Sancho-Vaello E., Marco-Marin C., Gougeard N., Fernandez-Murga L., Rufenacht V., Mustedanagic M., Rubio V.,Haberle J. (2016). Understanding N-acetyl-L-glutamate Synthase Deficiency: Mutational Spectrum, Impact of Clinical Mutations on Enzyme Functionality, and Structural Considerations. Hum Mutat 37(7): 679 – 94

Author's last version

Understanding N-acetyl-L-glutamate synthase deficiency: mutational spectrum, impact of clinical mutations on enzyme functionality, and structural considerations

Enea Sancho-Vaello,[1†] Clara Marco-Marín,[1,2†] Nadine Gougeard,[1,2] Leonor Fernández-Murga,[1,2] Véronique Rüfenacht,[3] Merima Mustedanagic,[3] Vicente Rubio[1,2*] and Johannes Häberle[3*]

[1]Instituto de Biomedicina de Valencia (IBV-CSIC), Valencia, Spain; [2]Group 739, CIBERER, ISCIII, Spain; [3]University Children's Hospital and Children's Research Center, Zurich, Switzerland

[†,*] These authors contributed equally to this work

Correspondence to:

Johannes Häberle, Kinderspital Zürich, Steinwiesstrasse 75, 8032, Zürich, Switzerland, Email: johannes.haeberle@kispi.uzh.ch

Vicente Rubio, Instituto de Biomedicina de Valencia, C/ Jaime Roig 11, Valencia-46010, Spain. Email: rubio@ibv.csic.es

Contract grant sponsors: Valencian Government (Prometeo II/2014/029); Science Department of the Spanish Government (BFU2011-30407 and BFU2014-58229-P) and Swiss National Science Foundation (grants 310030_127184 and 310030_153196).

ABSTRACT: N-acetyl-L-glutamate synthase (NAGS) deficiency (NAGSD), the rarest urea cycle defect, is clinically indistinguishable from carbamoyl phosphate synthetase 1 deficiency, rendering the identification of *NAGS* gene mutations key for differentiation, which is crucial, as only NAGSD has substitutive therapy. Over the last 13 years we have identified 43 patients from 33 families with *NAGS* mutations, of which 14 were novel. Overall, 36 *NAGS* mutations have been found so far in 56 patients from 42 families of which 76% are homozygous for the mutant allele. 61% of mutations are missense changes. Lack or decrease of NAGS protein is predicted for ~1/3 of mutations. Missense mutations frequency is inhomogeneous along *NAGS*: null for exon 1, but six in exon 6, which reflects the paramount substrate binding/catalytic role of the C-terminal domain (GNAT domain). Correspondingly, phenotypes associated with missense mutations mapping in the GNAT domain are more severe than phenotypes of AAK domain-mapping missense mutations. Enzyme activity and stability assays with twelve mutations introduced into pure recombinant *Pseudomonas aeruginosa* NAGS, together with *in silico* structural analysis, support the pathogenic role of most NAGSD-associated mutations found. The disease-causing mechanisms appear to be, from higher to lower frequency, decreased solubility/stability, aberrant kinetics/catalysis and altered arginine modulation.

Keywords: Urea cycle diseases, inborn errors, acetylglutamate synthase, site-directed mutagenesis, *argA, NAGS.*

Introduction

N-acetyl-glutamate synthase deficiency (NAGSD; OMIM #237310), the last reported classical urea cycle deficiency (UCD), can lead to life-threatening hyperammonemia of neonatal onset [Bachmann et al., 1981; Caldovic et al., 2003] or less severe late-onset presentations [Elpeleg, et al., 1990; Plecko, et al., 1998; Caldovic et al., 2005]. By making N-acetylglutamate (NAG) from L-glutamate and acetyl coenzyme A (AcCoA) N-acetyl-glutamate synthase (NAGS; EC 2.3.1.1; OMIM *608300) plays a pivotal regulatory role in ammonia detoxification [Caldovic and Tuchman, 2003]. NAG is the essential allosteric activator of carbamoyl phosphate synthetase 1 (CPS1; OMIM *608307), the first and rate limiting enzyme of the urea cycle [Cohen et al., 1956; Rubio et al., 1981]. As CPS1 is inactive in the absence of NAG [Rubio et al., 1983], NAGSD is clinically indistinguishable from CPS1 deficiency, presenting with hyperammonemia associated with low plasma citrulline and without increased urinary orotate excretion [Häberle et al., 2012].

While the *NAGS* gene remained unknown, very few NAGSD patients were reported [Bachmann et al., 1981,1988; Burlina et al., 1992; Elpeleg et al., 1990; Forget et al., 1999; Guffon et al., 1995; Plecko et al., 1998; Vockley et al., 1992], possibly because diagnosis was based on difficult enzyme activity determinations in liver tissue. However, with the identification of mouse and human *NAGS* genes [Caldovic et al., 2002a,b; Elpeleg et al. 2002; Häberle et al., 2003b], the way was opened for genetic confirmation of NAGSD. Genetic analysis was greatly facilitated by the fact that the human *NAGS* gene (HGNC ID:17996; gene reference sequence NG_008106.1; mRNA reference sequence: NM_153006.2; protein reference sequence NP_694551.1; Uniprot KB Q8N159) comprises only seven coding exons (Fig. 1) and six very short intronic sequences. This has led to a steady increase in the number of patients identified with this autosomal (17q21.31) recessive disorder. Thus since 2002-2003, and following an earlier compilation [Caldovic et al., 2007], we now compile 36

mutations affecting 56 patients in 42 families (Fig. 1, Tables 1 and 2); of the 25 of these mutations identified by our group, 14 are reported here for the first time (bold-type face in Tables 1 and 2). In the present report, we try to infer information on pathogenicity and severity associated to each mutation (Tables 1 and 2, Fig. 1). This is important particularly for missense mutations (the majority in NAGSD, see below), given the finding of a number of sequence variants in the *NAGS* gene ([Mitchell et al., 2009]; and ExAC database, http://exac.broadinstitute.org/gene/ENSG00000161653). Furthermore, in a recent knockout murine model of NAGSD [Senkevitch et al., 2012], citrulline, in addition to the orphan drug N-carbamyl-L-glutamate (NCG; an NAG analogue) [Rubio and Grisolia, 1981a,b], was required to rescue the animals. This raises the possibility of different drug therapies for complete and for partial NAGSD, underlining the need to distinguish mutations causing total lack of NAGS activity from those causing only partial deficiency.

In the present work, in addition to utilizing the usual pathogenicity-predicting *in silico* servers, we experimentally test the impact of the missense mutations on solubility, activity and stability of *Pseudomonas aeruginosa* NAGS (PaNAGS), used herein as a surrogate of human NAGS (HuNAGS) [Sancho-Vaello et al., 2008, 2009 & 2012]. Unlike HuNAGS, PaNAGS can be produced recombinantly in large amounts and in pure and stable form [Sancho-Vaello et al., 2008, 2009 & 2012], which opens the way to site-directed mutagenesis and to the experimental characterization of the effects of the mutations introduced. Similarly to the mammalian enzyme (Fig. 1) [Caldovic et al., 2002a,b], PaNAGS is composed of an N-terminal amino acid kinase (AAK) domain and a C-terminal catalytic GNAT domain (where GNAT stands for GCN5-related N-acetyltransferases) (Fig. 1). PaNAGS exhibits substantial, although relatively low, sequence similarity with HuNAGS along its entire sequence (Fig. 2), and the K_m values for its two substrates (see below and [Sancho-Vaello et al., 2008]) are not far from those of HuNAGS [Caldovic et al. 2006; Coudé et al., 1981]. Both enzymes have the

same allosteric effector, the amino acid L-arginine, which is an activator of HuNAGS and an inhibitor of bacterial NAGSs (including PaNAGS) [Bachmann et al., 1982; Caldovic et al., 2002a, Haas et al., 1972; Marvil and Leisinger, 1977; Shigesada and Tatibana, 1971]. Existing evidence indicates that arginine binds to an equivalent site in the same location of the AAK domains of human and bacterial NAGSs [Haskins et al., 2008; Sancho-Vaello et al., 2008]. The positive or negative effects of arginine on the activity of these enzymes do not appear to result from intrinsically different processes, given the gradual evolution of arginine from NAGS inhibitor to activator with the transition from marine vertebrates to mammals [Haskins et al., 2008]. Furthermore, changes in the linker that connects the AAK and the GNAT domains in PaNAGS can convert arginine from an inhibitor into a modest activator of this enzyme [Sancho-Vaello et al., 2009 & 2012]. On these bases, we also studied the susceptibility to arginine of different PaNAGS mutants in the hope that this information could be valuable for understanding whether a given mutation would render the human enzyme insensitive to this amino acid effector.

As previously indicated [Sancho-Vaello et al., 2008], a limitation for using PaNAGS as a model of HuNAGS in site-directed mutagenesis studies is that the residue mutated should be conserved or conservatively replaced in the two enzymes. This limitation has largely determined the choice of missense mutations examined experimentally here (Fig. 2, mutations highlighted in red above the sequence alignment): for the AAK domain, seven of the nine NAGSD-associated missense mutations mapping therein, plus one mutation found in heterozygosis in a control human DNA; for the GNAT domain, four mutations, which represent 1/3 of all NAGSD-associated missense mutations found in this domain.

In addition to these studies, we exploit the recently determined crystal structure of the isolated GNAT domain of HuNAGS [Zhao et al., 2013b] to gain further insight into the consequences of the mutations affecting this domain. The combination of the clinical

observations, *in silico* predictions and experimental determinations from present and past studies [Caldovic et al., 2005 & 2007; Schmidt et al., 2005], and structural inferences, allow a clearer picture to be built to understand NAGSD in molecular terms. Hereby, the disease-causing role of most of the mutations found in NAGSD patients is supported, and clear links between the type and localization of each mutation and the severity of its effects are provided.

Materials and methods

Sample accrual and mutation detection

Patients were considered to likely suffer from NAGSD based on [Häberle and Rubio, 2014] the clinical and family history, a positive response to NCG administration, liver enzyme testing (done in only few cases), and analysis of biochemical markers (typically hyperammonemia, low plasma citrulline, elevated plasma glutamine, lack of prominent alterations in the levels of other plasma amino acids, and normal biochemical markers of organic acidemias and fatty acid oxidation defects, with normal or low urinary orotate). Carbonic anhydrase Va deficiency had not yet been identified as a cause of hyperammonemia [van Karnebeek et al., 2014], therefore it was not investigated in this study. Since 2002, 38 such index-patients, mainly from Europe and the Middle East, had either their DNA or blood samples, or, in families 6, 27, 32, 40 and 41, also skin or liver biopsies, studied at either Münster University or Zurich University and found by sequencing to host non-synonymous mutations in the *NAGS* gene. Parental DNA was used for mutation confirmation whenever possible. Appropriate consent was obtained from all patients or their guardians in accordance with regulation at the time samples were collected. Unless indicated, when considering mutation frequencies, members of a family with the same mutation are considered here only once.

Mutation analysis in the novel patients of this study was carried out as reported on mRNA isolated from liver biopsies, fibroblasts or from the cultured lymphocytes also used for analyzing the *CPS1* gene [Kretz et al., 2012], and on blood or fibroblast genomic DNA [Häberle et al., 2003 a,b]. At the genomic DNA level, at least all seven coding exons and flanking intronic regions were sequenced. In the patients from families 12, 27 and 41, the enhancer region of the *NAGS* gene, all *NAGS* intronic regions and the *CPS1* gene were also sequenced. In the patients from families 6, 16, 19, 27, 32, 40 and 41, we also analyzed cDNA from cultured fibroblasts or from liver [Häberle et al., 2003a]. Mutations are reported according to the nomenclature suggested by the Mutalyzer software (https://mutalyzer.nl/), following also the most recent Human Genome Variation Society recommendations [den Dunnen and Antonarakis, 2000] (http://www.hgvs.org/mutnomen/recs.html). However, for space reasons, amino acid changes are given in Tables 1 and 2 and in the figures in 1-letter amino acid code. PaNAGS amino acid changes are experimentally generated and therefore parentheses are not used in their notation, as is also the case for HuNAGS mutations for which mRNA-derived information existed. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference coding sequence, with the initiation codon as codon 1. All the novel sequence variants of human *NAGS* have been submitted to the LOVD public database (http://chromium.lovd.nl/LOVD2/variants.php?action=search_unique&select_db=NAGS).

Production and assay of *P. aeruginosa* NAGS carrying the mutations found in NAGSD patients

We previously reported [Sancho-Vaello et al., 2008 and 2012] the cloning of the *P. aeruginosa argA* gene in pET22b (from Novagene), the expression in *Escherichia coli* BL21(DE3) of the encoded PaNAGS with a C-terminal $GSLEH_6$ extension, and the rapid

purification to homogeneity of the enzyme by centrifugal Ni-affinity chromatography. Site-directed mutagenesis was carried out with the Quickchange mutagenesis kit (Stratagene) according to the manufacturer's instructions; sequence of mutagenic primers is available on request. Presence of desired mutation and absence of unwanted sequence changes was verified by DNA sequencing. The same method was used to purify mutant and wild-type (WT) PaNAGS proteins [Sancho-Vaello et al., 2008 and 2012]. Immediately after purification, the WT or mutant enzyme forms were placed by centrifugal gel filtration (PD SpinTrap G-25 columns, GE Healthcare) in storage buffer at 4ºC (10 mM Na phosphate pH 7.0, 15% v/v glycerol, 1 mM EDTA, 1 mM dithiothreitol, 20 mM NaCl and 10 mM NAG) [Marvil and Leisinger, 1977]. Only the p.Ile201Asn PaNAGS mutant appeared too unstable for buffer exchange and was used immediately after production, dissolved in the elution buffer from the His-Spin Trap step.

Enzyme activity assays

NAGS activity was determined at 37ºC [Sancho-Vaello et al., 2012] as glutamate-dependent CoA release monitored colorimetrically at 412 nm with 5,5-dithio-bis(2-nitrobenzoic acid (Ellman's reagent) [Errey and Blanchard, 2005]. The standard assay mixture (20-200 μl) contained 0.2 M Tris-HCl pH 9, 4 mM acetyl coenzyme A (AcCoA) and 30 mM L-glutamate. For kinetic analysis, the concentration of one substrate was varied, whereas the other substrate was fixed at a concentration identical to that in the standard assay, unless indicated. Reactions were stopped after 10 min by the addition of 0.8 ml of 0.1 M Na phosphate pH 7 containing 0.2 mM Ellman's reagent. Results from at least duplicate assays were fitted with GraphPadPrism (GraphPad Software, San Diego, CA), to hyperbolic kinetics, either without or with substrate inhibition kinetics as previously reported [Sancho-Vaello et al., 2009]. One enzyme unit corresponds to the production of 1 μmol CoA per min.

PaNAGS stability assays using the thermofluor approach

The thermal stability of WT or mutant PaNAGS was investigated using a real-time PCR instrument (7500 model from Applied Biosystems) by monitoring the increase in SYPRO Orange (from Invitrogen, Carlsbad, CA) fluorescence upon gradual (1ºC/min) temperature increase [Vedadi et al., 2006]. This followed a reported protocol [Martinelli et al., 2012], except for the use of an enzyme concentration of 0.2 mg/ml in a solution of 0.15 M Tris-HCl pH 7.1, 75 mM imidazole, 75 mM NaCl, 0.15 mM dithiothreitol and a 1:1000 dilution of the SYPRO Orange commercial preparation.

Other techniques

Polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulphate (SDS-PAGE) was carried out with 15% polyacrylamide gels, using Coomassie Blue staining to visualize protein bands [Laemmli, 1970]. Protein concentrations were determined using a commercial reagent (from Bio-Rad) [Bradford, 1976] and bovine serum albumin as standard. The secondary structure of the AAK domain of HuNAGS was predicted with the SOPMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) [Geourjon and Deléage 1995], GOR V (http://gor.bb.iastate.edu/cdm/) [Sen et al., 2005], Jpred3 (http://www.compbio.dundee.ac.uk/www-jpred) [Cole et al., 2008] and PsiPred http://bioinf.cs.ucl.ac.uk/psipred) [McGuffin et al. 2000] servers. The predictions given for this AAK domain are the consensus for data from all these servers. Secondary structure elements for PaNAGS were obtained from the model provided by the Uniprot KB server, which was produced by Swiss Modeller. Figures representing protein structures were prepared using Pymol (http://www.pymol.org).

Results

Mutations found in NAGSD

Since we started mutation analysis of the human *NAGS* gene in 2002, we have identified 43 patients in 33 families with sequence changes in their *NAGS* genes and clinical presentation consistent with NAGSD (Figures 1 and 2, and Tables 1 and 2). Fourteen mutations are reported here for the first time, including six that should cause premature termination of translation and result in NAGS truncation [p.(Pro93Glnfs*18), p.(Gln331*), p.(Glu422*), p.Gly438Alafs*10, p.(Thr439Hisfs*52) due to c.1313dup, and p.Trp498*] and eight missense mutations [p.(Met167Val), p.(Pro260Leu), p.Thr264Met, p.(Ile291Asn), p.(Leu391Arg), p.(Ser398Cys), p.Gly457Asp and p.Tyr512Cys]. The latter changes affect residues with variable degrees of invariance: Ser398, conserved in vertebrate and fungi NAGS; Met167, Pro260, Leu391 and Tyr512, exclusively conserved in vertebrate NAGS; and Thr264, Ile291 and Gly457, non-conserved [Caldovic et al., 2007]. The conservation of the residues and/or the introduction of rather drastic amino acid changes support the disease-causing nature of these mutations.

By including the presently reported mutations and one mutation that affected the enhancer region, which markedly decreased the expression of the *NAGS* gene [Heibel et al 2011], 36 *NAGS* gene mutations have been found so far in 56 NAGSD patients from 42 families (Figure 1, Tables 1 and 2). Patients were homozygous for the mutation in 32 of the 42 families (76%), which suggests founder effects and endogamy as important disease determinants. In seven families, patients were compound heterozygotes for two mutations, and in three other families (12, 27 and 41, Tables 1 and 2) a second *NAGS* allele was not found even after sequencing all the introns and the enhancer region. No aberrant mRNAs (except for the presence of the mutation identified in the genomic DNA) was observed in fibroblasts or lymphocytes for the patients of families 27 and 41 (supplementary Fig. S1). A

dominant negative effect was excluded for the mutation found in family 41, since the patient's healthy father carried the mutation in heterozygosis with the normal allele. However, the unavailability of parental DNA prevented the exclusion of this possibility in families 12 and 27 if the mutations had arisen "de novo" in the patients. The possibilities of somatic mosaicism (so that the patients of families 12, 27 and 41 could have a second mutation exclusively in their livers) or of liver-restricted epigenetic silencing of the normal allele (as reported for muscle in the case of the *RYR1* gene in core myopathies [Zhou et al., 2006]) were not excluded, since no liver DNA or mRNA was available for study.

The larger number of patients and mutations than in a previous compilation [Caldovic et al., 2007] provides strong foundation for the conclusion that the majority of the mutations found in NAGSD (approximately 60% in our series) are missense changes (Table 1). Nevertheless, one such change (c.1450T>C) fell on the penultimate base of exon 6 and could in principle affect splicing at the exon 6/intron 6 donor site (this possibility is reflected in Fig. 1), although analysis of the sequence with GENSCAN (http://genes.mit.edu/GENSCAN.html) does not predict a splice aberration. Small insertions and deletions, nonsense changes and pure splice aberrations occur with lesser frequency (occurrences of 7, 4 and 2 mutations, corresponding to 19%, 11% and 6% of all mutations) and should cause obligate or predominant (splice aberrations) enzyme truncations. Indeed, they were associated with neonatal presentations (Tables 1 and 2) considered to reflect the existence of little, if any, residual enzyme activity. Mutations that result in protein truncation (in approximately 1/3 of the cohort), due to either nonsense or small indel/frameshift mutations, are likely to lead to absence of protein expression (patients from all the families in Table 2, except family 41) due to nonsense-mediated RNA decay or degradation of the truncated NAGS protein. The enzyme protein may also be decreased by 50% when truncating mutations are in heterozygosis with a non-truncating mutation (the cases of families 1, 26 and possibly of family 41). Even if these

truncating mutations did not cause a reduction in the number of enzyme molecules, all of them, including the most distal one, p.Trp498*, would inactivate NAGS because of the loss of all or a part of the glutamate and/or the AcCoA binding sites (Fig. 1; residues making the binding sites for the substrates [Zhao et al., 2013b] are marked with triangles).

Most missense mutations are predicted to be deleterious by PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2) [Adzhubei et al. 2010] and/or MutPred (http://mutpred.mutdb.org) [Li et al., 2009] servers (Table 1). Sequence variants p.(Met167Val), p.(Arg509Gln), p.Tyr512Cys and p.(Ala518Thr) have been found in the ExAC database of genetic variation with extremely low allele frequencies (≤0.0041 %; for 61,000-121,000 alleles analyzed) and never as homozygotes.

Distribution of the missense mutations along the NAGS protein

The normalized frequency (per 100 amino acids) of missense mutations in the GNAT domain is nearly double that in the AAK domain (Fig. 3A), particularly when missense mutations that fall on CpG dinucleotides (these dinucleotides exhibit increased mutation frequency) are not considered. This difference between the two domains in the frequency of disease-associated missense mutations is even more obvious if the number of families with mutations is considered (that is, one mutation is counted *n* times if present in *n* families). This leads to the conclusion that the GNAT domain is more important for enzyme function than the AAK domain, as expected given the fact that the GNAT domain binds the substrates and hosts the catalytic machinery, being able to catalyze the entire reaction by itself [Sancho-Vaello et al., 2012; Zhao et al., 2013b].

The inhomogeneous distribution of missense mutations along the *NAGS* gene is even more obvious when the normalized frequency of missense mutations per exon is compared (Fig. 3B). This frequency is nil for exon 1, most of which encodes the mitochondrial targeting

peptide (known from other proteins to be tolerant to mutations [Häberle et al., 2011]) and the variable region, which is unessential as shown by expression studies [Caldovic et al., 2002a,b; Schmidt et al., 2005]. The frequencies of missense mutations for exons 2-4, encoding most of the AAK domain, are relatively low compared with those for exons 5 and 6, which encode the GNAT domain, particularly when the mutations that fall on CpG dinucleotides are excluded (Fig. 3B). The normalized missense mutation frequency peaks in exon 6 (Fig. 3B), the exon that encodes the most central part of the GNAT domain, including most of the glutamate and AcCoA binding sites (Fig. 1). This difference in frequency per exon is even more marked when the number of families, rather than the number of mutations, is considered (Fig. 3C). Actually, there are two regions where the missense mutations that affect several families cluster in exon 6 (Fig. 1): three mutations at codons 430, 431 and 433 affect five families, and seven families carry the p.(Trp484Arg) mutation.

The increased cohort of NAGSD patients and families renders stronger the previous inference [Caldovic et al., 2007] that missense mutations that affect the GNAT domain are generally more deleterious than those that affect the AAK domain. Thus, of the missense mutations found in patients with genotypes that were clinically informative because of homozygosity or coexistence with a highly deleterious second *NAGS* allele (Table 1), the mutations that mapped in the GNAT domain generally caused neonatal presentations (Figs. 1 and 3D). In contrast, except for p.(Pro260Leu), all the clinically informative mutations that mapped in the AAK domain caused late-onset disease (Figs. 1 and 3D) and, therefore, might not have completely abolished enzyme activity.

Decreased enzyme solubility as a major determinant of disease-causation by NAGSD-associated missense mutations

Given the difficulties in obtaining stable and active recombinant HuNAGS (our own

unpublished results; and [Zhao et al., 2013b]), we used PaNAGS to experimentally test the impact of the different missense mutations found in NAGSD patients on NAGS functionality and stability. To identify homologous residues in HuNAGS and PaNAGS, a structure-based alignment of these enzymes was prepared (Fig. 2) by exploiting the recently reported structure of HuNAGS GNAT domain [Zhao et al., 2013b] and the predicted secondary structure elements of PaNAGS and of the AAK domain of HuNAGS (Fig. 2). This alignment was made more robust by the inclusion of the sequences of three structurally well-characterized close homologues of HuNAGS. These are NAGS from *Neisseria gonorrhoeae* (NgNAGS) [Min et al., 2009; Shi et al., 2008], NAGS/NAG kinase (a bifunctional enzyme) from the bacterium *Maricaulis maris* (MmNAGS/K) [Shi et al. 2011], and NAG kinase from yeast (ScNAGK; this enzyme has the same domain composition and important sequence similarity with human and yeast NAGSs) [de Cima et al., 2012]. Based on this optimized alignment, 12 mutations found in NAGSD patients, which affected conserved or conservatively replaced residues in HuNAGS and PaNAGS (Fig. 2, in red above the alignment) were introduced into PaNAGS by site-directed mutagenesis.

Except for mutations p.Leu50Val, p.Gly146Cys and p.Leu353Val, [that correspond respectively to HuNAGS mutations p.(Met167Val), p.(Gly236Cys) and p.(Leu442Val)], all the other mutations introduced into PaNAGS impaired enzyme solubility relative to the WT enzyme. This fact was revealed by studying the partitioning of the recombinant NAGS protein between the supernatant and precipitate when the initial crude extract of *E. coli* expressing a given PaNAGS mutant was centrifuged (Fig. 4A). This lack of solubility, which may reflect misfolding or stabilization of an unfolded state leading to aggregation, was virtually complete for the mutants p.Leu56Glu and p.Leu342Pro, and was also quite marked for the p.Pro170Leu mutant [respective HuNAGS counterparts, p.(Val173Glu), p.Leu430Pro and p.(Pro260Leu)]. These findings agree with previous reports of very low activity for the p.(Val173Glu) and

p.Leu430Pro mutants [Caldovic et al., 2005; Schmidt et al., 2005] in assays in which *E. coli* was used to express the recombinant human enzyme forms carrying these mutations. Decreased/low solubility could be a paramount disease-causing factor for mutations p.Thr264Met, p.(Ala279Thr) and p.(Trp484Arg), but not for p.(Ile291Asn), p.(Val350Ile) and p.Gly457Asp. The PaNAGS mutations that corresponded to the first three of these human mutations, p.Ser174Met, p.Met189Thr and p.Trp397Arg, markedly decreased PaNAGS solubility, whereas p.Ile201Asn, p.Val258Ile and p.Gly368Asp (that respectively correspond to the last three of these human mutants) had a similar solubility to WT PaNAGS (Fig. 4B). In any case, these findings indicate that decreased enzyme solubility, possibly reflecting aggregation, may be an important mechanism of disease causation for a substantial number or even for a large fraction of NAGSD-associated missense mutations.

Influence of the mutations on enzyme activity

We succeeded in purifying most mutant PaNAGS recombinant proteins to homogeneity (Fig. 4B); the yields of the p.Pro170Leu, p.Ser174Met and p.Val258Ile PaNAGS mutant proteins were low and they could not be purified to complete homogeneity, while p.Leu56Glu and p.Leu342Pro had the lowest yields probably due to their low solubility (Fig. 4B).

When the activity of the enzyme preparations obtained after the purification process was tested (Fig. 4B, figures below the gels), it was nil for mutants p.Leu56Glu and p.Leu342Pro, which correspond to HuNAGS mutants p.(Val173Glu) and p.Leu430Pro, as expected from the lack of soluble NAGS in the corresponding preparations. However, two other mutant forms which host amino acid changes in the GNAT domain were purified in soluble form and good yield, but, nevertheless, one was inactive [p.Gly368Asp mutant, corresponding to the p.Gly457Asp HuNAGS mutation] and the other exhibited very low

activity [p.Trp397Arg; HuNAGS counterpart, p.(Trp484Arg)]. A very low activity of the p.(Trp484Arg) mutant would agree with the prior finding of <5% activity in a crude extract of *E. coli* that expressed this mutant form of the human enzyme [Schmidt et al., 2005]. In contrast, the chemically minor change represented by the p.Leu353Val substitution [corresponding to p.(Leu442Val) in HuNAGS] had no substantial effect on enzyme activity (Fig. 4B), agreeing with the report of 50-75% activity [Caldovic et al. 2007], relative to WT, for the recombinant human enzyme that carries this mutation, expressed in *E. coli*. Essentially normal activity was exhibited also by five enzyme forms which carried the mutations that mapped in the AAK domain p.Leu50Val, p.Pro170Leu, p.Met189Thr, p.Ile201Asn and p.Val258Ile (Fig. 4B) and which corresponded, respectively, to the human mutations, p.(Met167Val), p.(Pro260Leu), p.(Ala279Thr), p.(Ile291Asn) and p.(Val350Ile)]. However, for two other mutations that mapped in this domain, p.Gly146Cys and p.Ser174Met [respective HuNAGS counterparts, p.(Gly236Cys) and p.Thr264Met], enzyme activity appeared substantially decreased even when assayed at increased L-glutamate concentrations (Fig. 4B).

We gained further insight on the effects of the different mutations by performing kinetic studies in which both substrates were varied (Table 3 and Figs. 4C,D). For all but three of the active mutants this analysis failed to reveal large decreases in apparent V_{max} or major increases in apparent K_m values for one or the other substrate. However, dramatic effects were observed (Table 3 and Figs. 4C,D) with the p.Trp397Arg mutation [corresponding to the HuNAGS p.(Trp484Arg) mutant], for which the velocity extrapolated at infinite concentration of both substrates was only ~10% of that of the WT enzyme, whereas the apparent K_m values for L-glutamate and AcCoA were increased ~90-fold and ~45-fold, respectively. In line with our previous findings [Sancho-Vaello et al., 2008], AAK domain mutation p.Gly146Cys [corresponding to the p.(Gly236Cys) mutation found in heterozygosis in a control human

DNA] [Caldovic et al., 2007] increased the apparent K_m^{Glu} (~30-fold; Table 3 and Fig. 4C). Another AAK domain mutant, p.Ser174Met [HuNAGS counterpart, p.Thr264Met], also considerably increased the apparent K_m^{Glu} (~20-fold, Fig. 4C). No mutant, except p.Trp397Arg, exhibited a drastic increase in K_m for AcCoA. However, two mutants, p.Pro170Leu and p.Ser174Met [corresponding, respectively, to p.(Pro260Leu) and p.Thr264Met of HuNAGS] exhibited substrate inhibition by AcCoA (Table 3; Fig. 4D). However, given the relatively high AcCoA concentrations needed for inhibition, it is uncertain whether this kinetic aberration would be significant *in vivo*.

As already indicated, arginine is an allosteric effector of bacterial NAGSs, including PaNAGS, binding to a site [De Cima et al., 2012; Min et al., 2009; Zhao et al., 2013a] that appears to be equivalent to the binding site for the L-arginine activator of human NAGS [Haskins et al., 2008; Sancho-Vaello et al., 2008]. All the GNAT domain mutants tested here that exhibited substantial activity were inhibited by arginine (Fig. 4B). This finding agrees with a previous observation [Caldovic et al., 2007] of sensitivity to arginine of one of the corresponding GNAT domain mutant forms of HuNAGS, p.(Leu442Val), when expressed recombinantly in *E. coli*. Contrasting with the observations with the GNAT domain mutations, arginine inhibition was abolished or strongly decreased in three of the AAK domain mutants studied here [p.Leu50Val, p.Gly146Cys and p.Ser174Met, corresponding to HuNAGS p.(Met167Val), p.(Gly236Cys) and p.Thr264Met], (Fig. 4B). These observations highlight the key role of the AAK domain in the mediation of arginine modulation of the whole NAGS [Sancho-Vaello et al., 2012; Zhao et al., 2013b].

Influence of the mutations on the thermal stability of the enzyme

Three of the mutants studied here, p.Ile201Asn, p.Val258Ile and p.Leu353Val [respective HuNAGS mutations, p.(Ile291Asn), p.(Val350Ile) and p.(Leu442Val)], exhibited

little or no decrease in solubility (Fig. 4A). They did not present either drastic decreases in enzyme activity (Fig. 4B) or major changes in the kinetic parameters for the substrates (Table 3), nor were they found to have lost their sensitivity to arginine modulation (Fig. 4B). We explored the influence of these mutations on the thermal stability of PaNAGS with the thermofluor approach (Fig. 4E), in which protein unfolding with increasing temperature is monitored as fluorescence increased in mixtures of the protein and added fluorophore. We also carried out this analysis with mutants p.Pro170Leu and p.Met189Thr [respective HuNAGS counterparts, p.(Pro260Leu) and p.(Ala279Thr)]. The thermofluor analysis results (Fig. 4E) were virtually identical for the WT form and mutants p.Met189Thr and p.Leu353Val, whereas the p.Val258Ile mutant exhibited a modest increase in the temperature required for unfolding (approximately 3ºC higher than for the WT enzyme). In contrast, the thermal stabilities of mutants p.Pro170Leu and p.Ile201Asn were drastically decreased, with temperatures giving half-maximal increase in fluorescence approximately 15-20ºC lower than for the WT enzyme. These findings strongly suggest that for these last two mutants [and, correspondingly, for the HuNAGS mutants p.(Pro260Leu) and p.(Ile291Asn)] decreased thermal stability is likely to be an important pathogenic determinant.

Discussion

In this work, we compile 22 missense mutations and 13 largely truncation-causing mutations of other types, including 14 novel mutations, found in patients with presentations labeled as NAGSD. The need for liver tissue for NAGS activity assay and the difficulty of this assay render mutation identification the practical standard for diagnosing NAGSD [Häberle et al., 2012]. Although such diagnosis can be supported by the positive effect of treating the patient with NCG [Guffon et al., 2005], the possibility that NCG could cause

improvement in hyperammonemias due to other UCDs (https://clinicaltrials.gov/ct2/show/NCT00843921?term=carglumic+acid&rank=3) cannot be excluded. Therefore, it is particularly important to ascertain whether the mutations found in putative NAGSD patients are disease-causing or not. In this respect, all the truncating mutations compiled here, including the most distal one, p.Trp498* (see below), are surely pathogenic, as the catalytic machinery, being near the enzyme C-terminus, is completely or partially deleted by these truncations. It is more difficult to ascertain whether missense mutations, which account for approximately 60% of all mutations reported here, are disease-causing. A disease-causing role was supported for many (~70%) of these mutations by the unanimous predictions of deleterious effects by two commonly used pathogenicity prediction servers (Table 1). The absence of most of these mutations from sequencing databases compiling ~60,000-120,000 *NAGS* genes (ExAC database) also agrees with their disease-causing role. Even for the few of these mutations that also appear in sequencing databases, the frequency of their appearance is extremely low and they have never been found in homozygosis (http://exac.broadinstitute.org/gene/ENSG00000161653) except in patients with presentations labeled as NAGSD, two traits that again are expected for pathogenic variants.

A more direct approach used here to infer the pathogenicity of individual missense mutations was the experimental assay of their effects on PaNAGS, a bacterial NAGS which, unlike HuNAGS, is easily produced in recombinant form and can be used in site-directed mutagenesis studies [Sancho-Vaello et al., 2008, 2009 & 2012]. The comparisons of the effects of the mutations p.Leu56Glu, p.Val258Ile, p.Leu342Pro, p.Leu353Val and p.Trp397Arg on PaNAGS with the effects reported [Caldovic et al., 2005 and 2007; Schmidt et al., 2005] for the corresponding HuNAGS mutations p.(Val173Glu), p.(Val350Ile), p.Leu430Pro, p.(Leu442Val) and p.(Trp484Arg) on the human enzyme support the use of PaNAGS as a model of HuNAGS. Nevertheless, a limitation of the present studies is that the

expression of single mutations cannot reveal the potential interactions of two different mutations in compound heterozygotes. Similarly, the use of the isolated protein may eliminate the influence of other variables, such as the diverse genetic backgrounds of human patients. Despite these limitations, the PaNAGS expression system has proven very useful, particularly to understand the effects of the mutations mapping in the AAK domain. Thanks to this approach, we infer that decreased solubility and/or decreased stability can be important elements by which AAK domain mutations cause NAGSD. Thus, a combination of extremely low solubility (Fig. 4A) and of markedly decreased thermal stability of the small fraction of the enzyme that remained soluble (Fig. 4E) can explain the neonatal presentation in the patient who is homozygote for the mutation p.(Pro260Leu) (family 5, Table 1). In turn, the late-onset and mild phenotypes of the four patients who are homozygotes for the mutation p.(Ile291Asn) (family 8, Table 1) agree with the observations of decreased thermal stability (Fig. 4E) and modestly decreased solubility of the corresponding PaNAGS mutant (p.Ile201Asn; Fig. 4A). These aberrations are detrimental for the enzyme but should not completely abolish its activity. The late-onset phenotype in the patient who was a compound heterozygote for the p.(Val173Glu) and p.(Thr431Ile) mutations (family 2, Table 1) can be explained by the residual enzymatic activity of p.(Thr431Ile), since p.(Val173Glu) is a null allele ([Caldovic et al., 2005] and present data with the p.Leu56Glu PaNAGS mutant). Indeed, the recombinantly expressed p.(Thr431Ile) HuNAGS mutant was reported to exhibit 31% activity in the presence of arginine [Caldovic et al., 2007]. In agreement with the proposal that p.(Thr431Ile) does not cause enzyme inactivation, this mutation was found in another late-onset patient with a different genotype (family 9, Table 1).

As already indicated, decreased solubility may also underlie the effects of mutations p.Thr264Met and p.(Ala279Thr) (Fig. 4A), which, because of the substantial amount of soluble protein, would be expected to give late-onset presentations in homozygous patients, as

observed (families 6 and 7, Table 1). The more drastic phenotype associated with mutation p.Thr264Met than with p.(Ala279Thr) (Table 1) could parallel increased K_m for L-glutamate (Table 2, Fig. 4C) and decreased sensitivity to arginine with the first but not with the second of these mutations, judging by the observations made with the corresponding PaNAGS mutations p.Ser174Met and p.Met189Thr (Figs. 4B and 4C). Indeed, lack of arginine activation may be a plausible pathogenic mechanism for some hypomorphic AAK domain mutations, as for the p.(Met167Val) mutation, whose bacterial counterpart was not inhibited by arginine (Fig. 4B), and which was associated with a null second disease allele and with a clinical presentation that would be consistent with partial deficiency (family 1, Table 1).

Overall, the use of the PaNAGS system provided strong disease clues for all but one of the experimentally studied mutations, p.(Val350Ile), found in a patient with very mild disease and a compound heterozygote genotype [p.(Val350Ile)/p.(Leu442Val); family 10, table 1]. The corresponding PaNAGS mutation, p.Val258Ile, caused a modest decrease in enzyme solubility, which was the only substantially abnormal feature observed (Fig. 4 and Table 2). It is uncertain whether this is enough to account for the 4-5% residual activity reported for the recombinant human enzyme carrying the p.(Val350Ile) mutation [Caldovic et al., 2007], although little detail and no evidence on the purity of the enzyme was given in these assays. In any case, the very mild phenotype would be justified by the very small or no effect of the second disease allele found in this patient, p.(Leu442Val), judging by the results with the corresponding PaNAGS mutant, p.Leu353Val (Fig. 4, Table 3) or with the recombinant HuNAGS (50-75% activity; no purity recorded and no details given) [Caldovic et al., 2007]).

To summarize the effects of the AAK domain mutations, the majority of those compiled here are clearly disease-causing. This conclusion can also be extended to the p.(Cys200Arg) mutation, which could not be examined here using PaNAGS, but has been

reported to be associated with <5% enzyme activity (relative to WT) in crude extracts of *E. coli* expressing recombinant HuNAGS [Schmidt et al., 2005]. Decreased solubility and/or reduced thermal stability appear the most important potential disease determinants for clinical AAK domain mutations, although a negative effect on the K_m^{Glu} or on the modulation by arginine appear also potential disease determinants.

Compared with the mutations that mapped in the AAK domain, the NAGSD-associated mutations that mapped in the GNAT domain were less amenable to testing with the PaNAGS model system given the lack of conservation in both HuNAGS and PaNAGS of most residues hosting these mutations (Fig. 2). Only four mutations reported in this domain were studied using PaNAGS. The results with these four mutations also support for this domain an important role of decreased enzyme solubility in disease causation (Fig. 4A). This effect ranged from extreme to modest for the mutations that corresponded to the HuNAGS mutations p.Leu430Pro, p.(Trp484Arg) and p.Gly457Asp (in decreasing order of importance of the effects); the first two have been associated with <5% enzyme activity in crude extracts of *E. coli* expressing recombinant HuNAGS [Schmidt et al., 2005]. Reflecting the catalytic nature of the GNAT domain, loss of activity or a marked decrease in V_{max} was observed for two PaNAGS mutations mapping in this domain, p.Gly368Asp and p.Trp397Arg, corresponding, respectively, to HuNAGS mutations p.Gly457Asp and p.(Trp484)Arg (Figs. 4B-D and Table 3). Similarly reflecting substrate binding, major increases in the K_m values for both substrates were observed with the mutation that corresponded to HuNAGS p.(Trp484Arg) (Table 3 and Figs. 4C,D). The derangements for three of the four tested mutations were so important that they justify complete or nearly complete deficiency, thus supporting the observation of particularly severe effects of the mutations that map in this domain ([Caldovic et al., 2007] and this report). However, no derangement was observed with the bacterial counterpart of the fourth mutant, p.(Leu442Val), as already mentioned above.

A crystal structure of the His_6-tagged human GNAT domain forming a dimer was recently determined ([Zhao et al., 2013b]; Protein DataBank file 4K30). This recombinantly produced domain exhibited an abnormally low k_{cat} (~20-fold lower than the complete enzyme produced and assayed in the same laboratory) and it exhibited a low K_m^{Glu} which was possibly a consequence of the decreased k_{cat} [Zhao et al., 2013b]. Nevertheless, this structure (Fig. 5A) is presently the best existing guide to rationalize the effects of GNAT domain mutations on structural bases. For example, the p.Leu430Pro mutation can alter/decrease protein stability because it could disrupt/break the ß3'strand and disturb the central sheet of the GNAT domain sandwich. Mutations p.Gly457Asp and p.(Trp484Arg) both affect helices that flank active sites, which could lead to a significant decrease, or even absence of enzymatic activity. Gly457 is the last residue (shown underlined) of the (Arg/Gln)-Xaa-Xaa-Gly-Xaa-(Gly/Ala) AcCoA binding consensus for GCN5-related N-acetyltransferases [Zhao et al., 2013b]. Indeed, in the crystalline complex with CoA of the homologous enzyme NgNAGS, the corresponding residue, Gly369, closely interacted with the pyrophosphate moiety of CoA, forming a hydrogen bond with it [Shi et al., 2008]. In turn, in this crystalline complex of NgNAGS with CoA, the residue corresponding to Trp484 of HuNAGS, Trp398, held on its flat indolic ring the *tert*-isobutyl group of the pantoic acid moiety of AcCoA, accounting for the strong effect on AcCoA binding observed here with the PaNAGS counterpart of the p.(Trp484Arg) HuNAGS mutation. In contrast, the mildness of the effects of the p.(Leu442Val) mutation can be understood on the basis of the structure of the human GNAT domain, given the small change in the size of a hydrophobic side-chain of the NAG binding site (Fig. 5B) which is oriented towards the methyl group of the N-acetyl moiety of NAG.

The human GNAT domain structure also predicted clear-cut effects of some GNAT domain mutations not tested in the PaNAGS expression system, which would explain their involvement in NAGSD. Mutations p.(Leu391Arg), p.(Ser410Pro), p.(Arg414Pro) and

p.(Glu433Asp) were expected to be disease-causing by triggering improper protein folding. The first of these changes replaces a constant Leu/Ile by a larger and charged polar arginine at the hydrophobic nest that glues the first α-helix to the central β-sheet (Fig. 5A). Nevertheless, this mutation could not be inactivating because it coexisted with the drastic p.(Trp484Arg) mutation in the late-onset patients of family 11 (Table 1). The p.(Ser410Pro) mutation was likely to break helix α2' (Fig. 5A), explaining the neonatal phenotype of the patients who were homozygotes for this mutation (families 13-14, Table 1), as well as the finding of <5% activity in crude extracts of *E. coli* expressing recombinant HuNAGS with the p.(Ser410Pro) mutation [Schmidt et al., 2005]. Mutations p.(Arg414Pro) and p.(Glu433Asp), by respectively eliminating the guanidinium group of arginine and by shortening the glutamate side-chain, should also trigger misfolding or favour an unfolded state and impair L-glutamate binding since the His517-Glu433-Arg414-Asp443-Lys444-NAG ion-pair network (Fig. 5C) would be disrupted. This network glues together the central sheet elements β4', β2', β3' and the C-terminal helix, connecting them to the substrate glutamate [Zhao et al., 2013b]. Indeed, the patients of families 15 and 18, which were homozygotes for these mutations (Table 1) had very severe clinical presentations.

The structure of the human GNAT domain (Fig. 5A) suggests that p.(Thr431Ile) should cause less marked destabilization than its neighbor mutation p.Leu430Pro, which would account for the milder effects proposed here for the p.(Thr431Ile) mutation (see above). The replacement of Ala518 by the larger and more polar threonine in p.(Ala518Thr) may displace the NAG-binding Arg474 side-chain, with which Ala518 interacts (Fig. 5A), and could possibly distort the NAG binding site. The patients of families 28 and 29 (Table 1), who are homozygotes for the mutation p.(Ala518Thr), presented manifestations that were consistent with severe, but not with complete, deficiency, which could be reflected in the low activity (<5%) detected for this mutant in crude expression assays of HuNAGS in *E. coli*

[Schmidt et al., 2005]. The impacts of mutations p.(Arg509Gln) and p.Tyr512Cys are harder to predict. They affect residues that are linked mutually by intersubunit salt bridges that would be abolished by these mutations (Fig. 5A). The nearly normal activity (74-83% of WT) reported for the recombinant human enzyme with the p.(Arg509Gln) mutation [Caldovic et al., 2007] suggests not-too-drastic consequences of breaking these salt bridges between Arg509 and Tyr512. Nevertheless, salt bridge abolition by these mutations might facilitate dimer dissociation and abnormal positioning of the terminal helix (to which Arg509 and Tyr512 belong) and the ensuing C-terminal part of the enzyme, which could affect somewhat NAG binding since Phe525, near the C-terminus, belongs to the NAG binding site (Fig. 5A). The effects of the conservative p.(Ser398Cys) substitution are also difficult to explain given the similarities of the thiol and alcohol groups. One interesting possibility concerns the fact that Ser398 is expected to be close to the thiol group of CoA (Fig. 5A), which raises the possibility of disulfide formation, which would block the active site. In this case the clinical observations did not provide hints as to the severity of the effects of this mutation since this allele was in heterozygosis with the normal allele, and no second mutation was identified in the patient of family 12 that hosted the p.(Ser398Cys) mutation. Nor was a second mutation found in the patients of families 27 and 41, affected by mutations p.Tyr512Cys and p.Trp498*, respectively. It is unclear whether a second defective allele was missed despite enhancer and full gene sequencing or whether other possibilities such as negative dominance [not excluded for the p.(Ser398Cys) and p.Tyr512Cys mutations] or somatic mosaicism (not excluded for any of the three families with a single mutant allele) could cause the disease.

In summary, a combination of approaches ranging from the clinical observations to structure-based predictions and including functional studies of experimental nature provided clues to advance in our understanding of NAGSD. Yet efforts towards HuNAGS production in stabilized and highly pure form, and to determine the crystal structure of the complete

enzyme molecule, should not be stopped by the present level of understanding. The use of the pure human enzyme will allow the experimental testing of the effects of all the mutations with a higher degree of confidence than extrapolations based on a homologous bacterial enzyme as is the present system. Detailed knowledge of the complete HuNAGS architecture would add an extra level of structure-based predictions of the effects of the mutations.

Acknowledgements. The mutation analysis of NAGSD was kindly supported at JH laboratory by Orphan Europe Recordati. This does not involve any conflict of interest. We are grateful to colleagues in many countries who referred patients' samples for diagnostic purposes.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods **7**:248-249.

Bachmann C, Brandis M, Weissenbarth-Riedel E, Burghard R, Colombo JP. 1988. N-acetylglutamate synthetase deficiency, a second patient. J Inherit Metab Dis. 11:191-193.

Bachmann C, Krähenbühl S, Colombo JP. 1982. Purification and properties of acetyl-CoA:L-glutamate N-acetyltransferase from human liver. Biochem J 205:123-127.

Bachmann C, Krähenbühl S, Colombo JP, Schubiger G, Jaggi KH, Tonz O. 1981. N-acetylglutamate synthetase deficiency: a disorder of ammonia detoxication. N Engl J Med 304:543.

Bradford MM. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem.

72:248-254.

Burlina AB, Bachmann C, Wermuth B, Bordugo A, Ferrari V, Colombo JP, Zacchello F. 1992. Partial N-acetylglutamate synthetase deficiency: a new case with uncontrollable movement disorders. J Inherit Metab Dis 15:395-398.

Caldovic L, Lopez GY, Haskins N, Panglao M, Shi D, Morizono H, Tuchman M. 2006. Biochemical properties of recombinant human and mouse N-acetylglutamate synthase. Mol Genet Metab 87:226-232.

Caldovic L, Morizono H, Panglao MG, Cheng SF, Packman S, Tuchman M. 2003. Null mutations in the N-acetylglutamate synthase gene associated with acute neonatal disease and hyperammonemia. Hum Genet 112:364-368.

Caldovic L, Morizono H, Panglao MG, Gallegos R, Yu X, Shi D, Malamy MH, Allewell NM, Tuchman M. 2002a. Cloning and expression of the human N-acetylglutamate synthase gene. Biochem Biophys Res Commun 299: 581-586.

Caldovic L, Morizono H, Panglao MG, Lopez GY, Shi D, Summar ML, Tuchman M. 2005. Late onset N-acetylglutamate synthase deficiency caused by hypomorphic alleles. Hum Mutat 25:293-298.

Caldovic L, Morizono H, Tuchman M. 2007. Mutations and polymorphisms in the human N-acetylglutamate synthase (NAGS) gene. Hum Mutat 28:754-759.

Caldovic L, Morizono H, Yu X, Thompson M, Shi D, Gallegos R, Alewell NM, Malamy MH, Tuchman M. 2002b. Identification, cloning and expression of the mouse N-acetylglutamate synthase gene. Biochem J 364: 825-831.

Caldovic L, Tuchman M. 2003. N-acetylglutamate and its changing role through evolution. Biochem J 372:279-290.

Cohen PP, Hall LM, Metzenberg RL. 1956. Isolation and characterization of a naturally occurring stimulator of citrulline biosynthesis. Nature 178:1468-1469.

Cole C, Barber JD, Barton GJ. 2008. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 35 (suppl. 2):W197-W201

Coudé FX, Grimber G, Parvy P, Kamoun P. 1981. N-Acetyl glutamate synthetase in human liver: regulation of activity by L-arginine and N-acetylglutamate. Biochem Biophys Res Commun 102:1016-1020.

de Cima S, Gil-Ortiz F, Crabeel M, Fita I, Rubio V. 2012. Insight on an arginine synthesis metabolon from the tetrameric structure of yeast acetylglutamate kinase. PLoS One. 7:e34734.

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7-12.

Elpeleg ON, Colombo JP, Amir N, Bachmann C, Hurvitz H. 1990. Late-onset form of partial N-acetylglutamate synthetase deficiency. Eur J Pediatr 149:634-636.

Elpeleg O, Shaag A, Ben-Shalom E, Schmid T, Bachmann C. 2002. N-acetylglutamate synthase deficiency and the treatment of hyperammonemic encephalopathy. Ann Neurol 52:845-849.

Errey JC, Blanchard JS. 2005. Functional characterization of a novel ArgA from *Mycobacterium tuberculosis*. J Bacteriol 187:3039-3044.

Forget PP, van Oosterhout M, Bakker JA, Wermuth B, Vles JS, Spaapen LJ. 1999. Partial N-acetyl-glutamate synthetase deficiency masquerading as a valproic acid-induced Reye-like syndrome. Acta Paediatr 88:1409-1411.

Geourjon C, Deléage G. 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput Appl Biosci 11:681-684.

Gessler P, Buchal P, Schwenk HU, Wermuth B. 2010. Favourable long-term outcome after immediate treatment of neonatal hyperammonemia due to N-acetylglutamate synthase

deficiency. Eur J Pediatr 169:197-199.

Guffon N, Schiff M, Cheillan D, Wermuth B, Häberle J, Vianey-Saban C. 2005. Neonatal hyperammonemia: the N-carbamoyl-L-glutamic acid test. J Pediatr. 147:260-262.

Guffon N, Vianey-Saban C, Bourgeois J, Rabier D, Colombo JP, Guibaud P. 1995. A new neonatal case of N-acetylglutamate synthase deficiency treated by carbamylglutamate. J Inherit Metab Dis 18:61-65.

Haas D, Kurer V, Leisinger T. 1972. N-acetylglutamate synthetase of *Pseudomonas aeruginosa*. An assay in vitro and feedback inhibition by arginine. Eur J Biochem 31:290-295.

Häberle J, Boddaert N, Burlina A, Chakrapani A, Dixon M, Huemer M, Karall D, Martinelli D, Crespo PS, Santer R, Servais A, Valayannopoulos V et al. 2012. Suggested guidelines for the diagnosis and management of urea cycle disorders. Orphanet J Rare Dis 7:32.

Häberle J, Denecke J, Schmidt E, Koch HG. 2003a. Diagnosis of N-acetylglutamate synthase deficiency by use of cultured fibroblasts and avoidance of nonsense-mediated mRNA decay. J Inherit Metab Dis 26:601-605.

Häberle J, Koch HG. 2004. Genetic approach to prenatal diagnosis in urea cycle defects. Prenat Diagn 24:378-383.

Häberle J, Rubio V. 2014. Chapter 4: Hyperammonemia and related disorders. In: Blau N, Duran R, Gibson KM, Blaskovics M, Dionisi-Vici C, editors. *Physician's Guide to the diagnosis, treatment and follow-up of Inherited Metabolic Diseases*. Heidelberg, Springer-Verlag, p. 67-72.

Häberle J, Schmidt E, Pauli S, Kreuder JG, Plecko B, Galler A, Wermuth B, Harms E, Koch HG. 2003b. Mutation analysis in patients with N-acetylglutamate synthase deficiency. Hum Mutat 21:593-597.

Häberle J, Shchelochkov OA, Wand J, Katsonis P, Hall L, Reiss S, Eeds A, Willis A, Yadav M, Summar S, Urea Cycle Disorders Consortium, Lichtarge O, Rubio V, Wong LJ, Summar M. 2011. Molecular defects in human carbamoyl phosphate synthetase I: Mutational spectrum, diagnostic and protein structure considerations. Hum Mutat 32:579-589.

Haskins N, Panglao M, Qu Q, Majumdar H, Cabrera-Luque J, Morizono H, Tuchman M, Caldovic L. 2008. Inversion of allosteric effect of arginine on N-acetylglutamate synthase, a molecular marker for evolution of tetrapods. BMC Biochem 9:24.

Heibel SK, Ah Mew N, Caldovic L, Daikhin Y, Yudkoff M, Tuchman M. 2011. N-carbamylglutamate enhancement of ureagenesis leads to discovery of a novel deleterious mutation in a newly defined enhancer of the NAGS gene and to effective therapy. Hum Mutat. 32:1153-1160.

Kretz R, Hu L, Wettstein V, Leiteritz D, Häberle J. 2012. Phytohemagglutinin stimulation of lymphocytes improves mutation analysis of carbamoylphosphate synthetase 1. Mol Genet Metab 106:375-378.

Kuchler G, Rabier D, Poggi-Travert F, Meyer-Gast D, Bardet J, Drouin V, Cadoudal M, Saudubray JM. 1996. Therapeutic use of carbamylglutamate in the case of carbamoyl-phosphate synthetase deficiency. J Inherit Metab Dis. 19:220-222.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25: 2744-2750.

Martinelli D, Häberle J, Rubio V, Giunta C, Hausser I, Carrozzo R, Gougeard N, Marco-Marín C, Goffredo BM, Meschini MC, Bevivino E, Boenzi S et al. 2012.

Understanding pyrroline-5-carboxylate synthetase deficiency: clinical, molecular, functional, and expression studies, structure-based analysis, and novel therapy with arginine. J Inherit Metab Dis; 35:761-776.

Marvil DK, Leisinger T. 1977. N-acetylglutamate synthase of *Escherichia coli*: purification, characterization, and molecular properties. J Biol Chem 252:3295-3303.

McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server. Bioinformatics 16:404-405.

Min L, Jin Z, Caldovic L, Morizono H, Allewell NM, Tuchman M, Shi D. 2009. Mechanism of allosteric inhibition of N-acetyl-L-glutamate synthase by L-arginine. J Biol Chem 284:4873-4880

Mitchell S, Ellingson C, Coyne T, Hall L, Neill M, Christian N, Higham C, Dobrowolski SF, Tuchman M, Summar M; Urea Cycle Disorder Consortium. 2009. Genetic variation in the urea cycle: a model resource for investigating key candidate genes for common diseases. Hum Mutat. 2009;30:56-60.

Nordenstrom A, Halldin M, Hallberg B, Alm J. 2007. A trial with N-carbamylglutamate may not detect all patients with NAGS deficiency and neonatal onset. J Inherit Metab Dis 30:400.

Plecko B, Erwa W, Wermuth B. 1998. Partial N-acetylglutamate synthetase deficiency in a 13-year-old girl: diagnosis and response to treatment with N-carbamylglutamate. Eur J Pediatr 157:996-998.

Rubio V, Britton HG, Grisolia S. 1983. Mitochondrial carbamoyl phosphate synthetase activity in the absence of N-acetyl-L-glutamate. Mechanism of activation by this cofactor. Eur J Biochem 134:337-343.

Rubio V, Grisolia S. 1981a. Human carbamoylphosphate synthetase I. Enzyme 26:233-239.

Rubio V, Grisolía S. 1981b. Treating urea cycle defects. Nature 292:496.

Rubio V, Ramponi G, Grisolia S. 1981. Carbamoyl phosphate synthetase I of human liver. Purification, some properties and immunological cross-reactivity with the rat liver enzyme. Biochim Biophys Acta 659:150-160.

Sancho-Vaello E, Fernandez-Murga ML, Rubio V. 2008. Site-directed mutagenesis studies of acetylglutamate synthase delineate the site for the arginine inhibitor. FEBS Lett 582:1081-1086.

Sancho-Vaello E, Fernandez-Murga ML, Rubio V. 2009. Mechanism of arginine regulation of acetylglutamate synthase, the first enzyme of arginine synthesis. FEBS Lett 583:202-206.

Sancho-Vaello E, Fernández-Murga ML, Rubio V. 2012. Functional dissection of N-acetylglutamate synthase (ArgA) of *Pseudomonas aeruginosa* and restoration of its ancestral N-acetylglutamate kinase activity. J Bacteriol. 194:2791-2801.

Schmidt E, Nuoffer JM, Häberle J, Pauli S, Guffon N, Vianey-Saban C, Wermuth B, Koch HG. 2005. Identification of novel mutations of the human N-acetylglutamate synthase gene and their functional investigation by expression studies. Biochim Biophys Acta 1740:54-59.

Sen TZ, Jernigan RL, Garnier J, Kloczkowski A. 2005. GOR V server for protein secondary structure prediction. Bioinformatics 21:2787-2788.

Senkevitch E, Cabrera-Luque J, Morizono H, Caldovic L, Tuchman M. 2012. A novel biochemically salvageable animal model of hyperammonemia devoid of N-acetylglutamate synthase. Mol Genet Metab 106:160-168.

Shi D, Li Y, Cabrera-Luque J, Jin Z, Yu X, Zhao G, Haskins N, Allewell NM, Tuchman M. 2011. A novel N-acetylglutamate synthase architecture revealed by the crystal structure of the bifunctional enzyme from *Maricaulis maris*. PLoS One 6:e28825

Shi D, Sagar V, Jin Z, Yu X, Caldovic L, Morizono H, Allewell NM, Tuchman M. 2008. The

crystal structure of N-acetyl-L-glutamate synthase from *Neisseria gonorrhoeae* provides insights into mechanisms of catalysis and regulation. J Biol Chem 283: 7176-7184.

Shigesada K, Tatibana M. 1971. Enzymatic synthesis of acetylglutamate by mammalian liver preparations and its stimulation by arginine. Biochem Biophys Res Commun 44:1117-1124.

Van Karnebeek CD, Sly WS, Ross CJ, Salvarinova R, Yaplito-Lee J, Santra S, Shyr C, Horvath GA, Eydoux P, Lehman AM, Bernard V, Newlove T, et al. 2014. Mitochondrial carbonic anhydrase VA deficiency resulting from *CA5A* alterations presents with hyperammonemia in early childhood. Am J Hum Genet 94:453-461.

Van Leynseele A, Jansen A, Goyens P, Martens G, Peeters S, Jonckheere A, De Meirleir L. 2014. Early treatment of a child with NAGS deficiency using N-carbamyl glutamate results in a normal neurological outcome. Eur J Pediatr 173:1635-1638.

Vedadi M, Niesen FH, Allali-Hassani A, Fedorov OY, Finerty PJ Jr, Wasney GA,Yeung R, Arrowsmith C, Ball LJ, Berglund H, Hui R, Marsden BD et al. 2006. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. Proc Natl Acad Sci U S A. 103: 15835-15840.

Vockley J, Vockley CM, Lin SP, Tuchman M, Wu TC, Lin CY, Seashore MR. 1992. Normal N-acetylglutamate concentration measured in liver from a new patient with N-acetylglutamate synthetase deficiency: physiologic and biochemical implications. Biochem Med Metab Biol 47:38-46.

Zhao G, Haskins N, Jin Z, M Allewell N, Tuchman M, Shi D. 2013a. Structure of N-acetyl-L-glutamate synthase/kinase from Maricaulis maris with the allosteric inhibitor L-arginine bound. Biochem Biophys Res Commun 437:585-590.

Zhao G, Jin Z, Allewell NM, Tuchman M, Shi D. 2013b. Crystal structure of the N-acetyltransferase domain of human N-acetyl-L-glutamate synthase in complex with N-acetyl-L-glutamate provides insights into its catalytic and regulatory mechanisms. PLoS One. 8:e70369.

Zhou H, Brockington M, Jungbluth H, Monk D, Stanier P, Sewry CA, Moore GE, Muntoni F. 2006. Epigenetic allele silencing unveils recessive *RYR1* mutations in core myopathies. Am J Hum Genet 79:859-868.

Legends to Figures.

Figure 1. Distribution of mutations among the seven NAGS exons and the protein polypeptide. An enhancer mutation, which mapped much upstream from the coding sequence (c.-3064C>A), found in homozygosis in one patient and strongly hampering *NAGS* gene expression [Heibel et al., 2011], is not included. The *NAGS* gene coding sequence is schematized as a horizontal bar at the top, with segments therein corresponding to individual exons and numbered, with odd exons shadowed. An asterisk marks the end of the coding sequence in the incomplete exon 7. A parallel bar at the bottom schematizes the enzyme polypeptide in exact correspondence with the exons encoding it. The mitochondrial signal peptide, and the variable, amino acid kinase (AAK) and acetyltransferase (GNAT) domains, as well as the linker between the last two domains are shown in different colors, with figures below the bar giving amino acid numbers at their boundaries [Caldovic et al., 2002a,b]. The α-helices and β-strands predicted for the AAK domain (see Materials and Methods; asterisks mark three elements with non-unanimous predictions) or observed in the crystal structure of the GNAT domain [Zhao et al., 2013b], are respectively shown as striped or solid rectangles of proportional length to the length of their sequences. They are named as in Fig. 2. The β-strands belonging to the central sheet of the AAK domain are boxed. Banners identify the mutations (in single-letter amino acid code) found in NAGSD patients (Tables 1 and 2) where they map in the cDNA sequence and along the protein sequence. The codes to identify the different types of mutations, whether the mutation causes neonatal- or late-onset presentations (when the data allow to infer this information), and whether the mutation is novel (blue letters) or was already reported, are provided in the figure itself. The p.G236C missense change was found in heterozygosis in a control DNA. The positions of the residues shown to bind NAG or inferred by superimposition with other structures to be involved in the binding

of CoA [Zhao et al., 2013b] or of arginine [De Cima et al., 2012] are marked with white, grey or light blue triangles, respectively. A grey rectangle marks the (Arg/Gln)-Xaa-Xaa-Gly-Xaa-(Gly/Ala) consensus AcCoA binding motif (see the text).

Figure 2. Secondary structure-assisted alignment of the sequences of PaNAGS and HuNAGS with those of *Neisseria gonorrhoeae* NAGS (NgNAGS), NAGS/NAG kinase (a bifunctional enzyme) from the bacterium *Maricaulis maris* (MmNAGS/K) and yeast NAG kinase (ScNAGK). ScNAGK has the same domain composition and substantial sequence similarity with human and yeast NAGSs, and very similar structure to MmNAGS/K [de Cima et al., 2012; Shi et al., 2011]. The sequences are those given in UniProt KB files P22567 and Q8N159 (http://www.uniprot.org) for PaNAGS and HuNAGS, respectively, and those given in the Protein Databank (PDB, http://www.rcsb.org/pdb/home/home.do) files 3D2M, 3S6H and 4AB7 for NgNAGS, MmNAGS/K and and ScNAGK, respectively. The secondary structure elements are symbolized as rectangles and arrows for α helices and β strands, respectively, placed immediately below the corresponding sequence. They are named or numbered for the AAK domain as in *P. aeruginosa* NAGK (Ramon-Maiques et al., 2006) and for the GNAT domain as in [De Cima et al., 2012], starting from 1' (strands) or A' (helices) with *prime* signs, giving above the alignment the elements of the GNAT domain of NgNAGS and below it those of this domain of ScNAGK, MmNAGS/K and HuNAGK. Secondary structure elements of PaNAGS are those in the structural model (http://www.proteinmodelportal.org/query/uniprot/P22567) built by Swissprot when using as templates the structures of the AAK domain of NgNAGS (PDB file 3D2MA [Shi et al., 2009]; 40% sequence identity) and of the GNAT domain *of Vibrio parahaemolyticus* NAGS (PDB 3E0KA; 53% sequence identity). The secondary structure of the AAK domain of HuNAGS has been predicted from the sequence (see Materials and Methods). All other

secondary structure elements are those in the experimental crystal structures (see above PDB files; [de Cima et al., 2012; Shi et al. 2009 & 2011]). Amino acid identities or conservation with HuNAGS are illustrated by white lettering on a grey background. Amino acid substitutions found in patients with NAGSD are indicated by showing above the sequences, boxed, the variant amino acid (in red when experimentally studied here in PaNAGS).

Figure 3. Evidence for stronger clinical impact of missense mutations affecting the GNAT domain than of those affecting the AAK domain. **(A)** Plot of the number of missense mutations (left bars) or of families with missense mutations (right bars) mapping in one or the other domain of HuNAGS. Data are normalized per 100 amino acids, and are given for either all the missense mutations or all the families carrying missense mutations (indicated as total), or after subtracting the mutations that fell on CpG dinucleotides (indicated as -CpG), to correct for the increased mutational frequency inherent to these dinucleotides. **(B)** Normalized frequency of missense mutations per exon in NAGSD. The thin trace corresponds to all the mutations, whereas the thick trace gives the mutations that did not fall on CpG dinucleotides. On top, a bar schematizes the coding sequence, in the scale given in the X-axis, showing the exons and superimposing on them the regions encoding the various enzyme domains (V, variable domain) in different shades of white or grey. Vertical lines below the bar mark missense mutations. **(C)** Same as (B) for the number of families who carried mutations. No lines with mutations are shown below the exon structure bar at the top. **(D)** Number of missense mutations that mapped in each domain, classified according to the type of presentation they caused. The figure shows only the mutations for which the existing data are informative.

Figure 4. Use of recombinant PaNAGS to test the effects of missense mutations (shown in

single letter amino acid code) found in NAGSD. Gels correspond to Coomassie-stained SDS-PAGE, where black arrowpoints mark the position of PaNAGS. **St**, molecular mass markers, with masses indicated in kDa. **WT**, wild-type PaNAGS. **(A)** SDS-PAGE of soluble (S) and insoluble (I) fractions of the sonicated and centrifuged crude extracts of *E. coli* expressing the indicated (in black lettering) WT or mutant form of PaNAGS (abbreviated Paer). To facilitate translation of the results to human mutants, the corresponding HuNAGS (abbreviated Hum) mutations are shown in grey lettering below the PaNAGS mutation. Precipitates were reconstituted to the original volume of extract and equal volumes of supernatant and precipitate were applied to each well. **(B)** SDS-PAGE of the preparations obtained after purification for the WT and the indicated mutant forms (again shown in black and grey for the PaNAGS and HuNAGS corresponding mutations). The mutations mapping in the GNAT domain are boxed. Below gels, NAGS enzymatic activity (mean ± SE) determined as indicated in Materials and Methods, in the presence of 30 mM glutamate (unless indicated) and 4 mM AcCoA, in the absence or presence of 1 mM arginine. [a]Indicates that the concentration of L-glutamate used was 300 mM. **(C)** and **(D)** Substrate kinetics for WT and mutant forms of PaNAGS, as indicated. Conditions and kinetic data are given in Table 3. Insets expand the Y-axis for the W397R mutant. **(E)** Thermofluor experiments monitoring increase in fluorescence with temperature as a fraction of the maximum increase for WT PaNAGS and the indicated mutants. The broken lines mark a change in fluorescence corresponding to 50% of the maximum ($T_{0.5}$). For further details, see Materials and Methods.

Figure 5. Mapping of residues affected by NAGSD-associated mutations in the crystal structure of the GNAT domain of HuNAGS. The GNAT domain structure is that determined with bound NAG (in sticks representation) by [Zhao et al., 2013b] (PDB entry 4K30). **(A)** Cartoon representation of the entire domain together with the C-terminal part of the adjacent

GNAT domain in the dimer (in green string representation) and a coenzyme A molecule superimposed essentially as in [Zhao et al., 2013b] (shown also in sticks). Secondary structure elements are labelled as at the bottom of the alignment for the GNAT domain in Fig. 2. Some relevant side-chains are shown and labelled. Spheres localize residues (labelled in single letter code) for which NAGSD-associated mutations have been identified (red, cyan and grey, mutations giving respectively neonatal-, late-onset or unknown presentations). The side chains of the mutations-hosting residues Arg509, Tyr512 and Ala518 are shown in sticks representation. The first two are illustrated (in the color of the corresponding main-chain, since no clinical severity information exists for them), to show their involvement in intersubunit salt bridges, while the side chain of Ala518 is shown in blue to indicate that the reported mutation at this residue was associated with partial deficiency. **(B)** Surface representation of the NAG binding site with NAG (in sticks) bound to it, with the surface corresponding to Leu442 coloured magenta and labelled. **(C)** Residues that participate in the ion pair network involving NAG are shown with side chains in sticks. Mutations are shown in parentheses and coloured red because of their neonatal presentation. Ion pairs are shown as dashed lines.

Table 1. Genotypes of families and NAGSD patients with missense mutations of the *NAGS* gene in at least one allele.

Family #	Number of affected patients	Allele	Genotype	Exon/ Intron	Protein or splicing aberration	NAGS domain	Onset and clinical data	Outcome	PolyPhen-2 prediction	MutPred prediction	
										g-score	Hypothesis
1[a]	**1**	**1** **2**	**c.499A>G** **c.278delC**	**E2** **E1**	**p.(M167V)** **p.(P93Qfs*18)**	**AAK** **AAK**	**Vomiting, poor feeding, episodic confusion as neonate, diagnosed at m6**	**Normal at y20 with treatment**	**Probably damaging**	**0.398**	
2[b]	1	1 2	c.518T>A c.1292C>T	E2 E6	p.(V173E) p.(T431I)	AAK GNAT	Combativeness, confusion and seizures postoperatively at y33, when hyperammonemia was documented; had had transient coma at y5	Death from cerebral edema after few days	Possibly damaging Probably damaging	0.895 0.848	
3[c] 4[d]	2 1	1 2	c.598T>C c.598T>C	E2	p.(C200R) p.(C200R)	AAK	Fam 3. Index: hypotonia (m2.5), later hepatomegaly, failure to thrive, hyperammonemia Older sibling: genetic diagnosis, asymptomatic. Fam 4. Onset at m3, hypotonia, develop. delay	No data	Possibly damaging	0.818	
Control DNA[e]		1 2	*c.706G>T* c.706C	E3	p.(G236C) p.G236	AAK	Heterozygous mutation reported in a control sample. No further information.	No data	Probably damaging	0.890	
5[a]	**1**	**1** **2**	**c.779C>T** **c.779C>T**	**E3**	**p.(P260L)** **p.(P260L)**	**AAK**	**Neonatal hyperammonemia**	**No data**	**Possibly damaging**	**0.818**	
6[a]	**1**	**1** **2**	***c.791C>T*** ***c.791C>T***	**E3**	**p.T264M** **p.T264M**	**AAK**	**Onset at d20, vomiting, failure to thrive, hepatomegaly, recurrent asymptomatic hyperammonemia**	**Normal at y5 with benzoate, citrulline, protein restriction**	**Possibly damaging**	**0.493**	
7[f,g]	1	1 2	*c.835G>A* *c.835G>A*	E3	p.(A279T) p.(A279T)	AAK	Onset at m13, vomiting, hypotonia, somnolence	Normal cogn. function & neurol. findings at y24, NCG-treated	Benign	0.477	
8[a]	**4**	**1** **2**	**c.872T>A** **c.872T>A**	**E3**	**p.(I291N)** **p.(I291N)**	**AAK**	**All four patients had late-onset and mild course**	**Normal without NCG**	**Possibly damaging**	**0.822**	**C: disorder gain**
9[e]	1	1 2	c.935T>C c.1292C>T	E4 E6	p.(L312P) p.(T431I)	AAK GNAT	Onset at y27, during pregnancy, with seizures and coma	No data	Probably damaging Probably damaging	0.869 0.848	VC: stability loss
10[e]	1	1 2	*c.1048G>A*[h] c.1324C>G	E4 E6	p.(V350I) p.(L442V)	AAK GNAT	Late-onset at y40 with migraine, nausea, vomiting, lethargy, ataxia, coma	Normal at y57	Benign Probably damaging	0.733 0.912	
11[a]	**1**	**1** 2	**c.1172T>G** c.1450T>C[i]	**E5** **E6**	**p.(L391R)** p.(W484R)	**GNAT** GNAT	**Onset m2 severe hyperammonemia. Elder brother dead at 2y (no material or diagnosis)**	**Stable during 6-month NCG treatment; death after stopping NCG**	**Probably damaging** Probably damaging	**0.753** 0.969	VC: disorder gain
12[a]	**1**	**1** **2**	**c.1192A>T** **Not found**	**E5**	**p.(S398C)** **Not found**	**GNAT**	**Onset y15, with UTI, somnolence, decreased conscience, peak ammonia 350 µmol/L**	**Normal**	**Probably damaging**	**0.639**	
13[c] 14[a]	2 1	1 2	c.1228T>C c.1228T>C	E5	p.(S410P) p.(S410P)	GNAT	Fam 13. Index: onset d2-4, poor feeding, seizures, lethargy. Younger sibling: NCG-treated before symptoms developed. Fam 14. Onset d3, vomiting, lethargy.	Fam 13. Index normal y1. Sibling well, treated Fam 14. Death at d20, no NCG	Probably damaging	0.833	
15[j]	2	1 2	*c.1241G>C* *c.1241G>C*	E5	p.(R414P) p.(R414P)	GNAT	Index: onset at d2-3, poor feeding, somnolence, seizures. Younger sibling: NCG-treated before symptoms developed	Index: at y2.5 normal psychomotor develop. Sibling: no data	Probably damaging	0.541	

16[f] 17[a]	1 1	1 2	c.1289T>C c.1289T>C	E6	p.L430P p.L430P	GNAT	Fam. 16. Onset d2, with hyperammonemic encephalopathy Fam. 17. Neonatal onset	Fam. 16. Death d4 with cerebral edema	Probably damaging	0.929	C: stability loss
18[f]	1	1 2	c. 1299G>C c.1299G>C	E6	p.(E433D) p.(E433D)	GNAT	Onset d6, poor feeding and lethargy	Died at m3	Benign	0.929	
19[a]	**1**	**1** **2**	**c.1370G>A** **c.1370G>A**	**E6**	**p.G457D** **p.G457D**	**GNAT**	**Neonatal onset**	**Died at d5 with hyperammonemia**	**Probably damaging**	**0.633**	
20[a,f] 21[a] 22[a] 23[a] 24[a] 25[k]	3 1 1 1 1 1	1 2	c.1450T>C[i] c.1450T>C[i]	E6	p.(W484R) p.(W484R)	GNAT	Fam 20. Index: onset d3 hyperammonemic coma. Sibling and relative: genetic demonstration and NCG-treated before symptoms developed. Fam. 21. Onset on d2, poor feeding, seizures. Fam. 22. Neonatal onset with hyperammonemia. Fam. 23. Onset d3; standard treatment for hyperammonemia without NCG. Fam. 24. Neonatal onset with hyperammonemia. Fam. 25. Hyperammonemia at m1, diagnosed at m22 while under treatment with diet and NCG.	Fam.20. Index dead m6.Sibling, relatives: no data. Fam. 21. Psychomotor delay and spasticity. NCG treatment. Fam. 22-24. No data. Fam. 25. Alive at m22 while treated.	Probably damaging	0.969	VC: disorder gain
26[b]	2	1 2	*c.1526G>A* c.1097-1G>C[l]	E7 I4	p.(R509Q) IVS4-1G>C	GNAT GNAT	Index patient: onset w4, vomiting, irritability, lethargy, many hyperammonemic episodes, diagnosis confirmed at y18. Sibling: onset at y9, lethargy, anorexia, vomiting, seizures.	Index: Normal cognitive development. Sibling: No data.	Probably damaging	0.826	
27[a]	**1**	**1** **2**	**c.1535A>G** **Not found**	**E7**	**p.Y512C** **Not found**	**GNAT**	**Reye-like syndrome at 16 years with hyperammonemia**	**Normal**	**Probably damaging**	0.583	
28[c] 29[a]	1 1	1 2	*c.1552G>A* *c.1552G>A*	E7	p.(A518T) p.(A518T)	GNAT	Fam.28: Onset d2, hypotonia, poor feeding, lethargy. Fam.29: Investigated at y4 when presenting with convulsions and microcephaly.	Fam. 28: No data. Fam. 29: Alive at y4, microcephaly.	Probably damaging	**0.831**	

GenBank reference sequences for the human *NAGS* gene and for its mRNA, NG_008106.1 and NM_153006.2, respectively. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon reference sequence, with the initiation codon as codon 1. Novel families with new mutations reported here are highlighted in bold-type. When the mutation affects a CpG island the nucleotide change is shown in italic type. AAK, amino acid kinase domain, corresponding to amino acids 95-372; GNAT, acetyltransferase domain (amino acids 376-534). A *d*, *w*, *m* or *y* followed by a number indicates the day, week, month or year of life, respectively (for example, *d2* means day 2 of life). PolyPhen-2 (HumVar-trained dataset) grades the probability of a damaging effect of an amino acid substitution, as *probably damaging*, *possibly damaging* and *benign*. MutPred gave a *g* score corresponding to the probability that a given amino-acid substitution was deleterious/disease-associated. When indicated, this server was very confident (VC) or confident (C) that the indicated changes caused loss of stability or gain of disorder.

[a]This study

[b] [Caldovic et al., 2005].

[c] [Schmidt et al., 2005].

[d]This patient was erroneously described as CPS1 deficient in [Kuchler et al., 1996].

[e] [Caldovic et al., 2007].

[f] [Häberle et al., 2003b].

[g] [Plecko et al., 1998].

[h]This mutation was probably erroneously reported as c.1050G>C in [Caldovic et al., 2007].

[i]This mutation changes a T by a C at the penultimate base of exon 6 at the exon 6/intron 6 donor splice site. Thus, it might also affect splicing of exons 6-7 (as indicated by the yellow/blue flag in Fig. 1.

[j] [Nordenstrom et al., 2007].

[k] [Van Leynseele et al., 2013].

[l]This mutation was probably erroneously reported as c.1096G>C in [Caldovic et al., 2005]

Table 2. Genotypes of families and NAGSD patients hosting mutations causing NAGS truncation.

Family #	Number of affected patients	Allele	Nucleotide changes	Exon/ Intron	Protein changes	Domain loss		Onset, clinical data and outcome
						AAK	GNAT	
30[a] 31[b]	1 1	1 2	c.545delC c.545delC	E2	p.(A182Vfs*23) p.(A182Vfs*23)	75%	Total	Family 30. Neonatal death with hyperammonemic coma. Family 31. Onset at 24 hours, hyperammonemic coma, died at 72 hours
32[c,d]	1	1 2	c.916-2A>T c.1307dup	I3 E6	p.V306Gfs*26 p.T439Hfs*52	25% No loss	Total 60%	Onset day 3, hyperammonemic encephalopathy, death on day 6
33[d,e] 34[f]	2 1	1 2	c.971G>A c.971G>A	E4	p.(W324*) p.(W324*)	20%	Total	Family 33. Index: onset day 4, death day 22 hyperammonemic encephalopathy. Sibling: onset day 2, treated but unstable until NCG started at 3 months. Normal neurological status at 13 years Family 34. Onset day 1, death on day 4
35[b]	**2**	**1** **2**	**c.991C>T** **c.991C>T**	**E4**	**p.(Q331*)** **p.(Q331*)**	**15%**	**Total**	**Index: neonatal onset, hyperammonemic coma, death on day 7** **Sibling: neonatal onset, hyperammonemic coma, death on day 19**
36[f]	2	1 2	*c.1025delG* *c.1025delG*	E4	p.(R342Pfs*50) p.(R342Pfs*50)	9%	Total	Index: onset on day 2, liver transplantation at 8 months, no information on neurological status. Sibling: onset on day 2
37[g]	2	1 2	c.1036dup c.1036dup	E4	p.(H346Pfs*10) p.(H346Pfs*10)	7%	Total	Index: onset day 4, treated but unstable until NCG started at 36 months. Suffers neurological sequelae. Sibling: prospective treatment from birth, but NCG only started at 3 months. Has neurological sequelae
38[b]	**1**	**1** **2**	***c.1264G>T*** ***c.1264G>T***	**E5**	**p.(E422*)** **p.(E422*)**	**No loss**	**70%**	**No clinical information**
39[b]	**1**	**1** **2**	**c.1313dup** **c.1313dup**	E6	p.(T439Hfs*52) p.(T439Hfs*52)	No loss	60%	**Onset on day 2 with hyperammonemia requiring dialysis; alive under treatment including NCG.**
40[b]	**2**	**1** **2**	**c.1313delG** **c.1313delG**	**E6**	**p.G438Afs*10** **p.G438Afs*10**	**No loss**	**60%**	**Index: Onset day 2 with vomiting and feeding problems, hyperammonemia was diagnosed and treated with hemodiafiltration. Normal development under treatment with NCG at 1 year.** **Elder sibling: onset first few days, death in early neonatal period, no ammonia measurements done.**
41[b]	**1**	**1** **2**	**c.1494G>A** **not found**	**E7**	**p.W498*** **not found**	**No loss**	**23%**	**Neonatal onset with hyperammonemia (max. 368 µmol/L). Metabolically stable with NCG.**

Genbank reference sequences for the human *NAGS* gene and for its mRNA, NG_008106.1 and NM_153006.2, respectively. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. New families with novel genotypes reported here are highlighted in bold-type. When the mutation affects a CpG island the nucleotide change is shown in italic type. AAK, amino acid kinase domain (amino acids 95-372); GNAT, acetyltransferase domain (amino acids 376-534).

[a] [Häberle and Koch, 2004]

[b] This study

[c] [Häberle et al., 2003a]

[d] [Häberle et al., 2003b]

[e] [Gessler et al., 2010]

[f] [Caldovic et al., 2003]

[g] [Elpeleg et al., 2002]

Table 3. Influence of NAGSD missense mutations, introduced into PaNAGS, on the substrate kinetics of this enzyme

HuNAGS missense mutation	Corresponding PaNAGS mutation	Acetyl Coenzyme A			Glutamate		
		$V^{[AcCoA]=\infty}$ U mg^{-1}	K_m^{App} mM	K_i mM	$V^{[Glu]=\infty}$ U mg^{-1}	K_m^{App} mM	K_i mM
Wild type		81 ± 1	0.11 ± 0.01	---	136 ± 20	5.2 ± 1.4	72 ± 36
AAK domain mutations							
p.(Met167Val)	p.Leu50Val	52 ± 1	0.05 ± 0.01	---	78 ± 3	3.0 ± 0.3	84 ± 8
p.(Val173Glu)	p.Leu56Glu	< 0.1[a]	---	---	< 0.1[a]	---	---
p.(Gly236Cys)	p.Gly146Cys[b]	43 ± 1	0.11 ± 0.01	---	46 ± 1	170 ± 13	---
p.(Pro260Leu)	p.Pro170Leu	56 ± 3	0.05 ± 0.01	9.5 ± 2.6	57 ± 1	5.7 ± 0.3	---
p.Thr264Met	p.Ser174Met[b]	46 ± 2	0.09 ± 0.02	10.5 ± 2.8	42 ± 2	93.1 ± 9.0[c]	---
p.(Ala279Thr)	p.Met189Thr	64 ± 1	0.06 ± 0.01	---	55 ± 2	4.3 ± 0.5	136 ± 21
p.(Ile291Asn)	p.Ile201Asn	64 ± 1	0.04 ± 0.01	---	96 ± 7	9.4 ± 1.3	155 ± 35
p.(Val350Ile)	p.Val258Ile	60 ± 3	0.09 ± 0.03	---	70 ± 3	6.3 ± 0.6	107 ± 12
GNAT domain mutations							
p.Leu430Pro	p.Leu342Pro	< 0.1[a]	---	---	< 0.1[a]	---	---
p.(Leu442Val)	p.Leu353Val	73 ± 1	0.08 ± 0.01	---	101 ± 6	5.9 ± 0.8	222 ± 53
p.Gly457Asp	p.Gly368Asp	< 0.1[a]	---	---	< 0.1[a]	---	---
p.(Trp484Arg)	p.Trp397Arg[b]	3.7 ± 0.8	4.8 ± 1.6	---	3.7 ± 0.3	478 ± 88	---

PaNAGS is the product of the *ArgA* gene (Genbank locus *PA5204;* gene ID: 879475) of *P. aeruginosa* strain PAO1 (genome sequence in Genbank, NC_002516.2). Nucleotide numbering uses +1 as the A of the ATG translation initiation codon, with the initiation codon as codon 1. Kinetic parameters were determined as indicated in Materials and Methods and are given as mean ± SE. K_m^{app} denote K_m values determined at fixed L-glutamate or AcCoA concentrations of 30mM (unless indicated) or 4 mM, respectively. K_i, inhibition constant for the indicated substrate for substrate inhibition [Sancho-Vaello et al., 2009].

[a]Detection limit

[b]AcCoA kinetic parameters determined in the presence of 300mM fixed L-glutamate.

$^{c}S_{0.5}$ value (the concentration of L-glutamate at which the velocity was 50% of $V^{[Glu]=\infty}$), since the dependency of the velocity on the concentration of glutamate fitted best sigmoidal kinetics with a Hill number $H = 1.8 \pm 0.2$.

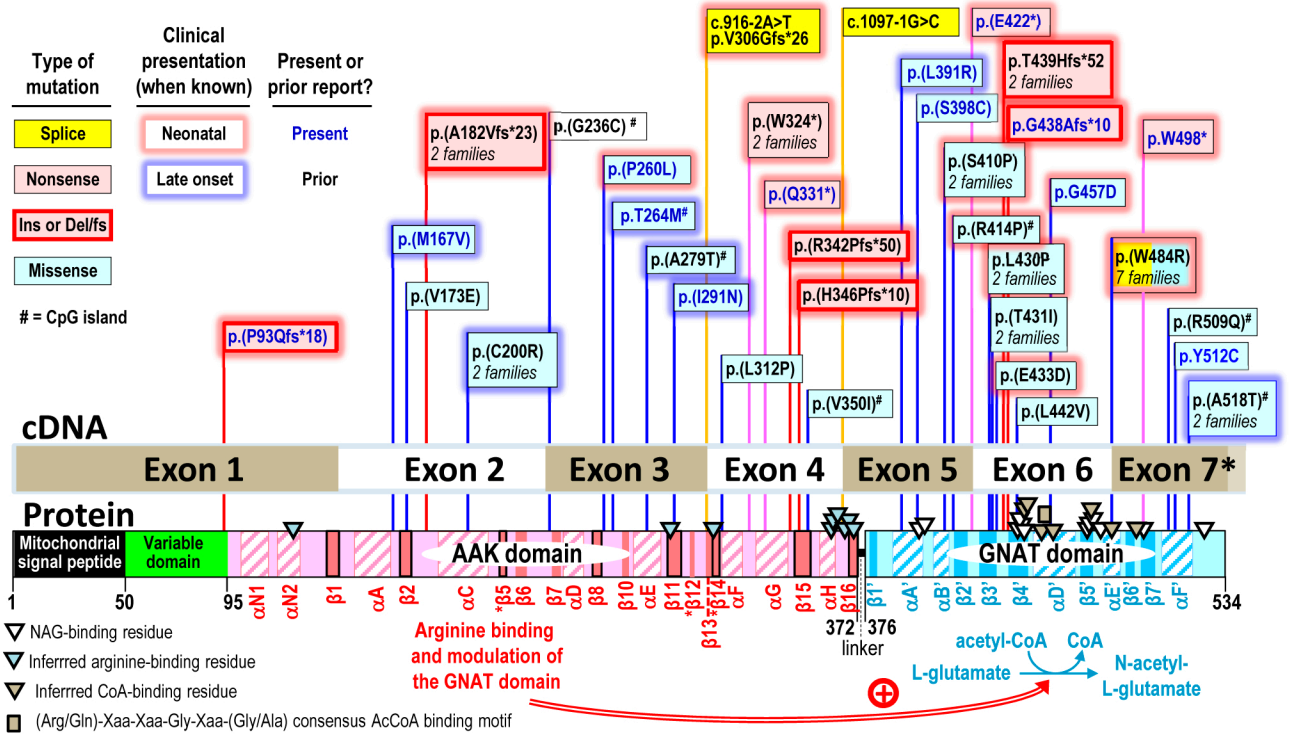
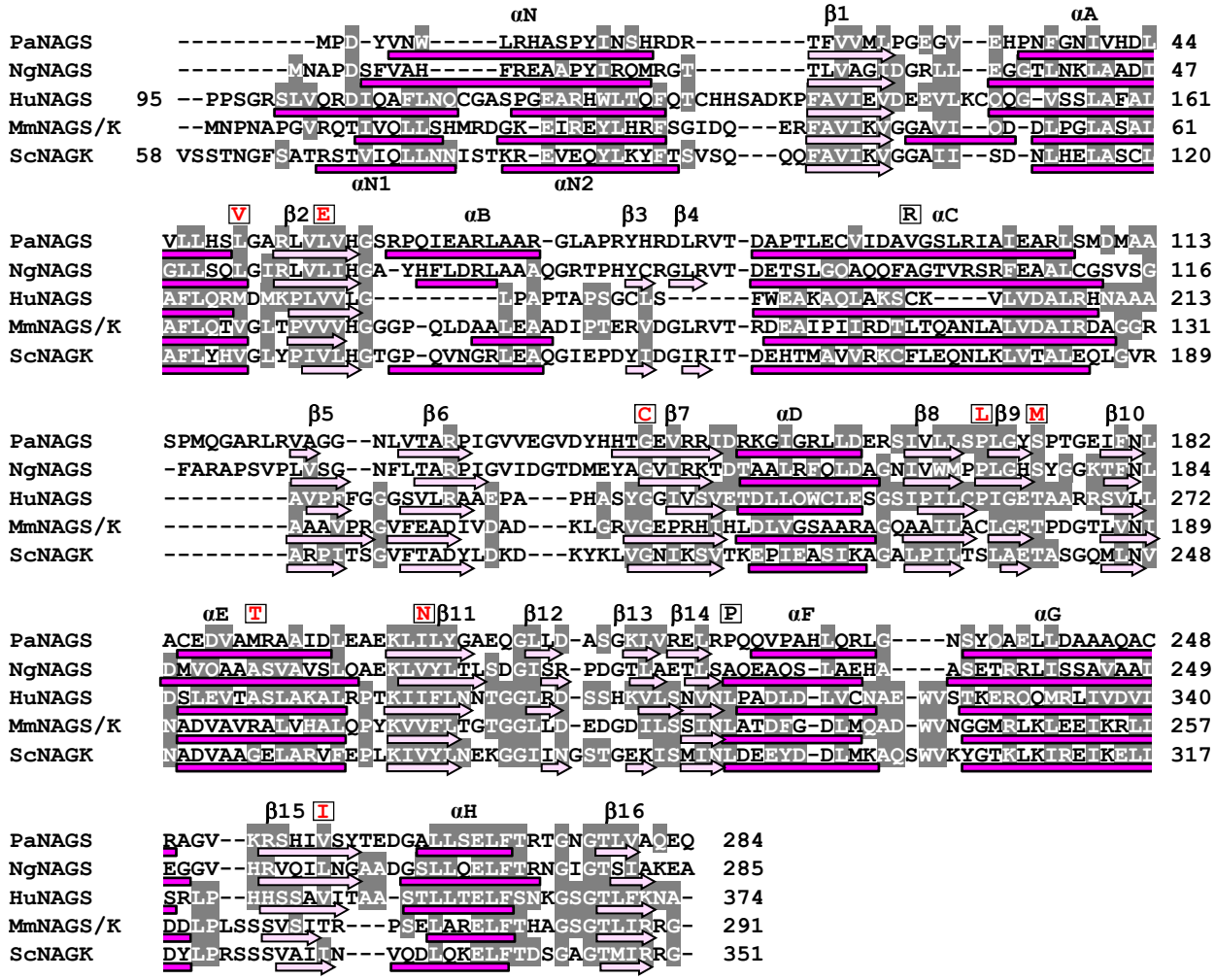

Type of mutation
Splice
Nonsense
Ins or Del/fs
Missense
= CpG island
Clinical presentation (when known)
Neonatal
Late onset
Present or prior report?
Present
Prior
p.(P93Qfs*18)
p.(M167V)
p.(V173E)
p.(A182Vfs*23) 2 families
p.(C200R) 2 families
p.(G236C) #
p.(P260L)
p.T264M#
p.(A279T)#
p.(I291N)
c.916-2A>T p.V306Gfs*26
p.(W324*) 2 families
p.(Q331*)
p.(L312P)
p.(R342Pfs*50)
p.(H346Pfs*10)
p.(V350I)#
c.1097-1G>C
p.(L391R)
p.(S398C)
p.(E422*)
p.T439Hfs*52 2 families
p.G438Afs*10
p.(S410P) 2 families
p.(R414P)#
p.G457D
p.L430P 2 families
p.(T431I) 2 families
p.(E433D)
p.(L442V)
p.W498*
p.(W484R) 7 families
p.(R509Q)#
p.Y512C
p.(A518T)# 2 families
cDNA
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7*
Protein
Mitochondrial signal peptide
Variable domain
AAK domain
GNAT domain
1
50
95
372
376
534
linker
NAG-binding residue
Inferrred arginine-binding residue
Inferrred CoA-binding residue
(Arg/Gln)-Xaa-Xaa-Gly-Xaa-(Gly/Ala) consensus AcCoA binding motif
Arginine binding and modulation of the GNAT domain
acetyl-CoA
CoA
L-glutamate
N-acetyl-L-glutamate

Fig. 1

AMINO ACID KINASE (AAK) DOMAIN



ACETYLTRANSFERASE (GNAT) DOMAIN

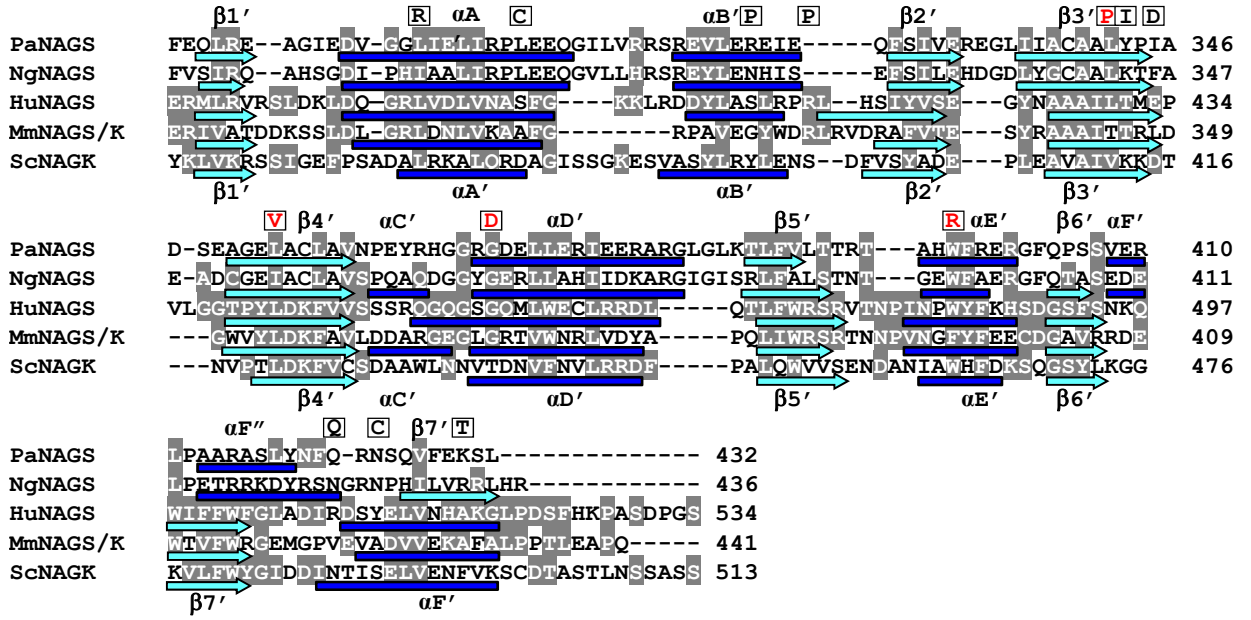


Fig. 2

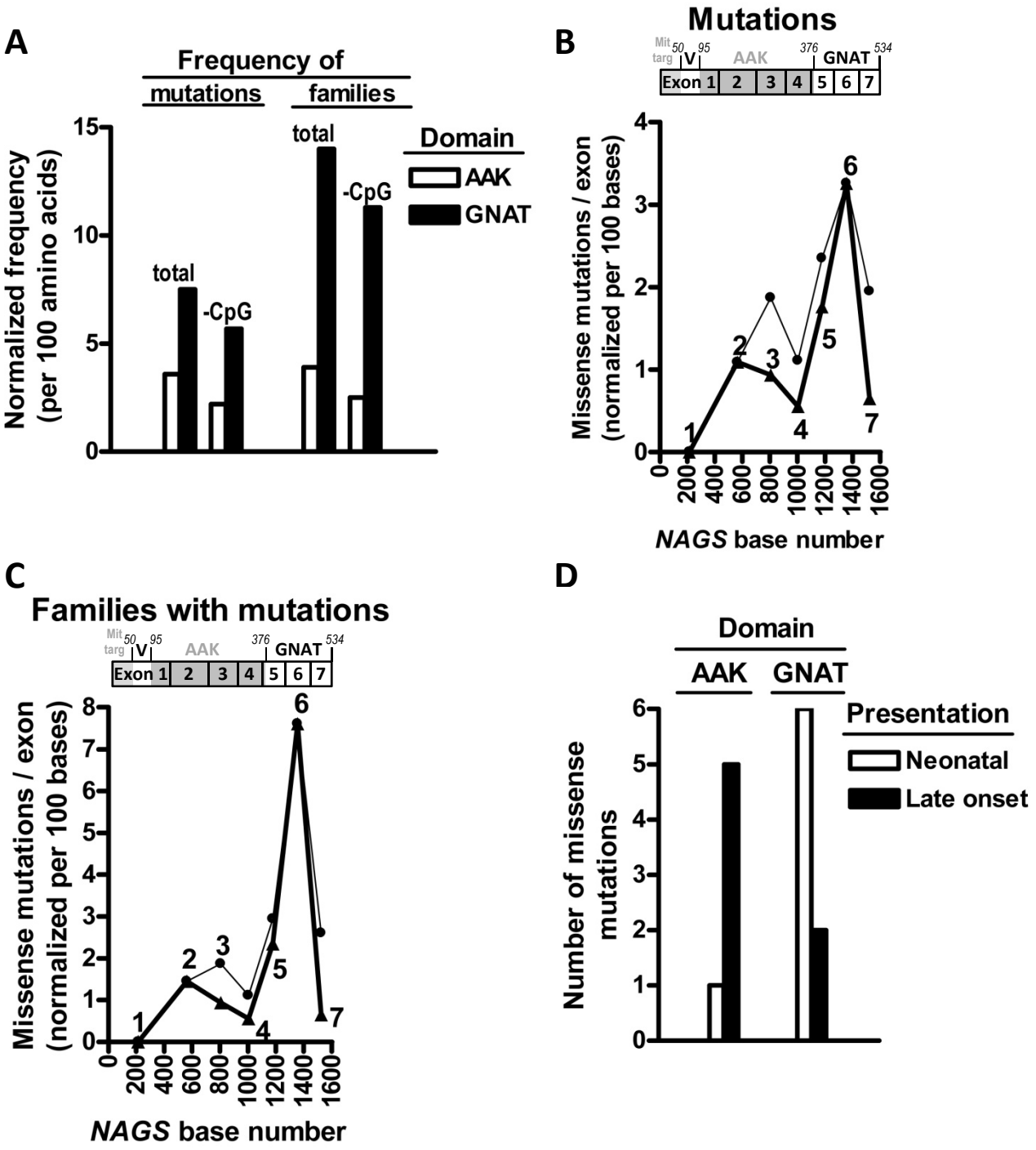
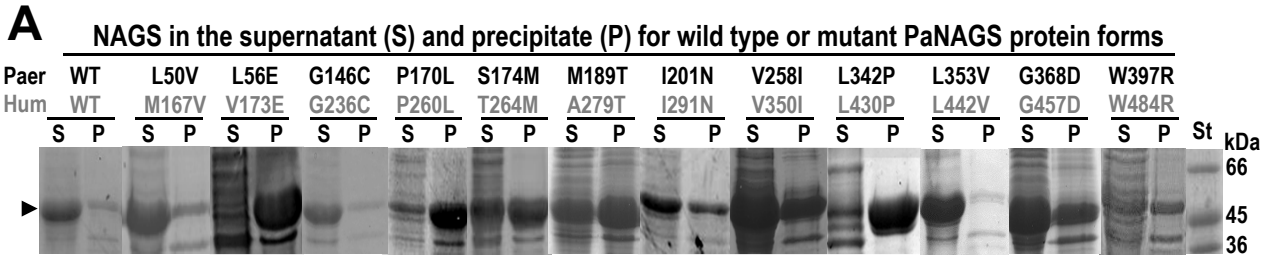
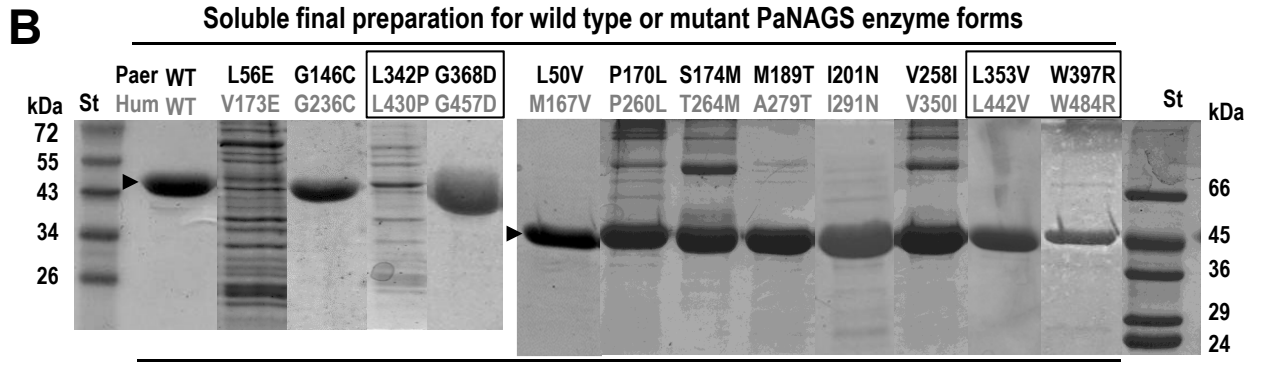

A
Frequency of
mutations
families
Normalized frequency
(per 100 amino acids)
15
10
5
0
total
-CpG
total
-CpG
Domain
AAK
GNAT
B
Mutations
Mit targ
50
V
95
AAK
376
GNAT
534
Exon 1 2 3 4 5 6 7
Missense mutations / exon
(normalized per 100 bases)
4
3
2
1
0
0 200 400 600 800 1000 1200 1400 1600
NAGS base number
C
Families with mutations
Mit targ
50
V
95
AAK
376
GNAT
534
Exon 1 2 3 4 5 6 7
Missense mutations / exon
(normalized per 100 bases)
8
7
6
5
4
3
2
1
0
0 200 400 600 800 1000 1200 1400 1600
NAGS base number
D
Domain
AAK
GNAT
Number of missense
mutations
6
5
4
3
2
1
0
Presentation
Neonatal
Late onset

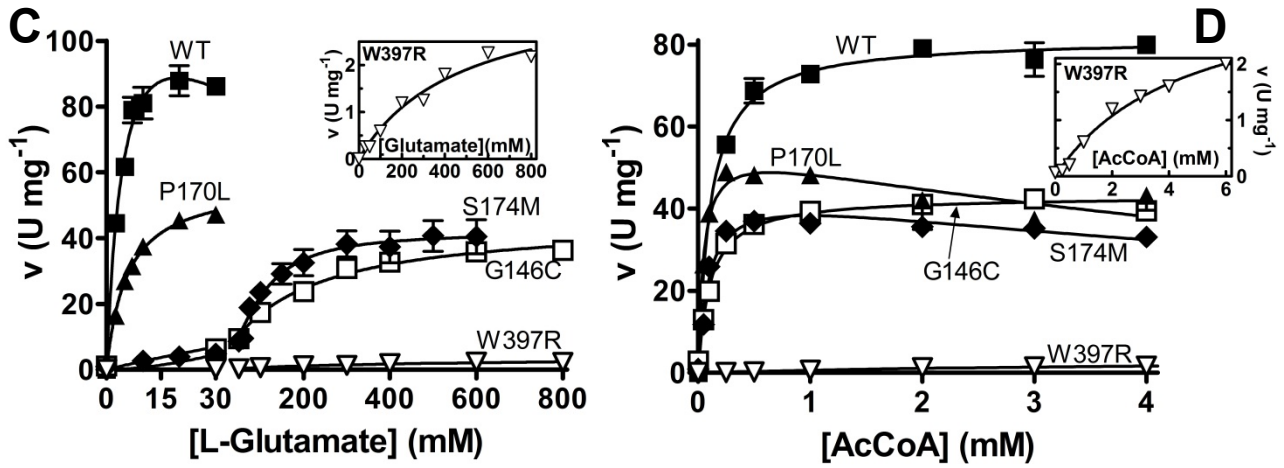
Fig. 3





No Arg	67.9 ±4.9	< 0.1	38.7[a] ±2.0	< 0.1	< 0.1	51.3 ±0.6	49.1 ±2.1	34.6[a] ±2.3	47.5 ±4.9	62.8 ±0.7	54.9 ±2.7	70.8 ±1.4	1.7 ±0.2
+ 1 mM Arg	4.3 ±0.4	nt	39.9 ±1.7	nt	nt	50.6 ±0.7	1.5 ±0.6	24.1 ±0.1	3.7 ±0.3	4.3 ±0.5	6.3 ±1.8	3.2 ±0.1	0.69 ±0.01

NAGS activity (U/mg)



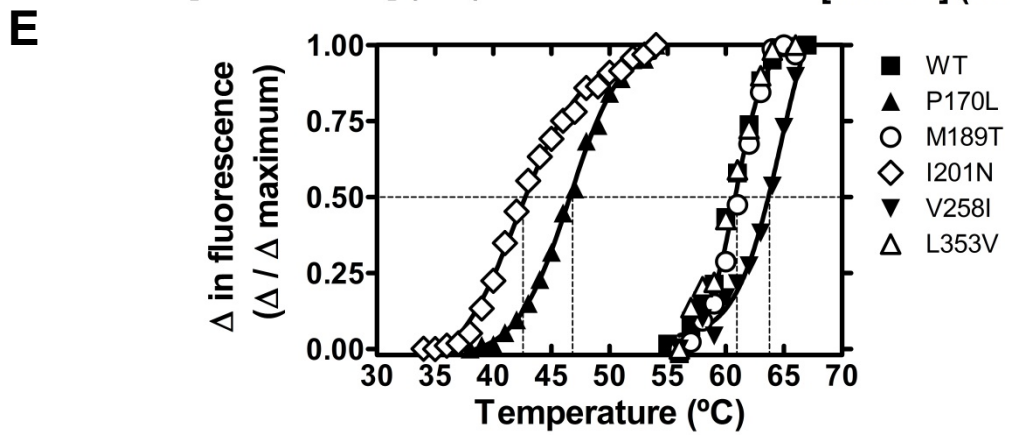


Fig. 4

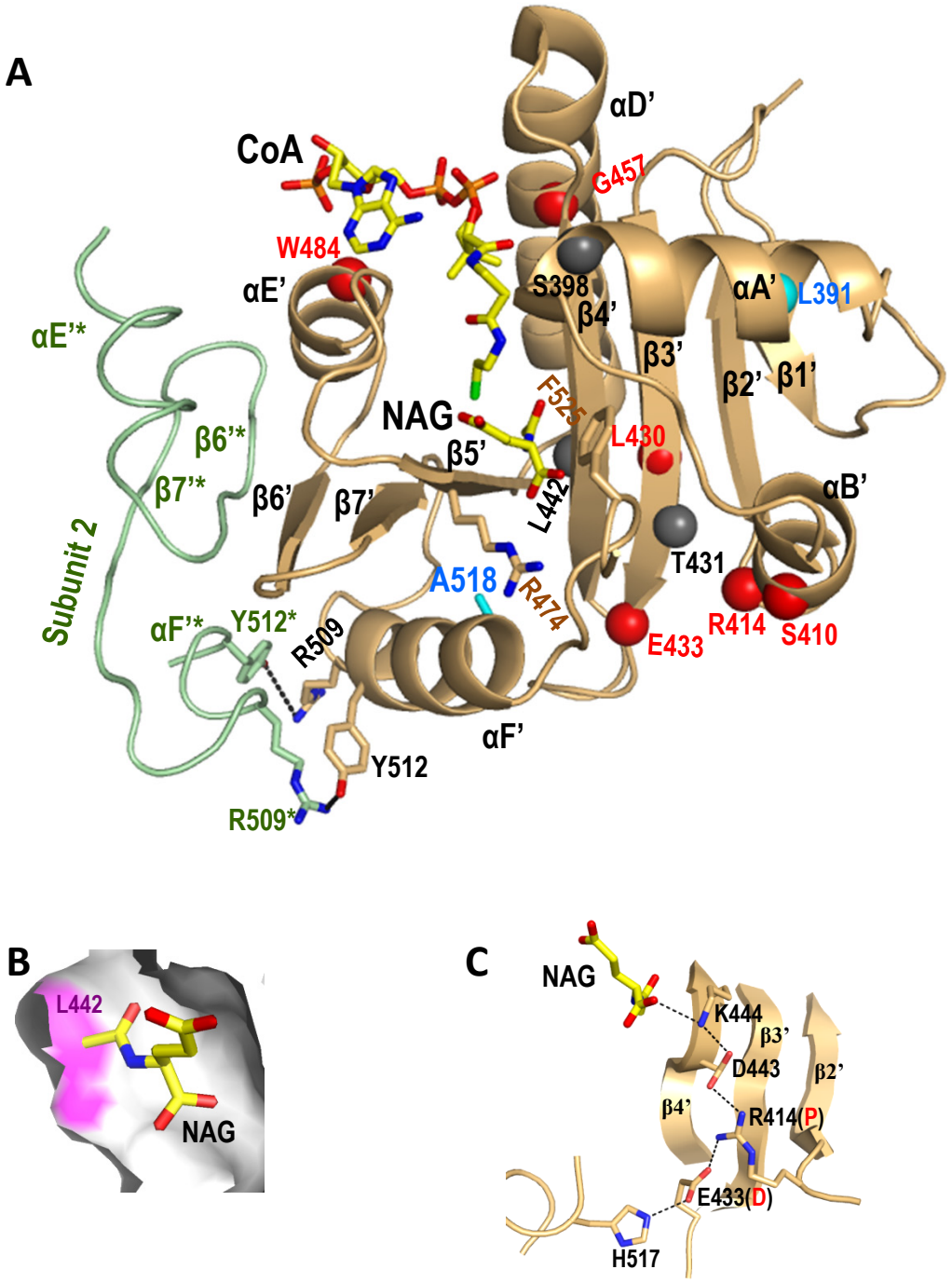
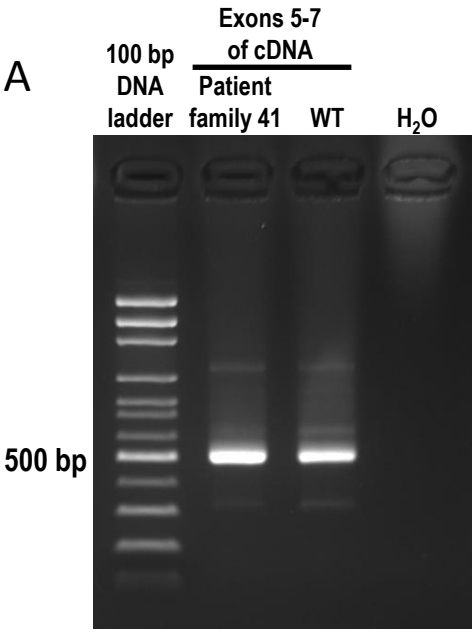

A
αD'
CoA
G457
W484
αE'
S398
αA'
L391
β4'
αE'*
β3'
β2'
β1'
F525
NAG
β6'*
β5'
L430
β7'*
β6'
β7'
αB'
L442
Subunit 2
T431
A518
R474
αF'*
Y512*
R509
E433
R414
S410
αF'
Y512
R509*
B
L442
NAG
C
NAG
K444
β3'
D443
β2'
β4'
R414(P)
E433(D)
H517

Fig. 5



Supplementary Figure S1. Data obtained with fibroblasts-derived cDNA and blood genomic DNA (gDNA) in patients from families 41 and 27. **(A)** Normal size of the product obtained after PCR-amplification of exons 5 to 7 of patient 41 using as template the cDNA from fibroblasts of this patient. Other gel tracks include, as indicated, DNA size standards, the product obtained in identical way from a control individual (WT) and from template-free amplification (H_2O; to prove the absence of contaminating mRNA or DNA). **(B, C)** Sanger sequencing of the amplified exons 5-7 of the cDNA **(B)** and of the corresponding region of genomic DNA **(C)** from patient 41 reveals the coexistence at base c.1494 of the G>A transition (that should cause the protein truncation p.Trp498*) and of the normal allele (indicated below the sequencing trace). The grey shadowing and the asterisk highlight the position of a SNP reported in the ExAC database (details given below the sequences), and the cyan shadowing and the arrow highlight the mutation found in this patient. **(D, E)** Sequencing of the relevant region from the cDNA and gDNA from the patient of family 27 shows the coexistence in this patient, at both mRNA and genomic DNA levels, of the c.1535A>G transition (causing the amino acid substitution p.Tyr512Cys) and the wild type allele (indicated below the sequences and highlighted with an arrow).