provides for the binding of the different amino acid substrates—namely, aspartate or ornithine. Thus, the equatorial domain appears to have diverged to a greater extent than the polar domain. Other divergent protein families, such as the serine proteases (30), RNases (31), and α/β barrel enzymes (32), possess similar functional diversification without altering their respective main chain backbone configurations.

Even though the equatorial domain exhibits diversification of sequence and secondary structure, a detailed comparison of predicted secondary structures indicates that regions of similarity do exist. Some of the predicted structures are known to deviate from the observations of the crystallographic data. For example, S9 and S10 of ATCase are poorly predicted (Fig. 3). It has been proposed that this region of the equatorial domain experiences structural perturbations as a consequence of substrate binding (18). Thus, the failure of Chou–Fasman analysis to accurately predict secondary structures within this region may not be a consequence of poor analysis, but of the difficulty in evaluating internal structural forces resulting in a *metastable* condition. Similar secondary structures are predicted for both ATCase and OTCase and may be indicative that this region in OTCase is also subject to transition upon binding of substrate.

In the absence of crystallographic data, the importance of the various regions of length polymorphism cannot be evaluated in detail. Although such regions may have little significance with respect to differing substrate specificity, the larger insertions or deletions may provide important functional specificities to OTCase. The evaluation of these differences by comparison of sequence and structure will be important in understanding the different molecular components that determine substrate specificity.

This research has been supported by the Robert A. Welch Foundation (A-915), the National Institutes of Health (NIHMS-5R01GM29152), and the National Science Foundation (PCM-8021983).

1. Hoover, T. A., Roof, W. D., Foltermann, K. F., O'Donovan, G. A., Bencini, D. A. & Wild, J. R. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2462–2466.
2. Konigsberg, W. H. & Henderson, L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2467–2471.
3. Bencini, D. A., Houghton, J. E., Hoover, T. A., Foltermann, K. F., Wild, J. R. & O'Donovan, G. A. (1983) *Nucleic Acids Res.* **11**, 8509–8518.
4. Gerhart, J. C. & Schachman, H. K. (1965) *Biochemistry* **4**, 1054–1062.
5. Gerhart, J. C. (1970) *Curr. Top. Cell. Regul.* **2**, 275–325.
6. Roof, W. D., Foltermann, K. F. & Wild, J. R. (1982) *Mol. Gen. Genet.* **187**, 391–400.
7. Bachmann, B. J. (1983) *Microbiol. Rev.* **47**, 180–230.
8. Glansdorff, N., Sand, G. & Verhoef, C. (1967) *Mutat. Res.* **4**, 743–751.
9. Legrain, C., Stalon, V. & Glansdorff, N. (1976) *J. Bacteriol.* **128**, 35–38.
10. Legrain, C., Halleux, P., Stalon, V. & Glansdorff, N. (1972) *Eur. J. Biochem.* **27**, 93–102.
11. Piette, J., Cunin, R., van Vliet, R., Charlier, D., Crabeel, M., Ota, Y. & Glansdorff, N. (1982) *EMBO J.* **1**, 853–857.
12. Gigot, D., Glansdorff, N., Legrain, C., Piérard, A., Stalon, V., Konigsberg, W., Caplier, I., Strosberg, A. D. & Hervé, G. (1977) *FEBS Lett.* **81**, 28–32.
13. Riley, M. & Anilionis, A. (1978) *Annu. Rev. Microbiol.* **32**, 519–560.
14. York, M. K. & Stodolsky, M. (1981) *Mol. Gen. Genet.* **181**, 230–240.
15. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–222.
16. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245.
17. Monaco, H. L., Crawford, J. L. & Lipscomb, W. N. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5276–5280.
18. Honzatko, R. B., Crawford, J. L., Monaco, H. L., Ladner, J. E., Edwards, B. F. P., Evans, D. R., Warren, S. G., Wiley, D. C., Ladner, R. C. & Lipscomb, W. N. (1982) *J. Mol. Biol.* **160**, 219–263.
19. Honzatko, R. B. & Lipscomb, W. N. (1982) *J. Mol. Biol.* **160**, 265–286.
20. Craik, C. S., Rutter, W. J. & Fletterick, R. (1983) *Science* **220**, 1125–1129.
21. Kempe, T. D. & Stark, G. R. (1975) *J. Biol. Chem.* **250**, 6861–6869.
22. Landfear, S. M., Evans, D. R. & Lipscomb, W. N. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2654–2658.
23. Wall, K. A., Flatgaard, J. E., Gibbons, I. & Schachman, H. K. (1979) *J. Biol. Chem.* **254**, 11910–11916.
24. Lauritzen, A. M., Landfear, S. M. & Lipscomb, W. N. (1980) *J. Biol. Chem.* **255**, 602–607.
25. Lauritzen, A. M. & Lipscomb, W. N. (1982) *J. Biol. Chem.* **257**, 1312–1319.
26. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York), p. 226.
27. Sachsenheimer, W. & Schulz, G. E. (1977) *J. Mol. Biol.* **114**, 23–36.
28. Rao, S. T. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76**, 241–256.
29. Pai, E. F., Sachsenheimer, W., Schirmer, R. H. & Schulz, G. E. (1977) *J. Mol. Biol.* **114**, 37–45.
30. Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G. & Thiessen, W. E. (1975) *Nature (London)* **257**, 758–763.
31. Nakamura, K. T., Iwahashi, K., Yamamoto, Y., Iitaka, Y., Yoshida, N. & Mitsui, Y. (1982) *Nature (London)* **299**, 564–566.
32. Lebioda, L., Hatada, M. H., Tulinsky, A. & Mavridis, I. M. (1982) *J. Mol. Biol.* **162**, 445–458.