On the Evolution of Arginases and Related Enzymes

C.A. Ouzounis,¹ N.C. Kyrpides²

¹ European Molecular Biology Laboratory Meyerhofstrasse 1, Postfach 10.2209, D-69117 Heidelberg, Germany ² Institute of Molecular Biology and Biotechnology, Heraklion, Greece

Received: 2 August 1993 / Revised: 1 December 1993

Abstract. Sequence analysis of the arginase/agmatine ureohydrolase family, important enzymes in arginine/agmatine metabolism and the urea cycle, reveals the similarity of arginases to formiminoglutamate hydrolase (*hutG*) in *Klebsiella aerogenes* and to a previously unidentified open reading frame adjacent to the HMf locus of the archaebacterium *Methanothermus fervidus*. The gene structure and distribution of these homologous proteins across primary kingdoms suggest that this family is another example of a primordial enzyme possibly present in the universal common ancestor and that can be used as phylogenetic marker.

Key words: Arginase — Agmatine ureohydrolase — Formiminoglutamate hydrolase — Protein evolution — Evolution of metabolism — Archaebacteria — *Methanothermus fervidus* — Phylogeny — Genome evolution

Introduction

Arginine catabolism is a process of interest both in the physiology and the evolution of metabolism because of the multiplicity of pathways through which it is accomplished (Cunin et al. 1986) and the relationships of the enzymes that participate in various catalytic steps. Arginases (EC 3.5.3.1) are urea-cycle enzymes catalyzing the hydrolysis of arginine to urea and ornithine and are known to be present in various eubacterial groups (Cunin et al. 1986). Arginases are also known to occur in a variety of eukaryotic taxa, including fungi

Correspondence to: C.A. Ouzounis

(Davis 1986). Arginase sequences from yeast (Sumrada and Cooper 1984), human (Haragushi et al. 1987), rat (Kawamoto et al. 1987), and *Xenopus* (Xu et al. 1993) are now available. The sequence of an arginase gene in an *Agrobacterium* Ti plasmid (Schrell et al. 1989) suggests a common origin for these enzymes (\sim 30% sequence identity).

Journal

of Molecular Evolution

© Springer-Verlag New York Inc. 1994

The sequence of the *speB* gene in *Escherichia coli* encoding agmatine ureohydrolase (EC 3.5.3.11) was shown to be homologous to arginases (Szumanski and Boyle 1990). Agmatine ureohydrolase catalyzes the second step in the arginine decarboxylase pathway, the hydrolysis of agmatine to urea and putrescine (Abdelal 1979).

Therefore, the wide distribution of this protein family in prokaryotes and eukaryotes, as well as the diversity of their function, suggests the universal occurrence of its members in all primary kingdoms. In a continued quest for phylogenetic markers (Ouzounis and Sander 1992) and the reconstruction of the genome composition of the universal common ancestor (Woese 1987), we have now identified two more homologs of this family in *Klebsiella aerogenes* and the methanogenic archaebacterium *Methanothermus fervidus*.

Materials and Methods

Sequences were primarily collected from the SWISS-PROT protein sequence database (Bairoch and Boeckmann 1991). In addition, DNA sequence databases were searched for the identification of recently deposited sequences not yet present in protein sequence databases (Bork et al. 1992). Sequence identifiers are reported as from SWISS-PROT (Fig. 1). Database searches were performed using the programs

9 9 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	144414666 14441466 144114669 144114699966622 1441146699966652	8 4413 8 4414 8 44146 8 44146666666666666666666666666666666666	8 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	
	DEGNANNE -	N H00000000 I E<0000000	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	arginase family. <i>Numbering</i> corresponds to the actual sequence positions of the proteins; <i>each line</i> contains 95 positions. The esidues are esidues at a particular position, on a similarity basis. <i>Boxed</i> residues correspond to the consensus. The 13 invariant residues are quences are given provisional names, as they are not yet contained in SWISS-PROT (marked accordingly). Protein sequence AB, clavulanic acid biosynthesis enzyme; HUTG, formiminoglutamate hydrolase; YHMF, ORF in the Hmf locus. Species name <i>time tamefaciens</i> , plasmid PTIC58; STRCL, <i>Steptomyces clavuligerus</i> ; KLEAE, <i>Klebsiella aerogenes</i> ; METFE, <i>Methanother</i> -courtesy of Peter Rice, EMBL).
OPORTERNA -	王文子(11) (1) (1) (1) (1) (1) (1) (1) (1) (1)		1 : 1 : 24 : 24 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :	re contains 95 positions. ' The 13 invariant residues cordingly). Protein seque the Hmf locus. Species na thes; METFE, Methanot enes; METFE, Methanot
1 - 1 200×4000	0000000000000 000000000000000000000000	A AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	· · · · · · · · · · · · · · · · · · ·	osit res pec Aetl
	· NAHPHNHU ·	ДО1 I I I I I I I I I I I I I I I I I I I	· · · Þ Þ · · · · ·	95 po Prot FE, A
	XXEPAAAHJX ' XAXEEEEA	HADAAHHAA I		ns 9 Noar Iocu Iocu
成:K>>>減: 、 第:1KKK口: 、	HEMHHHHHH	CHXAAXB IZ I		ain 13 in 13 in 11 mf 10 mf
				contains ae 13 invi ardingly). e Hmf lo tes; MET
ច្ច>១៧២២ · · ·	* HHHHHHHH * HHHHHHHH H - SALLER - SALLER			ine r The scco gen
20:244000;;;;		LETER T	XX · XXAFR · · XR · XXXJ> · ·	proteins; each line to the consensus. Th PROT (marked acco ; YHMF, ORF in th , Klebsiella aerogen ,
י ו גואי בי בי אים	EXESSEZ :		AA:JEXMA: .	ea ark OF
2021 · · · 33 · ·		AUNDERENSE R HATEXEN	AAXFFFJFW I VHOOCOLX I	ins; (m MF, siel
жы <mark>лалала</mark> 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			L ROCCOSPACI	the
· · · · · · · · · · · · · · · · · · ·	A A E O C C O O O C	AUZUUAZAI		d to PR d to E, k
			MADAAAAAA	f th pone ISS Cola: EAI
	EHJHHHYFE		AUDAPAACOA	ns c SW KL
> HAAAAHA : .	HOODYDXKA		HEANNAGAR ·	ition in in in in in
ALLELLIN ALLELLIN ALLELLIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLELIN ALLENN ALE	A LALANA ANA ANA ANA ANA ANA ANA ANA ANA	CACAMAMAA	AAAFFFFFFA '	pos ned ger
AQUHAXXO · ·			ANALHENDR (ence positions of the residues correspond ontained in SWISS-J oglutamate hydrolase davuligerus; KLEAE,
		HEXXXXX		sequence positions of the oxed residues correspond yet contained in SWJSS-F yet containate hydrolase iminoglutamate hydrolase ces clavuligerus; KLEAE,
HERENES,	A A COCOCOCO A A COCOCOCOCOCO A A A A A A A A A A A A A A A A A A A			l seque Boxed t yet c mimin yces cl
040>>>>0	PHHAAAASA '	NEMIJIJIH I	OHNXXXHII I	actual aasis. <i>E</i> aasis. <i>E</i> actual actual actual actual
AUTORNU AUTORNU AUTORNU AUTORNU	* Щининии мооооооооооооооооооооооооооооооооооо	TONTANNA		a actual sequence positio basis. <i>Boxed</i> residues con are not yet contained in 'G, formiminoglutamate' teptomyces clavuligerus;
		RUBYYYXX ·	ANAXXNXXX -	, S ₂
		· П ж ж д ж ж о ·	医尿管管室 (1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	onds to the actu similarity basis. s, as they are no s, as thUTG, fo yme; HUTG, fo STRCL, Steptom
ACCNNNNO · ·		IDEGGGGRA I		ond es, a STI
		I ACONONA I		ily. Numbering corresponds to the particular position, on a similarity b given provisional names, as they a ic acid biosynthesis enzyme; HUT icras, plasmid PTIC58; STRCL, Ste Peter Rice, EMBL).
	SZUGBAZUD	12)))))))))	L SURVERSE	con al n TTC ITC
	AXXXXXORX '	* KKKKKKKKK F0000000 -	· · · · · · · · · · · · · · · · · · ·	ily. Numbering con particular position, given provisional nic acid biosynthesi riens, plasmid PTIC Peter Rice, EMBL)
<u> </u>	JJZAAXXX '	ннрададна о россссоос	о <mark>Соссосу</mark> жжи	r po ovivo smi
47 X X X X X X X X X X X X X X X X X X X	RUQAAFUJA KOBQGAXBO (HHHHHHMMM H	DDAXXXMDK -	Vumbe sular p n prov id bio plasm Rice,
SYNKKBKB: -	HHMMVACNA VAVVVVV V	AVANUHUMAN AVANUHUMAN	· CEPHKKAGX	nily. A partic given given ric aci ciens, Peter
N N N N N N N N N N N N N N N N N N N	ANNARANNA ANNARANNA	HAMAAAAAA '	れた間間間回回回 いいれる思想である。	arginase family. esidues at a parti quences are give AB, clavulanic a <i>rium tumefaciens</i> (courtesy of Pete (courtesy of Pete
OOK · · · NK · ·	L LLALLANA AMEHOOMAX	0.0.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1	· HHTTTHHTTT	arginase fam residues at a equences are .AB, clavular .AB, clavular <i>trium tumefac</i> (courtesy of
2001, 622, 7 701, 192, 192, 193, 193, 193, 193, 193, 193, 193, 193	SHONNFFFE		エンほししれしはし、メヘダメメファしみ。	arginase residues a equences AB, clavv rium tume (courtesy
あある こうちょう うちまん こうかん うち かん かん うち うちょう ひろう うう	A PARKANAN A	IIIAAAUAH I	жындараан жылараан	arg resi "AB <i>riur</i> (cou
110:::02:: 220:::02::	AXF7>>>ゴコ ·	1 13333447 1	0000×××⊐₽-1 ×	the ing ss so so of the officer offic
「「「「」」と、「」、「」、「」、「」、「」、「」、「」、「」、「」、「」、「」、「」、「」、	0001140407 060114008	、、、SSS開発を用 、 、、「FFFFF111FF ま	년7년5년616262 · 1711111111111111111111111111111111111	of ase
ι⊱ι κ ζιιιικζι ι ιαβ∟ιιικΟι ι		1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	HAANNAKAA .	bers pron drol Ag
INTINITY INTI	I I RUNAD I		>48>>0>>4	ntly Strej T5, Strej
0<01.01.01.01.01.01 A.M.M.M.1.01.01.01.01.01.01.01.01.01.01.01.01.0	· · · >>>>0 · · ·	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	PHZPPPPHZ P	y P. GR
424	I I I MHZUZ I I)) U U U K G (H)	AAAAAGGGGGG AAAAAGGGGGG AAAAAGGGGGGGGGG	nin fre & au ine ;; A ed b
>>	· · · · · · · · × · · · · · · · · · · ·	第3日、100日の10日、10日、10日、10日、10日、10日、10日、10日、10日、10日、	* AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	the nost mat mat vivevi: srate
NO:	1,1,1,1,1,0,0,0,1,1,1 1,1,1,1,1,0,0,0,1,1,1	AHYDJMMFK - Komxxx000 -	A RHAHRAKONS A	for ix n Xer Xer is la gena
		HEREFURA		tent The PEB vopu
		PER>>>=>>=>>=>>= >=======================		gnn l fro sk. ⁷ sk. ⁷ Sl Xer re w
царияния Даранияния Даранияния				, ali atec <i>teric</i> nase Ta, Jgu
E	NEGRAGO			he f
24	FK< <hhh< td=""><td></td><td><u>⊨ณ⊧งงงงง</u>ษ ง</td><td>Sequence alignment for the nine members of the is calculated from six most frequently occurring ith an <i>asterisk</i>. The <i>Xenopus</i> and <i>Streptomyces</i> s RGI, arginase; SPEB, agmatine ureohydrolase; CI arginase; SPEB, agmatine ureohydrolase; CI arginase; The figure was generated by PRETTYPLOT das. The figure was generated by PRETTYPLOT</td></hhh<>		<u>⊨ณ⊧งงงงง</u> ษ ง	Sequence alignment for the nine members of the is calculated from six most frequently occurring ith an <i>asterisk</i> . The <i>Xenopus</i> and <i>Streptomyces</i> s RGI, arginase; SPEB, agmatine ureohydrolase; CI arginase; SPEB, agmatine ureohydrolase; CI arginase; The figure was generated by PRETTYPLOT das. The figure was generated by PRETTYPLOT
strater strate	sus sus sus sus sus sus sus sus sus sus	stratte stratte sae sae	oli frecli man man sst rt5 eae sae	N A N
Speb_Bcoli Clab/Strc1 Yhmf_Metfe Argi_Human Argi_Aat Argi_Yaeit Argi_Yeast Argi_Yeast Argi_Agrt5 Hutg_Klaea Consensus	Speb_Ecoli Speb_Ecoli Yhm Ketfe Argi_Human Argi_Human Argi_Kenla Argi_Yeast Argi_Yeast Argi_Agrt5 Hutg_Klease Hutg_Klease	Speb_Ecoli Yhmf_Mcffe Yrmf_Mcffe Argi_Muman Argi_Mara Argi_Yaat Argi_Yaat Argi_Yaat Argi_Yaat Argi_Yaat Argi_Yaat Consensus	Speb_Ecoli Clab/Strcl Yhm Ere Argi_Human Argi_Rat Argi_Rat Argi/Xenla Argi_Agrt5 Hutg_Kleae Consensus	Fig. 1. consens marked names: convent <i>mus fer</i>
Speb_] Yhmf_J Argi_j Argi_ Hutg_ Hutg_ Cons	Spel Spel Argi Argi Argi Hutg	Spel Argi Argi Hutç	C Argan	Fig. mar mar mus mus

FASTA (Pearson and Lipman 1988) and BLAST (Altschul et al. 1990). Multiple sequence alignment was performed using PILEUP (Devereux et al. 1984) and CLUSTAL (Higgins and Sharp 1989). Profile searches with the family (Gribskov et al. 1987) revealed no remote members.

Results and Discussion

Sequence database searches with the arginase family as a probe reveal the homology of agmatine ureohydrolase to arginases, as was noted before (Szumanski and Boyle 1990). These enzymes hydrolyze their substrates (arginine, agmatine—respectively) and both produce urea. The level of similarity suggests a common evolutionary origin and subsequent divergence, resulting in different substrate specificities. The presence of two homologous enzymes, arginase and agmatine ureohydrolase, in two different arginine catabolism pathways, the arginase and the arginine decarboxylase pathways, respectively (Abdelal 1979), raises questions on the origins and evolution of these pathways.

Interestingly, a similarity between this protein family and formiminoglutamate hydrolase (*hutG* gene) from *Klebsiella aerogenes* (Schwacha and Bender 1990) was also identified. Formiminoglutamate hydrolase catalyzes the hydrolysis of formiminoglutamate to glutamate and formamide. This similarity, not reported before, between this protein and any member of the family is around 24% sequence identity over its full length, suggesting a common evolutionary origin.

A second new member was also identified by these searches, a protein coded by an open reading frame (ORF) in the HMf locus of the archaebacterium *Methanothermus fervidus* (Sandman et al. 1990). *Methanothermus fervidus* is a distinct methanogenic species, being the member of a single family, Methanothermaceae (Jones et al. 1987). The HMf gene codes for a DNA-binding protein with similarity to eukaryotic histones (Sandman et al. 1990). The open reading frame present in the opposite strand codes for a protein of 285 amino acids, shown here to be homologous to arginases. Given the functional diversity within the family, we cannot assign accurately a function by sequence similarity, although this protein is most similar to agmatine ureohydrolase.

A sequence from *Streptomyces clavuligerus* recently submitted to the database (Aidoo et al. 1993) is highly similar to agmatine ureohydrolase (*speB*) of *Escherichia coli*. At present, the function of this gene is not known, although it is possible that it is involved in clavulanic acid biosynthesis (according to authors). Based on the high sequence similarity, we predict that this gene codes for an agmatine ureohydrolase or a related enzyme in *Streptomyces clavuligerus*.

A multiple alignment has been constructed for all members of this expanded family (Fig. 1). The invariant residues are now 13, after the identification of three new members, and are distributed throughout the pro-

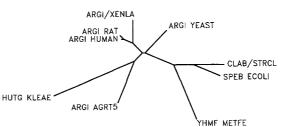


Fig. 2. Unrooted tree of the family. The tree was calculated using CLUSTAL (Higgins and Sharp 1989) and drawn by DRAWTREE (Felsenstein 1988; courtesy of Joseph Felsenstein). Due to the variety of species and functions involved, the tree is intended to represent sequence relationships and does not imply a phylogenetic tree. Naming conventions as in Fig. 1.

tein. These residues include four aspartates, three glycines, two histidines, and one arginine, which might be involved in binding and catalysis. The updated multiple alignment might provide a basis for further experimentation and functional analysis of this enzyme family.

An unrooted tree was constructed on the basis of pairwise sequence similarity as a distance measure. It is clear that eukaryotic arginases cluster together, and form one cluster (Fig. 2), with the eubacterial arginase and formiminoglutamate hydrolase clearly distinct from the agmatine ureohydrolase pair and the archaebacterial enzyme.

Finally, gene organization can be the basis for inferences of homology at a higher-than-molecular level. In the Agrobacterium Ti plasmid, arginase is adjacent to two genes for nopaline oxidase, preceding a protein of predicted molecular weight of 40 kD and ornithine decarboxylase (Schrell et al. 1989). In Escherichia coli, agmatine ureohydrolase is downstream of arginine decarboxylase (Moore and Boyle 1990). Finally, in Klebsiella aerogenes, formiminoglutamate hydrolase is upstream of a possible repressor of the hut cluster, and urocanase (Schwacha and Bender 1990). All these clusters are involved in arginine catabolism and related processes. However, no clear pattern of gene organization emerges. For the other members, no information on gene structure is available.

There has been a revival of the reconstruction of the genome of the universal common ancestor, on the basis of homologies of the extant molecular and biological species (Woese 1987). Here we show another case where a family is present throughout the primary kingdoms of life. We can therefore presume that a member of this family was already part of the genome of the universal common ancestor, before the major divergence into archaebacteria, eubacteria, and eukarya (Woese 1987).

References

Abdelal AT (1979) Arginine catabolism by microorganisms. Annu Rev Microbiol 33:139-168

- Aidoo KA, Wong A, Alexander DC, Rittammer RA, Jensen SE (1993) Cloning and sequencing of a gene from *Streptomyces clavuligerus* with possible involvement in clavulanic acid biosynthesis. EMBL databank, accession number L15470 (unpublished)
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
- Bairoch A, Boeckmann B (1991) The SWISS-PROT protein sequence databank. Nucleic Acids Res 19:2247–2249
- Bork P, Ouzounis C, Sander C, Scharf M, Schneider R, Sonnhammer E (1992) Comprehensive sequence analysis of the 182 predicted open reading frames of yeast chromosome III. Protein Sci 1:1677–1690
- Cunin R, Glansdorff N, Piérard A, Stalon V (1986) Biosynthesis and metabolism of arginine in bacteria. Microbiol Rev 50:314-352
- Davis RH (1986) Compartmental and regulatory mechanisms in the arginine pathways of *Neurospora crassa* and *Saccharomyces cerevisiae*. Microbiol Rev 50:280–313
- Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395
- Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565
- Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci USA 84:4355–4358
- Haraguchi Y, Takiguchi M, Amaya Y, Kawamoto S, Matsuda I, Mori M (1987) Molecular cloning and nucleotide sequence of cDNA for human liver arginase. Proc Natl Acad Sci USA 84:412–415
- Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5:151-153
- Jones WJ, Nagle DP Jr, Whitman WB (1987) Methanogens and the diversity of archaebacteria. Microbiol Rev 51:135–177

- Kawamoto S, Amaya Y, Murakami K, Tokunaga F, Iwanaga S, Kobayashi K, Saheki T, Kimura S, Mori M (1987) Complete nucleotide sequence of cDNA and deduced amino acid sequence of rat liver arginase. J Biol Chem 262:6280–6283
- Moore RC, Boyle SM (1990) Nucleotide sequence and analysis of the speA gene encoding biosynthetic arginine decarboxylase in Escherichia coli. J Bacteriol 172:4631–4640
- Ouzounis C, Sander C (1992) TFIIB, an evolutionary link between transcription machineries of archaebacteria and eukaryotes. Cell 71:189–190
- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:3338–3342
- Sandman K, Krzycki JA, Dobrinski B, Lurz R, Reeve JN (1990) HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proc Natl Acad Sci USA 87:5788–5791
- Schrell A, Alt-Moerbe J, Lanz T, Schroeder J (1989) Arginase of Agrobacterium Ti plasmid C58. DNA sequence, properties, and comparison with eucaryotic enzymes. Eur J Biochem 184:635–641
- Schwacha A, Bender RA (1990) Nucleotide sequence of the gene encoding the repressor for the histidine utilization genes of *Klebsiella* aerogenes. J Bacteriol 172:5477–5481
- Sumrada RA, Cooper TG (1984) Nucleotide sequence of the Saccharomyces cerevisiae arginase gene (CAR1) and its transcription under various physiological conditions. J Bacteriol 160:1078–1087
- Szumanski MBW, Boyle SM (1990) Analysis and sequence of the *speB* gene encoding agmatine ureohydrolase, a putrescine biosynthetic enzyme in *Escherichia coli*. J Bacteriol 172:538–547
- Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221-271
- Xu Q, Baker BS, Tata JR (1993) Developmental and hormonal regulation of the *Xenopus* liver-type arginase gene. Eur J Biochem 211:891–898