Structure of unligated aspartate carbamoyltransferase of *Escherichia coli* at 2.6-Å resolution

(allosteric enzymes/protein structure/crystallography)

HENG-MING KE*, RICHARD B. HONZATKO†, AND WILLIAM N. LIPSCOMB*

*Gibbs Chemical Laboratory, Harvard University, Cambridge, MA 02138; and †Department of Biochemistry and Biophysics, Iowa State University, Ames, IA 50011

Contributed by William N. Lipscomb, March 19, 1984

ABSTRACT The three-dimensional structure of the allosteric enzyme aspartate carbamoyltransferase (EC 2.1.3.2) has been refined to a crystallographic *R*-factor of 0.24 at 2.6-Å resolution in the space group *P*321, where *a* and *b* are 122.1 Å and *c* is 142.2 Å. This structure is isomorphous to the form of the enzyme complexed to the allosteric inhibitor cytidine triphosphate. All sources of sequence information have been evaluated against the electron density. The corrected amino acid sequences of the catalytic and regulatory proteins have been incorporated in the model, and three regions in the active site are described: (i) near arginine-105, histidine-134, and arginine-167, (ii) near lysine-232 and arginine-229, and (iii) near lysine-83 and lysine-84.

The allosteric enzyme aspartate carbamoyltransferase (carbamoylphosphate:L-aspartate carbamoyltransferase, EC 2.1.3.2) from *Escherichia coli* catalyzes the formation of carbamoyl-L-aspartate from L-aspartate and carbamoyl phosphate. The enzyme consists of catalytic trimers *c*₃ and regulatory dimers *r*₂ in a structure *c*₆*r*₂ having *D*₃ symmetry (1–3). Substrates induce very large conformational changes in the structure of the enzyme (4) which are influenced by the allosteric inhibitor CTP and the allosteric activator ATP (5–7).

Three-dimensional structures have been reported (8) for the T (less active) form of the enzyme in the unligated form (space group *R*32) and in the CTP-ligated form (space group *P*321). We report here the structure to a resolution of 2.6 Å of the unligated T form in the space group *P*321, isomorphous with the CTP form. The improved structure provides two new crystallographic environments for the unligated *rc* unit. In addition, we have rebuilt the model with the use of a primary structure, which is based on DNA sequences, in both the *c* and *r* chains (2, 9–11). Although no change occurs in the trace of either the *c* or the *r* chain, the insertion of residues after 220 of the *c* chain (9, 10) has required the relocation of amino acid side chains from 217 to 261. In addition, the insertion of a lysine at 129 increases the number of residues in the *r* chain from 152 (2) to 153 (11) and requires a small revision in the three-dimensional structure.

MATERIALS AND METHODS

Aspartate carbamoyltransferase of *E. coli* was purified by the method of Gerhart and Holoubek (12), as described elsewhere (8). Single crystals of the enzyme were grown at room temperature by dialysis of a solution containing 5–10 mg of protein per ml against a solution containing 100 mM maleic acid, 10 mM Tris·HCl, 1 mM 2-mercaptoethanol, 0.2 mM EDTA, and 0.3 mM NaN₃ at pH 5.8. After 3–7 days, plate-like crystals formed with typical dimensions of 1.2 × 1.0 × 0.2 mm. These crystals are hexagonal, in the space group

*P*321, and had unit cell dimensions of *a* = *b* = 122.1 Å and *c* = 142.2 Å.

X-ray diffraction data were collected by the oscillation method, using an Elliott GX6 rotating anode unit operated at about 39 kV and 19 mA, and using a 100-μm focusing cup. Each oscillation photograph was taken over a 1° interval at an exposure time of 4 hr. The collimator size was 0.35 mm. All films were scanned on an Optronics P-1000 Photoscan, which was interfaced with a PDP 11/34 computer (Digital Equipment). The software was developed by Crawford (13). The three films of a pack were scaled by the method of Fox and Holmes (14), and separate film packs were merged by Rae's method (15). A total of 78,896 reflections were merged to yield 26,912 unique reflections.

The initial atomic model was developed from the two T structures previously described (8) for the unligated enzyme in *R*32 and the CTP-ligated form in *P*321, as modified by the revised sequences of the *c* and *r* chains. The model was introduced into the interactive program BILDER (16) as modified by R. C. Ladner for use on a Vax 11/780 computer interfaced with an Evans and Sutherland Picture System II. The asymmetric unit (*c*₂*r*₂) was refined by restrained reciprocal space least-squares methods (17). Derivatives of structure factor were calculated from the difference map by the method of Jack and Levitt (18). In all structure factor and electron density calculations, the fast Fourier transform routines of Ten Eyck (19, 20) were used. To correct small regions, those residues were omitted, and maps were made using $2F_o - F_c$ and $F_o - F_c$ as coefficients with calculated phases, in which F_o and F_c are the magnitudes of the observed and calculated structure factors, respectively. We maintained the geometry within 0.02 Å or less of idealized values at all stages of refinement, as an aid in localizing less refined regions. The present value of the crystallographic *R*-factor is 0.24, and the root mean square deviations of the observed bond distances and angles from the idealized values are 0.024 Å and 4°, respectively.

RESULTS AND DISCUSSION

Five sources exist for sequence information pertaining to aspartate carbamoyltransferase from *E. coli* [DNA sequences of Schachman *et al.* (11) and Hoover *et al.* (9), amino acid sequences of Konigsberg and Henderson (10) and Weber (2), and our electron density maps]. In general all sources of information are now in good agreement; we comment here on the few discrepancies. The amino acid sequence of Konigsberg disagrees with the two DNA sequences in the region of 259–261 (Ala-Asn-Met for the DNA results and Met-Asn-Ala for the results from protein sequencing). The DNA version accounts for the electron density more favorably than does the amino acid sequence of Konigsberg. The DNA sequences of the catalytic chains disagree only in positions 60,

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: *c*, catalytic chain; *r*, regulatory chain; PALA, *N*-phosphonacetyl-L-aspartate.

Table 1. Summary of geometry for unliganded aspartate carbamoyltransferase, space group *P*321

Distances (rms Å)	
Bond	0.024
Angle	0.066
Planar	0.076
Planarity, rms Å	0.031
Chiral, rms Å	0.281
Nonbonded contacts, rms Å	
Single	0.297
Multiple	0.438
Hydrogen bonding	0.352
Torsional angle, rms degrees	
Planar	7.2
Staggered	24.9
Orthonormal	23.6

149, and 220, with Schachman *et al.* reporting Glu-60, Glu-149, and Ala-220, and with Hoover *et al.* reporting Gln-60, Gln-149, and Val-220. Point mutations may account for these differences in the DNA sequences. On the basis of electrostatic and steric interactions of the present model for aspartate carbamoyltransferase, either of the alternatives for DNA sequencing is reasonable.

The two regulatory chain sequences [protein sequences of Weber (2) and the DNA sequence of Schachman *et al.* (11)] differ at 21 locations. The bulk of these are discrepancies in the assignment of the functional group of amide or acid to side chains. However, in three instances the Weber sequence has pairs of amino acids inverted with respect to the Schachman *et al.* sequence and, in addition, the Weber sequence has one less amino acid (position 129 of the latter sequence) than does the Schachman *et al.* DNA sequence. In these three inversions and one insertion we have been able to fit the Schachman *et al.* sequence to the electron density with success equal to or greater than that of fits that employed the Weber sequence.

In a number of instances the chemical sequences call for large side chains where the electron density is insufficient. We attribute these discrepancies to conformational disorder in side chains, which effectively spreads the electron density to the noise level of the map. Regions of uncertainty in the model are (using the sequence numbering of Schachman *et*

FIG. 1. View of aspartate carbamoyltransferase along the threefold axis. The three twofold axes lie in the plane of the drawing.

al.) residues 76–80 and 239–244 of the catalytic chain, and residues 8–11, 53–58, and 129–130 of the regulatory chain. For each of the sequences above, the trace of the main chain is clear, but density for the side chains is weak. The modeling of the solvent structure for aspartate carbamoyltransferase may lower the noise level of the map sufficiently to remove ambiguities in our interpretation of the electron density. Although the refinement is not complete, we feel the current model is a significant improvement over former models of the T state (8); these coordinates have been deposited in the Protein Data Bank at Brookhaven. In Table 1, we present geometric parameters that are output from the computer routines of Hendrickson and Konnert (17).

For reference, the arrangement of the six catalytic chains and six regulatory chains is shown in Fig. 1. The lower c_3r_3 is shown in Fig. 2 as C4-C5-C6-R4-R5-R6 in the same relative

FIG. 2. An α -carbon trace of a c_3r_3 unit of the molecule, looking down on threefold axis, which is marked by a triangle in the middle. The enclosed chain is C4. The other two c chains in clockwise order are C5 and C6. The large circle in C4 represents the binding site for pyrophosphate. The two-domain structure of the r chains is apparent. The three outermost domains of R4, R5, and R6 bind CTP, while the adjacent domains contain the Zn site as small circles to the left of the Zn label.

FIG. 3. Interface region of C1-R4-C4, looking down on the *z* axis. The fragments indicated with double lines are from the lower catalytic chain (C4); residues Asp-236 to Asn-242 are from the upper catalytic chain (C1); the other single-line trace is from a neighboring regulatory chain, R4 (Cys-109 to Cys-114 and Glu-142 to Phe-145).

orientations as in Fig. 1. This view allows one to look into the active site, the large unlabeled circle within the enclosed *c* chain, where pyrophosphate binds. The molecular center is at the triangle, which indicates a threefold molecular axis perpendicular to the plane of Fig. 2, and which lies in the central cavity of the hexameric enzyme c_6r_6 .

In more detail, the central cavity is made up of six catalytic chambers. Most of the side chains defining a given catalytic chamber belong to that catalytic chain, except that residues 80-87 are contributed by an adjacent catalytic chain. In addition, the important tyrosine residues 165 and 240 are located at the edge of the catalytic chamber of the chain to which they are bound, but they extend in the direction of an adjacent catalytic chamber. Both of these tyrosine residues are near the C1-R4-C4 interface (8) (Fig. 3). This interface involves Lys-164 and Tyr-165 from C4, Asp-236 to Asn-242 from C1, Pro-110 to Asn-113 from R4, and Lys-143 to Phe-145, also from R4 (Fig. 3).

Turning now to the active site cavity, we show the pyrophosphate site at which one oxygen-bridged PO_3 overlaps with the phosphate site, which is also shown (Fig. 4). These two sites are added from the earlier study (21) without changing the original positions of the binding ligands, Arg-54, Arg-105, and Arg-167. However, the detailed binding has been described previously (21). His-134 is also in this immediate region.

Another region of potential binding of substrates is near Arg-229 and Lys-232. A third general region is near Lys-83

and Lys-84. The flexibility of these long side chains makes a precise assignment to sites a bit risky, but it is clear that if carbamoyl phosphate were to bind at the phosphate site and aspartate to the region near Lys-83 and Lys-84 they would be much too far apart to form products. One reasonable possibility is to assign at least three binding and catalytic sites and to regard Lys-83 and -84 as a recognition site rather than a catalytic site. It has been commented previously (8) that the presence of three sites supports the position taken by Heyde (22) in her discussion of binding studies of carbamoyl phosphate, aspartate, succinate, and *N*-phosphonacetyl-L-aspartate (PALA). Her results indicate that (i) both PALA and aspartate bind, (ii) this binding is noncompetitive, (iii) PALA competes with carbamoyl phosphate, and (iv) PALA decreases binding of succinate but does not decrease binding of aspartate. The most straightforward interpretation is that three sites are required—one for each end of PALA, and a third site. Further study and location of these binding sites are required to establish whether succinate is indeed a good analogue for aspartate and whether PALA is bound abnormally—i.e., not as a true substrate analogue.

The relocation of residues 217-261, as required by the new sequence of the catalytic chain, resolves some earlier problems in the structure-function relationship. For example, the amino-terminal group of Lys-232 (formerly Lys-224) is now in the active site, whereas it had previously (8) been placed some 23 Å away from the functional group of Lys-83 and Lys-84.

FIG. 4. A catalytic chamber, looking down on the *z* axis. The residues marked with double lines belong to C4, while the others are from C6. Residue Cys-47 is to the right of the chamber. Binding sites for pyrophosphate (P_2O_7) and phosphate (PO_4) are represented by the circles in the chamber. The triangle represents the position of the threefold axis. Comparison with Fig. 3 shows the spatial relationship of the chamber to the C1-R4-C4 interface.

Chemical modification of Lys-83, Lys-84, and Lys-232 by D- and L-bromosuccinate indicates that modification of any one of them causes loss of catalytic activity (23). The small suppression of alkylation by bromosuccinate of Lys-83 by carbamoyl phosphate raises a question as to whether the productive binding site for carbamoyl phosphate is near Lys-83 or deep in the pyrophosphate pocket. On the other hand, this effect of carbamoyl phosphate may be less direct. Again this question will have to be settled in future structural studies of binding of substrate analogues. Further refinement, in which water molecules near the protein have been included, is presently characterized by an *R* value of 0.21 while maintaining geometric constraints as described above.

Note Added in Proof. A similar refinement to a value of *R* of 0.21 has been completed on the crystals of aspartate carbamoyltransferase bound to its allosteric inhibitor, cytidine triphosphate. This refinement clarifies the binding mode of CTP to the regulatory site. Coordinates for both unligated enzyme and the CTP-ligated enzyme have been deposited at the Brookhaven Protein Data Bank (Brookhaven, Long Island, NY).

We thank the National Institutes of Health for Grant GM06920 and the National Science Foundation for Grant PCM-77-11398 in support of the computing facilities.

1. Wiley, D. C. & Lipscomb, W. N. (1968) *Nature (London)* **218**, 1119–1121.
2. Weber, K. (1968) *Nature (London)* **218**, 1116–1119.
3. Wiley, D. C., Evans, D. R., Warren, S. G., MacMurray, C. H., Edwards, B. F. P., Franks, W. A. & Lipscomb, W. N. (1972) *Cold Spring Harbor Symp. Quant. Biol.* **36**, 285–290.
4. Ladner, J. E., Kitchell, J. P., Honzatko, R. B., Ke, H. M., Volz, K. W., Kalb (Gilboa), A. J., Ladner, R. C. & Lipscomb, W. N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3125–3128.
5. Gerhart, J. C. & Pardee, A. B. (1962) *J. Biol. Chem.* **237**, 891–896.
6. Bethell, M. R., Smith, K. E., White, J. S. & Jones, M. E. (1968) *Proc. Natl. Acad. Sci. USA* **60**, 1442–1449.
7. Gerhart, J. C. & Pardee, A. B. (1964) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **23**, 727–735.
8. 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.
9. 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.
10. Konigsberg, W. H. & Henderson, L. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2467–2471.
11. Schachman, H. K., Pauze, C. D., Navre, M., Karels, M. J., Wu, L. & Yang, Y. R. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 115–119.
12. Gerhart, J. C. & Holoubek, H. (1967) *J. Biol. Chem.* **242**, 2886–2892.
13. Crawford, J. L. (1976) Dissertation (Harvard Univ., Cambridge, MA).
14. Fox, G. C. & Holmes, K. C. (1966) *Acta Crystallogr.* **20**, 886–891.
15. Rae, A. D. (1965) *Acta Crystallogr.* **19**, 683–684.
16. Diamond, R. (1966) *Acta Crystallogr.* **21**, 253–266.
17. Hendrickson, W. A. & Konnert, J. (1981) in *Biomolecular Structure, Function, Conformation and Evolution*, ed. Srinivasan, R. (Pergamon, London), Vol. 1, pp. 43–47.
18. Jack, A. & Levitt, M. (1978) *Acta Crystallogr. Sect. A* **34**, 931–935.
19. Ten Eyck, L. F. (1973) *Acta Crystallogr. Sect. A* **29**, 182–191.
20. Ten Eyck, L. F. (1977) *Acta Crystallogr. Sect. A* **33**, 486–492.
21. Honzatko, R. B. & Lipscomb, W. N. (1982) *J. Mol. Biol.* **160**, 265–286.
22. Heyde, E. (1976) *Biochim. Biophys. Acta* **452**, 81–88.
23. Lauritzen, A. M. & Lipscomb, W. N. (1982) *J. Biol. Chem.* **257**, 1312–1319.