ELSEVIER

FEBS Letters 339 (1994) 101–107

FEBS 13642

FEBS LETTERS

# The amino acid sequences of human and pig L-arginine:glycine amidinotransferase

Andreas Humm, Robert Huber, Karlheinz Mann*

*Max-Planck-Institut für Biochemie, Am Klopferspitz, D-82152 Martinsried, Germany*

Received 21 December 1993

**Abstract**

We have isolated and sequenced the L-arginine:glycine amidinotransferase of pig kidney mitochondria. Due to endogenous proteolysis, the purified molecules showed some heterogeneity at the N terminus. The longest form recovered had 386 amino acids. Part of the pig amidinotransferase sequence information was used to isolate cDNA clones coding for the human enzyme. The deduced amino acid of the human amidinotransferase was 37 amino acids longer due to the presence of a signal sequence. The mature proteins were 94% identical to each other and 36% identical to the sequences of bacterial L-arginine:inosamine phosphate amidinotransferases.

*Key words:* L-Arginine:glycine amidinotransferase; Amino acid sequence; Creatine

## 1. Introduction

L-Arginine:glycine amidinotransferase (EC 2.1.4.1) catalyzes the transfer of a guanido group from arginine to glycine. The resultant guanidoacetic acid is the immediate precursor of creatine [1]. Creatine and its phosphorylated form play an important role in the energy metabolism of muscle and nerve tissues, acting as a dynamic reservoir of high-energy phosphate, which buffers the rapid fluctuations of the ADP/ATP ratio during muscle and nerve action [2]. Interestingly, creatine is not synthesized in those tissues that have the largest pools of creatine and creatine phosphate but must be taken up from the blood. The major sites of creatine biosynthesis are pancreas, kidneys and liver, where the amidinotransferase seems to be located in the mitochondria of cells [3–5]. It has been suggested that amidinotransferase activity in these tissues is regulated in a number of ways including induction by growth hormone and thyroxine [6], inhibition of the enzyme by ornithine [7] and repression of its synthesis by creatine [8]. Transfer of guanido groups is also known in *Streptomyces* bacteria where L-arginine:inosamine phosphate amidinotransferases catalyze two reactions in streptomycin biosynthesis [1]. The amino acid sequences of the bacterial enzymes have been derived by cDNA sequencing [9–11]. In contrast, no sequence has been reported for eukaryatic amidinotransferases. The present report fills this gap.

## 2. Materials and methods

### 2.1. Isolation of porcine amidinotransferase

Fresh porcine kidney cortex was cut in small pieces and suspended in ice-cold homogenization buffer (250 mM saccharose, 10 mM triethanolamine, 10 mM EDTA, 10 mM glycine, 5 mM DTT, 1 mM dipyridyl, 1 mM 1,10-phenanthroline, 1 mM PMSF, 1 mM iodoacetamide, 1 mM iodoacetate, 0.02 mM TLCK, 0.04 mM TPCK, 0.016 mM APMSF, 0.5 mM *p*-chloromercuribenzoate, 2 mM sodium citrate, 0.0065 mM bestatin, pH 7.5) while still in the slaughter house. The cortex pieces were then homogenized in a Waring blender for 2 min and filtrated through gauze. This homogenate was then centrifuged at 940 × *g* for 10 min. The supernatant was centrifuged at 1,760 × *g* for 15 min and the resulting supernatant was finally centrifuged at 11,000 × *g* for 90 min. The pellet was resuspended in the same buffer and the last centrifugation was repeated. The final mitochondrial fraction, resuspended in homogenization buffer, was then centrifuged in a discontinuous gradient of 55 ml, 20 ml and 5 ml of 33, 44 and 50% saccharose in 10 mM triethanolamine, pH 7.5, and protease inhibitor mixture (10 mM EDTA, 2 mM DTT, 0.1 mM dipyridyl, 0.1 mM phenanthroline, 0.5 mM PMSF, 0.5 mM iodoacetamide, 0.5 mM iodoacetate, 0.02 mM TLCK, 0.04 mM TPCK, 0.005 mM APMSF, 0.2 mM *p*-chloromercuribenzoate, 1 mM sodium citrate, 0.0065 mM bestatin) in 100 ml centrifuge tubes at 54,000 × *g* for 11 h. The mitochondrial fraction collected from the 33/44% interface was then brought to 0.6 M NaCl and sonicated 2 × 5 min in a Branson sonicator model 250 at setting 8 with a large tip. The sonicated homogenate was then centrifuged at 35,000 × *g* for 2 h. The supernatant was applied to a phenyl sepharose column (19 × 1.4 cm) equilibrated in 50 mM Tris, pH 7.5, containing 0.6 M NaCl and a 1:1 dilution of the protease inhibitor mixture. Bound protein was eluted with a linear gradient from 0.6 to 0.05 M NaCl in the same buffer. Fractions containing amidinotransferase activity were diluted with the same volume of water and further separated on a hydroxyapatite column (30 × 1.4 cm) in 30 mM potassium phosphate, pH 7.5, containing the protease inhibitor mixture in a 1:50 dilution, with

*Corresponding author. Fax: (49) (89) 8578 2422.

*Abbreviations*: APMSF, 4-aminophenyl methanesulfonyl fluoride; DTT, dithiothreitol; PMSF, phenylmethanesulfonyl fluoride; TLCK, tosyl lysine chloromethyl ketone; TPCK, tosyl phenylalanine chloromethyl ketone.


*SSDI* 0014-5793(94)00026-R

a linear gradient up to 0.5 M potassium phosphate. The same dilution of protease inhibitors was used in all subsequent steps. Fractions containing amidinotransferase activity were brought to pH 8.5 with Tris and dialyzed against 20 mM Tris, pH 8.5, containing the usual protease inhibitors. The enzyme solution was then further separated by chromatography on DEAE-cellulose (Whatman DE-52) in 25 mM Tris, pH 8.5, and protease inhibitors, with a gradient from 0 to 0.5 M NaCl. The enzyme was concentrated in an Amicon ultrafiltration cell using a YM 30 membrane and finally purified by size exclusion chromatography on a Superose 12 column (Pharmacia) equilibrated in 150 mM potassium phosphate, pH 7.5, with protease inhibitor mixture. All purification steps were carried out in the cold room, except gel filtration. Each step was monitored by determining the enzymatic activity [12] and the protein concentration [13]. SDS-PAGE was as in [14].

*2.2. Isolation of peptides and sequence analysis*

The purified protein was reduced and pyridylethylated under denaturing conditions [15]. The reaction mixture was then dialysed against 0.2 M ammonium hydrogen carbonate. During dialysis the protein precipitated. A sample of this suspension was treated with lysyl endopeptidase (Wako Chemicals, Neuss, Germany) at an enzyme/substrate ratio of 1:100 (w/w) at 23°C for 16 h. The reaction was terminated by acidification with a drop of concentrated trifluoroacetic acid and 1:20 dilution with 0.1% trifluoroacetic acid. Soluble and insoluble fragments were separated by centrifugation with an Eppendorf bench top centrifuge. The soluble peptides were separated by reversed phase HPLC (Vydac C18, 4.6 mm × 250 mm column) using 0.1% trifluoroacetic acid (A) and 0.1% trifluoroacetic acid in 70% acetonitrile (B) with a gradient from 0% to 60% B in 160 min at a flow rate of 0.25 ml/min. The same procedures were subsequently used with all cleavage mixtures. The pelleted fragments were dissolved in 8 M urea and diluted with 0.2 M ammonium hydrogen carbonate to 1 M urea before adding lysyl endopeptidase as above. The insoluble fragments obtained by this second treatment with the same enzyme were dissolved in concentrated formic acid and sequenced. The soluble peptides were separated by HPLC. Another portion of the pyridylethylated protein was suspended in 0.2 ammonium hydrogen carbonate containing 20% acetonitrile and cleaved with endoproteinase Asp-N (sequencing grade, Boehringer Mannheim, Germany) at an enzyme/substrate ratio of 1:50 at 23°C. After 8 h the same amount of enzyme was added once again and the reaction was allowed to continue for a further 16 h. The peptides soluble in 0.1% trifluoroacetic acid were separated by HPLC. The insoluble fragments were divided in three portions, one of which was sequenced. The others were suspended in 0.2 M ammonium hydrogen carbonate and cleaved with thermolysin (Merck, Darmstadt, Germany) for 3 h or chymotrypsin (sequencing grade, Boehringer Mannheim) for 2 h both at 30°C with an enzyme/substrate ratio of 1:50. The peptides were separated by HPLC. Another sample of the amidinotransferase was cleaved in 7 M guanidine. HCl dissolved in 75% formic acid at 37°C for 60 h [16]. The reaction mixture was then diluted and soluble and insoluble fragments separated as above. A sample of the insoluble fragments was sequenced while another part was further cleaved, either in 75% formic acid by addition of a small crystal of CNBr and incubation at 23°C for 24 h under nitrogen in the dark, or with endoproteinase Asp-N in 20% acetonitrile as above. Peptides insoluble in 0.1% trifluoroacetic acid were sequenced, the soluble peptides were separated by HPLC. Sequence analysis of soluble and insoluble peptides was done by Edman degradation using Applied Biosystems sequencers 470A and 473A according to the manufacturer's instructions.

*2.3. Isolation of human cDNA and sequence analysis*

A 1047-bp fragment coding for a partial arginine:glycine amidinotransferase sequence was isolated and amplified by PCR from a λUni-Zap cDNA library (Stratagene Cloning Systems) constructed with mRNA from human kidney carcinoma cells using primers synthesized according to the porcine amino acid sequences 32–41 and 371–380 (Fig. 2). An *Xho*I and a *Pst*I restriction site sequence were introduced at the 5′ and at the 3′ end (Fig. 3). PCR was used with 50 pmol primers in 50 μl standard buffer with 200 μM deoxynucleotides (Perkin-Elmer 480), and 3 U Taq polymerase. 30 cycles were run with an annealing temperature of 55°C. The amplified fragment was restricted with *Xho*I and *Pst*I and cloned into pUC19 (Gibco BRL). Labeled fragment was produced by random priming [17,18] using a purchased kit (United States Biochemical Corp.). The labeled fragment was used to screen the λUni-Zap library according to established procedures [19]. 5 positive phage clones (AT1 to AT5) were subcloned into pBluescript SK⁻ (Stratagene). Nucleotide sequencing was performed with Sequenase (United States Biochemical Corp.) or using an Applied Biosystems sequencer 373A with the Prism Ready Reaction Dye Deoxy Terminator Cycle Sequencing kit. Both strands were sequenced to obtain complete nucleotide sequences of clones AT1 and AT5.

## 3. Results and discussion

### *3.1. Isolation and sequence analysis of porcine amidino-transferase*

First attempts to isolate the amidinotransferase were undertaken with homogenized kidney tissue. This yielded molecules with heavily truncated N-terminal ends. Only purification of the enzyme from isolated mitochondria yielded two major N-terminal amino acid sequences (Figs. 1 and 2). These sequences were confirmed later with peptides and are indicated in Fig. 2 above the first row of amino acid sequence. Because we were not able to prevent this proteolytic degradation during purification of the enzyme with protease inhibitors, we do not know with certainty whether the longest form we recovered, and which is shown in Fig. 2, actually represents the intact molecule. The purification procedure described above resulted in a 280-fold enrichment of amidinotransferase activity with a final yield of 8%. SDS-PAGE (Fig. 1) showed two closely spaced bands of $M_r$ 46,000. This heterogeneity was probably due to partial N-terminal degradation.

Sequence analysis was hampered by the low solubility of the denatured and pyridylethylated or carboxymethylated enzyme in solvents suitable for proteolytic degradation. A first cleavage experiment with lysyl endopeptidase yielded some peptides which were soluble in re-

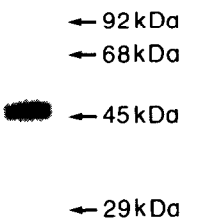


Fig. 1. SDS-PAGE of purified pig kidney arginine:glycine amidinotransferase on a 12% gel. The positions of marker proteins are shown on the left (phosphorylase b, bovine serum albumin, ovalbumin and carbonic anhydrase).

versed phase HPLC starting solvent (numbered K peptides in Fig. 2). Sequence analysis of the insoluble fragments showed them to be a very heterogeneous mixture. A second digest in the presence of urea yielded some more soluble peptides (K peptides with small letters in Fig. 2) and also an insoluble residue which yielded two major sequences (K1ins and K2ins, Fig. 2). Next, the amidinotransferase was exhaustively digested with endoproteinase Asp-N. Apart from the soluble peptides (D peptides in Fig. 2), this yielded an insoluble residue which could be solubilized by degradation with either thermolysin or chymotrypsin. These rather unspecific proteases cleaved the substrate into a heterogeneous mixture of mostly very short peptides. However, sequence analysis of the major peaks after reversed phase HPLC separation of the peptides yielded some valuable information (TH and CH peptides in Fig. 2). The sequence data available at that time indicated a favorable number and distribution of Asp–Pro bonds which could be cleaved by limited acid cleavage in 75% formic acid

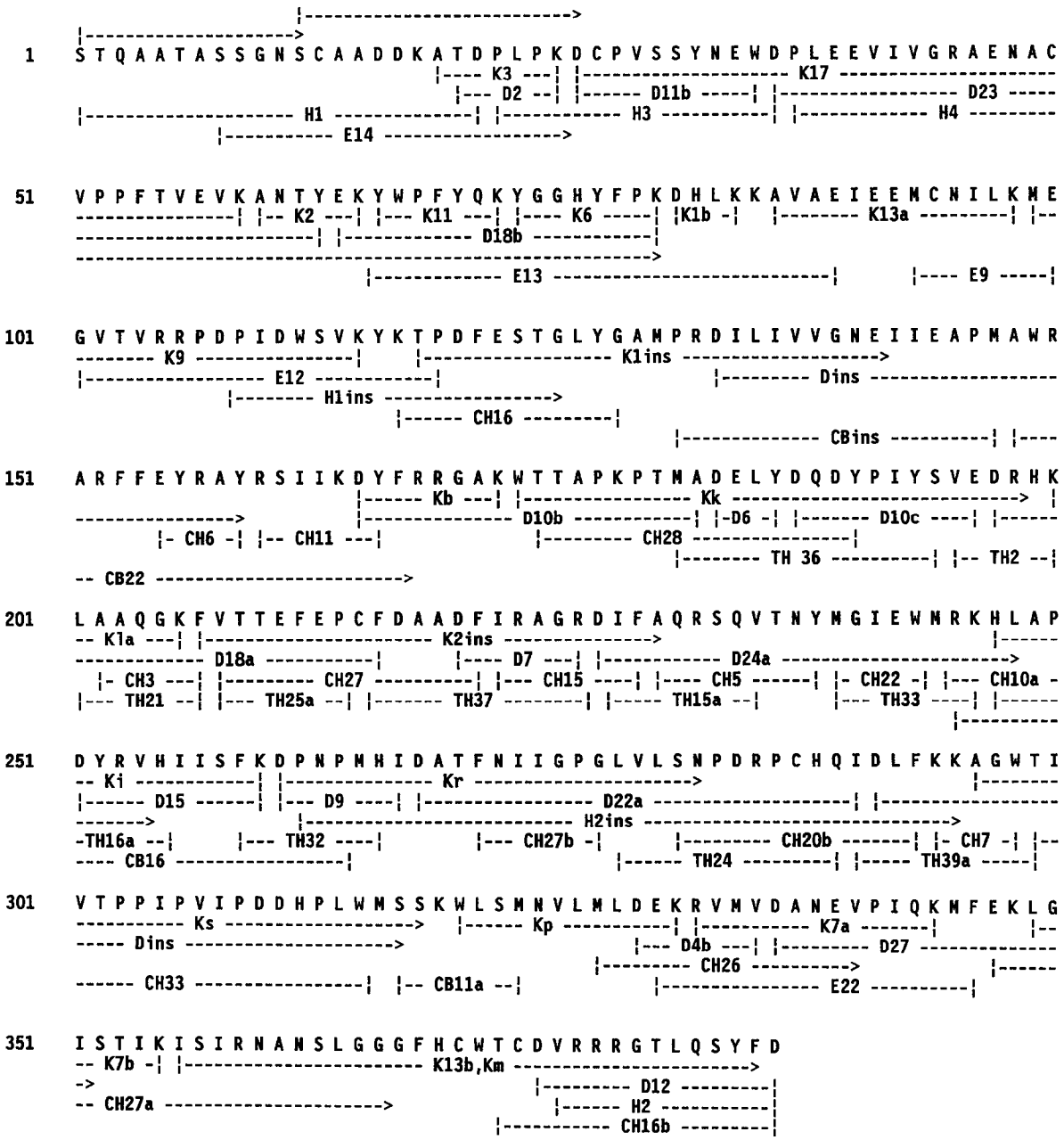


Fig. 2. Amino acid sequence of pig kidney arginine:glycine amidinotransferase. Dashed lines identify positions sequenced by Edman degradation and are terminated by a vertical line in case of complete sequence determination or by an arrow in the case of partial sequence determination. The lines above the first row of sequence indicate the major N-terminal sequences found. Individual peptides were generated by cleavage with cyanogen bromide (CB), chymotrypsin (CH), endoproteinase Asp-N (D), formic acid (H), lysyl endopeptidase (K), or thermolysin (TH) and are distinguished by numbers or small letters according to different chromatographic pools. ins means insoluble in RP-HPLC starting solvent.

containing 7 M guanidine. The cleavage of the substrate was nearly complete and highly specific. Apart from the Asp–Pro bonds only an Asp–Val bond (374/375, Fig. 2) was cleaved to some extent. The reaction mixture contained the peptides H1 to H4 soluble in HPLC starting solvent and the insoluble larger fragments H1ins and H2ins (Fig. 2). The latter were further cleaved by CNBr or endoproteinase Asp-N to yield the insoluble peptides Dins and CBins (Fig. 2). Screening of the soluble CNBr-peptides (CB peptides in Fig. 2) eventually yielded the final sequences needed to assemble the complete amino acid sequence of porcine arginine:glycine amidinotransferase.

The porcine arginine:glycine amidotransferase contained 386 amino acids giving a molecular weight of 44,188, which is in good agreement with the apparent molecular weight obtained by gel electrophoresis (Fig. 1). No signs of post-translational modification of amino acids were found. A search for known sequence motifs (using the Prosite Dictionary of Dr. A. Bairoch, Geneva) did not yield any meaningful result with the possible exception of the presence of an incomplete mitochondrial energy transfer proteins signature [20,21] between positions 78 and 88. This sequence motif was only found in mitochondrial membrane proteins, where it occurs several times. The presence of one incomplete copy of this motif in the amidinotransferase is probably insignificant.

### 3.2. *Isolation of human cDNA clones and sequence analysis*

When it became clear that we could not isolate a homogeneous population of porcine amidinotransferase molecules suitable for crystallization, we began experiments aiming at the recombinant production of the enzyme. A fragment of cDNA was isolated from a human kidney carcinoma cDNA library by PCR using oligonucleotide primers synthesized according to the porcine protein sequences 32–41 and 371–380 (Fig. 3). This fragment of 1047 bp was labeled and used to identify longer clones in the same library. 5 positive clones were isolated, the inserts subcloned and characterized by partial sequencing. The deduced amino acid sequences were highly similar to sequences of the porcine enzyme. 4 clones were near full-length (AT1–AT4), 1 started with nucleotide 166 of the complete clones (AT5). The complete nucleotide sequence of a near full-length clone (AT1) and the derived amino acid sequence are shown in Fig. 4. The human cDNA-derived protein sequence is 37 amino acids longer than the longest porcine form. This is clearly due to the presence of a leader peptide (Figs. 4 and 5). However, it is difficult to predict the

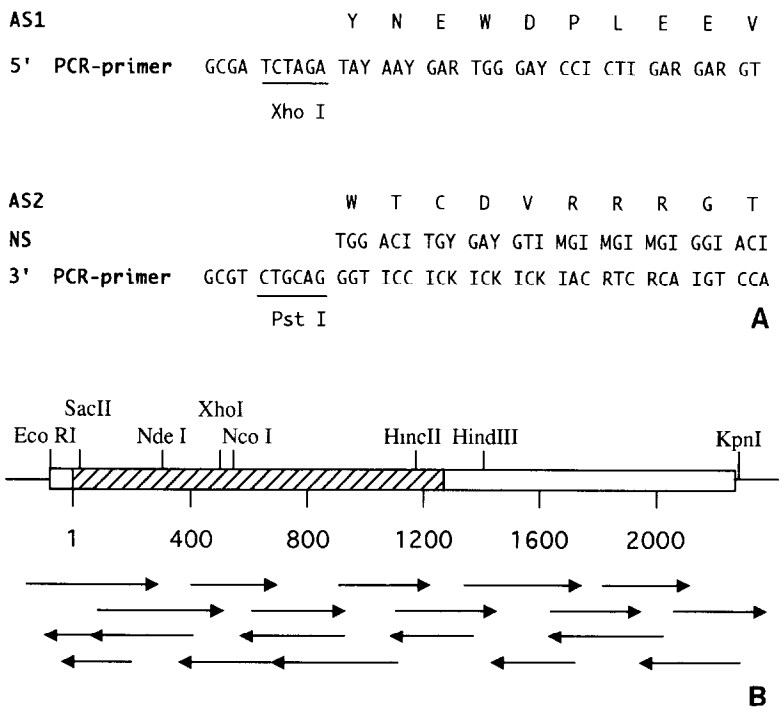


Fig. 3 (A) Oligonucleotide primers used for PCR.AS1, amino acids 32–41 of the pig amidinotransferase; AS2, amino acids 371–380. NS, nucleotide sequence deduced from AS2. The 3' primer is the inverse complement of this sequence. I, inosine; R, purine; Y, pyrimidine; K, bases G or T; M, bases A or C. Both primers were synthesized with a restriction site as indicated. (B) Partial restriction map and sequencing strategy used to obtain the nucleotide sequence of a near full-length human arginine:glycine amidinotransferase cDNA clone. Arrows indicate sequenced regions and direction of sequencing. The hatched area indicates the position of the open reading frame.

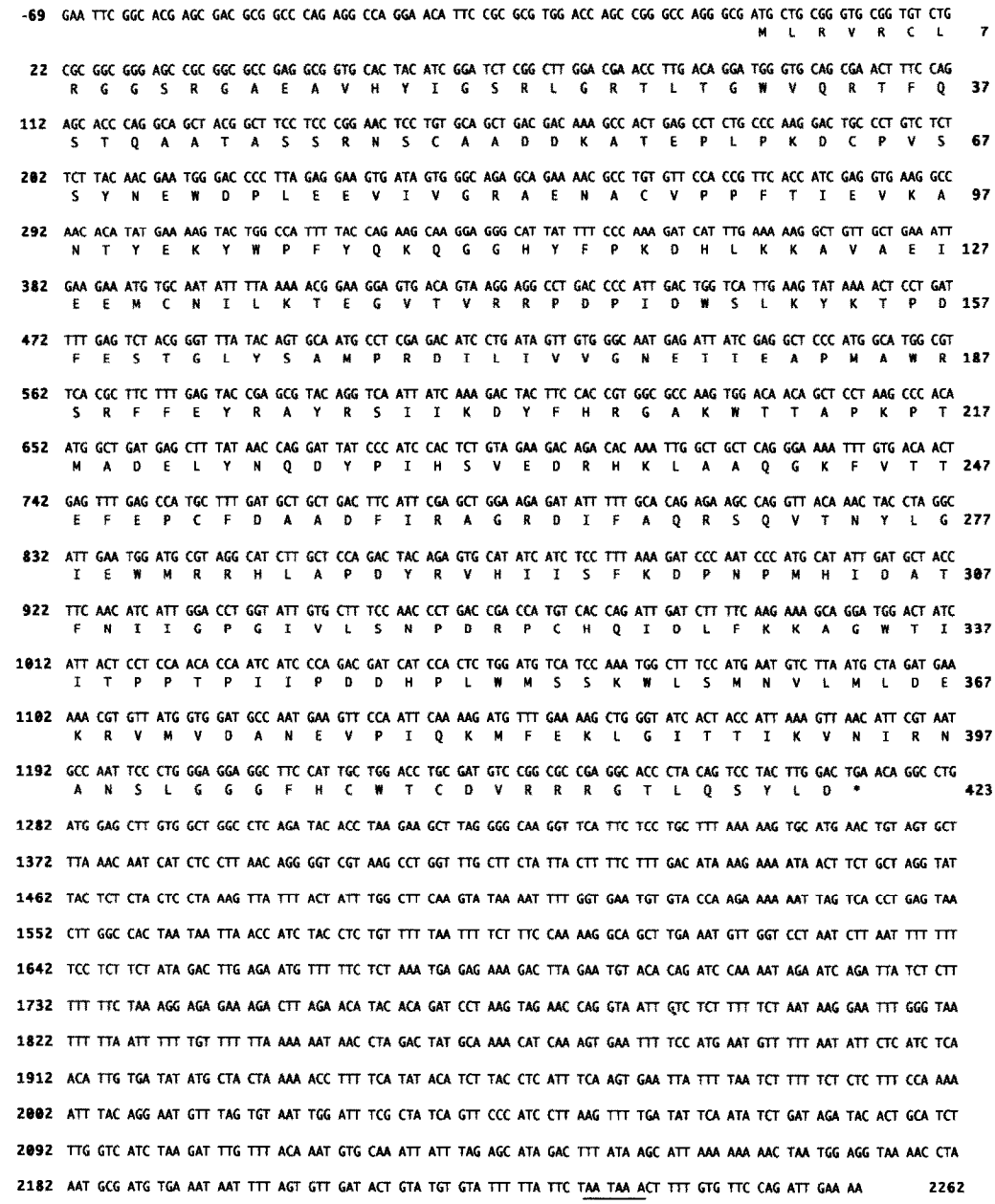


Fig. 4. Nucleotide sequence and deduced amino acid sequence of a human L-arginine:glycine amidinotransferase cDNA near full-length clone. Nucleotides are numbered on the left, amino acids on the right. The polyadenylation sequence is underlined.

exact cleavage site between leader sequence and mature protein. Recently a survey of N-terminal proteolytic cleavage sites in mitochondrial precursor proteins has identified a motif consisting of an arginine at position −10, a hydrophobic amino acid at position −8 and a serine, threonine or glycine at position −5 relative to the mature amino terminus [22]. The N-terminal region of the human precursor molecule contains exactly this motif with Arg-34, Phe-36 and Thr-39. However, with the cleavage site predicted in this way, between Thr-43 and Ala-44, the mature human enzyme would be 6 amino acids shorter than the longest porcine form. Another, although not perfect, motif may include Arg-26 and Leu-28 but would lack the canonical residue at position −5. This imperfect motif would predict cleavage between Thr-35 and Phe-36 resulting in a mature protein 2 amino acids longer than the longest porcine species found in mitochondrial preparations. Thus the exact location of the precursor cleavage site remains unclear. Leaving aside the putative leader sequence, the human and porcine amidinotransferases are 94% identical, with most of the changes being conservative amino acid replacements (Fig. 5).

### 3.3. Comparison to other proteins

When compared with the FASTA program [23] to

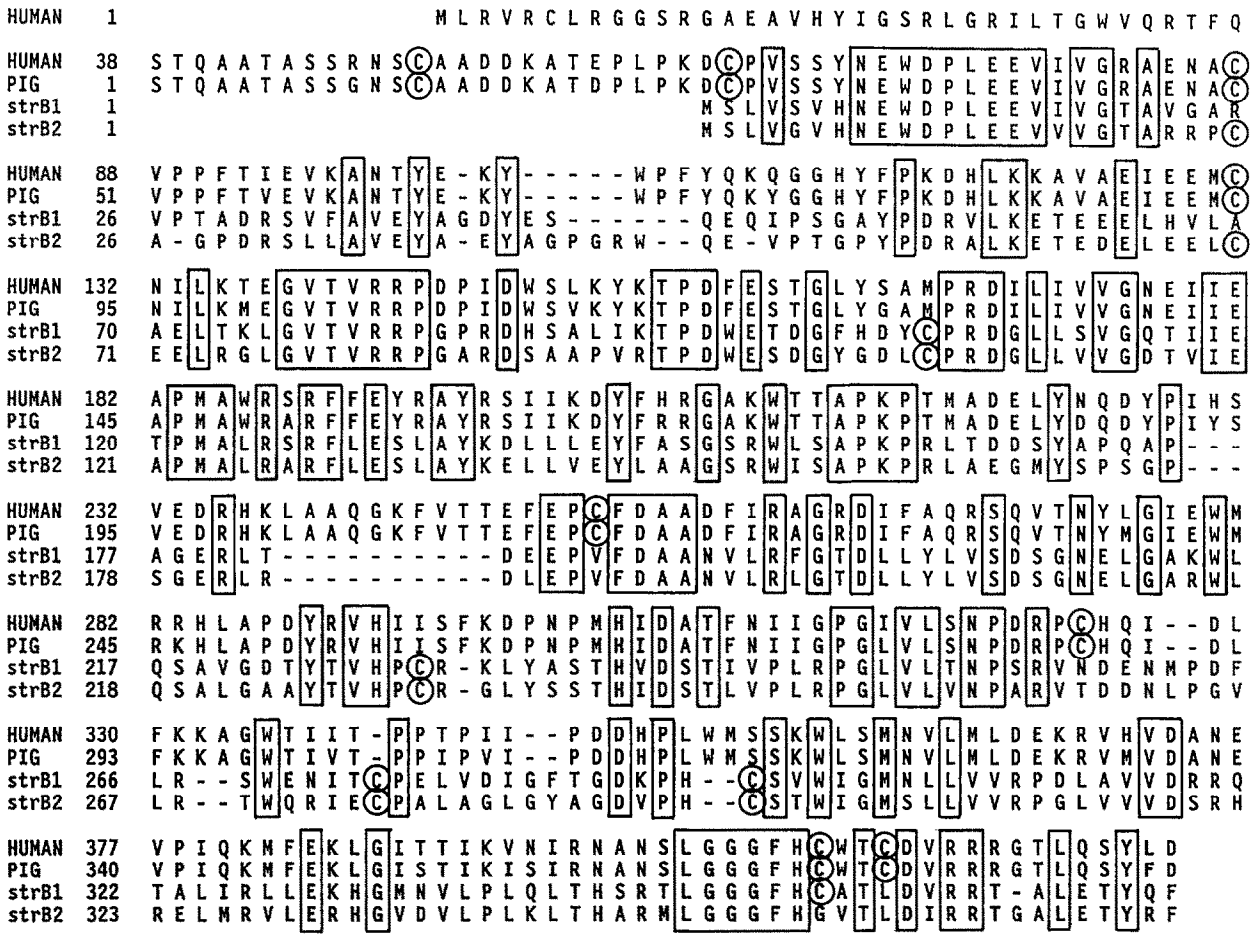


Fig. 5. Alignment of mammalian arginine:glycine amidinotransferase sequences to bacterial arginine:inosamine phosphate amidinotransferases strB1 [10] and strB2 [11]. Boxes enclose amino acids identical in all 4 sequences. Cysteines are in circles.

known protein sequences compiled in the MIPSX (Martinsried Institute for Protein Sequences) data collection, the mammalian enzymes were found to be very similar to the bacterial inosamine-phosphate amidinotransferases [9–11]. After introduction of some gaps to accommodate the different length of the proteins, about 36% of the amino acid residues were identical between the mammalian and the somewhat shorter bacterial proteins (Fig. 5). About 15% of the changes were conservative amino acid replacements which maintained the general properties (i.e. hydrophobic, acidic or basic) in all four sequences.

In both mammalian and bacterial amidinotransferases, a cysteine residue has been identified which is involved in the transfer of the amidine from arginine to the respective acceptor molecule by formation of a covalent isothiourea intermediate [24]. Therefore it was interesting to compare the cysteine pattern of both types of enzyme (Fig. 5). While the mature mammalian amidinotransferases have 8 cysteines, the bacterial enzymes have only 3 (strB1) and 6 (strB2). Only 1 (strB1) or 2 (strB2) of these coincide with cysteines in the mammalian enzyme sequences. Because the STRB2 protein actually did not show the amidinotransferase activity [11] the STRB1 protein with demonstrated enzymatic activity is probably more relevant for this comparison. The strB1 sequence shows only 1 cysteine (C352, Fig. 5) coinciding with a cysteine in the mammalian sequences. The strB2 sequence does not have a cysteine in this position or in the immediate neighbourhoud. Thus this cysteine would be a prime candidate for the putative active site cysteine. Since the active cysteine can be labeled radioactively [24] it should be possible to locate it experimentally within the sequences presented in the present report.

*Acknowledgements*: We thank Prof. A. Ullrich for providing us with the human cDNA library and W. Strasshofer for technical assistance.

## References

[1] Walker, J.B. (1973) in: The Enzymes, 3rd edition, vol. 9, part B, (P.D. Boyer, Ed.) pp. 497–509, Academic Press, New York.
[2] Walker, J.B. (1979) Adv. Enzymol. 50, 177–242.
[3] Magri, E., Balboni, G. and Grazi, E. (1975) FEBS Lett. 55, 91–93.
[4] Grazi, E., Magri, E. and Balboni, G. (1975) Eur. J. Biochem. 60, 431–436.

[5] Walker, J.B. and Hannan, J.K. (1976) Biochemistry 15, 2519–2522.
[6] McGuire, D.M., Tormanen, C.D., Segal, I.S. and van Pilsum, J.F. (1980) J. Biol. Chem. 255, 1152–1159.
[7] Sipilä, I. (1980) Biochim. Biophys. Acta 613, 79–84.
[8] McGuire, D.M., Gross, M.D., van Pilsum, J.F. and Towle, H.C. (1984) J. Biol. Chem. 259, 12034–12038.
[9] Tohyama, H., Okami, Y. and Umezawa, H. (1987) Nucleic Acids Res. 15, 1819–1833.
[10] Distler, J., Ebert, A., Mansouri, K., Pissowotzki, K., Stockmann, M. and Piepersberg, W. (1987) Nucleic Acids Res. 15, 8041–8056.
[11] Pissowotzki, K., Mansouri, K. and Piepersberg, W. (1991) Mol. Gen. Genet. 231, 113–123.
[12] van Pilsum, J.F., Taylor, D., Zakis, B. and McCormick, P. (1970) Anal. Biochem. 35, 277–286.
[13] Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951) J. Biol. Chem. 193, 265–275.
[14] Schägger, H. and von Jagow, G. (1987) Anal. Biochem. 166, 368–379.
[15] Henschen, A. (1986) in: Advanced Methods in Protein Microsequence Analysis (Wittmann-Liebold, B., Salnikow, J. and Erdmann, V.A., Eds.) pp. 244–255, Springer, Berlin.
[16] Jauregui-Adell, J. and Marti, J. (1975) Anal. Biochem. 69, 468–473.
[17] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[18] Feinberg, A.P. and Vogelstein, B (1984) Anal. Biochem. 137, 266–269.
[19] Strauss, W.M. (1990) in: Current Protocols in Molecular Biology (Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K., Eds.) vol. 1, chap. 6.3, Wiley, New York.
[20] Klingenberg, M. (1990) Trends Biochem. Sci. 15, 108–112.
[21] Walker, J.E. (1992) Curr. Opin. Struct. Biol. 2, 519–526.
[22] Hendrick, J.P., Hodges, P.E. and Rosenberg, L.E. (1989) Proc. Natl. Acad. Sci. USA 86, 4056–4060.
[23] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[24] Grazi, E. and Rossi, N. (1968) J. Biol. Chem. 243, 539–542.