Structure of aspartoacylase, the brain enzyme impaired in Canavan disease

Eduard Bitto, Craig A. Bingman, Gary E. Wesenberg, Jason G. McCoy, and George N. Phillips, Jr.*

Center for Eukaryotic Structural Genomics, University of Wisconsin, Madison, WI 53706-1544

Edited by Gregory A. Petsko, Brandeis University, Waltham, MA, and approved November 8, 2006 (received for review September 6, 2006)

Aspartoacylase catalyzes hydrolysis of *N*-acetyl-L-aspartate to aspartate and acetate in the vertebrate brain. Deficiency in this activity leads to spongiform degeneration of the white matter of the brain and is the established cause of Canavan disease, a fatal progressive leukodystrophy affecting young children. We present crystal structures of recombinant human and rat aspartoacylase refined to 2.8- and 1.8-Å resolution, respectively. The structures revealed that the N-terminal domain of aspartoacylase adopts a protein fold similar to that of zinc-dependent hydrolases related to carboxypeptidases A. The catalytic site of aspartoacylase shows close structural similarity to those of carboxypeptidases despite only 10–13% sequence identity between these proteins. About 100 C-terminal residues of aspartoacylase form a globular domain with a two-stranded β-sheet linker that wraps around the N-terminal domain. The long channel leading to the active site is formed by the interface of the N- and C-terminal domains. The C-terminal domain is positioned in a way that prevents productive binding of polypeptides in the active site. The structures revealed that residues 158–164 may undergo a conformational change that results in opening and partial closing of the channel entrance. We hypothesize that the catalytic mechanism of aspartoacylase is closely analogous to that of carboxypeptidases. We identify residues involved in zinc coordination, and propose which residues may be involved in substrate binding and catalysis. The structures also provide a structural framework necessary for understanding the deleterious effects of many missense mutations of human aspartoacylase.

N-acetyl-L-aspartate | x-ray structure | zinc-dependent hydrolase

N-acetyl-L-aspartate (NAA) is one of the most abundant amino acid derivatives found in the vertebrate brain, second only to glutamate (1–3). Proton magnetic resonance spectroscopic imaging of spatial variations in the concentration of NAA within the brain allows noninvasive diagnostics of several central nervous system diseases, including multiple sclerosis and Alzheimer's disease (3, 4). NAA is synthesized from L-aspartate and acetyl-CoA in the neuronal mitochondria of brain gray matter (5). NAA is released from neurons to the cerebrospinal fluid and then transported to oligodendrocytes (6, 7). Although the exact role of NAA in the brain remains a matter of investigations (1, 2, 8–11), it is believed that proper metabolism of this compound in the brain is important for correct development and maintenance of the white matter (12, 13).

The enzyme hydrolyzing NAA, aspartoacylase (ASPA, also known as aminoacylase 2), is found primarily in oligodendrocytes of white matter (14). ASPA has been purified to homogeneity from bovine brain (15). The enzyme has been proposed to act as an esterase by means of a catalytic Ser–His–Glu triad (16). However, recent sequence analyses have suggested that the enzyme belongs to the zinc–carboxypeptidase family (17). It has since been established that ASPA binds zinc, and that this metal is necessary for activity (18).

A deficiency in ASPA is the established cause of Canavan disease (CD, OMIM no. 271900), a fatal progressive neurodegenerative disorder characterized by dysmyelination and spongiform degeneration of white matter in children (19–22). Clinically, CD is characterized by megalocephaly, hypotonia of neck muscles that results in poor head control, muscle rigidity, developmental delay of motor and verbal skills, visual impairment, and feeding difficulties (20–24). Lack of ASPA activity leads to a dramatic increase of NAA concentration in the brain, spinal fluid, blood, and urine (19, 25). Measurement of elevated NAA levels in urine is used to confirm the diagnosis of CD (24). A mouse knockout model deficient in ASPA activity (13) and a rat model with a natural deletion of the aspartoacylase gene (26) have been recently developed and serve as important tools in understanding biochemical pathologies related to CD (10, 27–29).

CD is a recessive autosomal disorder. The carrier frequency reported for the two most common mutations among Ashkenazi Jews is high; estimates range between 1:40 to 1:59 (30–32). More than 50 mutations in ASPA, including numerous deletions, missense mutations, and premature terminations, have been so far described in patients of diverse ethnic origins (16, 32–43).

Here we report the x-ray structures of human ASPA (hASPA, 313 aa, UniProt entry P45381) and rat ASPA (rASPA, 312 aa, UniProt entry Q6AZ03), identify important residues and structural features responsible for the substrate specificity and selectivity, propose a mechanism of action, and provide structural insights into molecular pathologies of selected missense mutations. The structure of ASPA represents a sequence-to-structure target (with $<30\%$ sequence identity to any structure in the Protein Data Bank at the time of selection and deposition) determined under the National Institutes of Health Protein Structure Initiative.

Results

Structure Quality. The structure of rASPA was determined by two wavelength anomalous diffraction and refined to a resolution of 1.8 Å. rASPA has 86% sequence identity and 95% sequence similarity to hASPA. The structure of hASPA was solved by molecular replacement using rASPA as a search model and refined to 2.8 Å. Data-collection, phasing, and refinement statistics are summarized in Table 1. The final models describe residues 4–310 of two rASPA molecules present in the asymmetric unit and residues 9–310 of two hASPA molecules. The rASPA and hASPA molecules superpose with a rmsd of ≈0.63 Å for ≈2,300 corresponding protein atoms. The only notable difference between rASPA and hASPA involves residues 157–163, which undergo a crank-shaft-like movement with $>180°$ rotation of Leu-161, which is displaced as much as 10 Å

Author contributions: E.B., C.A.B., and G.N.P. designed research; E.B. and C.A.B. performed research; E.B., C.A.B., G.E.W., J.G.M., and G.N.P. analyzed data; E.B., C.A.B., G.E.W., J.G.M., and G.N.P. wrote the paper; and G.E.W. deposited coordinates.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: NAA, *N*-acetyl-L-aspartate; ASPA, aspartoacylase; hASPA, human ASPA; rASPA, rat ASPA; CD, Canavan disease.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PBD ID codes 2GU2 and 2I3C).

See Commentary on page 399.

*To whom correspondence should be addressed. E-mail: phillips@biochem.wisc.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0607817104/DC1.

Table 1. Crystal parameters, x-ray data collection, phasing, and refinement statistics

	rASPA Se-Peak	rASPA Se-HRem	hASPA
Space group	C2		$P4_22_12$
Unit-cell parameters, Å, °	a=92.6, b=135.8, c=54.0, β=101.5		a=145.6, c=103.4
Data-collection statistics			
Wavelength, Å	0.97928	0.95373	0.97926
Energy, eV	12,661	13,000	12,661
Resolution range, Å	41.75–1.80 (1.84–1.80)	40.49–1.80 (1.84–1.80)	48.72–2.80 (2.90–2.80)
No. of reflections, measured/unique*	410,617/58,208	430,447/59,112	363,519/27,367
Completeness, %	97.5 (86.6)	97.6 (87.1)	97.9 (81.4)
R_{merge}†	0.093 (0.470)	0.087 (0.500)	0.164 (0.582)
Redundancy	7.1 (4.2)	7.3 (4.4)	13.3 (6.2)
Mean I/σ (I)	10.0 (2.2)	9.9 (2.0)	10.6 (1.9)
Phasing statistics‡			
Phasing power,isomorphous/anomalous	0.0/1.90	1.99/0.76	
Mean FOM, centric/acentric	0.45/0.63		
Rcullis, isomorphous/anomalous	0.0/0.616	0.635/0.793	
Refinement and model statistics			
Resolution range	41.75–1.80 (1.85–1.80)		48.72–2.8 (2.87–2.8)
No. of reflections, work/test	58,208/2,955		25,931/1,374
R_{cryst}§	0.149 (0.208)		0.195 (0.354)
R_{free}¶	0.194 (0.268)		0.243 (0.419)
rmsd bonds, Å	0.020		0.016
rmsd angles, °	1.616		1.551
ESU from R_{free}, Å	0.114		0.305
Average B factor, Å^2	21.8		47.7
No. of protein molecules/all atoms	2/5,509		2/4,938
No. of waters/ions	558/2 Zn^{2+}, 5 sulfates		36/2 Zn^{2+}, 8 phosphates
Ramachandran plot by PROCHECK, %			
Most favorable region	91.0		86.9
Additional allowed region	8.6		12.5
Generously allowed region	0.4		0.6
PBD ID code	2gu2		2i3c

*Values in parentheses are for the highest resolution shell.
†$R_{merge} = \Sigma_h\Sigma_i|I_i(h) - \langle I(h)\rangle|/\Sigma_h\Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h)\rangle$ is the mean intensity of the reflection.
‡Phasing by SHARP in 40.5– 2.4 Å resolution range.
§$R_{cryst} = \Sigma_h||F_{obs}| - |F_{calc}||/\Sigma_h|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively.
¶R_{free} was calculated as R_{cryst} using ≈5.0% of the randomly selected unique reflections that were omitted from structure refinement.

[supporting information (SI) Fig. 5]. It should be noted that the structure of rASPA presented here contains missense mutations in two surface residues (E128G and R131G). We did not find any significant differences in the affected region between folds of rASPA and hASPA; therefore, we believe that the mutations are benign. Throughout the paper, we will use residue numbering that corresponds to hASPA.

Protein Fold of ASPA. The structure of ASPA presents two distinct domains: the N-terminal domain (N-domain) formed by residues 1–212 and the C-terminal domain (C-domain) formed by residues 213–313 (Fig. 1*A*). Several N- and C-terminal residues bring the two domains of ASPA together in a two-stranded β-sheet anchor formed by β-strands β1 and β13. Based on the CATH classification (44), the N-domain belongs to the $\alpha\beta$-class of proteins in a three-layer $\alpha\beta\alpha$-sandwich architecture with aminopeptidase topology. The N-domain is formed by the central six-stranded mixed β-sheet surrounded by eight helices of variable size and multiple connecting loops. The C-domain belongs to the "mainly β" class of proteins. Antiparallel β-strands β8 and β12 form a linker that wraps around the N-domain and leads to the globular portion of the C-domain (Fig. 1*A*).

Conserved Residues of ASPA. The Conserved Domain Database (48) classifies ASPA as a member of the succinylglutamate desuccinylase/aspartoacylase family (AstE-AspA, Pfam04952, $E = 8 \times 10^{-52}$). There are four identical residues among the 33 seed members of the family: Asn-70, Asp-114, His-116, and Glu-178. Additional highly conserved residues include His-21, Gly-22, Glu-24, Asn-54, Arg-63, Arg-71, and Phe-73. Most of these conserved residues cluster within the area that likely represents the active site of ASPA (Fig. 1*A*, blue sticks).

Oligomeric State of ASPA. The two rASPA molecules present in the asymmetric unit of the rASPA crystal form an apparent dimer. Similarly, hASPA forms a closely analogous dimer in crystals in the different space group (Table 1). Dimers of hASPA and rASPA superpose with a rmsd of 0.83 Å for 4,328 aligned atoms. The dimer interface buries ≈1,200 Å^2 of solvent accessible area and involves 12 hydrogen bonds and two salt bridges (45). Gel filtration of the rASPA sample suggested the presence of a species with a molecular weight of 84 kDa (data not shown). The rASPA dimer has an expected molecular weight of 71 kDa. The observed difference of 13 kDa may be related to a larger apparent hydrodynamic radius (and thus lower mobility) of the ellipsoid-shaped dimer of rASPA (Fig. 1*B*). Altogether, our data and other observations (18, 46) suggest that recombinant ASPA forms a dimer *in vitro*. However, a recent report has indicated that the active form of native rASPA is, in fact, a monomer (47).

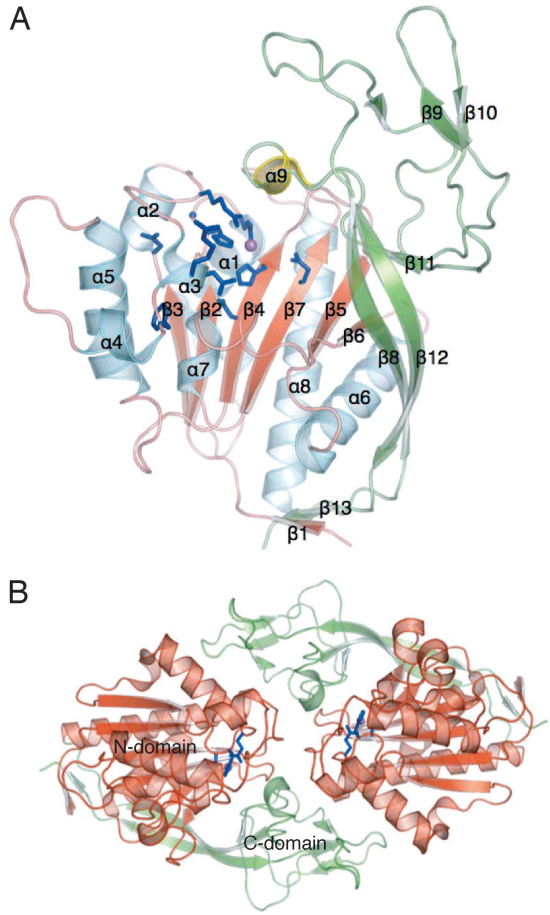

Fig. 1. Ribbon diagrams of the rASPA monomer and dimer. (*A*) N-domain of rASPA is color-coded in cyan and red. C-domain is color-coded in yellow and green. Residues His-21, Gly-22, Glu-24, Asn-54, Arg-63, Asn-70, Arg-71, Phe-73, Asp-114, His-116, and Glu-178 (blue sticks) are highly conserved in the AstE-AspA family and delineate the active site. Zn^{2+} is shown as a pink sphere. (*B*) The rASPA dimer observed in the asymmetric unit of the rASPA crystals is shown in ribbon representation. Both the N-domain (red) and C-domain (green) of the rASPA monomers are involved in formation of the dimer interface. Residues His-21, Glu-24, and His-116 (blue sticks) coordinate Zn^{2+} (pink sphere).

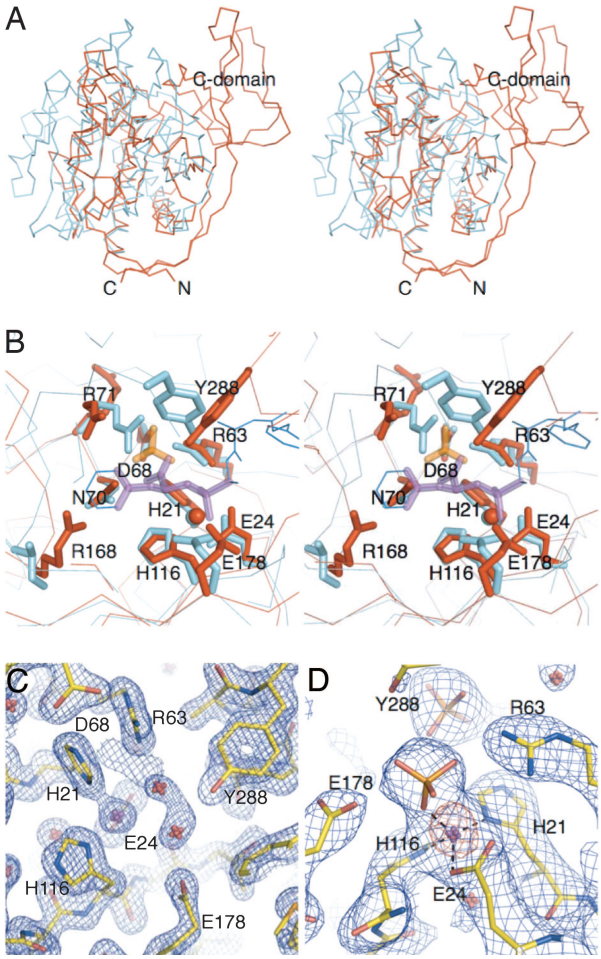

Fig. 2. Details of the ASPA active site. (*A*) A stereo representation of structural superposition of rASPA (red) and bovine carboxypeptidase A (cyan; PDB code 1m4l). (*B*) A stereo representation of structural superposition of the active sites of bovine pancreatic carboxypeptidase A (cyan, PDB code 6cpa) and rASPA (red). Red and cyan lines represent the C_α-trace of the enzymes. Important residues involved in Zn^{2+} (sphere) coordination, substrate binding, and catalysis are shown in sticks and are annotated for hASPA (corresponding carboxypeptidase A residues are listed in Table 2). A transition state inhibitor (blue lines) is bound in the active site of carboxypeptidase A. Selected atoms of this inhibitor that topologically correspond to those of putative NAA transition state are highlighted (magenta) for clarity. Sulfate bound in the active site of rASPA is show in orange sticks. Arg-71 of rASPA adopts two alternate conformations. (*C*) The $2F_o - F_c$ electron density map (blue) of rASPA is shown at contour level of 1.8σ. rASPA residues are shown in sticks, the metal ion (magenta) coordinated in the active site and waters (red) are depicted by three-dimensional crosses. (*D*) The $2F_o - F_c$ electron density map (blue) of hASPA is shown at contour level of 1.5σ. The anomalous difference map (red) is shown at contour level of 6σ. Black dashed lines represent coordination bonds from protein residues and phosphate to Zn^{2+} (magenta) bound in the active site of hASPA.

Structural Homology Search. The structural homology search by the VAST server (49) established similarity between the N-domain of ASPA and a range of zinc-dependent hydrolases related to carboxypeptidase A. A superposition of rASPA and bovine carboxypeptidase A (50) is presented in Fig. 2*A*. The proteins show the same connectivity between multiple aligned secondary structure elements. Two major topological differences between the proteins are also obvious: (*i*) Carboxypeptidases do not possess a domain corresponding to the C-domain of ASPA. (*ii*) The N terminus of mature carboxypeptidase A is longer by ≈60 residues compared with ASPA and extends the central β-sheet by two β-strands. In ASPA, the N- and C-terminal residues form a parallel two-stranded β-sheet that stabilizes the overall fold of this protein.

Active Site of ASPA. A detailed comparison of the active sites of several carboxypeptidases with the structurally corresponding region of rASPA revealed a remarkable similarity between these proteins. Fig. 2*B* presents the structural overlay of rASPA and bovine pancreatic carboxypetidase A in complex with a phosphonate transition-state inhibitor (51). Highlighted in magenta are selected atoms of the phosphonate inhibitor that could topologically correspond to those of the putative transition state of the ASPA substrate, NAA. We believe that this structural overlay provides a very useful framework for understanding of NAA binding in ASPA.

Zinc Coordination. Inductively coupled plasma metal analysis of the rASPA and hASPA samples used for crystallization revealed the presence of 0.15 mol of zinc and 0.53 mol of nickel per mol of protein, and 0.05 mol of zinc and 0.11 mol of nickel per mol of protein, respectively. Crystals of hASPA were soaked during cryoprotection in solutions that contained 0.5 mM $ZnSO_4$. The structures of both rASPA and hASPA revealed that the conserved residues His-21, Glu-24, and His-116 coordinate a metal ion (Fig. 2*C*). Analysis of anomalous signal in the hASPA diffraction data

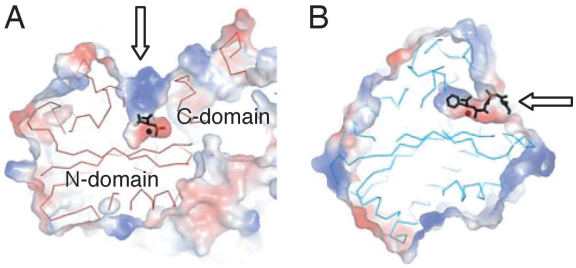

Fig. 3. Comparison of the substrate binding cavities. Lines represent C_α-traces of the rASPA (red) and carboxypeptidase A (cyan). Both proteins were superposed and are shown in the same orientation. Solvent accessible electrostatic surfaces of both proteins are clipped to reveal internal pockets responsible for binding of substrates (black sticks, see Fig. 2*B* for details). Arrows indicate direction in which substrates enter binding pockets of both proteins. Zinc ions are shown as spheres. (*A*) The substrate binding cavity is formed at the interface of the N- and C-domains of rASPA. (*B*) The substrate binding cavity recognizes the terminal hydrophobic and a portion of the penultimate residue of the substrate in bovine pancreatic carboxypeptidase A (PDB ID code 6cpa).

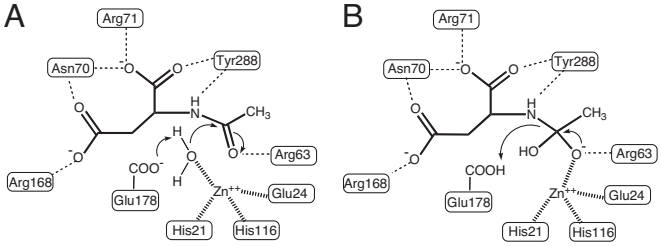

Fig. 4. Proposed mechanism of action of ASPA.

confirmed the presence of an anomalously scattering atom in the active site (Fig. 2*D*). Diffraction experiments performed both 100 eV below and above the K-edge of zinc established the identity of this atom as zinc (data not shown). No additional zinc sites were observed in the active site. In addition to protein residues, a phosphate was coordinated to Zn^{2+} in the hASPA structure (Fig. 2*D*).

Access to the Active Site. In carboxypeptidase A, the active site is accessible to large substrates and a deep cavity in the protein accommodates the bulky C-terminal residue of polypeptides. In ASPA, the C-domain sterically hinders access to the active site from the same direction. Instead, the N-domain and C-domain of ASPA form a deep narrow channel that leads to the active site. The access channel is formed in a direction ≈90° away from the one seen in carboxypetidase A (Fig. 3). The entry of the channel shows a distinct positive electrostatic potential (due to the presence of Arg-71, Lys-228, Lys-291, and Lys-292) that may help to guide a negatively charged substrate to the active site.

Discussion

Substrate Binding. Active site residues of ASPA that could be involved in the substrate binding include Arg-63, Asn-70, Arg-71, Arg-168, and Tyr-288. The putative roles of these and other selected residues of ASPA, which we base on the close structural alignment of the rASPA active site to that of carboxypetidase A, are summarized in Table 2. It is of interest that many of these residues were found to be mutated in patients with CD, thus further stressing their importance in hASPA and corroborating our hypothesis about the close analogy of substrate binding in carboxypeptidases and ASPA. Arg-168 is well positioned to provide specificity for a negatively charged β-carboxyl of NAA. The corresponding residue in bovine carboxypeptidase A, which specifically hydrolyzes polypeptide chains with large hydrophobic C-terminal residues, is Ile-255. Carboxypeptidases B, which are specific for positively charged C-terminal residues, have a negatively charged aspartate (Asp-255) at this position. Tyr-288

Table 2. Proposed role of selected residues of hASPA catalytic site

Residue in hASPA	Conservation in AstE-AspA family	Corresponding residue in carboxypeptidase A	Proposed role*	Familial mutation in humans
His-21	High	His-69	Coordinates zinc	His21Pro
Glu-24	High	Glu-72	Coordinates zinc	Glu24Gly
Arg-63	High	Arg-127	Stabilizes substrate binding. Interacts with the terminal carboxyl of NAA	
Asp-68	Minimal	Asp-142	Participates in formation of the hydrogen bond network of the active site. Forms hydrogen bonds to His-21, Arg-63, and backbone of Asn-70	Asp68Ala
Asn-70	Absolute	Asn-144	Stabilizes substrate binding. Forms a hydrogen bond with the terminal carboxyl of NAA	
Arg-71	High	Arg-145	Stabilizes substrate binding. May guide NAA to the active site, then interact with the terminal carboxyl of NAA	Arg71His
Asp-114	Absolute	Ser-194	Participates in formation of the hydrogen bond network of the active site. Forms hydrogen bonds to His-116 and Thr-20	Asp114Glu
His-116	Absolute	His-196	Coordinates zinc	
Arg-168	Minimal	Ile-255	Stabilizes substrate binding. Provides specificity for the β-carboxyl of NAA	Arg168Glu, Arg168His, Arg168Cys
Glu-178	Absolute	Glu-270	The general acid/base residue. Activates catalytic water coordinated to zinc, protonates leaving group during collapse of the tetrahedral intermediate	
Tyr-288	NA	Tyr-248 (functional)	Stabilizes substrate binding. Forms a hydrogen bond with the terminal carboxyl of NAA. May act as a general acid during collapse of the tetrahedral intermediate	Tyr288Cys

*Proposed roles are based on the structural alignment of the active sites of rASPA and carboxypeptidase A. See text for details.

Table 3. Categorization of missense mutations leading to CD

No.	Type	Mutations	Possible structural explanation of defect for a selected residue
1	Zinc coordination/substrate binding	H21P, E24G, R71H, R168E, R168H, R168C, Y288C	E24G: loss of residue coordinating zinc
2	Alteration of hydrogen bonding network	A57T, D68A, D114E, D114Y, **E285A***, A287T	E285A: loss of the active site hydrogen network that may lead to destabilization of interaction between the N- and C-domain
3	Hydrophobic core	I16T, G27R, I143T, C152R, C152Y, C152W, M195R, Y231C, P280L, P280S, F295S, **A305E**	A305E: local unfolding due to placement of bulky, charged residue into the hydrophobic core. Mutation likely results in an unstable protein that undergoes degradation *in vivo.*
4	Early/late termination	Y109X†, Q184X, E214X, C218X, **Y231X,** X314W	Y231X: loss of the C-domain. Protein is likely degraded *in vivo.*
5	Apparent dimer formation‡	P183H, V186F, D249V	D249V: destabilization of dimer interface due to loss of salt-bridge to Arg-187 of the other monomer
6	Defect of unclear origin	G123E, P181T, K213E, H244R, G274R	H244R: surface residue affected, problem not clear

*Mutations shown in bold represent the most common mutations.
†X represents a termination codon.
‡Dimerization of rASPA is not required for activity (47).

is well positioned to functionally mimic a conserved Tyr-248 found in carboxypeptidases. The phenyl oxygen of Tyr-288 of rASPA resides within hydrogen bonding distance of the putative terminal carboxyl of NAA and the amide nitrogen of NAA (Fig. 2*B*). Interestingly, this residue is located in the C-domain, which is not present in carboxypeptidases, and enters the active site from the opposite direction to Tyr-248 in carboxypeptidases.

Mechanism of Action of ASPA. The structures of rASPA and hASPA establish that ASPA shows structural similarities to a range of carboxypeptidases from different species despite very low sequence identity (10–15%). The structures thus provide structural confirmation that ASPA is a member of the zinc hydrolase superfamily (52). In addition, the structures revealed a remarkable similarity of the active site to that of carboxypeptidase A. This allows for the formulation of hypotheses about what residues are involved in substrate binding and/or catalysis. We hypothesize that this enzyme may follow a "promoted-water pathway" mechanism widely accepted for carboxypeptidase A (53, 54). If this hypothesis is correct, we expect that Glu-178 acts as a general base that activates a water molecule coordinated to the catalytic zinc ion. The activated hydroxyl attacks the carbonyl group of NAA polarized by Arg-63 and possibly by Zn(II) (Fig. 4*A*). The negative charge on the oxygen of the resulting tetrahedral intermediate is stabilized by coordination to Zn(II) and electrostatic interaction with Arg-63 (Fig. 4*B*). In the last step of the catalysis, the tetrahedral intermediate collapses, the carbonyl group is reformed and aspartate is eliminated as the leaving group. Glu-178 (or possibly Tyr-288) acts as a general acid during this step and provides a proton to the leaving group (Fig. 4*B*). We propose that phosphate the coordinated to Zn^{2+} in the active site of hASPA resembles the initial attack complex, where the coordinated oxygen of phosphate mimics the activated hydroxyl.

Role of the C-Domain. The structure of ASPA provides important insight into the role of its C-domain. The central roles of this domain seem to be selecting against larger polypeptides and providing specificity for the acetyl group of NAA. A narrow channel that leads to the active site is formed at the interface of the N- and C-domains of ASPA. Due to the geometry of the active site, a polypeptide that enters this channel cannot bind productively to undergo hydrolysis. This is because the C-domain sterically hinders access to the active site in a direction from which the enzyme would have to bind the penultimate residue of the polypeptide (Fig. 3). ASPA can therefore reside and function in the cytosol of glial cells without interfering with protein degradation pathways. In the case of carboxypeptidases, a similar logistic issue is solved by synthesis of inactive zymogens. Interestingly, procarboxypeptidases are inhibited by a mechanism similar to the one just described for ASPA: the globular portion of the N-terminal prodomain sterically hinders the access of polypeptides to the active site. The prodomain is connected to the rest of the protein by a long linker helix. (The procarboxypeptidases are activated in the duodenum by trypsin, which cleaves off the prodomain at the end of the linker helix.) A structural overlay of the human procarboxypeptidase A2 (55) and rASPA revealed that the globular portion of the prodomain reaches the area above the active site from a direction opposite to that of the C-domain of rASPA.

Determinants of Substrate Specificity. The C-domain forms a pocket that accommodates the acetyl group of NAA and puts severe restriction on the size of this portion of substrate (Fig. 3*A*). The tight pocket is formed by residues Thr-118, Gln-184 (His-183 in rASPA), Phe-282, Glu-285, Ala-287, and Tyr-288. Compounds with acyl groups longer than acetate (i.e., short or long fatty acids) are unlikely to bind productively in the active site of ASPA. In fact, it has been reported that substrate derivatives with larger groups including *N*-*t*-butylacyl, *t*-butoxyacyl, and guanidine groups are very poor substrates of hASPA (46). hASPA also shows a high specificity for the aspartate portion of NAA (46). Arg-168, which may be responsible for this specificity, may form a hydrogen bond to and provide electrostatic stabilization of the β-carboxyl group of NAA. The loss of one or both of these interactions can explain a loss of activity toward *N*-acetylasparagine, -valine, -leucine, and -alanine (46). A lack of activity against *N*-acetylglutamate may be related to a shift of the whole compound in the active site leading to a binding nonconducive to efficient catalysis. Interestingly, three familial mutations have been reported that target Arg-168 (Table 2).

Structural Insights into the Pathologies of hASPA Mutants. Approximately 60% of known missense mutations of hASPA are located within the N-domain. The most common mutations, A305E, Y231X (X represents a termination codon), and E285A, are all located in the C-domain. Table 3 categorizes currently known missense mutations of hASPA into six categories and provides hypotheses about the molecular defect of selected mutations. Mapping of these mutations onto the structure of ASPA is presented in SI Fig. 6.

Materials and Methods

Protein Expression and Purification. The genes encoding rASPA and hASPA were cloned, and selenomethionine-labeled proteins were expressed and purified following the standard Center for Eukary-

otic Structural Genomics pipeline protocols for cloning (56), protein expression (57), protein purification (58), and overall bioinformatics management (59). Protein purification of heterologously expressed, nonglycosylated proteins from *Escherichia coli* did not involve refolding steps and resulted in 18.6 and 2.5 mg of recombinant rASPA and hASPA proteins, respectively.

Structure Determination. Crystals of rASPA were grown at 293 K by the hanging drop method from a 10 mg·ml^{-1} protein solution mixed with an equal amount of well solution containing 1.9 M ammonium sulfate, 100 mM HEPPS, pH 8.5. Crystals of hASPA were grown using well solution containing 56 mM NaH_2PO_4 and 1344 mM K_2HPO_4 and were cryoprotected in well solutions supplemented with 0.5 mM $ZnSO_4$ and up to 25% (v/v) ethylene glycol. X-ray diffraction data for rASPA were collected at the General Medicine and Cancer Institutes Collaborative Access Team 23-ID-D beamline at the Advanced Photon Source at Argonne National Laboratory. Data for hASPA were collected at Southeast Regional Collaborative Access Team 22-ID beamline at the same institution. The diffraction images were integrated and scaled using HKL2000 (60). The selenium substructure of selenomethionine-labeled rASPA was determined by using HySS (61). The structure was phased using autoSHARP (62) with the help of auxiliary programs from CCP4 suite (63). Initial phase information was improved by density modification of phases from SHARP with 2-fold noncrystallographic averaging, based on positions of anomalous scatterers as implemented in RESOLVE (64). The starting model build by ARP/wARP (65) was completed in cycles of iterative manual building in COOT (66) and refinement in REFMAC5 (67). The structure of hASPA was solved by molecular replacement in MOLREP (68) using rASPA as a search model and refined using a similar protocol to that described for rASPA. The figures were prepared by using PyMOL (69).

We thank the members of the Center for Eukaryotic Structural Genomics team. This work was supported by National Institutes of Health/National Institute for General Medical Sciences Grants P50 GM64598 and U54 GM074901 (J. L. Markley, PI). Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under contract No.W-31–109-ENG-38. General Medicine and Cancer Institutes Collaborative Access Team has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Science (Y1-GM-1104). Data were also collected at Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions may be found at www.ser-cat.org/members.html.

1. Baslow MH (2003) *Neurochem Res* 28:941–953.
2. Clark JF, Doepke A, Filosa JA, Wardle RL, Lu A, Meeker TJ, Pyne-Geithman GJ (2006) *Med Hypotheses* 67:506–512.
3. Tsai G, Coyle JT (1995) *Prog Neurobiol* 46:531–540.
4. Gujar SK, Maheshwari S, Bjorkman-Burtscher I, Sundgren PC (2005) *J Neuroophthalmol* 25:217–226.
5. Madhavarao CN, Chinopoulos C, Chandrasekaran K, Namboodiri MA (2003) *J Neurochem* 86:824–835.
6. Huang W, Wang H, Kekuda R, Fei YJ, Friedrich A, Wang J, Conway SJ, Cameron RS, Leibach FH, Ganapathy V (2000) *J Pharmacol Exp Ther* 295:392–403.
7. Sager TN, Thomsen C, Valsborg JS, Laursen H, Hansen AJ (1999) *J Neurochem* 73:807–811.
8. Baslow MH (1999) *J Inherit Metab Dis* 22:99–101.
9. Baslow MH (2002) *Neurochem Int* 40:295–300.
10. Madhavarao CN, Arun P, Moffett JR, Szucs S, Surendran S, Matalon R, Garbern J, Hristova D, Johnson A, Jiang W, Namboodiri MA (2005) *Proc Natl Acad Sci USA* 102:5221–5226.
11. Namboodiri AM, Peethambaran A, Mathew R, Sambhu PA, Hershfield J, Moffett JR, Madhavarao CN (2006) *Mol Cell Endocrinol* 252:216–223.
12. Chakraborty G, Mekala P, Yahya D, Wu G, Ledeen RW (2001) *J Neurochem* 78:736–745.
13. Matalon R, Rady PL, Platt KA, Skinner HB, Quast MJ, Campbell GA, Matalon K, Ceci JD, Tyring SK, Nehls M, *et al.* (2000) *J Gene Med* 2:165–175.
14. Madhavarao CN, Moffett JR, Moore RA, Viola RE, Namboodiri MA, Jacobowitz DM (2004) *J Comp Neurol* 472:318–329.
15. Kaul R, Casanova J, Johnson AB, Tang P, Matalon R (1991) *J Neurochem* 56:129–135.
16. Kaul R, Gao GP, Balamurugan K, Matalon R (1993) *Nat Genet* 5:118–123.
17. Makarova KS, Grishin NV (1999) *J Mol Biol* 292:11–17.
18. Le Coq J, An HJ, Lebrilla C, Viola RE (2006) *Biochemistry* 45:5878–5884.
19. Matalon R, Michals K, Sebesta D, Deanching M, Gashkoff P, Casanova J (1988) *Am J Med Genet* 29:463–471.
20. Matalon R, Michals K, Kaul R (1995) *J Pediatr* 127:511–517.
21. Matalon R, Michals-Matalon K (1999) *Neurochem Res* 24:507–513.
22. Surendran S, Michals-Matalon K, Quast MJ, Tyring SK, Wei J, Ezell EL, Matalon R (2003) *Mol Genet Metab* 80:74–80.
23. Baslow MH, Resnik TR (1997) *J Mol Neurosci* 9:109–125.
24. Matalon R (1997) *Genet Test* 1:21–25.
25. Jakobs C, ten Brink HJ, Langelaar SA, Zee T, Stellaard F, Macek M, Srsnova K, Srsen S, Kleijer WJ (1991) *J Inherit Metab Dis* 14:653–660.
26. Kitada K, Akimitsu T, Shigematsu Y, Kondo A, Maihara T, Yokoi N, Kuramoto T, Sasa M, Serikawa T (2000) *J Neurochem* 74:2512–2519.
27. Surendran S, Campbell GA, Tyring SK, Matalon R (2005) *Neurobiol Dis* 18:385–389.
28. Surendran S, Ezell EL, Quast MJ, Wei J, Tyring SK, Michals-Matalon K, Matalon R (2004) *Brain Res* 1016:268–271.
29. Surendran S, Szucs S, Tyring SK, Matalon R (2005) *Reprod Toxicol* 20:281–283.
30. Elpeleg ON, Anikster Y, Barash V, Branski D, Shaag A (1994) *Am J Hum Genet* 55:287–288.
31. Feigenbaum A, Moore R, Clarke J, Hewson S, Chitayat D, Ray PN, Stockley TL (2004) *Am J Med Genet A* 124:142–147.
32. Kaul R, Gao GP, Aloya M, Balamurugan K, Petrosky A, Michals K, Matalon R (1994) *Am J Hum Genet* 55:34–41.
33. Elpeleg ON, Shaag A (1999) *J Inherit Metab Dis* 22:531–534.
34. Kaul R, Gao GP, Matalon R, Aloya M, Su Q, Jin M, Johnson AB, Schutgens RB, Clarke JT (1996) *Am J Hum Genet* 59:95–102.
35. Kaul R, Gao GP, Michals K, Whelan DT, Levin S, Matalon R (1995) *Hum Mutat* 5:269–271.
36. Kobayashi K, Tsujino S, Ezoe T, Hamaguchi H, Nihei K, Sakuragawa N (1998) *Hum Mutat* Suppl 1:S308–S309.
37. Olsen TR, Tranebjaerg L, Kvittingen EA, Hagenfeldt L, Moller C, Nilssen O (2002) *J Med Genet* 39:e55.
38. Rady PL, Vargas T, Tyring SK, Matalon R, Langenbeck U (1999) *Am J Med Genet* 87:273–275.
39. Shaag A, Anikster Y, Christensen E, Glustein JZ, Fois A, Michelakakis H, Nigro F, Pronicka E, Ribes A, Zabot MT, *et al.* (1995) *Am J Hum Genet* 57:572–580.
40. Sistermans EA, de Coo RF, van Beerendonk HM, Poll-The BT, Kleijer WJ, van Oost BA (2000) *Eur J Hum Genet* 8:557–560.
41. Zeng BJ, Wang ZH, Ribeiro LA, Leone P, De Gasperi R, Kim SJ, Raghavan S, Ong E, Pastores GM, Kolodny EH (2002) *J Inherit Metab Dis* 25:557–570.
42. Janson CG, Kolodny EH, Zeng BJ, Raghavan S, Pastores G, Torres P, Assadi M, McPhee S, Goldfarb O, Saslow B, *et al.* (2006) *Ann Neurol* 59:428–431.
43. Tacke U, Olbrich H, Sass JO, Fekete A, Horvath J, Ziyeh S, Kleijer WJ, Rolland MO, Fisher S, Payne S, *et al.* (2005) *Neuropediatrics* 36:252–255.
44. Pearl FM, Bennett CF, Bray JE, Harrison AP, Martin N, Shepherd A, Sillitoe I, Thornton J, Orengo CA (2003) *Nucleic Acids Res* 31:452–455.
45. Jones S, Thornton JM (1996) *Proc Natl Acad Sci USA* 93:13–20.
46. Moore RA, Le Coq J, Faehnle CR, Viola RE (2003) *Arch Biochem Biophys* 413:1–8.
47. Hershfield JR, Madhavarao CN, Moffett JR, Benjamins JA, Garbern JY, Namboodiri A (2006) *FASEB J* 20:2139–2141.
48. Marchler-Bauer A, Bryant SH (2004) *Nucleic Acids Res* 32:W327–W331.
49. Madej T, Gibrat JF, Bryant SH (1995) *Proteins* 23:356–369.
50. Kilshtain-Vardi A, Glick M, Greenblatt HM, Goldblum A, Shoham G (2003) *Acta Crystallogr D* 59:323–333.
51. Kim H, Lipscomb WN (1990) *Biochemistry* 29:5546–5555.
52. Wouters MA, Husain A (2001) *J Mol Biol* 314:1191–1207.
53. Christianson DW, Lipscomb WN (1989) *Acc Chem Res* 22:62–69.
54. Lipscomb WN (1980) *Proc Natl Acad Sci USA* 77:3875–3878.
55. Garcia-Saez I, Reverter D, Vendrell J, Aviles FX, Coll M (1997) *EMBO J* 16:6906–6913.
56. Thao S, Zhao Q, Kimball T, Steffen E, Blommel PG, Riters M, Newman CS, Fox BG, Wrobel RL (2004) *J Struct Funct Genomics* 5:267–276.
57. Sreenath HK, Bingman CA, Buchan BW, Seder KD, Burns BT, Geetha HV, Jeon WB, Vojtik FC, Aceti DJ, Frederick RO, *et al.* (2005) *Protein Expr Purif* 40:56–67.
58. Jeon WB, Aceti DJ, Bingman CA, Vojtik FC, Olson AC, Ellefson JM, McCombs JE, Sreenath HK, Blommel PG, Seder KD, *et al.* (2005) *J Struct Funct Genomics* 6:143–147.
59. Zolnai Z, Lee PT, Li J, Chapman MR, Newman CS, Phillips GN, Jr, Rayment I, Ulrich EL, Volkman BF, Markley JL (2003) *J Struct Funct Genomics* 4:11–23.
60. Otwinowski Z, Minor W (1997) *Methods Enzymol* 276:307–326.
61. Grosse-Kunstleve RW, Adams PD (2003) *Acta Crystallogr D* 59:1966–1973.
62. de la Fortelle E, Bricogne G (1997) *Methods Enzymol* 276:472–494.
63. Collaborative Computational Project No 4 (1994) *Acta Crystallogr D* 50:760–763.
64. Terwilliger TC (2000) *Acta Crystallogr D* 56:965–972.
65. Perrakis A, Morris R, Lamzin VS (1999) *Nat Struct Biol* 6:458–463.
66. Emsley P, Cowtan K (2004) *Acta Crystallogr D* 60:2126–2132.
67. Murshudov GN, Vagin AA, Dodson EJ (1997) *Acta Crystallogr D* 53:240–255.
68. Vagin A, Teplyakov A (1997) *J Appl Crystallogr* 30:1022–1025.
69. DeLano WL (2002) The PyMOL Molecular Graphics System (DeLano Scientific, San Carlos, CA).