Understanding carbamoyl phosphate synthetase (CPS1) deficiency by using the recombinantly purified human enzyme: effects of CPS1 mutations that concentrate in a central domain of unknown function

Carmen Díez-Fernández^a, Liyan Hu^b, Javier Cervera^{a,c}, Johannes Häberle^{b,*} and Vicente Rubio.^{a,c,}**

^aInstituto de Biomedicina de Valencia of the CSIC, Valencia, Spain

^bUniversity Children's Hospital Zurich and Children's Research Center, Zurich, Switzerland

^cGroup 739 of the Centro de Investigación Biomédica en Red sobre Enfermedades Raras (CIBERER) del Instituto de Salud Carlos III, Spain

*Correspondence to: J. Häberle, Kinderspital Zürich, Steinwiesstrasse 75, 8032 Zürich, Switzerland. Phone: +41 442667342. Fax: +41 442667167

** Correspondence to: V. Rubio, Instituto de Biomedicina de Valencia (IBV-CSIC), Jaume Roig 11, 46010 Valencia, Spain. Phone: +34 96 3391772. Fax: +34 96 3690800.
Email addresses: johannes.haeberle@kispi.uzh.ch (J. Häberle), rubio@ibv.csic.es (V. Rubio).

Abstract

Carbamoyl phosphate synthetase 1 deficiency (CPS1D) is an inborn error of the urea cycle that is due to mutations in the *CPS1* gene. In the first large repertory of mutations found in CPS1D, a small CPS1 domain of unknown function (called the UFSD) was found to host missense changes with high frequency, despite the fact that this domain does not host substrate-binding or catalytic machinery. We investigate here by *in vitro* expression studies using baculovirus/insect cells the reasons for the prominence of the UFSD in CPS1D, as well as the disease-causing roles and pathogenic mechanisms of the mutations affecting this domain. All but three of the 18 missense changes found thus far mapping in this domain in CPS1D patients drastically decreased the yield of pure CPS1, mainly because of decreased enzyme solubility, strongly suggesting misfolding as a major determinant of the mutations negative effects. In addition, the majority of the mutations also decreased from modestly to very drastically the specific activity of the fraction of the enzyme that remained soluble and that could be purified, apparently because they decreased $V_{\max}$. Substantial although not dramatic increases in K_m values for the substrates or for N-acetyl-L-glutamate were observed for only five mutations. Similarly, important thermal stability decreases were observed for three mutations. The results indicate a disease-causing role for all the mutations, due in most cases to the combined effects of the low enzyme level and the decreased activity. Our data strongly support the value of the present expression system for ascertaining the disease-causing potential of CPS1 mutations, provided that the CPS1 yield is monitored. The observed effects of the mutations have been rationalized on the basis of an existing structural model of CPS1. This model shows that the UFSD, which is in the middle of the 1462-residue multidomain CPS1 protein, plays a key integrating role for creating the CPS1 multidomain architecture leading us to propose here a denomination of "Integrating

Domain" for this CPS1 region. The majority of these 18 mutations distort the interaction of this domain with other CPS1 domains, in many cases by causing improper folding of structural elements of the Integrating Domain that play key roles in these interactions.

Key words: urea cycle diseases; CPS1 deficiency; hyperammonemia; inborn errors; CPS 1 structure; site-directed mutagenesis

1. Introduction

Primary CPS1 deficiency (CPS1D; MIM #237300), a recessively inherited urea cycle disease leading to frequently fatal hyperammonemia [1,2], is due to mutations in the *CPS1* gene. This gene, located in 2q35 [3] and being composed of 38 exons and 37 introns [4-6], with 4,500 coding nucleotides, encodes a 1500-residue proenzyme [7] that is synthesized in hepatocytes and enterocytes [8,9], and which, upon internalization to the mitochondrial matrix, yields after cleavage of its N-terminal 38 amino acids, the mature 1,462-amino acid multidomain (Fig. 1A) CPS1 protein [10-12] [E.C. 6.3.4.16].

CPS1 catalyzes the first step of the urea cycle ($2\text{ATP} + \text{NH}_3 + \text{HCO}_3^- \rightarrow 2\text{ADP} + \text{HPO}_4^{2-} + \text{NH}_2\text{CO}_2\text{PO}_3^{2-}$) [13], converting ammonia to carbamoyl phosphate (CP), a compound that is utilized by ornithine transcarbamylase (OTC) to make citrulline in the second reaction of the urea cycle. The three-step CPS1 reaction (Fig. 1B) includes two analogous ATP-dependent phosphorylations, of bicarbonate and carbamate, and an intervening step of carbamate synthesis from carboxyphosphate and ammonia [13]. To be active, CPS1 requires the presence of an essential allosteric activator, N-acetyl-L-glutamate (NAG) [10,14,15], made by NAG synthase from glutamate and acetyl-coenzyme A [16]. The rate of NAG synthesis heavily depends on glutamate concentration [17], and therefore NAG represents a switch for CPS1 activity, which is turned off when glutamate levels decrease, thus preventing excessive nitrogen draining by the urea cycle from an already low amino acid pool [18,19]. NAG activation is a unique property of CPS1 (and to a lesser extent of the piscine CPS1 homologue CPSIII) [20], not being shared by other CPSs, all of which are active in the absence of effectors and are insensitive to NAG [13,21,22]. NAG activation involves a functionally crucial cross-talk between NAG and phosphorylation sites. Thus, NAG vastly increases the

apparent affinity of CPS1 for ATP and for its ionic activators K^+ and Mg^{2+} [15] and, conversely, ATP, K^+ and Mg^{2+} greatly increase CPS1 affinity for NAG [23].

The complexities of the CPS1 reaction and regulation are backed by a sophisticated multidomain protein machinery [13], that is still imperfectly characterized, since the crystal structure of CPS1 has not been determined (except for its ~15 kDa C-terminal domain, representing only ~10% of the entire molecule) [24,25]. Nevertheless, a number of approaches including the use of limited proteolysis [26] support a domain composition for CPS1 that mirrors that of *Escherichia coli* CPS, the only CPS for which the structure is known [27], notwithstanding the fact that these two enzymes have very important differences [13,21]. These differences include the chain composition (a single chain in CPS1; two subunits in *E. coli* CPS), the already indicated differences in the requirement for NAG, and the fact that CPS1 uses ammonia with high efficiency and cannot utilize glutamine as an ammonia source, whereas bacterial CPS uses ammonia poorly but utilizes glutamine as an internal ammonia source. The small subunit of the bacterial enzyme binds and cleaves the glutamine, channeling the resulting ammonia to the large subunit, where the entire carbamoyl phosphate synthesis reaction from ammonia takes place [21].

As already mentioned, limited proteolysis and other studies support the similarity in domain composition and function of bacterial CPS and CPS1 (Fig. 1A). These studies evidenced [13,26,28] that CPS1 is composed of N- and C-terminal moieties of, respectively, ~40 and ~120 kDa, that correspond to the small and large subunits of *E. coli* CPS [12]. Since CPS1 uses no glutamine [10], no functions for the N-terminal ~40-kDa moiety are known other than possibly some enzyme stabilization and activation [26,29]. In contrast, the ~120 kDa moiety is known to host the catalytic and NAG-regulatory machinery [13,21], being composed of two ~60-kDa halves, each

one of them consisting of an N-terminal ~45-kDa phosphorylation domain followed by a ~15 kDa domain [13]. Both phosphorylation domains are homologous [12]. The one in the N-half phosphorylates bicarbonate and the other in the C-half phosphorylates carbamate (Fig. 1B) [30,31]. The C-terminal ~15-kDa domain (residues 1354-1500) of the C-half (the C-terminal domain of the enzyme) binds NAG and is crucial for activation, having been called the allosteric domain (ASD) [25,26,32], not hosting substrate sites or catalytic machinery. The other ~15 kDa domain (precise mass, 17.7 kDa; residues 822-973) connects both phosphorylation domains and its function is unclear since it has no substrate sites or catalytic components, having been called unknown function subdomain (UFSD) [26]. Interestingly, the UFSD has been found to host missense mutations in CPS1D patients (Fig. 1C,D) with rather high frequency [2], suggesting an important and until now unclear role of this domain.

CPS1D has been associated with a relatively large number of different, generally "private" mutations which occur in single families with very little recurrence [2]. Many of the >130 missense mutations reported in CPS1D remain to be proven responsible for the deficiency. The fact that they appear to be distributed non-homogeneously among the different gene exons even after correcting for the presence of CpG islands, suggests that some enzyme regions have a more important role on enzyme stability, folding or functionality than other regions, and that, therefore, mutations falling on these more important regions have higher repercussion than those falling in less critical regions of the protein. The UFSD domain, particularly its C-terminal half, may be such an important region since the relative occurrence of missense mutations (normalized per 100 nucleotides) in exons 22 and 23, which encode most of the C-terminal half of the UFSD and which have no CpG islands, approximately doubles the relative frequency of these mutations for the entire coding sequence of the enzyme ($p < 0.01$; χ^2 test) [2].

We exploit here our recent ability to produce pure mature human CPS1, either wild-type or with the desired mutations, in a baculovirus/insect cell system [26], to examine the disease-causing potential of all known UFSD missense mutations (n=18) found in CPS1D patients [2,4,33-37] and to clarify the role and importance of the UFSD. When expression was possible, we studied the properties of the purified mutant enzyme forms, comparing them with wild-type human CPS1. Our findings support the disease-causing role of the mutations reported to affect the UFSD, revealing a key role of the UFSD for proper enzyme folding and for the regulatory cross-talk between NAG and phosphorylation sites.

2. Materials and Methods

2.1 Patients and CPS1 Mutations

Table 1 lists 18 amino acid substitutions found in patients with CPS1D which are localized in the UFSD, giving some detail on the type of the clinical presentation in the patients carrying these mutations. Seventeen of the mutations were reported already [2,4,33-37], whereas one (mutation 13) is a novel change identified by one of us (J.H.) in a late onset CPS1D patient. We also report a patient with a neonatal presentation carrying an already reported mutation (p.Asp914His), a mutation for which there was no clinical information on the previously reported patient. In contrast, the new patient was known to have a neonatal presentation, indicating that the mutation was probably severe. The missense mutations in these two new patients were initially identified by mRNA studies in phytohemagglutinin-stimulated lymphocytes and then were confirmed by studies on genomic DNA as previously reported [37]. Samples were obtained with full informed consent of those entitled to give it, to perform molecular genetic diagnostics for clinical purposes. The PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) [38] and MutPred (<http://mutpred.mutdb.org/>) servers [39] were used to assess in silico the disease-causing potential of the different mutations. Amino-acid conservation was determined by ClustalW sequence alignment of either CPS1, CPSIII or other CPSs from 14, 6 and 24 species, respectively.

2.2 Production of CPS1 carrying the desired mutation

The indicated mutations were introduced into pFastBac-CPS1 [26] using the overlapping extension method (Quickchange kit from Stratagene) and the forward and reverse primers given in Supp. Table S1. The correctness of the constructs, the presence

of the desired mutation, and the absence of unwanted mutations were corroborated by sequencing.

Human mature liver CPS1, either wild-type or carrying the desired mutation, with the N-terminal mitochondrial targeting sequence replaced by the 28-residue N-terminal His₆-tag, MSYYHHHHHHHDYDIPTTENLYFQGAMDP, was expressed in a baculovirus/insect cell system and purified as previously described [26]. The same purification procedure (cell centrifugation, cell lysis, centrifugal clarification of the extract, Ni-affinity chromatography and centrifugal ultrafiltrative concentration [26]) was used for wild-type CPS1 and for the mutant enzyme forms. Protein in the final enzyme preparation was determined according to Bradford [40]. The CPS1 yield was determined from the total amount of protein in the enzyme preparation and the fraction of the protein corresponding to pure CPS1. The latter fraction was determined densitometrically from the band of approximately 163 kDa observed in Coomassie-stained 8% polyacrylamide gels after SDS-PAGE [41]. For this purpose, stained gel images were collected with a Fujifilm LAS-3000 Imager using transmitted white light, and the fraction of the protein migrating in the CPS1 band of ~163 kDa was determined with the Multi Gauge quantification software (from Fuji Film).

2.3 Enzyme activity assays

CPS1 activity was assayed at 37°C as carbamoyl phosphate production, monitored as citrulline by using OTC [26]. The standard assay mixture contained 50 mM glycyl-glycine pH 7.4, 70 mM KCl, 1 mM dithiothreitol, 20 mM MgSO₄, 5 mM ATP, 35 mM NH₄Cl, 50 mM KHCO₃, 10 mM NAG, 5 mM L-ornithine and 4 U/ml OTC. Specific activities refer to the pure enzyme (determined densitometrically, see section 2.2). When the concentration of a substrate was varied, other substrates were

kept at the concentrations given above, with MgSO_4 being in 20 mM excess over ATP. Data for variable substrate or NAG concentration were fitted to hyperbolae using the GraphPad Prism program (GraphPad Software, San Diego, CA). One enzyme unit makes 1 μmol citrulline per minute. Specific activities refer to pure CPS1, estimated by densitometry (see above).

2.4 Other assays

The thermal stability of CPS1 was monitored by incubating for 15 min at the specified temperature 0.5 mg protein/ml of the indicated CPS1 form (either wild-type or mutant), in a solution containing 50 mM glycyl-glycine pH 7.4, 2 mM dithiothreitol, 10% glycerol, 0.5 M NaCl and 20 mM imidazole. At the end of the incubation the solution was rapidly cooled at 0°C and enzyme activity was determined immediately in the standard assay at 37°C.

Western blotting was carried out after SDS-PAGE (section 2.2) as reported [26] using immunoluminescent detection (Super Signal West Pico Chemiluminiscent Substrate, Thermo Scientific).

The previously reported [42] atomic structure model of CPS1 (based on the experimental crystal structure of *E. coli* CPS) was kindly provided by B. Barcelona (Instituto de Biomedicina de Valencia) and is used here for structural analysis of the mutations effects. Pymol (DeLano Scientific; <http://www.pymol.org>) was used for visual analysis, for structural superimposition and for depicting protein structures.

3. Results

3.1. CPS1D mutations that affect the UFSD

Table 1 summarizes the 17 already reported [2,4,33-37] mutations reported in CPS1D patients that affect the UFSD, plus one more mutation (#13) falling in this domain which is reported here in a late onset patient. Thirteen of these mutations were found in patients with neonatal or with very severe presentations and thus might be expected to have drastic negative effects on CPS1 level or activity. Only mutations 4, 13, 14, 16 and 17 were found exclusively in late onset patients or in patients without appropriate clinical information to judge about their severity. The fact that two mutations affecting the same residue were reported for Arg850, Gly911 and Asp914, three residues that are invariant among all CPSs, strongly supports a damaging role of the corresponding six mutations (p.Arg850Cys, p.Arg850His, p.Gly911Glu, p.Gly911Val, p.Asp914His, and p.Asp914Gly), in agreement with the predictions of the Polyphen-2 and MutPred servers (Table 1). In another patient with a neonatal presentation, the p.Leu843Ser and p.Lys875Glu mutations, both mapping in the UFSD, coexisted in the same allele [4]. Of these, the p.Leu843Ser mutation appears more likely to be disease-causing, because of the high conservation of Leu843 and because Polyphen-2 and MutPred give p.Leu843Ser a higher degree of probability and a higher score for having a damaging role than they do for p.Lys875Glu. The high residue conservation and the Polyphen-2 and MutPred server predictions also support the disease-causing nature of the p.Arg932Thr and p.Leu958Pro mutations, found in another two patients with neonatal presentations, whereas a disease-causing role appears less certain for the p.Ser913Leu, p.Ser918Pro and p.Gly964Asp mutations, found in other neonatal patients, since the residues affected are less conserved and the Polyphen-2 and MutPred servers are non-unanimous in anticipating a disease-causing role (Table

1). The lack of unanimity in the predictions by both servers is possible, since they do not utilize the same traits [38,39] for assessing the disease-causing role of a given mutation.

The mutations p.Thr871Pro, p.Ile937Asn, p.Ala949Thr, p.Tyr959Cys and p.Tyr962Cys, reported in CPS1D patients with late presentations or with unreported presentation but with patient survival (suggesting less severe phenotype), affect invariant or virtually invariant residues and give high probability scores for being damaging in the Polyphen-2 and MutPred servers. Since the p.Ala949Thr mutation coexists with a null allele (p.Tyr89*) in the patient [2], it cannot abolish enzyme activity since otherwise the patient would have had a neonatal presentation. In principle, the same can be said for the p.Ile937Asn mutation, given the fact that the other *CPS1* allele found in the same patient corresponds to a very severe splice site aberration that should lead to exon 27 skipping, with frameshift and truncation. Since in this patient the *CPS1* allele hosting the p.Ile937Asn mutation also hosts a second missense change, p.Gly401Arg (Table 1), this second change cannot be expected to be highly deleterious either. In contrast, a drastic effect is likely for the p.Thr871Pro mutation since it coexists in the patient with the expectedly mild mutation (given the mildness of the substitution), p.Glu1194Asp [37].

3.2. UFSD mutations negatively influence the yield of CPS1

We previously used the baculovirus/insect cell system utilized here for producing twenty CPS1 forms carrying different missense changes identified in CPS1D patients and spread all over the CPS1 protein except the UFSD [26,43]. We also reported [25] the effects of another five non-CPS1D-associated but functionally important mutations affecting the C-terminal domain (ASD). Furthermore, in

unpublished studies, we have monitored the effects of ten more mutations (nine in the ASD) of which five had been identified in patients. Only three among these 35 experimentally tested mutations compromised drastically recombinant CPS1 production ([26], and data not shown). The present UFSD mutations yield an entirely different picture (Fig. 2A). Recombinant production of CPS1 protein was virtually abolished by six mutations and it was decreased 85-95% and ~75% by five and four mutations, respectively. Only for three of the eighteen mutations studied here (p.Arg850Cys, p.Arg850His and p.Ala949Thr) the yield of pure CPS1 was comparable to that of the wild-type enzyme. Interestingly, each one of the two mutations found in the same allele of one patient, p.Leu843Ser and p.Lys875Glu (Table 1), reduced the yield of CPS1 by nearly 90% (Fig. 2A), raising the possibility that their joint presence in the same protein molecule might reduce further CPS1 production.

The low yields observed with most mutants appear to be due to poor CPS1 polypeptide production and to gross misfolding. Thus, SDS-PAGE of the pellet (Fig. 2B, *P*) and the supernatant (Fig. 2B, *S*) obtained by centrifugation of the initial cell extract not only revealed for most mutants decreased CPS1 levels, but also showed that the CPS1 present tended to appear in the precipitate. Furthermore, a multiplicity of bands of lower molecular weight than complete CPS1, corresponding to CPS1 degradation products, were revealed by western blotting in the precipitates of these low-yield mutants (Fig. 2C).

In summary, the majority of the mutations found in CPS1D patients that affect the UFSD appear to disturb CPS1 folding and to enhance CPS1 degradation. The drastic decrease in enzyme production should be disease-causing at least for the seven mutations that reduce enzyme production by >90% (p.Thr871Pro, p.Gly911Glu, p.Gly911Val, p.Asp914His, p.Asp914Gly, p.Ser918Pro and p.Gly964Asp; Fig. 2A).

Indeed, six of these mutations were found in patients with neonatal or very severe presentations, and the other mutation (p.Thr871Pro), although found in a late onset patient, coexisted with an expectedly mild second allele and is likely to have by itself a severe effect (Table 1). On the other hand, the two mutations found in late onset patients that were anticipated to be mild (see section 3.1), p.Ala949Thr and p.Ile937Asn, were associated with substantial or normal yield of CPS1 (Fig. 2A).

3.3. Eight UFSD mutations strongly impair CPS1 activity

Although with most mutations little soluble CPS1 was produced, we always carried out the complete CPS1 purification protocol, observing by SDS-PAGE of the final preparation, with all the mutants (Figs. 2D,E), a 163 kDa-band corresponding to soluble CPS1 (confirmed by western blotting, data not shown). However, the purity of CPS1 decreased with the yield, being <10% of the protein (densitometric estimation) for the mutants with the lowest yield (Figs. 2D,E). Despite the presence of the CPS1 band, CPS1 activity was very low in the final preparations of the mutants p.Thr871Pro, p.Gly911Glu, p.Gly911Val, p.Asp914His, p.Asp914Gly, p.Ser918Pro, p.Leu958Pro and p.Gly964Asp. The estimated specific activity of pure soluble CPS1 was $\leq 6\%$ of wild-type for six of these mutations and $\leq 12\%$ for the other two (Fig. 3A). Assuming that the mutation-induced misfolding causes similar decreases in the production of soluble CPS1 protein in insect cells and in the liver, the low yields (Fig. 2A), combined with the decreased activities (Fig. 3A), lead to an expectation of very low residual activities in the liver (<1% of normal for homozygosity) for any of these eight mutant forms. The low residual activity further supports the disease-causing nature of these mutations and agrees with their observation in neonatal or very severely affected patients (Table 1).

Only the p.Thr871Pro mutation was observed in a late onset patient, but, as already indicated, the second allele in this patient was expectedly mild (p.Glu1194Asp) .

3.4. Activity and stability changes with another eight UFSD mutations

Eight of the ten mutations not dealt with in section 3.3 generally were associated with less drastic but substantial decreases in specific activity, and/or with decreased thermal stability of the enzyme (Fig. 3). Among these mutations, p.Ile937Asn caused the largest specific activity decrease (~90% decrease) (Fig. 3A), which combined with the observed 75% reduction in yield (Fig. 2A), would result (making the already indicated assumption that the yield in insect cells grossly corresponds to the yield in the liver) in ~3% residual activity for homozygosity, supporting disease-causation. Nevertheless, the role of the other mutation found in the same allele (Table 1), p.Gly401Arg, remains to be determined, although this mutation maps in the CPS1 N-terminal region of unknown function (Fig. 1) which hosts few CPS1D mutations [2], and it is predicted by Polyphen-2 and MutPred as benign or as having little probability of being disease-causing (not shown).

In the other case (Table 1) in which the same allele carried two missense mutations, p.Leu843Ser and p.Lys875Glu, the first of these mutations decreased activity by ~70% (Fig. 3A) and both mutations reduced ~90% the yield of CPS1, leading to the assumption of a liver activity of ~3% of normal for a null second allele (the case in the patient carrying these mutations [2]). This estimate might be optimistic considering the already mentioned possibility that the joint presence of both mutations in the same CPS1 polypeptide chain could reduce even further proper folding and CPS1 production, and also because the p.Lys875Glu mutation substantially decreased CPS1 thermal

stability (Fig. 3B). In summary, the data for this patient appear to account fully for its neonatal presentation.

Decreased thermal stability appears a major determinant of the effects of the p.Ala949Thr and p.Tyr959Cys mutations (Figs. 3B,C), being the only important aberration observed for the first of these two mutations. Thus, p.Ala949Thr was not associated with decreased yield (Fig. 2A) or with a large specific activity decrease (Fig. 3A), but it lowered the half-inactivation temperature of the enzyme by $\sim 8^{\circ}\text{C}$ (Fig. 3C), bringing it down to nearly 40°C , which is close to the physiological body temperature, leading to the expectation of an important reduction in enzyme half-life and thus of enzyme level in the tissue. Nevertheless, some residual activity should be expected, agreeing with the late onset presentation in the patient carrying this mutation (the other allele carried a null mutation, Table 1). A similar degree of reduction in thermal stability was also observed for the p.Tyr959Cys mutation (Fig. 3B,C). However, in this case the mutation also decreased importantly CPS1 yield (by 75%; Fig. 2A) and specific activity (by 70%; Fig. 3A), with the resultant final activity in the tissue possibly being decreased further by the accelerated thermal inactivation. The late onset presentation in this patient (Table 1) could be due to the residual activity from this mutation, together with that resulting from the mutation carried in the second *CPS1* allele (p.Pro1462Arg), which is not inactivating (our own unpublished data).

Although the p.Tyr962Cys mutation only reduced $\sim 50\%$ the specific activity of the soluble enzyme (Fig. 3A) without substantially decreasing thermal stability (Fig. 3B), it decreased $\sim 75\%$ CPS1 yield (Fig. 2A). The combination of these two detrimental effects might result in an activity in the tissue of $\sim 6\%$ of normal, provided that the second *CPS1* allele found in the patient carrying this mutation (p.Ile632Arg; #17, Table

1), which affects the bicarbonate phosphorylation domain, is inactivating. This residual activity would explain that patient #17 is alive.

The patients with the p.Ser913Leu or p.Arg932Thr mutations had neonatal presentations (Table 1) that must reflect a cumulative effect of modest to drastic but not extreme changes in yield or specific activity of CPS1 caused by these mutations. Thus, the p.Ser913Leu mutation decreased CPS1 yield and specific activity, respectively, to ~20% and ~30% of normal, and the p.Arg932Thr mutation caused respective decreases to ~12% and 40% of normal (Figs. 2A and 3A). A residual liver activity of ~3% of normal is conceivable for both mutations, if they coexist in the patients with null second *CPS1* alleles.

3.5. K_m effects are the major changes associated with mutations affecting Arg850.

Changes in yield (Fig. 2A), specific activity (Fig. 3A) or thermal stability (Fig. 3B) cannot account for the neonatal presentations observed in the two patients that carried mutations affecting Arg850, since for the p.Arg850His and p.Arg850Cys mutations thermal stability was essentially normal and the yields and specific activities were at least ~50% and ~25% of normal, respectively. However, since the specific activity assay used here utilizes saturating (for the wild-type enzyme) substrate concentrations, there might be room for K_m or K_a effects that could have negative consequences on enzyme activity at the generally low substrate concentrations present in the tissue [44,45]. This is certainly the case for the p.Arg850Cys mutation, which exhibits ~5-fold and ~3-fold increases in the K_m values for ATP and bicarbonate, respectively, and, even more importantly, a ~18-fold increase in the K_a value for NAG (Fig. 4A-C,E,G). Similarly, in the case of the p.Arg850His mutation the K_m for ATP and the K_a for NAG were increased 3-fold and 10-fold, respectively (Fig. 4A,B,F).

Although these kinetic constant changes are not large enough to decrease importantly the activity observed in the standard assay, they can lead to drastic reductions in the CPS1 activity *in vivo*, since, for example, NAG may be far from saturating even for the wild-type enzyme [17-19,44,45], and thus, much less saturating for these mutant enzyme forms.

Among the ten UFSD mutants in which CPS1 yield and activity permitted kinetic analysis (Figs. 4A-D), substantial K_m or K_a changes were rarely observed. In addition to the already indicated changes in the two mutants of Arg850, the p.Ile937Asn mutation modestly increased the K_a value for NAG (Fig. 4A) and the K_m for ammonia (Figs. 4D,H), and the p.Leu843Ser and p.Ser913Leu mutations increased the K_a for NAG (Fig. 4A). It is interesting that the K_a for NAG was increased in all cases that exhibited a kinetic change, suggesting a role of the UFSD in the cross-talk known to exist between the NAG site and the catalytic machinery of the enzyme [15,23]. In summary, while K_m or K_a changes do not appear to play a paramount role in decreasing enzyme activity with most UFSD mutants, they certainly do so with the two mutations affecting Arg850, for which they appear major determinants of the severe deficiency observed in the patients carrying these mutations.

3.6. UFSD mutations generally decrease V_{max} of the enzyme

With those mutations in which kinetic analysis was possible, it was observed that, except in the case of p.Lys875Glu, the mutations modestly decreased V_{max} (Fig. 5A) in a proportion that was similar to the decrease in specific activity (Fig. 3A). This was to be expected if the mutations affect mainly k_{cat} without producing very large K_m effects, since the standard activity assay utilizes saturating (for the wild-type enzyme) concentrations of the substrates and of NAG. A plot of specific activity versus V_{max}

(Fig. 5B) could be reasonably adjusted to linear regression passing through zero. This suggests that the very low activity of the mutants that were very little expressed and that, therefore, were not amenable to kinetic analysis, could be due to a $V_{\max}$ effect. Thus, the UFSD domain appears to influence the rate at which the enzyme catalyzes its complex, three-step reaction.

4. Discussion

The present results validate the baculovirus/insect cell expression system [46] for monitoring the consequences of CPS1D mutations. Such system is already the standard for the same task in Gaucher's disease [47,48] and has been used with other diseases [49-52]. It associates the abundant expression and possibility of purification that characterizes *E. coli* expression, with a high compatibility with human proteins, which frequently cannot be expressed in bacteria [46]. In the case of Gaucher's disease, the large experience (exemplified in [47, 48]) attests the concordance of the results in this expression system with the clinical phenotype. This concordance was prospectively proven with two polymorphisms of methylenetetrahydrofolate reductase [50], since the one (p.Ala222Val) shown to have negative effects on the enzyme when using baculovirus/insect cells, was later on proven by epidemiological and genome-wide association studies to be linked to increased plasma homocysteine levels [53]. Our experience with CPS1 also supports the faithfulness of this system in revealing the impact of *CPS1* mutations. Thus, in our prior studies [26] two *CPS1* polymorphisms believed to be trivial had no substantial effect on any CPS1 trait investigated (activity, stability, kinetic constants, gross estimation of yield), whereas two mutations affecting a catalytic domain and used as positive controls were strongly detrimental. Furthermore, the results with three CPS1D-associated mutations affecting the glutaminase-like subdomain (GSD) (Fig. 1A) of unknown function and of three additional mutations affecting the allosteric domain (ASD, Fig. 1A) accounted for disease-causation because of impaired activity, stability or NAG activation. These results, which followed earlier pilot studies using baculovirus/insect cell-expressed rat CPS1 (a surrogate of human CPS1) [43], are extended now with those for 18 mutations affecting the UFSD.

The present results confirm that the seventeen mutations reported earlier [2,4,33-37] and the one described here that affect the UFSD of CPS1 in patients with CPS1D are disease-causing. The observed effects of the mutations on the production, activity, thermal stability and kinetic constants of the recombinant enzyme match in most cases the severity of the clinical presentation. Our data illustrate in one case the fact that each one of two missense changes in the same allele (p.Leu843Ser and p.Lys875Glu) has a detrimental effect on the enzyme, raising the possibility that their combined presence in the same protein molecule could be even more detrimental.

A clear conclusion of our studies is that, at least for this domain and with this eukaryotic expression system, the determination of the yield of CPS1 protein is a key element in judging the disease-causing role of each mutation. We clearly show that the abolition or the drastic decrease of CPS1 production is a crucial mechanism of disease production by many missense mutations affecting the UFSD. Our findings point to decreased efficiency of proper folding as a key determinant of the poor production of the soluble protein. Although there are few *in vitro* expression studies for CPS1 [25,26,29,43], our prior investigations with 35 missense mutations mapping in other domains (with a predominance of mutations affecting the C-terminal ASD) do not highlight poor CPS1 production as a key determining element of the effects of the mutations affecting these other protein regions, since only with ~10% of the mutations was CPS1 production hampered or abolished. It therefore appears that a decrease in the efficiency of proper CPS1 folding is a characteristic and remarkable trait of CPS1D-associated mutations affecting the UFSD. Such trait may be the reason for the prominence of this domain in CPS1D, where missense mutations have been found in relatively high density [2].

The central position in the enzyme architecture and the intimate relations of the UFSD with other CPS1 domains [2,27,42] may account for the particular impact of the mutations affecting this domain on CPS1 folding. In the existing structural CPS1 models [2,42] based on the *E. coli* CPS structure [27], the UFSD makes very extensive contacts with both phosphorylation domains and, to a lesser extent, with the N-terminal moiety of the enzyme (the small subunit-like region) (Figs. 1C,D). This L-shaped domain (Fig. S1A) which embraces between its two arms the C subdomain of the bicarbonate phosphorylation domain (Fig. S1D), is sandwiched between the small subunit-like N-terminal moiety and the carbamate phosphorylation domain, two domains that lie respectively on top and below the plane defined by this L (Fig. 1D, right panel). Therefore, mutation-triggered UFSD misfolding can be expected to result in distorted relations between these other protein domains. The CPS1D-associated mutations affecting this domain cluster in the regions of interaction of the UFSD with other domains rather than in the exposed regions (Fig. 1D) that are involved in intermolecular interactions [27]. Therefore, the distribution of the mutations does not support the possibility of an increase in the interactions between the UFSDs of different enzyme molecules as the cause for the decreased solubility observed with many of these mutants. This agrees with our observation with *E. coli* CPS that the UFSD is not involved in dimer formation [54], and with our finding that human CPS1 exists mainly as monomers [10,26].

A detailed structural rationalization of the effects of the mutations studied here is included as Supplementary material. In summary, our results highlight a paramount integrating role of the UFSD for building the highly complex CPS1 architecture, revealing the key organizing function of this domain, and exemplifying the importance of this core structural element despite its lack of substrate binding and catalytic

machinery. Interestingly, a recent in silico study already proposed for this domain a key structural role in CPS1 [55]. On the basis of such key function of this domain, which is conserved in all forms of CPS (including the archaeal types, in which the C-terminal moiety is split into two complementary regions encoded by different genes), we would propose to call it in the future the "Integrating Domain" of CPS1. Our data fully account for the important representation of this domain in the database of missense mutations found in CPS1D [2], revealing that the main effects of the mutations in this domain are to negatively affect global CPS1 folding and architecture. We also document for some mutations and propose for other mutations that they have effects of greater or lower magnitude on the $V_{\max}$ of the enzyme. This effect could be due to gross defects on domain architecture, that, for example, could block the conduit of the activating allosteric NAG signal, or the tunnel [27] through which intermediates must flow between both phosphorylation centers; or it could derive from more subtle changes such as the hampering of the concerted opening [56,57] of the B subdomains of both phosphorylation domains to allow product release. Indeed, mutations in the UFSD could be in the signaling path between both phosphorylation centers since the UFSD contacts the B domain of the carbamate phosphorylation domain, which is believed to trigger the concerted opening [57], and it also contacts the active site of the bicarbonate phosphorylation domain, which is to be opened (Fig. 1D). Similarly, the UFSD appears to be involved in the cross-talk between the NAG site and the catalytic centers, since some UFSD mutations increase the K_a for NAG. Full understanding of this involvement will have to await the determination of the structural mechanism of NAG activation of CPS1.

Conflict of interest statement

The authors state that there is no conflict of interest.

Acknowledgements

We thank Belén Barcelona (IBV-CSIC, Valencia) for her structural CPS1 model, and María Pilar Alberó (Centro de Investigación Príncipe Felipe, Valencia) for initial data on p.Arg850Cys and p.Arg850His mutants. This work was supported by grants from the Fundación Alicia Koplowitz, the Valencian and Spanish governments (Prometeo 2009/051 and BFU2011-30407, respectively) and the Swiss National Science Foundation (grant 310030_127184). C.D-F was a FPU fellow of the Spanish Government and received from that Government a bursary for short-time work in Zurich.

References

- [1] S. Brusilow, A.L. Horwich, Urea Cycle Enzymes, in: C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle (Eds.), *The Metabolic & Molecular Bases of Inherited Disease*, 8th Edition, McGraw-Hill, New York, 2001, pp. 1909-1963.
- [2] J. Häberle, O.A. Shchelochkov, J. Wand, P. Katsonis, L. Hall, S. Reiss, A. Eeds, A. Willis, M. Yadav, S. Summar, O. Lichtarge, V. Rubio, L.J. Wong, M. Summar, Molecular defects in human carbamoyl phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations, *Hum Mutat.* 32 (2011) 579-589.
- [3] M.L. Summar, M.J. Dasouki, P.J. Schofield, M.R. Krishnamani, C. Vnencak-Jones, M. Tuchman, J. Mao, J.A Phillips 3rd, Physical and linkage mapping of human carbamyl phosphate synthetase I (CPS1) and reassignment from 2p to 2q35, *Cytogenet Cell Genet.* 71 (1995) 266-267.
- [4] J. Häberle, E. Schmidt, S. Pauli, B. Rapp, E. Christensen, B. Wermuth, H.G. Koch, Gene structure of human carbamylphosphate synthetase 1 and novel mutations in patients with neonatal onset, *Hum Mutat.* 21 (2003) 444.
- [5] M.L. Summar, L.D. Hall, A.M. Eeds, H.B. Hutcheson, A.N. Kuo, A.S. Willis, V. Rubio, M.K. Arvin, J.P Schofield, E.P. Dawson, Characterization of genomic structure and polymorphisms in the human carbamyl phosphate synthetase I gene, *Gene.* 311 (2003) 51-57.
- [6] S. Funghini, M.A. Donati, E. Pasquini, E. Zammarchi, A. Morrone, Structural organization of the human carbamyl phosphate synthetase I gene (CPS1) and identification of two novel genetic lesions, *Hum Mutat.* 22 (2003) 340-341.
- [7] Y. Haraguchi, T. Uchino, M. Takiguchi, F. Endo, M. Mori, I. Matsuda, Cloning and sequence of a cDNA encoding human carbamyl phosphate synthetase I: molecular analysis of hyperammonemia, *Gene.* 107 (1991) 335-340.

- [8] E.H. Van Beers, E.H. Rings, G. Posthuma, M.A. Dingemanse, J.A. Taminiau, H.S. Heymans, A.W. Einerhand, H.A. Büller, J. Dekker, Intestinal carbamoyl phosphate synthase I in human and rat. Expression during development shows species differences and mosaic expression in duodenum of both species, *J Histochem Cytochem.* 46 (1998) 231-240.
- [9] M.A. Neill, J. Aschner, F. Barr, M.L. Summar, Quantitative RT-PCR comparison of the urea and nitric oxide cycle gene transcripts in adult human tissues, *Mol Genet Metab.* 97 (2009) 121-127.
- [10] V. Rubio, G. Ramponi, S. Grisolia, Carbamoyl phosphate synthetase I of human liver. Purification, some properties and immunological cross-reactivity with the rat liver enzyme, *Biochim Biophys Acta.* 659 (1981) 150-160.
- [11] M. Mori, S. Miura, T. Morita, M. Takiguchi, M. Tatibana, Synthesis, intracellular transport and processing of mitochondrial urea cycle enzymes, *Adv Enzyme Regul.* 21 (1982) 121-132.
- [12] H. Nyunoya, K.E. Broglie, E.E. Widgren, C.J. Lusty, Characterization and derivation of the gene coding for mitochondrial carbamyl phosphate synthetase I of rat, *J Biol Chem.* 260 (1985) 9346-9356.
- [13] V. Rubio, Structure-function studies in carbamoyl phosphate synthetases, *Biochem Soc Trans.* 21 (1993) 198-202.
- [14] L.M. Hall, R.L. Metzenberg, P.P. Cohen, Isolation and characterization of a naturally occurring cofactor of carbamyl phosphate biosynthesis, *J Biol Chem.* 230 (1958) 1013-1021.
- [15] V. Rubio, H.G. Britton, S. Grisolia, Mitochondrial carbamoyl phosphate synthetase activity in the absence of N-acetyl-L-glutamate. Mechanism of activation by this cofactor, *Eur J Biochem.* 134 (1983) 337-343.

- [16] C. Bachmann, S. Krähenbühl, J.P. Colombo, Purification and properties of acetyl-CoA: L-glutamate N-acetyltransferase from human liver, *Biochem J.* 205 (1982) 123-127.
- [17] M. Tatibana, S. Kawamoto, T. Sonoda, M. Mori, Enzyme regulation of N-acetylglutamate synthesis in mouse and rat liver, *Adv Exp Med Biol.* 153 (1982) 207-216.
- [18] K. Shigesada, K. Aoyagi, M. Tatibana, Role of acetylglutamate in ureotelism. Variations in acetylglutamate level and its possible significance in control of urea synthesis in mammalian liver, *Eur J Biochem.* 85 (1978) 385-391.
- [19] P.M. Stewart, M. Walser, Short term regulation of ureagenesis, *J Biol Chem.* 255 (1980) 5270-5280.
- [20] P.M. Anderson, A glutamine- and N-acetyl-L-glutamate-dependent carbamyl phosphate synthetase activity in the teleost *Micropterus salmoides*, *Comp Biochem Physiol B.* 54 (1976) 261-263.
- [21] A. Meister, Mechanism and regulation of the glutamine-dependent carbamyl phosphate synthetase of *Escherichia coli*, *Adv Enzymol Relat Areas Mol Biol.* 62 (1989) 315-374.
- [22] X. Liu, H.I. Guy, D.R. Evans, Identification of the regulatory domain of the mammalian multifunctional protein CAD by the construction of an *Escherichia coli* hamster hybrid carbamyl-phosphate synthetase, *J Biol Chem.* 269 (1994) 27747-27755.
- [23] E. Alonso, V. Rubio, Binding of N-acetyl-L-glutamate to rat liver carbamoyl phosphate synthetase (ammonia), *Eur J Biochem.* 135 (1983) 331-337.
- [24] Y. Xie, K. Ihsanawati, S. Kishishita, K. Murayama, C. Takemoto, M. Shirozu, RIKEN Structural Genomics/Proteomics Initiative, Crystal structure of

MGS domain of carbamoyl-phosphate synthetase from *Homo sapiens*. (February-2009). Protein DataBank file 2YVQ, <http://www.rcsb.org/pdb>.

[25] S. Pekkala, A.I. Martinez, B. Barcelona, J. Gallego, E. Bendala, I. Yefimenko, V. Rubio, J. Cervera, Structural insight on the control of urea synthesis: identification of the binding site for N-acetyl-L-glutamate, the essential allosteric activator of mitochondrial carbamoyl phosphate synthetase, *Biochem J.* 424 (2009) 211-220.

[26] C. Diez-Fernandez, A.I. Martinez, S. Pekkala, B. Barcelona, I. Perez-Arellano, A. Guadalajara, M. Summar, J. Cervera, V. Rubio, Molecular characterization of carbamoyl-phosphate synthetase (CPS1) deficiency using human recombinant CPS1 as a key tool, *Hum Mutat.* 8 (2013) 1149-1159.

[27] J.B. Thoden, H.M. Holden, G. Wesenberg, F.M. Raushel, I. Rayment, Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product, *Biochemistry.* 36 (1997) 6305-6316.

[28] S.G. Powers-Lee, K. Corina, Domain structure of rat liver carbamoyl phosphate synthetase I, *J Biol Chem.* 261(1986) 15349-15352.

[29] V. Ahuja, S.G. Powers-Lee, Human carbamoyl-phosphate synthetase: insight into N-acetylglutamate interaction and the functional effects of a common single nucleotide polymorphism, *J Inherit Metab Dis.* 31 (2008) 481-491.

[30] E. Alonso, J. Cervera, A. García-España, E. Bendala, V. Rubio, Oxidative inactivation of carbamoyl phosphate synthetase (ammonia). Mechanism and sites of oxidation, degradation of the oxidized enzyme, and inactivation by glycerol, EDTA, and thiol protecting agents, *J Biol Chem.* 267 (1992) 4524-4532.

[31] E. Alonso, V. Rubio, Affinity cleavage of carbamoyl-phosphate synthetase I localizes regions of the enzyme interacting with the molecule of ATP that phosphorylates carbamate, *Eur J Biochem.* 229 (1995) 377-384.

- [32] L.B. Rodriguez-Aparicio, A.M. Guadalajara, V. Rubio, Physical location of the site for N-acetyl-L-glutamate, the allosteric activator of carbamoyl phosphate synthetase, in the 20-kilodalton COOH-terminal domain, *Biochemistry*. 28 (1989) 3070-3074.
- [33] Y. Wakutani, H. Nakayasu, T. Takeshima, M. Adachi, M. Kawataki, K. Kihira, H. Sawada, M. Bonno, H. Yamamoto, K. Nakashima, Mutational analysis of carbamoylphosphate synthetase I deficiency in three Japanese patients, *J. Inherit. Metab. Dis.* 27 (2004) 787–788.
- [34] A.M. Eeds, L.D. Hall, M. Yadav, A. Willis, S. Summar, A. Putnam, F. Barr, M.L. Summar, The frequent observation of evidence for nonsense-mediated decay in RNA from patients with carbamyl phosphate synthetase I deficiency, *Mol. Genet. Metab.* 89 (2006) 80–86.
- [35] K. Kurokawa, 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, *J. Hum. Genet.* 52 (2007) 349–354.
- [36] S. Funghini, J. Thusberg, M. Spada, S. Gasperini, R. Parini, L. Ventura, C. Meli, L. De Cosmo, M. Sibilio, S.D. Mooney, R. Guerrini, M.A. Donati, A. Morrone, Carbamoyl phosphate synthetase 1 deficiency in Italy: clinical and genetic findings in a heterogeneous cohort, *Gene*. 493 (2012) 228-234.
- [37] R. Kretz, L. Hu, V. Wettstein, D. Leiteritz, J. Häberle, Phytohemagglutinin stimulation of lymphocytes improves mutation analysis of carbamoylphosphate synthetase 1, *Mol Genet Metab.* 106 (2012) 375-388.

- [38] I.A. Adzhubei, S. Schmidt, L. Peshkin, V.E. Ramensky, A. Gerasimova, P. Bork, A.S. Kondrashov, S.R. Sunyaev, A method and server for predicting damaging missense mutations, *Nat Methods*. 7 (2010) 248-249.
- [39] B. Li, V.G. Krishnan, M.E. Mort, F. Xin, K.K. Kamati, D.N. Cooper, S.D. Mooney, P. Radivojac, Automated interference of molecular mechanisms of disease from amino acid substitutions, *Bioinformatics*. 25 (2009) 2744-2750.
- [40] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem*. 72 (1976) 248-54.
- [41] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature*. 227 (1970) 680-685.
- [42] A.I. Martínez, I. Pérez-Arellano, S. Pekkala, B. Barcelona, J. Cervera, Genetic, structural and biochemical basis of carbamoyl phosphate synthetase 1 deficiency, *Mol Genet Metab*. 101 (2010) 311-323.
- [43] S. Pekkala, A.I. Martinez, B. Barcelona, I. Yefimenko, U. Finckh, V. Rubio, J. Cervera, Understanding carbamoyl-phosphate synthetase I (CPS1) deficiency by using expression studies and structure-based analysis, *Hum Mutat*. 31 (2010) 801-808.
- [44] E. Alonso, J. Girbés, A. García-España, V. Rubio, Changes in urea cycle-related metabolites in the mouse after combined administration of valproic acid and an amino acid load, *Arch Biochem Biophys*. 272 (1989) 267-273.
- [45] A. García-España, E. Alonso, V. Rubio, Influence of anions on the activation of carbamoyl phosphate synthetase (ammonia) by acetylglutamate: implications for the activation of the enzyme in the mitochondria, *Arch Biochem Biophys*. 288 (1991) 414-420.

- [46] T.A. Kost, J.P. Condreay, D.L. Jarvis, Baculovirus as versatile vectors for protein expression in insect and mammalian cells, *Nat Biotechnol.* 23 (2005) 567-575.
- [47] M. Montfort, A. Chabás, L. Vilageliu, D. Grinberg, Functional analysis of 13 GBA mutant alleles identified in Gaucher disease patients: Pathogenic changes and "modifier" polymorphisms, *Hum Mutat.* 23 (2004) 567-575.
- [48] B. Liou, A. Kazimierczuk, M. Zhang, C.R. Scott, R.S. Hegde, G.A. Grabowski, Analyses of variant acid beta-glucosidases: effects of Gaucher disease mutations, *J Biol Chem.* 281 (2006) 4242-4253.
- [49] A.K. Das, J.Y. Lu, S.L. Hofmann, Biochemical analysis of mutations in palmitoyl-protein thioesterase causing infantile and late-onset forms of neuronal ceroid lipofuscinosis, *Hum Mol Genet.* 10 (2001) 1431-1439.
- [50] K. Yamada, Z. Chen, R. Rozen, R.G. Matthews, Effects of common polymorphisms on the properties of recombinant human ethylenetetrahydrofolate reductase, *Proc Natl Acad Sci U S A.* 98 (2001) 14853-14858.
- [51] T. Grau, N.O. Artemyev, T. Rosenberg, H. Dollfus, O.H. Haugen, E. Cumhur Sener, B. Jurklies, S. Andreasson, C. Kernstock, M. Larsen, E. Zrenner, B. Wissinger, S. Kohl, Decreased catalytic activity and altered activation properties of PDE6C mutants associated with autosomal recessive achromatopsia, *Hum Mol Genet.* 20 (2011) 719-730.
- [52] E. Ostergaard, R.J. Rodenburg, M. van den Brand, L.L. Thomsen, M. Duno, M. Batbayli, F. Wibbrand, L. Nijtmans, Respiratory chain complex I deficiency due to NDUFA12 mutations as a new cause of Leigh syndrome, *J Med Genet.* 48 (2011) 737-740.

- [53] K. Miyaki, Genetic polymorphisms in homocysteine metabolism and response to folate intake: a comprehensive strategy to elucidate useful genetic information. *J Epidemiol.* 20 (2010) 266-270.
- [54] P. Mora, V. Rubio, J. Cervera, Mechanism of oligomerization of *Escherichia coli* carbamoyl phosphate synthetase and modulation by the allosteric effectors. A site-directed mutagenesis study, *FEBS Lett.* 511 (2002) 6-10.
- [55] M. Lopes-Marques, G. Igrejas, A. Amorim, L. Azevedo. Human carbamoyl phosphate synthetase I (CPSI): insights on the structural role of the unknown function domains. *Biochem Biophys Res Commun.* 421 (2012) 409-412.
- [56] J.B. Thoden, G. Wesenberg, F.M. Raushel, H.M. Holden, Carbamoyl phosphate synthetase: closure of the B-domain as a result of nucleotide binding, *Biochemistry.* 38 (1999) 2347-2357.
- [57] I. Yefimenko, V. Fresquet, C. Marco-Marín, V. Rubio, J. Cervera, Understanding carbamoyl phosphate synthetase deficiency: impact of clinical mutations on enzyme functionality, *J Mol Biol.* 349 (2005) 127-141.

Legends to Figures

Fig. 1. Domain composition, function and architecture of CPS1. (A) Schematic linear representation in the CPS1 polypeptide of the different domains. The grey-shaded background bars schematize the 40-kDa N-terminal and the 120-kDa C-terminal moieties of the enzyme that correspond to the small and large subunits of *E. coli* CPS, respectively. The colored bars represent the different CPS1 domains as defined from sequence alignment with *E. coli* CPS. Their approximate masses, in kDa, and the domain start/end residue numbers are given, respectively, above and below each domain. The N-terminal mitochondrial targeting peptide that is removed upon internalization in the organelle is not represented. The domains are defined by their function (when known; ?, function unknown), showing below each of them the corresponding acronym (ISD, GSD, BPSD, UFSD, CPSD and ASD) as defined in [2]. (B) Reactional steps of the CPS1 reaction, shown under the domain catalyzing it, colored as the domain. The thick empty horizontal black arrow denotes the migration of carbamate from the bicarbonate phosphorylation site to the carbamate phosphorylation site. (C) Stereo view of structural model of C α trace of human CPS1 [42]. Each domain is colored differently and labeled in the same color. In the UFSD, the spheres mark the residues (numbered) hosting extremely drastic (red) or less drastic (blue) mutations found in CPS1D patients. Both ATP molecules and essential K⁺ ions are placed in their sites by superimposition of the *E. coli* CPS structure (Protein Databank file 1BXR; [48]). (D) Relations of the UFSD with other domains. Two views of the domain (cartoon representation, in yellow) are shown, highlighting the residues hosting the mutations (red and blue spheres as above) illustrating that the mutations cluster at the regions of contact with the other domains (labeled). Only the parts of these other

domains that contact the UFSD are shown, in surface representation, colored as in (A). In the right panel the ISD is shown as transparent surface to allow visualization of Ala949, which is clamped between the ISD, the GSD and the BPSD.

Fig. 2. Effects of UFSD domain mutations on CPS1 production. WT, wild-type recombinant enzyme. Mutations are represented in single-letter code. Arrowheads signal the position of the CPS1 band. (A) Yield of pure CPS1 protein (total protein $\times$ CPS1 purity determined densitometrically from Coomassie-stained SDS) per liter of cell culture, relative to the mean yield of pure wild-type CPS1 (2.7 mg/L cell culture). (B) Distribution of protein between the insoluble, P, and the soluble, S, fractions of the initial extracts of cells expressing the wild-type enzyme or the indicated mutants, analyzed by SDS-PAGE and Coomassie staining. The samples applied to each track correspond to identical volumes of original cell extract. (C) Western blotting and CPS1 immunostaining in the precipitates from the cell extracts for the indicated enzyme forms. (D, E) Coomassie-stained SDS-PAGE of the preparations obtained after purification of the wild-type and the different mutant enzyme forms as indicated. St, protein markers, with masses, in kDa, indicated at the side.

Fig. 3. Enzyme activity (A) and thermal stability (B, C) of wild-type or mutant forms of CPS1. The activity per mg of pure CPS1 is expressed relative to the corresponding activity of the pure wild-type enzyme (2.3 U/mg). (B,C) Fraction (as percentage) of the activity remaining for the indicated enzyme form after 15-min heating either at 46°C (B) or at the indicated temperatures (C).

Fig. 4. Effects of UFSD mutations on the K_m values of CPS1 for its three substrates and on the K_a value for NAG. (A-D) Histograms illustrating the changes in K_m or K_a values for the indicated CPS1 mutants and substrates or NAG, relative to the corresponding values for the wild-type enzyme. The curves on the right panels (E-H) illustrate the dependency of the reaction velocity on the concentration of each substrate or NAG for the wild-type enzyme or for the indicated mutants.

Fig. 5. Influence of UFSD mutations on V_{max} values for CPS1 activity. (A) Changes in V_{max} values for the different mutants, as a fraction of the value for wild-type CPS1. (B) Plot of the specific activity for each mutant as a function of the V_{max} for this same mutant. Relative values with respect to wild-type are given for both parameters. The regression line goes virtually through zero and gives a value of $r^2=0.842$.

Table 1. CPS1 missense mutations mapping in the UFSD, found in CPS1D patients

Mutation #	Amino acid change ^a	Presentation	Report	Amino acid in CPS			PolyPhen-2 prediction	MutPred prediction g-score
				I	III	Other		
1	p.Leu843Ser ^b	Neonatal	[4]	L/M	L	L/i/a	Probably damaging	0.92
2	p.Arg850Cys	Neonatal	[35]	R	R	R	Probably damaging	0.98
3	p.Arg850His	Neonatal	[2,33,37]				Probably damaging	0.98
4	p.Thr871Pro	Late. 2 nd allele: p.E1194D	[37]	T	T	T/s	Probably damaging	0.86
5	p.Lys875Glu ^b	Neonatal	[4]	K	K	Variab.	Possibly damaging	0.84
6	p.Gly911Glu	Neonatal	[2]	G	G	G	Probably damaging	0.87
7	p.Gly911Val	Neonatal	[34]				Probably damaging	0.97
8	p.Ser913Leu	Neonatal	[2]	S	S	S/a/g/d	Probably damaging	0.79
9	p.Asp914His	Pat. 1:Unknown Pat. 2 ^c : Neonatal	[2] Present	D	D	D	Probably damaging	0.95
10	p.Asp914Gly	Hyperammonemia, seizures. Homozyg. ^d	[2]				Probably damaging	0.92
11	p.Ser918Pro	Neonatal	[33]	S/G	G	A/s/e	Benign	0.93
12	p.Arg932Thr	Neonatal	[2]	R	R	R/k	Probably damaging	0.93
13	p.Ile937Asn	Late. ^e	Present	I	I	I/V/L	Probably damaging	0.83
14	p.Ala949Thr	Late. ^d 2 nd allele: p.Y89*	[2]	A	A	A/G	Probably damaging	0.89
15	p.Leu958Pro	Neonatal	[34]	L	L	L/M/F/ Y	Probably damaging	0.94
16	p.Tyr959Cys	Unknown. Alive ^d 2 nd allele: p.P1462R	[2]	Y	Y	Y	Probably damaging	0.94
17	p.Tyr962Cys	Unknown. Alive ^d 2 nd allele: p.I632R	[2]	Y	Y	Y/f	Probably damaging	0.92
18	p.Gly964Asp	Neonatal	[36]	G	G	Variab.	Probably damaging	0.37

Thin horizontal lines separate mutations affecting different amino acids. Unless indicated below, cDNA changes are to be found in [4,5,7]. Amino acids are shown in three-letter code for the mutant allele found in the UFSD, and in one-letter code for the second allele and for occurrence in the various CPSs, being in low case when found with low frequency. *Variab.* denotes the occurrence at a given position, in the indicated groups of CPSs, of >4 types of amino acids with no constant chemical characteristics (polar, apolar, charged, etc.). PolyPhen-2 grades the probability of a damaging effect of an amino-acid substitution, from higher to lower, as *Probably damaging*, *Possibly damaging*, and

Benign. MutPred gives a *g* score corresponding to the probability that a given amino-acid substitution was deleterious/disease-associated.

^aTranslation of the cDNA reference sequence NM_001875.4 (GenBank). Nucleotide 136 in this sequence is considered +1, since it is the A of the translation initiation codon.

^bThe changes p.Leu843Ser and p.Lys875Glu coexisted within the same allele.

^cNovel patient. The cDNA change is c.2740G>C, the same found in the previously reported patient with this amino acid change [4].

^dUnkown when originally reported. New data gathered on the patient.

^eNovel patient. The cDNA change is c.2810T>A. The changes p.Ile937Asn and p.Gly401Arg (c.1201G>C) coexisted within the same allele. The second allele carries the mutation c.3337 -3T>A/ c.3337 -2A>T, which should lead to exon 27 skipping, frameshift and premature termination.

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig.5

Fig. S1. Mapping of the clinical mutations of the UFSD domain in the structural model of CPS1. The figures are in cartoon representation. Residues that host clinical mutations are shown either as red (drastic) or blue spheres (less drastic), or detailing the side-chains (in blue-colored sticks). Residues are labeled in single letter code. CPS1 domains and secondary structure elements of the UFSD are labeled (from N- to C-terminus, α1 to α7; and β1 to β2). (A) Mapping of the mutations in the structural model of the isolated UFSD. (B-D) Details of the three regions of the UFSD that host clinical mutations, illustrating their relations with elements of the neighboring domains. In (C) the broken lines mark the ion pair network involving Asp914, Arg932 and Tyr962 that is lost when any of these three residues is mutated.