Prediction of the structure and function of AstA and AstB, the first two enzymes of the arginine succinyltransferase pathway of arginine catabolism

Hiroki Shirai[1], Kenji Mizuguchi*

Department of Biochemistry, University of Cambridge, Old Addenbrooks Site, 80 Tennis Court Road, Cambridge CB2 1GA, UK

Received 30 September 2003; revised 5 November 2003; accepted 6 November 2003

First published online 18 November 2003

Edited by Hans Eklund

Abstract **Arginine succinyltransferase and succinylarginine dihydrolase catalyze the first two steps of arginine catabolism by the arginine succinyltransferase pathway. This route is the only major arginine catabolic pathway in *Escherichia coli* including its pathogenic strains O157 and CFT073. We have used fold recognition tools and identified novel homologies between each of these two enzymes and proteins of known three-dimensional structure: arginine succinyltransferase belongs to the acyl-CoA *N*-acyltransferase superfamily and succinylarginine dihydrolase belongs to the amidinotransferase superfamily. These findings shed light on the structures, catalytic mechanisms and evolution of diverse enzymes involved in arginine catabolism.**

Key words: Fold recognition; FUGUE; Enzyme function; Protein structure; Metabolic pathway

1. Introduction

Arginine is used by various bacteria as nitrogen, carbon and energy sources. In *Escherichia coli*, the well characterized arginine decarboxylase (ADC) pathway accounts for only 3% of arginine catabolism, while the major contribution comes from the arginine succinyltransferase (AST) pathway [1]. On the AST pathway, arginine is first succinylated by arginine succinyltransferase AST (AstA; EC 2.3.1.109) to produce N^2-succinylarginine, which is then dihydrolized by succinylarginine dihydrolase (AstB; EC 3.-.-.-), yielding N^2-succinylornithine. Three other enzymes subsequently convert it into glutamate. Unlike *Pseudomonas aeruginosa*, which has at least four pathways of arginine catabolism, *E. coli*, including its pathogenic strains O157 and CFT073, appears to have only one major pathway. Thus, further insights into the structures and catalytic mechanisms of enzymes of the AST pathway will have important clinical implications and could eventually lead to novel therapeutic agents to block this unexploited metabolic pathway. In this paper, we combined several fold and homology recognition tools and identified novel homologies involving AstA and AstB; we showed evidence that AstA belongs to the acyl-CoA *N*-acyltransferase (Nat) superfamily and AstB belongs to the amidinotransferase (AT) superfamily.

*Corresponding author. Fax: (44)-1223-766002.
E-mail address: kenji@cryst.bioc.cam.ac.uk (K. Mizuguchi).

[1] Present address: Genomics Research, Molecular Medicine Laboratories, Institute for Drug Discovery Research, Yamanouchi Pharmaceutical Co., Ltd., 21 Miyukigaoka, Tsukuba-shi, Ibaraki 305-8585, Japan.

2. Materials and methods

The sequence-structure homology recognition software FUGUE [2] was run with default parameters for the sequences of *E. coli* AstA (SwissProt ASTA_ECOLI) and AstB (SwissProt ASTB_ECOLI). These sequences were also submitted to the consensus structure prediction server, the 3D-Jury system [3].

Once the relationship between AstA and the GCN5-related *N*-acetyltransferase (GNAT) family was identified, a new alignment (Fig. 1a) was generated as follows. Homologs of AstA were collected by running PSI-BLAST [4] with default parameters (converged at iteration two and collected 39 hits). The PSI-BLAST alignment was aligned, using FUGUE, against the structural profile of the HOMSTRAD [5] entry hs1i12a. An alignment of the members of the GNAT family was taken from [11] and combined with the FUGUE alignment manually. Multiple models for the three-dimensional (3D) structure of AstA (excluding residues on two long insertions) were built with Modeller [6] based on this alignment, and the best model selected using Verify3D [7] and JOY [8].

For the AstB-AT alignment (Fig. 2a), homologs of AstB were collected by running PSI-BLAST with default parameters (converged at iteration one and collected 23 hits). The PSI-BLAST alignment was aligned, using FUGUE, against the structural profile of the HOMSTRAD entries, AT and hs1h70a. An alignment of the members of the AT family was taken from [18] and combined with the FUGUE alignment manually. Multiple models for the 3D structure of AstB were built with Modeller [6] based on this alignment, and the best model selected using Verify3D [7] and JOY [8].

All the alignments were annotated with JOY [8] and minor adjustments were made manually using Seaview [9]. 3D structures were displayed and examined with Rasmol [10] and PyMOL (http://www.pymol.org).

3. Results and discussion

3.1. AstA is a novel member of the acyl-CoA N-acyltransferase superfamily

A search for homologs of AstA using the software, FUGUE [2], detected proteins belonging to the GNAT family (the top hit being *Bacillus subtilis* yqjY with a *Z*-score of 11.6; >99% confidence). The 3D-Jury system [3], which compares models from various independent structure prediction servers and selects the most abundant predictions, identified a member of the GNAT family, yeast glucosamine 6-phosphate *N*-acetyltransferase 1 (GNA1; PDB code:1I1D) with a score of 77.7 (>90% accuracy). A reverse PSI-BLAST [4] search starting from yeast GNA1 identified AstA with a moderate *E*-value of 0.12 after iteration five. A reasonable model for the

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)01314-0

a

```
motif          (--)<-------motifC------->(--)<----motifD--->(--)<--------------motifA------------->(--)<-------motifB------>(--)
conserved      (--)---+-----+-----------(--)------+---O----(--)---*---&-O----O--O-O----+----------(--)-#---------o--+OO----(--)
1I12 No.       (--)--10--------20--------(--)--70--------80-(--)-----100-------110-------120-------(--)------140-------150--(--)
1I12           ( 6)fyirrMeegdlegVteTlkvlt(38)IVDkrtetVAATGnI(11)LCGhIediAvNskyqgggLgklLIdqLvtiGfdyg( 4)ildcdeknvkfyekCgfsnag( 7)
AAC6_CITDI     ( 1)NYQIVNIAECSNYQLEAANILT(31)FGLLINNSLVGWIGL( 5)ETWELHPLVVRPDYQNKGIGKILLKELENRAREQG(33)IKNINKHPYEFYQKNGYYIVG(22)
PAT_STRHY      ( 7)ADIRRATEADMPAVCTIVNHYI(30)LVAEVDGEVAGIAYA( 9)DWTAESTVYVSPRHQRTGLGSTLYTHLLKSLEAQG( 7)IGLPNDPSVRMHEALGYAPRG(37)
ATDA_ECOLI     ( 5)VKLRPLEREDLRYVHQLDNNAS(31)FVVECDGEKAGLVEL( 9)HRRAEFQIIISPEYQGKGLATRAAKLAMDYGFTVL( 8)VDKENEKAIHIYRKLGFSVEG(42)
SNAT_HUMAN     (34)SEFRCLTPEDAVSAFEIEREAF(25)LGWFEEGCLVAFIIG(20)HIAHLHVLAVHRAFRQQGRGPILLWRYLHHLGSQP( 5)ALMCEDALVPFYERFSFHAVG(30)
Q94521         (56)YTIELIQPEDGEAVIAMLKTFF(32)KAVNKKGEIIGVFLN(48)LILDGKILSVDTNYRGLGIAGRLTERAYEYMRENG( 5)VLCSSHYSARVMEKLGFHEVF(41)
ACUA_BASCU     (37)LTAFRPPREQHEALVDIAGLPE( 3)IIARDGRTIVGYVTY(15)DLIELGAIEVAPDYRGCAVGKTLLTVSMMDEQMEN(59)GKNVSQESIEQFDRLRFYHRL( 2)
IAAT_AZOBR     (24)RYLGDLYPAESNHLLDLQTLAK( 5)LVARRSGTVVGCGAI( 6)GYGEVKRMFVQPTARCGQIGRRLLERIEDEARAAG( 7)TGVYQATRIALYRKQGFADRG(18)

ASTA_ECOLI No. (--)-------10--------20---(--)-------70------(--)120-------130-------140-------150--(--)230-------240--------(--)
ASTA_ECOLI     ( 1)MVIRPVERSDVSALMQLASKTG(38)LEDSETGTVAGICAI(41)GSSELCTLFLDPDWRKEGNGYLLSKSRFMFMAAFR(75)VHPQTAPARAVLEKEGFRYRN(96)
Q8ZPV1         ( 1)RVIRPVEHADIAALMQLAGKTG(38)LEDSETGEVGGICAI(41)GSSELCTLFLDPEWRKEGNGYLLSKSRFMFMAAFR(75)VHPQTAPARAVLEKEGFRYRH(95)
gi22987667     ( 1)IVVRVVQRGDVDALMRLAQETG(38)MEDTDTGDVAGVCGI(41)GYAEVCSLFLSPRYRTSGVGGLLSRSRFMFLAQFR(75)THSDTIPARRMLEAEGLRYEN(88)
ASTG_PSEAE     ( 0)MIVRPVTSADLPALIELARSTG(38)LED-DAGKVVGISAI(41)GNSELCSLFLHADHRSGLNGKLLSRARFLFIAEFR(75)VHPNTEPALAMLKAEGFSYQG(94)
Q885J8         ( 0)MIVRPVRSSDLPALIELARSTG(38)LEDDD-GKVVGISAI(41)GNSELCSLFLHSDSRSGLNGRLLSKARMLFIAEFP(75)VHADTEPALTMLKGEGFSYQG(94)
Q88IE3         ( 0)MIVRPVRSSDLPALIELARSGT(38)LENDE-GLVVGISAI(41)GNSELCSLFLRSDYRSGLNGRLLSRARMLFIAEFP(75)VHKDTEPALAMLKQEGFNYQG(94)
gi22987666     ( 1)LFVRPGRLADLDALEHMARTAQ(38)LEDASSGKLMGTASI(41)GKSRLAGYYIDPSLRGDAAAHLMSRARMMYIAANR(75)PDSKALLAYDIHLEEGFETDR(94)
ASTA_PSEAE     ( 1)LVMRPAQAADLPQVQRLAADSP(38)LEDSASGELVGCSAI(41)GNSLLTSFYVQRDLVQSVYAELNSRGRLLFMASHP(75)VHPRAQITFDILMREGFETDN(88)
Q885J9         ( 1)LVMRPAQMADLAEVQRLAADSP(38)LEDSESGKLVGCSGI(41)GNSLLTSFYVVPELVGTSWSELNSRGRLLFVASHP(75)VHPRAQITFDILMREGFETDH(88)
Q88IE2         ( 1)LVMRPAQMADLNDVQRMAADSP(38)LEDSETGKLAGCSAI(41)GNSLLTSFYVAPELVNSGWAELNSRGRLLFMAAHP(75)VHPRAQITFDILMREGFETEH(88)
```

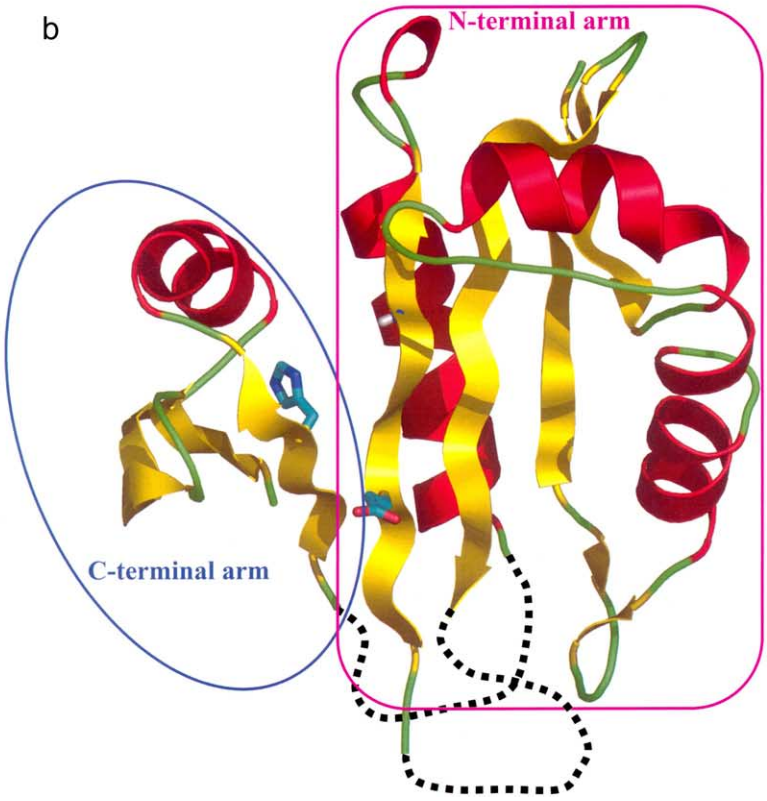
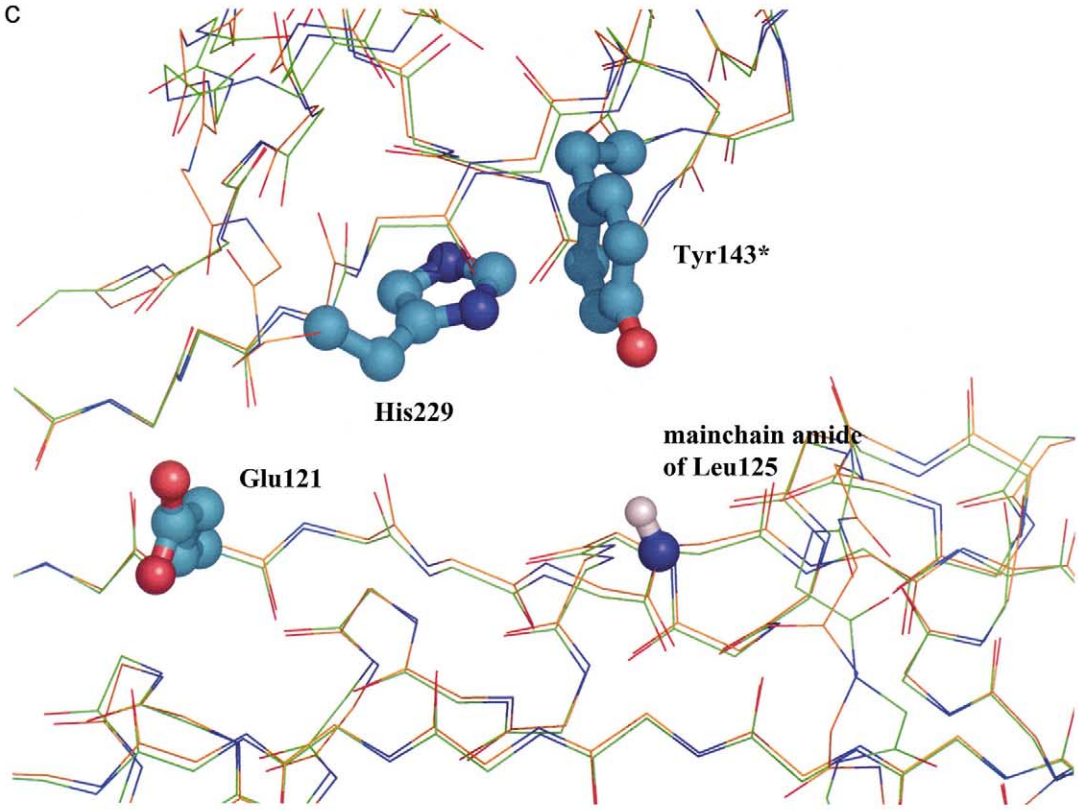


Fig. 1. (a) Multiple alignment of AstA and GNAT sequences. The first column gives the HOMSTRAD [5], SwissProt/TrEMBL [21] or NCBI gi identifiers for the sequences. Residue numbers of 1I12 and ASTA_ECOLI are indicated. The conserved blocks (motifs C, D, A and B), defined in [11] for the GNAT family, are mapped onto the alignment and only the regions that precisely correspond to these motifs are displayed. Numbers in parentheses between conserved blocks are insert lengths. The sequences given represent eight members of GNAT and 10 of AstA. The first member of GNAT, yeast glucosamine 6-phosphate *N*-acetyltransferase 1 (GNA1; PDB 1I12; HOMSTRAD hs1i12a), is of known structure and its structural features are annotated with JOY [8]. The formatting convention of JOY is as follows: red, α-helices; blue, β-strands; upper case letters, solvent inaccessible; lower case letters, solvent accessible; bold type, hydrogen bonds to mainchain amides; underlining, hydrogen bonds to mainchain carbonyls; italic, positive mainchain torsion angles φ. In column 1, α- and β-chains of hetero-oligomeric members of AstA are shown in blue and in red, respectively. The highly conserved residues of GNAT defined in [11] are shown in the second line. Seven of them are also conserved in AstA (large circles), with only one exception (small circle). An asterisk in the second line indicates the position of a general base, proposed for many members of GNAT. The residue whose mainchain amide polarizes the substrate carbonyl oxygen is marked with &. A conserved His residue of AstA is marked with #. Other residues that are highly conserved in AstA are marked with +. (b) Schematic representation of a model for the 3D structure of *E. coli* AstA. Putative catalytic residues are shown in stick representation. Backbone structure is shown as cartoons; α- or 3_{10}-helices are colored red, β-strands yellow and loops green. Dotted lines indicate the regions of long insertions between motifs D and A and between motifs A and B. Figure was drawn with the PyMOL molecular graphics system (http://www.pymol.org). (c) Superposition of the model for *E. coli* AstA and the structure of GNA1 (PDB code 1I12) near the catalytic site. Putative catalytic residues of AstA, as well as Tyr143 (marked with an asterisk) of GNA1, are shown in ball-and-stick representation. Hydrogen, nitrogen and oxygen atoms are colored white, blue and red, respectively. Mainchain carbon atoms of the AstA model are colored orange, those of GNA1 green and sidechain carbon atoms cyan. Figure is drawn with PyMOL. (d) Proposed catalytic reaction of AstA. Residue numbering is based on *E. coli* AstA.

3D structure of AstA was built (Fig. 1b,c); Verify3D [7] produced entirely positive 3D–1D compatibility scores except for the residues of long insertions, and examination with JOY [8] showed that the hydrophobic core of the protein was formed successfully.

GNAT is a large family of enzymes that catalyze the transfer of an acetyl group to a substrate, implicated in a variety of functions, from bacterial antibiotic resistance to circadian rhythms in mammals [11–13]. GNAT belongs to the Nat superfamily, which includes other acyl group transfer enzymes such as acyl-homoserinelactone synthase. These enzymes appear to share similar catalytic mechanisms [14] although the



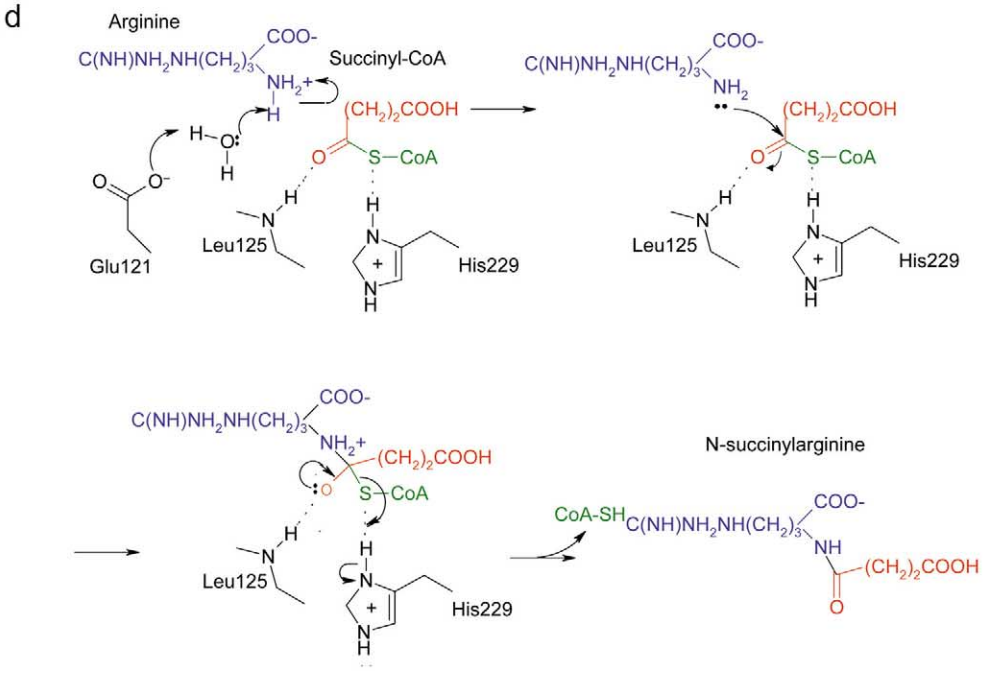


Fig. 1 (*Continued*).

types of acyl group transferred are different. Given that AstA and GNAT share conserved features (see below) but that they transfer different acyl groups, AstA can be regarded as a novel member of the Nat superfamily.

The 3D structure of the catalytic domain of Nat consists of a β-sheet surrounded by α-helices. The central β-sheet is curved and adopts a shape similar to that of the letter 'V', with the acyl-CoA-binding site located between the two arms [13] (Fig. 1b). Four sequence motifs (C, D, A and B) [11] have been defined for GNAT: motifs C, D, and A are located on the N-terminal arm and motif B on the C-terminal arm. Regions between these motifs generally adopt variable structures and sequences and allow long insertions and deletions. In the alignment of AstA and GNAT (Fig. 1a), the GNAT motifs are well conserved in AstA (see below). Although AstA has longer insertions between motifs D and A and between motifs A and B than those observed among the members of the GNAT family, these insertions are within the variable regions and would be placed on the opposite side of the putative substrate-binding site (Fig. 1b). Some members of the Nat superfamily also have inserted subdomains, not involved in the catalytic reaction.

Seven of the highly conserved residues in the GNAT motifs, as defined in [11], are also conserved in AstA (large circles in Fig. 1a) with only one exception (Tyr143 of yeast GNA1, marked with a small circle in Fig. 1a). Arg5 and Asp11 (*E. coli* AstA numbering) are perfectly conserved in AstA and the equivalent positions in GNAT are often occupied by Arg and Asp, respectively. These amino acid residues can form a salt bridge, which appears to orient the first α helix with respect to the central β sheet, as observed in the known structure of *yeast* GNA1. Other conserved residues in AstA, Glu242 and Arg/Lys142, could form an additional salt bridge, not found in GNAT, and stabilize the α helices in the N-terminal arm. The perfectly conserved Gly68 of AstA is mapped onto the fourth position of a class C β-hairpin [15] in the GNA1 structure. This position of a class C β-hairpin tends to adopt a left-handed α-helical conformation and prefers Gly, thus the local structure around this residue is likely to be conserved in AstA.

The GNAT enzymes are all thought to share similar catalytic mechanisms including a direct nucleophilic attack by a deprotonated amino group [12,13,16]. A general base is necessary to deprotonate the substrate amino group prior to acetylation and although candidates for the general base vary among members of GNAT, many are located in the position equivalent to Glu121 of AstA (*E. coli* AstA numbering, marked with an asterisk in Fig. 1a). Some homologs of AstA are known to act as hetero-oligomers, comprising α (shown in blue in column 1 of Fig. 1a) and β (shown in red in column 1 of Fig. 1a) chains. Other AstA homologs act as monomers and are shown in black. Glu121 is absolutely conserved in the monomeric enzymes and β-chains of AstA and therefore, is a strong candidate for the general base. α-chains, lacking Glu at this position, may be enzymatically inactive.

Catalytic enhancement of the chemical step may result from stabilization of a tetrahedral species and/or activation of leaving group (CoASH) departure [16]. Polarization of the thioester carbonyl group is important to enhance nucleophilic attack and to stabilize the reaction intermediate. In many members of GNAT, the backbone amide group of the position marked with '&' in Fig. 1a, fulfills this requirement by forming a hydrogen bond with the carbonyl oxygen of the acetyl group. In GNA1, the backbone amide group of the adjacent residue, Asp99, is also hydrogen-bonded to the carbonyl oxygen of the thioester. The alignment and model building suggest that a region near the putative catalytic site is well conserved between GNAT and AstA. Therefore, AstA can also enhance catalysis by forming hydrogen bonds between the succinyl-CoA carbonyl and one or more backbone amides around Leu125.

In GNAT, a Tyr residue, found in the position marked with a small circle in Fig. 1a, is sometimes proposed to serve as a catalytic acid facilitating CoASH departure [16]. In AstA, the equivalent residue (often Leu) cannot play the same role. However, there is a conserved His residue (His229, marked with # in Fig. 1a), which spatially lies close to this position (Fig. 1c). A model for the 3D structure of AstA suggests that His229 can act as the catalytic acid. There are no other conserved polar residues near the catalytic site.

Based on these observations, we propose that a catalytic reaction of *N*-succinylation by AstA proceeds in a manner similar to that for GNAT (Fig. 1d). Glu121 would act as a general base and transfer the proton of an amino group via a water molecule. His229 would function as a general acid. The reaction begins with a direct nucleophilic attack of deprotonated amine at the succinyl-CoA carbonyl (analogous to the acetyl-CoA carbonyl for GNAT). The succinyl-CoA carbonyl would be polarized via hydrogen bonds to the backbone amides around Leu125.

3.2. AstB is a novel member of the AT superfamily

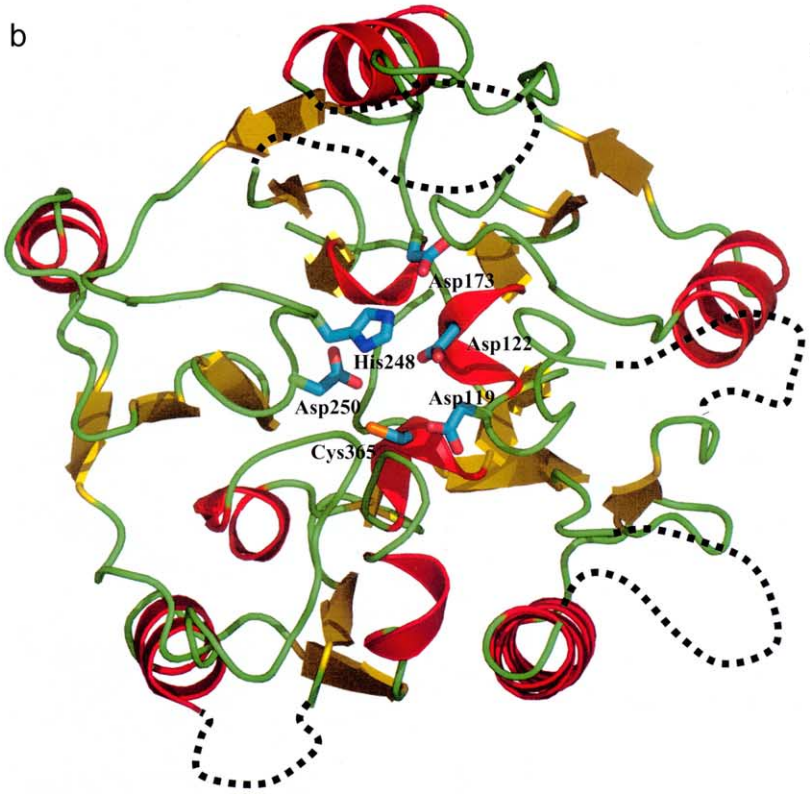
A FUGUE search for homologs of AstB produced the top hit to human AT (PDB code 1JDW) with a *Z*-score of 3.2 (> 50% confidence). The alignment included a long C-terminal overhang of the query sequence. When the C-terminal 70 residues were removed from the AstB sequence and a new search was carried out, FUGUE identified two significant hits; human AT with a *Z*-score of 6.2 (> 99% confidence) and dimethylarginine dimethylaminohydrolase (DDAH) (PDB code 1H70) with a *Z*-score of 5.02 (> 95% confidence). DDAH is a member of the AT superfamily [17,18]. 3D-Jury produced the top hit, inosamine-phosphate AT from *Streptomyces* sp. (PDB code 1BWD) with a score of 66.7 (> 90% accuracy) and all the other highest scoring predictions (originally reported by independent prediction servers) were members of the AT superfamily.

AT catalyzes the transamidination reaction of L-Arg to Gly and represents a diverse superfamily of enzymes, including peptidyl-arginine deaminase from *Porphyromonas gingivalis*, arginine deaminase (ADI) and DDAH, all of which modify the guanidino group of arginine or related molecules [17,18]. The 3D structure of AT is called an α/β propeller and is composed of (3_{10} or α-helix)–(β-strand)–(β-strand)–(α-helix)–(β-strand) units, arranged circularly around a pseudo 5-fold axis [19,20]. Almost all catalytically important residues (Asp170, Asp254, His303, Asp305, and Cys407; human AT numbering) are located in the $3_{10}/\alpha_1$–β_1–β_2 portion of each unit and interact with the substrate, which is accommodated in the cavity around the pseudo-5-fold axis. These residues are among the few that are conserved in all the members of this superfamily. The loop regions between the $3_{10}/\alpha_1$–β_1–β_2 and α_2-β_3 portions vary in sequence and structure and allow substantial insertions and deletions. By considering these unique

a

```
Catalytic   (---)--#-Oo----------(--)o-------------(--)o-o------------(--)-----------------(--)-***--o-----(--)
1JDW No.    (---)---170-------180-(--)----260-------(--)-----310-------(--)-360-------370---(--)-------410--(--)
1JDW        (164)SAMPRDILiVVgneIIE(72)DAADFIRAGrDIFA(35)HIDaTFNIIGpgiVL(39)WLSMNVLMLdekrVMVd(27)LGGGFHCWTCDV(11)
1BWD        (102)DYCPRDGLLSVgqtIIe(59)dAANVLRFGtDLLY(34)hVDsTIvPLrpgLVL(40)wIgMNLLVVrpdlAVVd(27)LGGgFhcATLDV(11)
1H70        ( 60)SVFVEDPVlCTsrCAII(37)EAGDIMmvgdhFyI(34)HLktgLAyLehnnLL(23)syAAnCIWVNe-rVIMP(27)IdGGVSSMSLrf( 0)

ASTB_ECOLI No.(--)-120-------130---(--)-----180------(--)250-------260--(--)----310-------320(--)---------370(--)
ASTB_ECOLI  (116)SADTLDGKVHLTVANLN(39)DEGAANHNRLGGHY(61)HNDVIAVSNRQVLFC(41)LFNSQLLSRDDGSMMLV(38)NGGGPACLRLRV(77)
Q8ZPU9      (116)SADALDGKVHLTVANLN(39)DEGAANHNRLGGEY(61)HNDVIAVSNRQVLFC(41)LFNSQLLSRNDGLMLLV(38)NGGGPACLRLRV(77)
Q88EI5      (116)SADTADGRVHFTAANLN(39)DEGAANHTRFCRAY(60)HNDVISVGNGEVLFY(41)LFNSQLLSREDGSMLLV(38)NGGGPACLRLRV(73)
ASTB_PSEAE  (116)SADTADGRVHFTAANLN(39)DEGAANHTRFCKDY(60)HNDVISVGNGEVLFH(41)LFNSQLLSKADGSMLLV(38)NGGGPACLRLRV(72)
Q885J6      (116)SADTADGRVHFTAANLN(39)DEGAANHTRFCQDY(60)HNDVIAVGNGEVLFY(41)LFNSQLLSRPDGSMLLI(38)NGGGPACLRLRV(72)
gi22987669  (116)SADTNDGRVHFTPANLC(39)DEGAANHTRFCEEY(60)HNDVIAVGNRNTLFC(41)LFNSQLLTRPDGKQVLV(38)NGGGPACLRLRV(71)
```



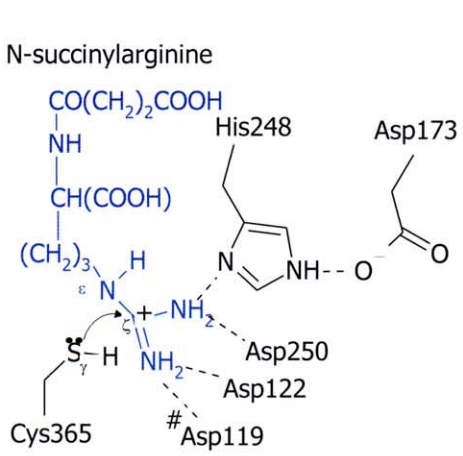


Fig. 2. (a) Multiple alignment of AstB and AT sequences. Alignment roughly covers the $3_{10}/\alpha_1$–β_1–β_2 portions of the AT structure. The first column gives the PDB, SwissProt/TrEMBL or NCBI gi identifiers for the sequences. Residue numbers of 1JDW and ASTB_ECOLI are indicated. Numbers in parentheses between conserved blocks are insert lengths. The sequences given represent two members of AT, DDAH (a distant homolog of AT) and six members of AstB. AT and DDAH are of known structure and their structural features are annotated with JOY. Shown at the top are: substrate binding and catalytic residues, common to both AT and DDAH (small circles), an additional substrate binding residue of DDAH (large circle) and an additional substrate binding residue of AT (underlined circle). The conserved Asp residue, marked with #, of AstB would be spatially close to these functional residues and is also likely to be involved in the substrate binding. The structural hallmark of AT, the three consecutive Gly residues near the catalytic Cys residue, is marked with asterisks. (b) Schematic representation of a model for the 3D structure of *E. coli* AstB. Backbone structure is shown as cartoons; α- or 3_{10}-helices are colored red, β-strand yellow and loops green. Black dotted lines indicate long insertion loops. Putative catalytic residues are shown in stick representation; sulfur, nitrogen, oxygen and carbon atoms are colored orange, blue, red and cyan, respectively. Figure is drawn with PyMOL. (c) Schematic drawing of substrate recognition and a nucleophilic attack of AstB. Residue numbering is based on *E. coli* AstB.

features of this fold and producing structure-aided alignments, the membership of DDAH to the AT superfamily had been predicted prior to its structure determination [17]. The alignment of AT and AstB (Fig. 2a) indicates precisely the same situation; AstB also catalyzes the modification of a guanidino group and all the functional residues, identified for the AT superfamily, are highly conserved in AstB and its homologs.

To summarize, the following evidence strongly suggests that AstB and AT are indeed related: (1) statistically significant hits from various independent structure prediction servers; (2) the conserved cluster of functional residues, highly unique in the AT superfamily; (3) conservation of structurally important residues, such as buried hydrophobic residues and buried hydrophilic residues hydrogen bonded with mainchain groups; and (4) a structural hallmark of three consecutive buried Gly residues that precede the catalytic Cys residue (Fig. 2a).

Proposed catalytic mechanisms for the AT superfamily resemble that of cysteine proteases and involve a nucleophilic attack by the sulfur atom (Sγ) of a Cys residue on the (dimethyl) guanidino carbon atom (Cζ) of the substrate [18]. The catalytic triad, consisting of Cys407, His303 and Asp254 (human AT numbering), is conserved in AT, DDAH and AstB. Therefore, it is likely that AstB uses essentially the same mechanism for bond cleavage in the substrate *N*-succinyl arginine.

Differences in the substrate-binding residues, however, may define which bonds are cleaved; DDAH cleaves the bond between Nη and Cζ, while AstB and AT cleave the bond between Nε and Cζ. Since AstB produces two ammonia and one carbon dioxide molecules without generating urea, the enzyme is likely to cleave at least one additional bond prior to this reaction. Asp170 and Asp305 of AT are involved in the binding of the substrate guanidino group. Both are conserved in AstB but Asp305 is not conserved in DDAH, which instead uses another Asp residue for substrate recognition (marked with a large circle in Fig. 2a). AstB has an additional conserved acidic residue, Asp119 (marked with a hash), which may determine AstB specific interactions (Fig. 2b,c). To elucidate exact catalytic mechanisms, experimental verification will be required.

3.3. Evolution of arginine catabolism pathways

To catabolize the small but energy-rich amino acid, arginine, nature has produced a diverse set of pathways, including the arginase and the ADI pathways. Every arginine catabolism pathway requires the modification of the guanidino group of arginine. Despite the diverse sequences of enzymes that catalyze the modification of guanidino groups, these proteins may have evolved from a relatively small number of common ancestors and thus exhibit a limited repertoire of structures and catalytic mechanisms. For example, agmatine ureohydrolase of the ADC pathway belongs to the arginase superfamily and we have shown here that AstB of the AST pathway and ADI both belong to the AT superfamily.

In enzyme evolution, two different scenarios can be imagined; a protein with a specific catalytic activity can gain a new substrate affinity, or an enzyme with a given substrate affinity can change its catalytic mechanism. GNAT is a superfamily of enzymes retaining the catalytic activity of acyl transfer, while members of the AT superfamily retain the affinity to arginine. To create a new pathway of arginine catabolism, the first enzyme, AstA, has evolved from an ancestor by keeping its catalytic mechanism while the second enzyme, AstB, has evolved by keeping its ancestor's substrate specificity.

3.5. Conclusions

In this paper, we used fold recognition methods and predicted the 3D structures and catalytic mechanisms of AstA and AstB, the first two enzymes of the AST pathway of arginine catabolism. These predictions should help design new molecules to block this important pathway in *E. coli* O157 and other pathogenic bacteria.

Acknowledgements: We thank Lucy Stebbings for comments on the manuscript. K.M. is a Wellcome Trust Research Career Development Fellow.

References

[1] Schneider, B.L., Kiupakis, A.K. and Reitzer, L.J. (1998) J. Bacteriol. 180, 4278–4286.
[2] Shi, J., Blundell, T.L. and Mizuguchi, K. (2001) J. Mol. Biol. 310, 243–257.
[3] Ginalski, K., Elofsson, A., Fischer, D. and Rychlewski, L. (2003) Bioinformatics 19, 1015–1018.
[4] Altschule, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[5] Mizuguchi, K., Dean, C.M., Blundell, T. and Overington, JP. (1998) Protein Sci. 7, 2469–2471.
[6] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[7] Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Nature 356, 83–85.
[8] Mizuguchi, K., Dean, C.M., Johnson, M.S., Blundell, T. and Overington, J.P. (1998) Bioinformatics 14, 617–623.
[9] Galtier, N., Gouy, M. and Gautier, C. (1996) Comput. Appl. Biosci. 12, 543–548.
[10] Bernstein, H.J. (2000) Trends Biochem. Sci. 9, 453–455.
[11] Neuwald, A.F. and Landsman, D. (1997) Trends Biochem. Sci. 22, 154–155.
[12] Marmorstein, R. (2001) J. Mol. Biol. 311, 433–444.
[13] He, H., Ding, Y., Bartlam, M., Sun, F., Le, Y., Qin, X., Tang, H., Zhang, R., Joachimiak, A., Liu, J., Zhao, N. and Rao, Z. (2003) J. Mol. Biol. 325, 1019–1030.
[14] Watson, W.T., Minogue, T.D., Val, D., Bodman, S. and Churchill, M. (2002) Mol. Cell 9, 685–694.
[15] Sibanda, B.L., Blundell, T.L. and Thornton, J.M. (1989) J. Mol. Biol. 206, 759–777.
[16] Wolf, E., Angelis, J.D., Khalil, E.M., Cole, P.A. and Burley, S.K. (2002) J. Mol. Biol. 317, 215–224.
[17] Shirai, H., Blundell, T.L. and Mizuguchi, K. (2001) Trends Biochem. Sci. 26, 465–468.
[18] Murray-Rust, J., Leiper, J., McAlister, M., Phelan, J., Tilley, S., Maria, J., Vallance, P. and McDonald, N. (2001) Nat. Struct. Biol. 8, 679–683.
[19] Humm, A., Fritsche, E., Steinbacher, S. and Huber, R. (1997) EMBO J. 16, 3373–3385.
[20] Fritsche, E., Bergner, A., Humm, A., Piepersberg, W. and Huber, R. (1998) Biochemistry 37, 17664–17672.
[21] Bairoch, A. and Apweiler, R. (1996) Nucleic Acids Res. 24, 21–25.