Human argininosuccinate lyase: A structural basis for intragenic complementation

MARY A. TURNER*, ALAN SIMPSON†‡, RODERICK R. McINNES§¶, and P. LYNNE HOWELL*••††

*Division of Biochemistry Research, Departments of Genetics and Pediatrics, and Program in Developmental Biology, Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada; †Laboratory of Molecular Biology and Imperial Cancer Research Fund Unit, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, United Kingdom; and Departments of **Biochemistry and Molecular and Medical Genetics, Faculty of Medicine, University of Toronto, Medical Sciences Building, Toronto, Ontario M5S 1A8, Canada

Communicated by Gregory A. Petsko, Brandeis University, Waltham, MA, June 17, 1997 (received for review December 3, 1996)

ABSTRACT Intragenic complementation has been observed at the argininosuccinate lyase (ASL) locus. Intragenic complementation is a phenomenon that occurs when a multimeric protein is formed from subunits produced by different mutant alleles of a gene. The resulting hybrid protein exhibits enzymatic activity that is greater than that found in the oligomeric proteins produced by each mutant allele alone. The mutations involved in the most successful complementation event observed in ASL deficiency were found to be an aspartate to glycine mutation at codon 87 of one allele (D87G) coupled with a glutamine to arginine mutation at codon 286 of the other (Q286R). To understand the structural basis of the Q286R:D87G intragenic complementation event at the ASL locus, we have determined the x-ray crystal structure of recombinant human ASL at 4.0 Å resolution. The structure has been refined to an R factor of 18.8%. Two monomers related by a noncrystallographic 2-fold axis comprise the asymmetric unit, and a crystallographic 2-fold axis of space group P3121 completes the tetramer. Each of the four active sites is composed of residues from three monomers. Structural mapping of the Q286R and D87G mutations indicate that both are near the active site and each is contributed by a different monomer. Thus when mutant monomers combine randomly such that one active site contains both mutations, it is required by molecular symmetry that another active site exists with no mutations. These “native” active sites give rise to the observed partial recovery of enzymatic activity.

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.

© 1997 by The National Academy of Sciences 0027-8424

Abbreviations: ASL, argininosuccinate lyase; TDIC, turkey δ I crystallin.

Data deposition: The coordinates and structure factors have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY, 11973 (accession nos. 1AOS and R1A0SSF, respectively).

†Present address: Purdue University, West Lafayette, IN 47907.

To whom reprint requests should be addressed. e-mail: howell@aragorn.psfs.sickkids.on.ca.

9063
FIG. 1. The three regions of high sequence conservation among members of the ASL superfamily. ASL, argininosuccinate lyase; D2C, δ II crystallin; fumarase, E. coli fumarase C; aspartase, E. coli aspartase; CMLE, P. putida 3-carboxy-cis,cis-muconate lactonizing enzyme; ADS, B. subtilis adenylsuccinase.

deficient strains and provides experimental proof of Crick and Orgel’s initial hypothesis of intragenic complementation (17). Evidence for complementation according to the scheme of statistical regeneration of wild-type active sites also has been observed in vitro for the homodimeric proteins thymidylate synthase (18), ribulose-bisphosphate carboxylase/oxygenase (19), glutathione reductase (20), and mercuric reductase (21) and the homotrimeric enzyme aspartate transcarbamoylase (22). As intragenic complementation has been implicated in other diseases, involving mutations in homomultimeric proteins such as propionic acidemia (23, 24) and methylmalonic aciduria (25), a logical extension to heteromultimeric proteins also may apply, making intragenic complementation a much overlooked source of phenotypic and biochemical variation in genetic disease.

EXPERIMENTAL METHODS

Crystallization and Data Collection. Recombinant human ASL was expressed in Escherichia coli, purified, and crystallized as described previously (26). All crystals were obtained by the hanging drop vapor diffusion method. Crystals have been grown from a 10–15 mg/ml protein solution in 50 mM phosphate buffer, 5 mM DDT at pH 7.1 with 1.1 M phosphate by the hanging drop vapor diffusion method. Crystals have been mounted on a Rigaku RU-200 rotating anode generator (37 room temperature using a Siemens multiwire detector as the precipitating agent. Intensity data were measured at 2.9 Å^3 Da^-1 giving rise to a solvent content of ~43%. A summary of the data collection statistics is presented in Table 1.

Structure Solution and Refinement. The structure of human ASL was solved by molecular replacement using the program XPLOR (29) with the coordinates of enzymatically inactive turkey δ I crystallin (TDIC) as the search model (30). Because the sequence of TDIC is unknown, Simpson et al. (30) have used the sequence of chicken δ I crystallin in their crystallographic model. ASL and chicken δ I crystallin share 64% amino acid sequence identity, whereas turkey and chicken are expected to share >90% sequence identity. Data between 8 and 4 Å resolution were used in all rotation and translational searches. Initial rotation searches and Patterson correlation refinement were carried out using coordinates of TDIC subunit A, which had lower average atomic temperature factors than the three other monomers in that structure. The search yielded two solutions with maximum rotation function values with unpruned TDIC side-chain coordinates. Subsequent translation function searches were inconclusive.

In addition to TDIC (30), the structure of E. coli fumarase (31) also has been determined. The three-dimensional fold of these proteins is identical, and both associate to form a tetramer in a similar manner with almost identical 222 symmetry. This tetramer association almost certainly represents a common feature of the superfamily, and therefore it is reasonable to expect that the arrangement of ASL monomers would be similar. As Vm calculations indicated the presence of two monomers in the asymmetric unit of the ASL crystal, a series of searches was carried out using various TDIC dimer arrangements. Three unique monomer combinations are possible, denoted as AB, AC, or AD. Each was used as a dimeric probe in a series of rotation searches. All rotation searches gave the same pair of solutions as that found in the previous searches using TDIC subunit A alone and after PC refinement, no dimer combination gave a significantly better result than the others. All translation searches were carried out in space groups P321 and P321. The solution was apparent as the highest peak in all translation searches at 22 SD above the mean in space group P321 for the AC dimer. This allowed the correct enantiomorphic space group (P321) to be determined. The final rotation/translational solution had an R factor of 39% (Rfree, ref. 32; 41%).

Rigid body refinement was carried out using XPLOR with the individual monomers refined as distinct bodies to give an Rfactor of 34.9% and Rfree of 36.3% for data [F > 2σ(F)] between 8 and 4 Å resolution. The initial model thus produced was examined with the graphics package O (33), and residues were mutated computationally to the correct ASL sequence. Rigorous testing of potential XPLOR simulated annealing refinement protocols was carried out by examination of the behavior of Rfree with the aim of making the best use of the relatively low-resolution data. Only those protocols where both the Rfree and Rfactor decreased were considered valid. As a result of these tests, the decision was made to truncate the data to 4.2 Å to

Table 1. Crystal and diffraction data

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Space group</th>
<th>P321</th>
<th>a = b = 104.59 Å</th>
<th>c = 183.32 Å, γ = 120°</th>
<th>Resolution 4.0 Å</th>
<th>Total number of reflections 23,049</th>
<th>Unique reflections 8,659</th>
<th>Completeness* 92.3%</th>
<th>Rmerge(1) 11.6%</th>
</tr>
</thead>
</table>

*Completeness quoted for all data, 69.4% of the data was >1σ(I). †Rmerge(1) = Σ(I − μ)/Σ(I) where I are the intensity measurements for symmetry-related reflections, and μ is the mean intensity for the reflection.
Table 2. Refinement statistics

<table>
<thead>
<tr>
<th>Refinement</th>
<th>Reflections, $F &gt; 2\sigma F$</th>
<th>6,474</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>$8-4.2$ Å</td>
<td></td>
</tr>
<tr>
<td>$R_{\text{factor}}$*</td>
<td>18.79%</td>
<td></td>
</tr>
<tr>
<td>$R_{\text{free}}$†</td>
<td>29.82%</td>
<td></td>
</tr>
<tr>
<td>Reflections, all data</td>
<td>6,919</td>
<td></td>
</tr>
<tr>
<td>$R$ factor, all data</td>
<td>21.16%</td>
<td></td>
</tr>
<tr>
<td>Overall $B$ factor, Å²</td>
<td>20.5</td>
<td></td>
</tr>
<tr>
<td>rms bonds, Å</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>rms angle, °</td>
<td>1.58</td>
<td></td>
</tr>
</tbody>
</table>

* $R_{\text{factor}} = \sum (|F_o| - |F_c|)/\sum |F_o|$, where $s$ refers to a subset of data not used in refinement comprising 8% of the data and selected from thin-resolution shells dispersed at intervals through the data set.

† $R_{\text{free}} = \sum (|F_o| - |F_{\text{calc}}|)/\sum |F_o|$, where $s$ refers to a subset of data not used in refinement comprising 8% of the data and selected from thin-resolution shells dispersed at intervals through the data set.

diminish the possibility of overfitting the structure due to the incompleteness of the data in the highest resolution shell. It also proved valid to choose an appropriate weighting scheme comparable to that implemented in the program PROLSQ (34) and based on the magnitude of the difference between observed and calculated structure factors for ranges of $\sin \theta/\lambda$. This strategy resulted in lower weights for the generally stronger (yet more poorly modeled) lowest-resolution reflections, which otherwise would dominate a refinement run. Strong noncrystallographic restraints (ncs-weight = 1,500 kcal/mol$^{-1}$Å$^{-2}$) were maintained during all stages of refinement. This resulted in a final rms difference between noncrystallographic symmetry-related monomers of 0.01 Å for all atoms. Several rounds of refinement were done with a simulated heat stage to 4,000 K followed by cooling to 300 K and several rounds of minimization. These rounds were alternated with sessions of rebuilding in XPLOR. Fig. 2 is a simulated annealing $2F_o - F_c$ omit map for residues 158–161. All atoms displayed in the map were removed from the structure before a round of simulated annealing refinement. The map is contoured at 1 $\sigma$.

RESULTS AND DISCUSSION

Overall Topology. As expected from the close sequence similarities between ASL and the delta crystallins (64–72% identity) and the ASL activity demonstrated by some delta crystallins, the overall three-dimensional structures are very similar with rms deviations in dimer $C_\alpha$ positions of 0.96 Å as compared with TDIC (30) and 0.99 Å vs. duck $\delta$ II crystallin (M.A.T., K. Dole, and P.L.H., unpublished work). Duck $\delta$ II crystallin is a 2.0-Å resolution structure of an enzymatically active delta crystallin (M.A.T., K. Dole, and P.L.H., unpublished work). Although ASL is a low-resolution structure, the simulated annealing omit trials and the general good quality of the electron density maps (see Fig. 2) allowed some significant differences in the backbone atoms of the TDIC and ASL to be modeled. Each monomer is approximately 100 Å long and can be divided into three predominantly helical domains, comprising residues 1–112, 112–362, and 371–436. Fig. 3 is a schematic three-dimensional diagram of the overall topology of the monomer. ASL is catalytically active as a tetramer. Fig. 4a is a schematic diagram showing the 222 symmetric arrangement of the four monomers in the tetramer. It is drawn perpendicular to the long axis of the helix bundle in domain 2 and viewed down another of the 2-fold axes. The formation of the tetramer can be described as the arrangement of two dimeric pairs where each dimer is formed through mainly hydrophobic interactions between helices h8, h11, and h12. The numbering of the helices is consistent with that described previously for TDIC (30). Less extensive interactions between dimers result in the formation of the tetramer with helix h12 of each monomer forming the primary, innermost interface in the 20-helix bundle of the tetramer.

Fig. 2. A simulated annealing $2F_o - F_c$ omit map for residues 158–161. All atoms displayed in the map were removed from the structure before a round of simulated annealing refinement. The map is contoured at 1 $\sigma$. 
Location of Active Site. The three regions of highly conserved sequence homology are believed, because of their intolerance to mutation, to participate in the formation of a general active site among superfamily members (30, 31). These regions are shown in Fig. 1 and are mapped onto the structure of the ASL monomer in Fig. 3. The spatial relationship between conserved regions becomes more obvious in the tetrameric arrangement of monomers in the active protein (Fig. 4a) where three different monomers contribute a section of conserved sequence to form a bowl-shaped indentation at each of the four “corners” of the ASL tetramer (Fig. 4a and b). The ASL residues that contribute to the active are similar to those found in the TDIC and the E. coli fumarase structure. Further evidence that these regions comprise the active site was provided by the crystal structure of E. coli fumarase C complexed with the inhibitors tungstate and 1,2,4,5-benzenetetracarboxylic acid (31) as the inhibitor binding occurs in this putative active site pocket. Tungstate and benzenetetracarboxylate are thought to mimic fumarate in the enzyme complex and, by extension, may be analogous to the fumarate/succinate moiety in an argininosuccinate complex with ASL.

Earlier biochemical studies of ASL (9) implicated a histidine and a carboxylic acid as residues directly involved in catalysis. Indeed, His-160 is found at the putative ASL active site and corresponds to His-188 in fumarase C, which forms a hydrogen bond to the inhibitors through the oxygen atoms of the oxyanion. The role of His-160 in the catalytic mechanism has not yet been established. The catalytic mechanism requires a proton abstractor, a proton donor, and stabilizers of the carbanion intermediate either through interactions with the carbanion itself or through neutralization of the negatively charged carboxylates of the substrate. The side chain of His-160 could perform any of these functions depending on its protonation state. The structure of the fumarase C/tungstate inhibitor complex suggests that this histidine is involved in bonding to and neutralizing a negatively charged carboxylate group. However, there are few other residues in the vicinity that seem capable of proton abstraction. Presumably histidine acts as a proton abstractor to form the carbanion, and cleavage is achieved upon subsequent protonation by an acid. In the ASL structure His-160 ND1 is in a position to form potential hydrogen bonds with Glu-294 OE1, Asn-289 OD1, and Pro-290 O. The proximity of Glu-294 and its strong hydrogen bond through OE1 with the His-160 ND1 also is observed in the TDIC and fumarase structures and led Weaver et al. (31) to suggest a type of charge relay, which increases the nucleophilicity of His-160, making it a good candidate for the role of proton abstractor. The strong hydrogen bond to Glu-294 renders His-160 less likely to carry a positive charge and therefore less likely to offer stabilization of the negatively charged intermediate.

The basic residue Lys-287 is the only positively charged amino acid in this region that is strictly conserved among all superfamily members and may be responsible for neutralizing one of the three negative charges residing on the carbanion...
dicarboxylate. Indeed, mutation of the corresponding residue (Lys-326) to arginine in aspartase eliminates catalytic activity by elevating $K_m$ (39). The neutral residue Asn-289 is also strictly conserved and may be available for hydrogen bonding to substrate. The role of Met-284 also has been speculated on (5) in light of absolute conservation of this residue in sequences of superfamily members; however, there is no experimental data to suggest its possible role. The side chain of this residue is oriented into the active site at a distance of 11 Å from the His-160 side chain in the present structure. Other positively charged residues in the vicinity of the putative carboxylate binding site are Arg-113, Lys-323, and His-388, which are conserved among ASL and δ crystallin sequences but not across the superfamily. Simpson et al. (30) speculate that Glu-294, because of its proximity to His-160, is a good candidate for proton donation, whereas Weaver et al. (31) suggest for fumarase that the proton donor may be Ser-112 or Ser-140 (corresponding to Ser-112 or Arg-113 of ASL). It is possible that the candidate proton donor is found in this part of the structure; however, only two residues, Asp-115 and Thr-119, are strictly conserved across the superfamily. Whereas Thr-119 is buried and somewhat distant from the active site, Asp-115 could form a number of potential hydrogen bonds with neighboring main chain atoms of Ser-112 and Glu-116 and would appear to be important in stabilizing the turn containing residues Ser-112 and Arg-113.

**Location of Q286 and D87.** The D87G and Q286R mutations associated with the most successful complementation event observed at the ASL locus yield tetrameric proteins with very low or no enzymatic activity, respectively. Gln-286, although located in the loop comprising the third conserved region between helices h11 and h12 (see Figs. 3 and 4b) is one of the least conserved residues across the superfamily (Fig. 1). The adjacent Lys-287 has been implicated in the reaction mechanism in light of its strict conservation throughout the superfamily and likely is involved in stabilizing the carbonyan intermediate by neutralizing a negative charge either on the carbanion or on one the carboxylate groups of the nascent fumarate. The effect of the glutamine → arginine mutation is difficult to explain in the absence of the structure of a complexed substrate or inhibitor molecule. The small size difference between glutamine and arginine is unlikely to affect enzymatic activity, because the glutamine side chain extends upward from the loop toward the solvent and away from the active site. It seems more likely that the additional positive charge perturbs the mechanism in some way. Replacement of the glutamine with a positively charged arginine may enable the formation of a salt bridge between Asp-367 or possibly Glu-399 of a different monomer (see Fig. 4b) resulting in a distortion and/or rigidification of the conserved loop, the flexibility of which may be required for activity.

The D87G point mutation results in a reduction of the specific enzymatic activity to approximately 5% that of wild type. Residue Asp-87 is located at the beginning of a helix in domain one (see Figs. 3 and 4b) and is in close proximity to His-89, a residue we postulate as being important for the binding of the substrate and...
in defining the substrate specificity of the enzyme (M. Abu-Abed, M.A.T., C. Slingsby, and P.L.H., unpublished work). Side-chain density is weak for residues in this region, implying some disorder due to flexibility; however, potential hydrogen bonds could be formed between Asp-87 and main chain of Thr-90 and between Asp-87 and Glu-86 (Å). This residue therefore could play a role in stabilizing the helix by hydrogen bonding to the nitrogen of the main chain of Thr-90 (i.e., N-capping this helix). This hydrogen bonding capability would be lost upon mutation to glycine as would the negative charge, both of which may play a role in enzyme structure stabilization or in formation of the enzyme/substrate complex.

**Complementation in Hybrid Tetramer.** Whereas the structure of ASL presented here is a low-resolution structure and the side-chain conformations are less precisely defined than they would be in a high-resolution structure, it is, however, obvious that the individual mutations, D87G and Q286R, which participate in most successful complementation event observed at the ASL locus, are located at different loci in the active site cleft (Fig. 3b) and thus detrimentally affect different aspects of the catalytic mechanism. Any active site containing one or the other or both mutations would be compromised. Because a monomer will contain only one of the point mutations, there are five statistically distinguishable monomer combinations that can be formed. These unique combinations can give rise to the following tetramers in D87G:Q286R polypeptide ratios 4:0, 3:1, 2:2, 1:3, and 0:4. Examination of each of the four active sites in the protein reveals that for any one monomer (e.g., monomer A), residues 87 and 286 are found in two different active sites (active sites 1 and 2 in Fig. 5). The second half of the diagram illustrates the results for the four active sites for each of the various possible tetramer combinations. When the set of all possible hybrid tetramer combinations is considered there is a degeneracy corresponding to the binomial distribution 1:4:6:4:1 to give an overall expected enzymatic activity of ~25% of that of the wild-type protein, assuming that zero activity is observed for each of the homotetrameric mutants and that the monomers combine in a randomly distributed manner. This expected level of activity is close to the 30% of wild-type activity seen in the reconstruction of the complementation event in COS cells (13), the difference between the two values being accounted for by the 5% activity exhibited by the D87G tetramer.

Although the exact roles of Q286 and D87 are still in question, the structure of ASL has provided experimental evidence in support of the statistical regeneration of the various possible tetramer combinations. In the homotetrameric mutants and that the monomers combine in a randomly distributed manner. This expected level of activity is close to the 30% of wild-type activity seen in the reconstruction of the complementation event in COS cells (13), the difference between the two values being accounted for by the 5% activity exhibited by the D87G tetramer.

Although the exact roles of Q286 and D87 are still in question, the structure of ASL has provided experimental evidence in support of the statistical regeneration of the various possible tetramer combinations. In the homotetrameric mutants and that the monomers combine in a randomly distributed manner. This expected level of activity is close to the 30% of wild-type activity seen in the reconstruction of the complementation event in COS cells (13), the difference between the two values being accounted for by the 5% activity exhibited by the D87G tetramer.