Biochemical and Biophysical Research Communications 421 (2012) 409–412

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Mini Review

Human carbamoyl phosphate synthetase I (CPSI): Insights on the structural role of the unknown function domains

Monica Lopes-Marques [a,b,*], Gilberto Igrejas [b,c], António Amorim [a,d], Luisa Azevedo [a,d,*]

[a] IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal
[b] University of Tras-os-Montes and Alto Douro, Institute for Biotechnology and Bioengineering, Center of Genomics and Biotechnology Vila Real, Portugal
[c] University of Tras-os-Montes and Alto Douro, Department of Genetics and Biotechnology Vila Real, Portugal
[d] FCUP – Faculty of Sciences of the University of Porto, Porto, Portugal

ARTICLE INFO

Article history:
Received 27 March 2012
Available online 10 April 2012

Keywords:
Urea cycle enzymes
Carbamoyl phosphate synthetase (CPS1)
Domain organization
Protein conservation
Escherichia coli CPS

ABSTRACT

Carbamoyl phosphate synthetase (CPS) is an ancient protein. In mammals it intervenes in the urea cycle. This enzyme is organized into six domains, three of which have no established role in the mammalian enzyme. Taking advantage of the high degree of conservation between the human and the *Escherichia coli* homologue a comparative study was carried out in order to infer about the biological role of these less characterized domains. We show that among the residues involved in the maintenance of quaternary structure of the *E. coli* enzyme, several are highly conserved between human and bacterial enzyme and match the homologous positions of the "unknown function" domains in human enzyme, suggesting they are involved in the structural stability of the human enzyme as they are in bacteria.

1. Introduction

Urea cycle disorders result from the impairment of this cycle at any point and the clarification of the functional effect of each genetic variant, individually or in combination, is relevant to establish phenotype–genotype relationships. In the case of ornithine transcarbamylase deficiency (OMIM 311250), the most common urea cycle defect [1], the structural effect of amino acid replacements may be assessed using the human enzyme structure [2]. However, for carbamoyl phosphate synthetase I (hsCPS) deficiency (OMIM 237300), no complete crystal structure for the human enzyme has been determined so far, hindering the analysis of the structural impact of genetic variants. In opposition, for the *Escherichia coli* enzyme (eCPS) a complete crystal structure has been determined and published [3]. The eCPS is encoded by two distinct genes, *carA* and *carB* [4] which encode the α-subunit (~42 kDa) and the β-subunit (~118 kDa), respectively that assemble into a α/β heterodimer [3] (Fig. 1A). On the other hand the hsCPS is encoded by a single gene and consequently is a monomer; this alternative structural organization resulted from an ancient gene fusion event [5] (Fig. 1A). Sequence comparison between hsCPS and eCPS reveals a high degree of sequence conservation and similar domain organization [6] though no particular function has been recognized for three of the matching domains in the human enzyme. In this study, comparative analyses combined with structural information were used to understand the role of these three domains in the maintenance of the predicted ternary structure of the human enzyme. By comparative analysis we observed that the hsCPS unknown function domains are spatially located in a region that corresponds to the α/β subunits interface in eCPS. Our data strongly suggests that these unknown function domains are essentially involved in the preservation of the correct tertiary structure of the hsCPS monomer and that genetic variants allocated to these domains may impact enzyme structure and function.

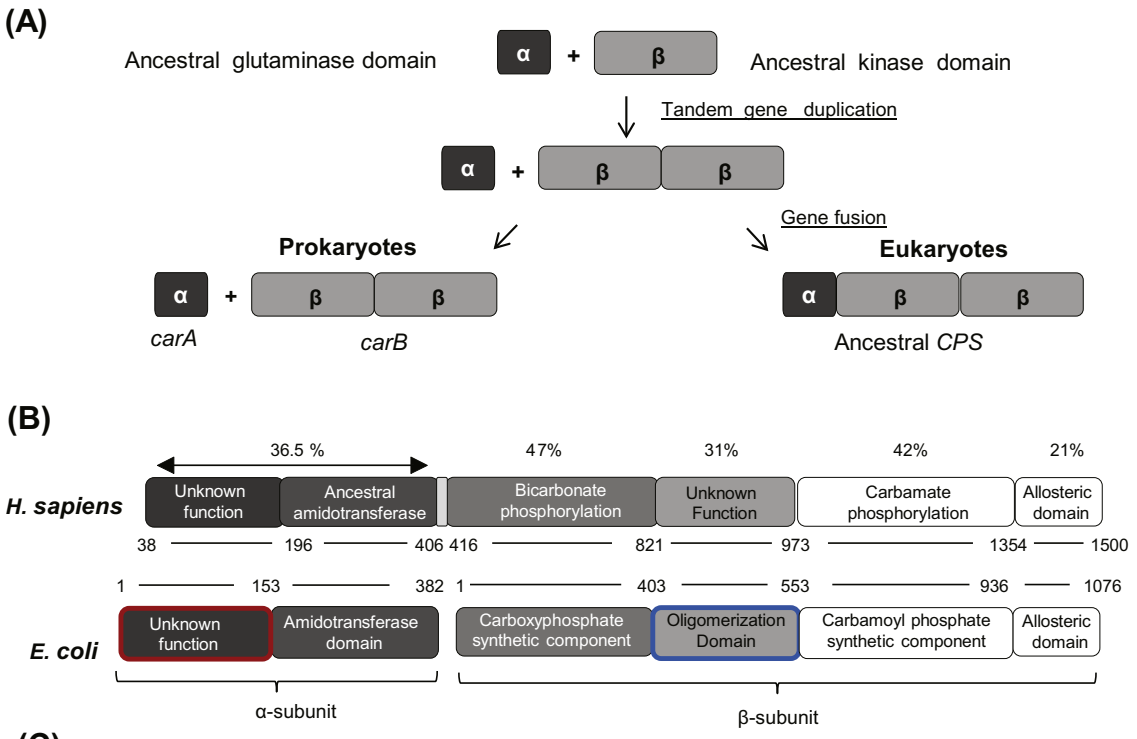
2. Materials and methods

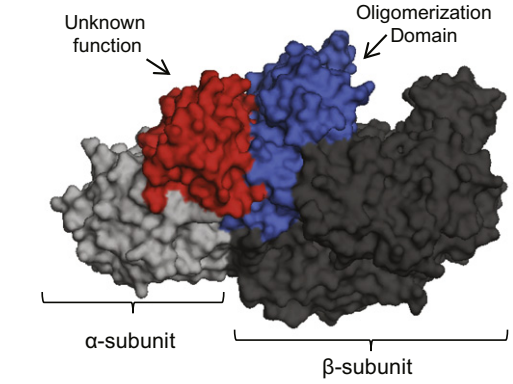
2.1. Conservation analysis and homology modelling

Amino acid sequences for the human (P31327) and bacterial CPS (P0A6F1 and P00968) were aligned using the ClustalW2 [7]. Conservation analysis and identification of the biologically important residues in hsCPS was performed using the ConSurf server [8]. This analysis was performed using 293 unique sequences resulting from the PSI-BLAST search and conservation scores calculated Bayesian method and the JTT

* Corresponding authors at: IPATIMUP, Population Genetics Group, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal. Fax: +351 22 55 70799.
E-mail addresses: monicaslm@hotmail.com (M. Lopes-Marques), lazevedo@ipatimup.pt (L. Azevedo).

0006-291X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.04.033





(D)

E. coli (β-subunit - α-subunit)	Human Corresponding residues
Asn227 - Val305	**Asn643** - Ala330
Gln254 - Asn62	**Gln670 - Asn103**
Gln254 - Tyr57	**Gln670** —Asn98
Thr257 - Arg93	Ser673 - Leu144
Asp258 - Ala89	Asn674- Tyr 140
Lys259 - Asp69	Ala675 - Thr110
Arg265 - Asp362	**Arg681 - Asp386**
Asn266-His361	Arg682- Ile385
Asp333 - Lys298	Pro748 - Leu323
Glu349 - Asn294	**Glu763 - Asn319**
Ser351 - Thr34	**Ser765 - Thr75**
Arg389 - Arg116	**Arg803 - Arg167**
Arg389 - Thr115	**Arg803 - Thr166**
Gln391-His59	Phe805-Ile100
Asp459 - Ser90	**Asp874** - Asn141
Trp461 - Asn62	**Trp876 - Asn103**
Arg494-Arg83	Glu909 - Leu134
Lys527 - Asp114	**Lys942 - Asp165**
Arg528 - **Arg116**	Gln943 - **Arg167**
Asp530 - Arg116	**Asp945 - Arg167**
Ala534 - Arg123	**Ala949 - Arg174**
Glu548 - Arg83	Asn963 - Leu134
Glu549 - **Asp114**	Gly964 - **Asp165**
Glu550 - **Asp114**	Gln965 - **Asp165**
Glu550 - Arg120	Gln965 - Lys171
Glu552 - Arg116	**Glu966 - Arg167**

(E)

E. coli β-subunit - β-subunit	Homologous residues in human
Leu415 - **Arg425**	Met829 - **Arg840**
Pro418 - Thr422	Glu832 - Leu837
Ser448 - Gly446	**Ser863** - Asn861
Leu421*-Leu447*	Asn836-Met862
Leu447*-Leu421*	Met862-Asn836

Fig. 1. (A) Schematic representation of the evolutionary history of the *CPS1* gene. (B) Schematic representation of domain organization using the domain limits previously reported [3]. Percentage of sequence pairwise identity held between human and bacterial homologous domains is shown above the corresponding boxes. (C) Localization in the eCPS structure (1JDB) of the "unknown function domain" (1–153a.a) in red and the oligomerization domain (403–553a.a) in blue, in light grey correspondes to the eCPS α-subunit in dark grey eCPS β-subunit. (D) α/β interacting pairs reported in eCPS and corresponding homologous residues in human CPSI. (E) β/β subunit interacting residue pairs reported in eCPS [3] and corresponding homologous residues in human CPSI. All conserved residues between the human and the bacterial enzyme are shown in bold. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model for substitution in proteins. Crystal structure of the eCPS [3] (1JDB) was used as a template to model the human CPS1 structure using the Modeller algorithm [9] following previous reported methodology [10].

3. Results and discussion

The *E. coli* CPS is composed of six major structural and functional domains: unknown function (1–153a.a); amidotransferase domain

(153–382a.a); carboxyphosphate synthetic component (1–402a.a); oligomerization domain (403–553a.a); carbamoyl phosphate synthetic component (553–936a.a) and the allosteric domain (936–1076a.a), where the activator IMP (inosine 5′ monophosphate) binds to the enzyme [3] (Fig. 1B).

Currently there is no specific function attributed to three of the hsCPS domains [6] (Fig. 1B). The first unknown function domain ranges from residues 38 to 196. The second unknown domain is located from residue 196 to 406 and corresponds to the ancestral amidotransferase domain in eCPS, which is inactive in the human enzyme as the primary nitrogen source for the hsCPS is ammonia while eCPS is able to uptake glutamine which is then processed it in the amidotransferase domain. The third domain with unknown function is located within residues 821–973 and is homologous to the oligomerization domain of *e*CPS.

The alignment between eCPS and hsCPS confirmed an overall pairwise identity of about 40% and such degree of conservation is within the threshold of sequence similarity satisfactory for structural homology deductions [11], therefore these values allow us to safely proceed towards homology modeling with a high degree of confidence.

The ancestral gene fusion of eukaryotic CPS (Fig. 1A) resulted in a fraction of residues from the unknown function domains in hsCPS to be buried within the monomer corresponding to the prokaryotic α/β heterodimer interface (Fig. 1C). Other residues are expected be located on the enzyme surface corresponding to the β/β dimer interface of the bacterial enzyme. In the eCPS a total of 23 residues of the β-subunit mediate contacts with 25 residues of the α-subunit in the α/β heterodimer interface region [3] (Fig. 1D, left column). The corresponding homologous residues were identified in human CPSI by pairwise sequence alignment with the bacterial enzyme (Fig. 1D, right column). Of these, 11 interacting pairs were found to be invariant between both enzymes (Gln^{670}–Asn^{103}; Arg^{681}–Asp^{386}; Glu^{763}–Asn^{319}; Ser^{765}–Thr^{75}; Arg^{803}–Arg^{167}; Arg^{803}–Thr^{166}; Trp^{876}–Asn^{103}; Lys^{942}–Asp^{165}; Asp^{945}–Arg^{167}; Ala^{949}–Arg^{174}; and Glu^{966}–Arg^{167}, human coordinates). In addition, the conservation score of these residues was estimated in the ConSurf analysis and the results revealed a high degree of conservation among the 293 homologues compared, suggesting a relevant structural role (Fig. 2A). Three of the residues with the highest conservation score, Asn^{103}, Arg^{167} and Arg^{803}, interact with more than one partner in a complex network of bonds (Fig. 2B), and their replacement is expected to be deleterious as previously reported such as the disease-associated Arg803Gly substitution [12]. These observations allude to the possibility that some residues identified in the human protein as belonging to the unknown function domains have, in fact, a functional role, possibly contributing to a correct fold and stability. In contrast, residues matching the β/β interface region of the *E. coli* enzyme are less conserved between bacteria and humans (Fig. 1E and 2B). This is in agreement

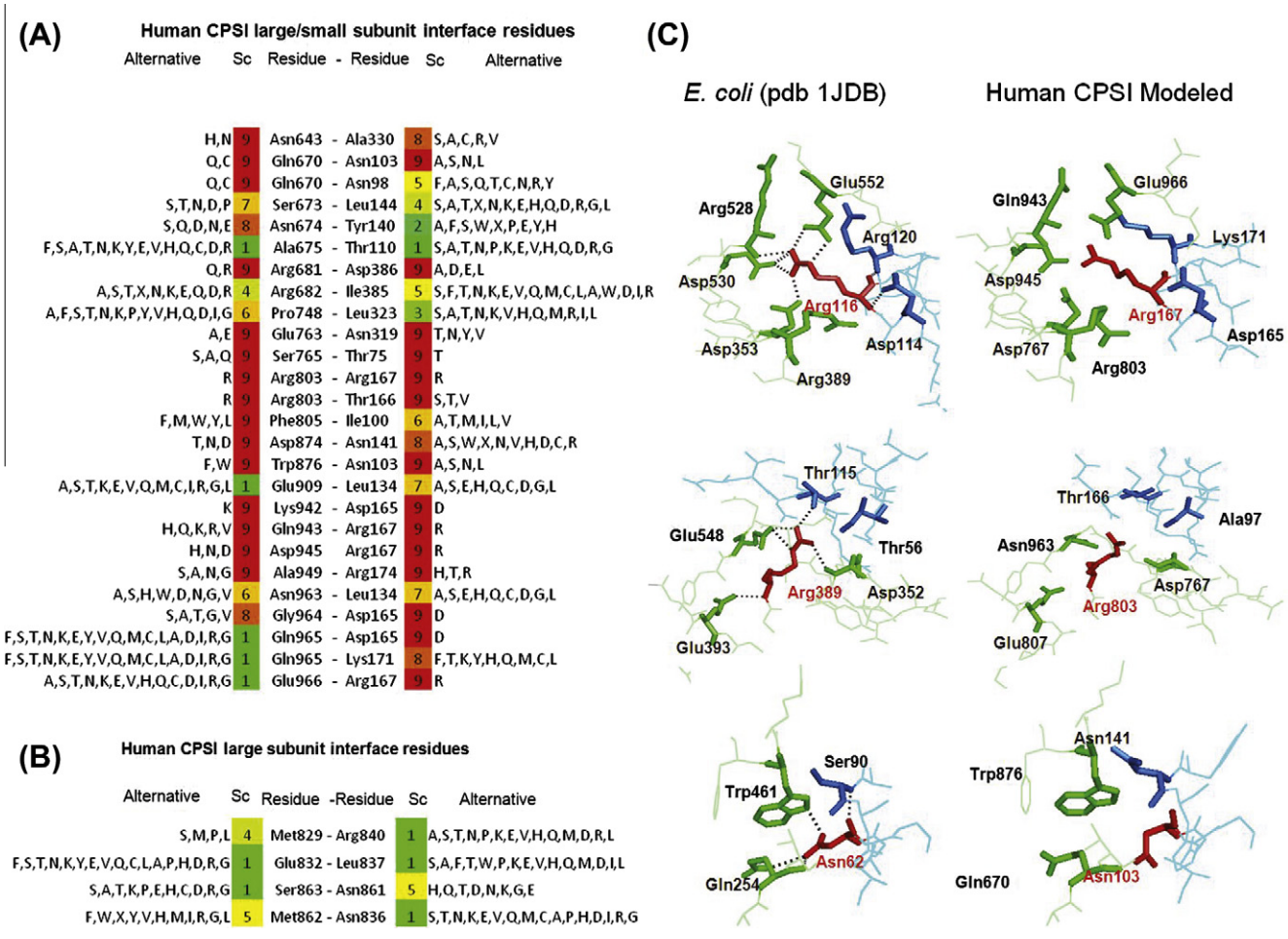


Fig. 2. (A) Cross-species conservation scores obtained in ConSurf in human CPSI. Conservation score of residues buried within the human CPS monomer (Sc 1, less conserved; 9, highly conserved). (B) Conservation score (Sc) of CPS residues predicted to mediate monomer-monomer interactions. (C) On the left, detailed view of bond network established by the highly conserved residues Arg116, Arg389 and Asn62 in the α/β interface of the eCPS (1JDB) and the corresponding homologous residues (Arg167, Arg803 and Asn103) in the human modeled CPSI. The *E. coli* Arg116 is conserved in human CPSI as well as most of the neighboring residues (Asp530, Asp353, and Glu552). The *E. coli* Arg389 is conserved in human CPSI as are three neighboring residues (Asp352, Glu393 and Glu548) with which Arg389 interacts directly. The *E. coli* Asn62 is conserved in the human enzyme as well as two (Trp461 and Gln254) of the three interacting partners.

with the observation that protein–protein interfaces are frequently less conserved than other structural segments, challenging the attempt to establish a conservation profile derived from the conservation pattern alone [13,14]. Therefore it is possible that distinct monomer–monomer interface solutions may have arisen in distinct phylogenetic groups during the evolution of the enzyme, an interesting possibility which waits the determination of the quaternary structure of the human enzyme. As a final remark, the high degree of conservation presented between the human and bacterial CPS allowed the identification of highly conserved residues in the hsCPS and the prediction of a functional role for these residues.

In light of the present findings we propose to term the unknown function domains as "structural domain I, II and III," respectively and anticipate that further experimental confirmation by the resolution of the complete structure of the human CPS will support this hypothesis. Nevertheless, additional functions may also be allocated to these unknown function domains. We assume that a better understanding of the protein structure will assist phenotype–genotype relationships whenever the disease-associated mutations occur within these unknown function domains such as the previously reported mutations Asn674Ile [15], Arg803Gly [12] and Ala949Thr and Asp165Gly [16].

Acknowledgments

The authors are grateful to Vicente Rubio for his valuable comments on this paper. The work was supported by the POPH-QREN-Tipologia 4.2-Scientific employment promotion funded by the European Social Fund and by national funds from the Ministry of Science, Technology and Higher Education (MCTES) to L.A., by research project PIC/IC/82794/2007 sponsored by Fundação para a Ciencia e a Tecnologia (FCT), COMPETE program and by FEDER funds. The Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) is an Associate Laboratory of the Portuguese MCTES and is partially supported by FCT.

References

[1] L. Azevedo, L. Stolnaja, E. Tietzeova, M. Hrebicek, E. Hruba, L. Vilarinho, A.N. Amorim, L. Dvorakova, New polymorphic sites within ornithine transcarbamylase gene: population genetics studies and implications for diagnosis, Molecular Genetics and Metabolism 78 (2003) 152–157.

[2] D. Shi, H. Morizono, Y. Ha, M. Aoyagi, M. Tuchman, N.M. Allewell, 1.85 A resolution crystal structure of human ornithine transcarbamoylase complexed with *N*-phosphonacetyl-l-ornithine, Journal of Biological Chemistry 273 (1998) 34247–34254.

[3] J.B. Thoden, F.M. Raushel, M.M. Benning, I. Rayment, H.M. Holden, The structure of carbamoyl phosphate synthetase determined to 2.1 A resolution, Acta Crystallographica Section D, Biological Crystallography (1999) 8–24.

[4] F.S. Lawson, R.L. Charlebois, J.A. Dillon, Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life, Molecular Biology and Evolution 13 (1996) 970–977.

[5] H. Nyunoya, K.E. Broglie, C.J. Lusty, The gene coding for carbamoyl-phosphate synthetase I was formed by fusion of an ancestral glutaminase gene and a synthetase gene, Proceedings of the National Academic Sciences United States of America 82 (1985) 2244–2246.

[6] A.I. Martinez, I. Perez-Arellano, S. Pekkala, B. Barcelona, J. Cervera, Genetic, structural and biochemical basis of carbamoyl phosphate synthetase 1 deficiency, Molecular Genetics and Metabolism 101 (2010) 311–323.

[7] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, Multiple sequence alignment with the Clustal series of programs, Nucleic Acids Research 31 (2003) 3497–3500.

[8] C. Berezin, F. Glaser, J. Rosenberg, I. Paz, T. Pupko, P. Fariselli, R. Casadio, N. Ben-Tal, ConSeq: the identification of functionally and structurally important residues in protein sequences, Bioinformatics 20 (2004) 1322–1324.

[9] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, Journal of Molecular Biology 234 (1993) 779–815.

[10] L. Azevedo, J. Carneiro, B. van Asch, A. Moleirinho, F. Pereira, A. Amorim, Epistatic interactions modulate the evolution of mammalian mitochondrial respiratory complex components, BMC Genomics 10 (2009) 266.

[11] C. Sander, R. Schneider, Database of homology-derived protein structures and the structural meaning of sequence alignment, Proteins: structure, Function, and Bioinformatics 9 (1991) 56–68.

[12] H. Ono, T. Suto, Y. Kinoshita, T. Sakano, T. Furue, T. Ohta, A case of carbamoyl phosphate synthetase 1 deficiency presenting symptoms at one month of age, Brain and Development 31 (2009) 779–781.

[13] D.R. Caffrey, S. Somaroo, J.D. Hughes, J. Mintseris, E.S. Huang, Are protein–protein interfaces more conserved in sequence than the rest of the protein surface?, Protein Science 13 (2004) 190–202

[14] Q.C. Zhang, D. Petrey, R. Norel, B.H. Honig, Protein interface conservation across structure space, Proceedings of the National Academy of Sciences 17 (2010) 10896–10901.

[15] K. Kurokawa, T.T. Yorifuji, M. Kawai, T. Momoi, H. Nagasaka, M. Takayanagi, K. Kobayashi, M. Yoshino, T. Kosho, M. Adachi, H. Otsuka, S. Yamamoto, T. Murata, A. Suenaga, T. Ishii, K. Terada, N. Shimura, K. Kiwaki, H. Shintaku, M. Yamakawa, H. Nakabayashi, Y. Wakutani, T. Nakahata, Molecular and clinical analyses of Japanese patients with carbamoylphosphate synthetase 1 (CPS1) deficiency, Journal of Human Genetics 52 (2007) 349–354.

[16] J. Häberle, O.A. Shchelochkov, J. Wang, P. Katsonis, L. Hall, S. Reiss, A. Eeds, A. Willis, M. Yadav, S. Summar, Urea Cycle Disorders Consortium, O. Lichtarge, V. Rubio, L.-J. Wong, M. Summar, Molecular defects in human carbamoyl phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations, Human Mutation 32 (2011) 579–589.