

Computational analysis of deleterious missense mutations in aspartoacylase that cause Canavan's disease

SREEVISHNUPRIYA K.¹, CHANDRASEKARAN P.¹, SENTHILKUMAR A.¹,
SETHUMADHAVAN R.¹, SHANTHI V.¹, DAISY P.², NISHA J.², RAMANATHAN K.^{1*}
& RAJASEKARAN R.^{1*}

¹Bioinformatics Division, School of Biosciences and Technology, Vellore Institute of Technology, Vellore 632014, Tamil Nadu, India;

²Department of Biotechnology, Holy Cross College, Trichy 620002, Tamil Nadu, India

Received August 8, 2012; accepted November 6, 2012

In this work, the most detrimental missense mutations of aspartoacylase that cause Canavan's disease were identified computationally and the substrate binding efficiencies of those missense mutations were analyzed. Out of 30 missense mutations, I-Mutant 2.0, SIFT and PolyPhen programs identified 22 variants that were less stable, deleterious and damaging respectively. Subsequently, modeling of these 22 variants was performed to understand the change in their conformations with respect to the native aspartoacylase by computing their root mean squared deviation (RMSD). Furthermore, the native protein and the 22 mutants were docked with the substrate NAA (N-Acetyl-Aspartic acid) to explain the substrate binding efficiencies of those detrimental missense mutations. Among the 22 mutants, the docking studies identified that 15 mutants caused lower binding affinity for NAA than the native protein. Finally, normal mode analysis determined that the loss of binding affinity of these 15 mutants was caused by altered flexibility in the amino acids that bind to NAA compared with the native protein. Thus, the present study showed that the majority of the substrate-binding amino acids in those 15 mutants displayed loss of flexibility, which could be the theoretical explanation of decreased binding affinity between the mutant aspartoacylases and NAA.

missense mutation, Canavan's disease, aspartoacylase, NAA, flexibility

Citation: Sreevishnupriya K, Chandrasekaran P, Senthilkumar A, *et al.* Computational analysis of deleterious missense mutations in aspartoacylase that cause Canavan's disease. *Sci China Life Sci*, 2012, 55: 1109–1119, doi: 10.1007/s11427-012-4406-8

Canavan's disease (CD) (OMIM #271900) is an inherited, fatal, autosomal recessive form of leukodystrophy [1]. More than 50 mutations in the aspartoacylase (ASPA) gene, including numerous deletions, missense mutations, and premature terminations, have been described in patients of diverse ethnic origins. The mutations cause ASPA deficiency and accumulation of N-acetyl aspartic acid (NAA) in the brain [2,3]. ASPA is primarily found in oligodendrocytes of the white matter and was originally proposed to be a member of the esterase family by the presence of a catalytic Ser-His-Glu triad [4,5]. However, subsequent alignment

studies showed few similarities between ASPA and the esterase family [6,7]. The enzyme deficiency in CD interferes with the normal hydrolysis of NAA, which results in disruption of myelin and spongy degeneration of the white matter of the brain [8–11]. The clinical features of the disease are macrocephaly, head lag, progressive-severe mental retardation, and hypotonia in early life, which later changes to spasticity. Symptoms of CD appear in early infancy and typically progress very rapidly [12,13]. Structural analysis of ASPA revealed that the N-terminal domain adopted a protein fold similar to that of zinc-dependent hydrolases related to carboxypeptidase A. The catalytic site of ASPA

*Corresponding author (email: rrajasekaran@vit.ac.in; kramanathan@vit.ac.in)

showed close structural similarity to those of carboxypeptidases despite only 10%–13% sequence identity between these proteins. About 100 residues in the C-terminal region of ASPA form a globular domain with a two-stranded β -sheet linker that wraps around the N-terminal domain. It was proposed that the deacetylation of NAA follows a carboxypeptidase-type mechanism for the hydrolysis of the amide bond of the substrate, based on the highly ordered binding of a potent inhibitor. However, there has been no report on the structure of any Michaelis complexes involving an ASPA and its substrate. 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 such that it prevents productive binding of polypeptides in the active site [14]. Aspartoacylase had been a member of the carboxypeptidase-A family. 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. This explains most loss-of-function aspartoacylase mutations associated with Canavan's disease [15]. Canavan's disease may occur in any ethnic group, it had been more frequent among Ashkenazi Jews from eastern Poland, Lithuania, and western Russia, and among Saudi Arabians [16]. Identifying the disease-associated missense mutation had been a challenging task for genetic disorder research. Therefore, we attempted to investigate the mutants of ASPA using a computational protocol that we devised for the analysis of BRCA1, CDKN2A and SMAD4 [17–19]. The computational protocol was used to identify the detrimental missense mutations in ASPA protein and we proposed a model structure for the mutants. The substrate, NAA, was then docked with both the native protein and ASPA mutants to determine the binding effect and the nature of the flexibility in the binding pockets, which explained the decreased binding efficiency of these missense mutations.

1 Materials and methods

1.1 Datasets

The protein sequence and variants (single amino acid polymorphisms/missense mutations/point mutations) of ASPA were obtained from the Swissprot database available at <http://www.expasy.ch/sprot/>. The subsection of each Swissprot entry provides information on polymorphic variants, some of which polymorphic variants may be disease(s)-associated by causing defects in a given protein; most of them were nsSNPs (non-synonymous SNPs) in the gene sequence and SAPs (single amino acid polymorphisms) in the protein sequence [20–22]. The 3D Cartesian coordinates of ASPA and its complex were obtained from Protein Data Bank with PDB IDs 2I3C and 2O4H [23] for *in silico* mutation modeling and docking studies based on detrimental

point mutants.

1.2 Predicting stability changes caused by SAPs using support vector machine (I-Mutant 2.0)

We used the program I-Mutant2.0 available at <http://gpcr.biocomp.unibo.it/cgi/predictors/IMutant2.0/I-Mutant2.0.cgi>. I-Mutant2.0 is a support vector machine (SMV) based tool for the automatic prediction of protein stability changes caused by single point mutations. I-Mutant2.0 predictions were performed starting either from the protein structure or, more importantly, from the protein sequence [24]. This program was trained and tested on a dataset derived from ProTherm [25], which is the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability caused by mutations under different conditions. The output files show the predicted free energy change value or sign ($\Delta\Delta G$), which was calculated from the unfolding Gibbs free energy value of the mutated protein minus the unfolding Gibbs free energy value of the native protein (kcal mol^{-1}). Positive $\Delta\Delta G$ values meant that the mutated protein has higher stability and negative values indicate lower stability.

1.3 Analysis of functional consequences of point mutations by a sequence homology-based method (SIFT)

We used the program SIFT (available at <http://blocks.fhrc.org/sift/SIFT.html>) [26], specifically to detect deleterious single amino acid polymorphisms. SIFT is a sequence homology-based tool, which presumes that important amino acids will be conserved in a protein family; therefore, changes at well-conserved positions tend to be predicted as deleterious [27]. Queries are submitted in the form of protein sequences. SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that, for given a protein sequence, (i) searches for similar sequences, (ii) chooses closely related sequences that may share similar function, (iii) obtains the multiple alignment of these chosen sequences, and (iv) calculates normalized probabilities for all possible substitutions at each position from the alignment. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious and those greater than or equal to the cutoff are predicted to be tolerated [26]. The cutoff value in SIFT program was tolerance index of ≥ 0.05 . The higher the tolerance index, the less functional impact a particular amino acid substitution would be likely to have.

1.4 Simulation for functional change in a point mutant by structure homology-based method (PolyPhen)

Analyzing the damage caused by point mutations at the

structural level is considered very important to understand the functional activity of the protein. We used the server PolyPhen [28] which is available at <http://coot.embl.de/PolyPhen/> for this purpose. Input options for the PolyPhen server are protein sequence, SWALL database ID or accession number, together with the sequence position of two amino acid variants. The query is submitted in the form of a protein sequence with a mutational position and two amino acid variants. Sequence-based characterization of the substitution site, profile analysis of homologous sequences, and mapping of the substitution site to known protein 3D structures are the parameters taken into account by PolyPhen server to calculate the score. It calculates position-specific independent counts (PSIC) scores for each of the two variants and then computes the PSIC scores difference between them. The higher the PSIC score difference, the higher the functional impact a particular amino acid substitution would be likely to have.

1.5 Modeling SAAP locations on protein structure to compute the RMSD

Structure analysis was performed to evaluate the structural deviation between native proteins and mutant proteins by means of root mean square deviation (RMSD). We used the web resource Protein Data Bank [23] and the single amino acid polymorphism database (SAAPdb) [29] to identify the 3D structure of ASPA (PDB ID: 2I3C). We also confirmed the mutation position and the mutation residue in PDB ID 2I3C. The mutation was performed *in silico* using the SWISSPDB viewer, and NOMAD-Ref server performed the energy minimization for 3D structures [30]. This server uses Gromacs as the default force field for energy minimization, based on the methods of steepest descent, conjugate gradient, and limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) methods [31]. We used the conjugate gradient method to minimize the energy of the 3D structure of ASPA. To optimize the 3D structure of ASPA, we used the ifold server [32] for simulated annealing, which is based on discrete molecular dynamics and is one of the fastest strategies for simulating protein dynamics. This server efficiently samples the vast conformational space of biomolecules in both length and time scales. Divergence of the mutant structure from the native structure could be caused by substitutions, deletions and insertions [33] and the deviation between the two structures could alter the functional activity [34] with respect to binding efficiency of the inhibitors, which was evaluated by their RMSD values.

1.6 Computation of total energy and stabilizing residues

Total energy is one of the parameter that can indicate the stability between native and mutant modeled structures, and could be computed by the GROMOS96 force field that is

embedded in the SWISSPDB viewer. Note that molecular mechanics or force field methods use classical type models to predict the energy of the molecule as a function of its conformation. This allows prediction of equilibrium geometries, transition states and relative energies between conformers or between different molecules. Molecular mechanics expresses the total energy as a sum of Taylor series expansions for the stretches for every pair of bonded atoms, and adds additional potential energy terms contributed by bending, torsional energy, van der Waals energy, and electrostatics [35]. Thus the total energy calculation could be considered as reliable parameter for understanding the stability of protein molecules with the aid of Force field (Gromos96 and Gromacs). Performing energy minimization and simulated annealing removes steric clashes and to obtain the best stable conformation [36]. Finally, the total energy was computed for native and mutant ASPAs by the GROMOS force field. Moreover, the total energy of the native structure was considered as a reference point for comparing the total energy of mutant structures for stability analysis. In addition, identifying the stabilizing residues for both the native and mutant structures represented a significant parameter for understanding their stability. Hence, we used the server SRide [37] to identify the stabilizing residues in the native and mutant protein models. Stabilizing residues were computed using parameters such as surrounding hydrophobicity, long-range order, stabilization center, and conservation score [37].

1.7 Identification of binding sites and computation of atomic contact energy (ACE) between ASPA and its substrate

To compute the ACE between ASPA and its substrate, we submitted the PDB ID: 2O4H, a complex of ASPA with N-phosphonomethyl-L-aspartate, into the ligand contact tool (LCT) program available at <http://firedb.bioinfo.cnio.es/Php/Contact.php> [39]. This server calculates contacts between the binding amino acid residues (active site) of ASPA with N-phosphonomethyl-L-aspartate with default parameters. N-phosphonomethyl-L-aspartate is a stable tetrahedral intermediate analog of NAA. Hence, the SMILES string was collected for the NAA molecule from PubChem, a database maintained at NCBI [38] and submitted it to CORINA (www.molecular-networks.com/online_demos/corina_demo.html) to construct the 3D structure of the substrate (NAA).

We unbound the N-phosphonomethyl-L-aspartate from the ASPA of the PDB ID: 2O4H to perform point mutations on ASPA using the SWISSPDB viewer to perform energy minimization by NOMAD-Ref and simulated annealing by ifold. Finally, we used the program PatchDock for docking the native and mutant ASPA with NAA to compute the ACE by using additional option of binding residue parameter. The underlying principle of this server is based on mo-

lecular shape representation, surface patch matching plus filtering and scoring [40]. It finds docking transformations that yield good molecular shape complementarity. Such transformations, when applied, induce both wide interface areas and small amounts of steric clashes. A wide interface ensured that include several matched local features of the docked molecules that have complementary characteristics were included. The PatchDock algorithm divides the Connolly dot surface representation [41] of the molecules into concave, convex and flat patches. Then, complementary patches are matched to generate candidate transformations. Each candidate transformation is further evaluated by a scoring function that considers both geometric fit and atomic desolvation energy [42,43]. Finally, an RMSD clustering was applied to the candidate solutions to discard redundant solutions. The main reason behind Patch Dock's high efficiency is its fast transformational search, which is driven by local feature matching rather than by brute force searching of the six-dimensional transformation spaces. It further speeds up the computational processing time using advanced data structures and spatial pattern detection techniques, such as geometric hashing and pose clustering.

1.8 Exploring the flexibility of binding pocket by normal mode analysis

A quantitative measure of the atomic motions in proteins could be obtained from the mean square fluctuations of the atoms relative to their average positions. These could be related to the B-factor [44,45]. Analysis of B-factors, therefore, could provide fresh insights into protein dynamics, the flexibility of amino acids, and protein stability [46]. Protein flexibility is important for protein function and for rational drug design [47]. In addition, the flexibility of certain amino acids in a protein is useful for various types of interactions. Moreover, the flexibility of amino acids in the binding pocket is considered a significant parameter for understanding the binding efficiency. In fact, loss of flexibility impairs the binding effect [48] and vice versa [18]. Hence, this can be analyzed by the B-factor, which is computed from the mean-square displacement $\langle R^2 \rangle$ of the lowest-frequency normal mode using the ElNémo server [49].

2 Results and discussion

2.1 The SAP data set from Swissprot

The ASPA protein and 30 variants, namely, I16T, H21P, E24G, G27R, A57T, D68A, D114E, D114Y, G123E, I143T, C152R, C152W, C152Y, R168C, R168H, P181T, P183H, V186F, M195R, Y231C, H244R, D249V, G274R, P280L, P280S, E285A, A287T, F295S, A305E and C310G investigated in this work were retrieved from the Swissprot database [20–22].

2.2 Identification of functional variants by I-mutant 2.0

Of the 30 variants, 28 variants were found to be less stable using the I-Mutant 2.0 server (Table 1) [24]. Among these 28 variants, four variants showed a $\Delta\Delta G$ value <-2.0 . Fifteen variants showed a $\Delta\Delta G$ value <-1.0 and nine variants showed a $\Delta\Delta G$ value >-1.0 as depicted in Table 1.

Of the 28 variants that showed a negative $\Delta\Delta G$, four variants (E24G, D68A, D249V and E285A) changed their negatively charged amino acid to non-polar amino acid, four variants (I16T, A57T, I143T and A287T) changed from non-polar to polar and two variants (P280L and C310G) changed from polar to non-polar. Two variants (G27R and G274R) changed from non-polar to positively charged, two variants (G123E and A305E) changed from non-polar to negatively charged and two variants (C152R and P183H) changed from polar to positively charged. Two variants (H21P and R168C) changed from positive to polar, two variants, C152W and C152Y changed from polar to aromatic and two variants (P181T and P280S) retained its polar property. Two variants H244R and R168H retained its positively charged property followed by variants viz., V186F, F295S and D114Y which changed amino acids from non-polar to aromatic, aromatic to polar and aromatic to negative charged amino acid respectively. One more variant D114E retained its negatively charged amino acid. Indeed, by considering only amino acid substitution based on physico-chemical properties, we could not be able to identify the detrimental effect. Rather, by considering the sequence conservation along with the above said properties could have more advantages and reliable to find out the detrimental effect of missense mutations [48].

2.3 Deleterious single point mutants identified by the SIFT program

The degree of conservation of a particular position in a protein was determined using sequence homology based tool SIFT [26]. The protein sequences of the 30 variants were submitted to SIFT to determine their tolerance indices. As the tolerance level increases, the functional influence of the amino acid substitution decreases and vice versa.

Among the 30 variants, 27 variants were found to be deleterious, having tolerance index scores of ≤ 0.05 (Table 1). Among these 27 variants, 15 variants showed a very high deleterious tolerance index score of 0.00. Eight variants had a tolerance index score of 0.01, one variant G274R had a tolerance index score of 0.02, two variants had tolerance index scores of 0.03, and one had a tolerance index score of 0.05 (Table 1). Interestingly, 25 deleterious variants identified by SIFT also were seen to be less stable by the I-Mutant 2.0 server.

Table 1 List of functionally significant mutants predicted to be by I-Mutant 2.0, SIFT and PolyPhen^{a)}

AA change	$\Delta\Delta G$	Tolerance index	PSIC SD	Population prevalence references
I16T	-1.8	0.01	2.062	[12,49]
H21P	-0.89	0	3.869	[51]
E24G	-1.29	0	2.931	[3]
G27R	-1.04	0	2.409	[12,49]
A57T	-0.65	0	2.147	[51]
D68A	-1.58	0	3.014	[3]
D114E	-1.1	0.06	2.12	[12]
D114Y	-0.97	0.01	3.128	[52]
G123E	-0.62	0	2.79	[12]
I143T	-3.26	0.01	1.788	[8]
C152R	-1.04	0.01	2.599	[6]
C152W	-1.1	0	2.842	[12,49]
C152Y	-0.44	0.01	2.575	[12]
R168C	-1.46	0.01	0.101	[12]
R168H	-1.42	0.05	1.869	[51]
P181T	-1.96	0	2.589	[51]
P183H	-1.94	0.06	1.104	[12]
V186F	-1.52	0.01	0.818	[49]
M195R	0.00	0.03	0.999	[49]
Y231C	0.53	0	2.61	[53,55]
H244R	-0.26	0	3.286	[3]
D249V	-0.07	0	2.363	[3,52]
G274R	-2.23	0.02	2.278	[50]
P280L	-3.64	0	3.321	[12]
P280S	-3.64	0	2.798	[12]
E285A	-1.07	0	2.43	[12,53-55]
A287T	-1.33	0.01	1.000	[49]
F295S	-1.94	0	2.412	[5,49]
A305E	-0.2	0.03	1.246	[5,12,49,56]
C310G	-0.42	0.54	2.808	[12]

a) Letters in bold indicate mutants predicted to be less stable, deleterious and damaging by I-Mutant 2.0, SIFT and PolyPhen respectively.

2.4 Damaging single point mutations identified by the PolyPhen server

Structural level alterations were determined by PolyPhen program. Protein sequence with mutational position and amino acid variants associated with the 30 single point mutants were submitted to the PolyPhen server [28]. A PSIC score difference of 1.1 and above was considered to be damaging. It could be seen from Table 1 that, out of 30 variants, 25 were considered to be damaging by PolyPhen. These variants also exhibited a PSIC score difference from 1.246 to 3.869. It was to be noted that all the variants that were considered to be damaging by PolyPhen except for D114E and C310G were also identified as deleterious according to the SIFT server. Variant Y231C, which was found to be deleterious by SIFT and damaging by PolyPhen, was found to be stable by I-Mutant2.0. More specifically, this particular mutant position (231) was considered deleterious for causing Canavan's disease [55]. In Y231C, we have analyzed the mutation from Tyrosine to Cysteine at position 231; however, the Y231X mutant was considered to be detrimental by experimental studies [7], where X de-

notes termination (truncation or stop codon) [57]. As this mutant was not identified as deleterious by all three programs (SIFT, Polyphen and I-Mutant2.0), it was not considered for subsequent structural analysis.

2.5 Rational consideration of detrimental point mutations

We rationally considered the 22 most potential detrimental point mutations (I16T, H21P, E24G, G27R, A57T, D68A, D114Y, G123E, I143T, C152R, C152W, C152Y, R168H, P181T, H244R, D249V, G274R, P280L, P280S, E285A, F295S and A305E) for further course of investigations because they were commonly found to be less stable, deleterious, and damaging by the I-Mutant2.0, SIFT and PolyPhen servers respectively [24,26,29]. We considered the statistical accuracy of these three programs, I-Mutant improves the quality of the prediction of the free energy change caused by single point protein mutations by adopting a hypothesis of thermodynamic reversibility of the existing experimental data. The accuracy of prediction for

sequence and structure based values were 78% and 84% with correlation coefficient of 0.56 and 0.69, respectively [59]. SIFT correctly predicted 69% of the substitutions associated with the disease that affect protein function. PolyPhen-2 evaluates rare alleles at loci potentially involved in complex phenotypes, densely mapped regions identified by genome-wide association studies, and analyses natural selection from sequence data, where even mildly deleterious alleles must be treated as damaging. PolyPhen-2 was reported to achieve a rate of true positive predictions of 92% [58–60]. To obtain precise and accurate measures of the detrimental effect of our variants, comprehensive parameters of all these three programs could be more significant than individual tool parameters. Hence, we further investigated these detrimental missense mutations by structural analysis.

2.6 Computing the RMSD by modeling of mutant structures

The available structure of ASPA is PDB ID 2I3C. The mutational position and amino acid variants were mapped onto 2I3C native structure. Mutations at a specified position were performed *in silico* by SWISSPDB viewer independently to obtain a modeled structure. NOMAD-Ref server [30] and ifold server [32] performed the energy minimizations and stimulated annealing respectively, for both native structure and the 22 mutant modeled structures.

To determine the deviation between the native structure and the mutants, we superimposed the native structures with all 22 mutant modeled structures and calculated the RMSD. The higher the RMSD value, the more deviation there is between the native and mutant structure, which in turn changes the binding efficiency with the substrate because of deviation in the 3D space of the binding residues of ASPA. Table 2 shows the RMSD values for native structure with each mutant modeled structure. Table 2 shows that, two mutants, P280L and D249V exhibited a high RMSD >2.00 Å, 16 mutants exhibited an RMSD >1.00 Å and four mu-

tants (C152Y, P181T, E285S and F295S) exhibited a low RMSD value <1.00 Å. Figure 1A shows the superimposed structure of the native protein with mutant P280L, which has an RMSD of 2.13 Å as an illustrative example.

2.7 Application of GROMOS 96 and SRIDE for native structure and mutant modeled structures

The total energy was calculated for both native and mutant structures. Table 2 shows that total energy of native structure was $-18195.142 \text{ kcal mol}^{-1}$, whereas the 22 mutant structures all had slightly higher total energies compared with the native structure. Note that the higher the total energy, the lesser the stability and vice versa. We then used the SRide server [37] to identify the stabilizing residues of both the native structure and the mutant modeled structures (Table 2). The native structure has 20 stabilizing residues whereas on the other hand, the mutant structures have between 14 and 19 stabilizing residues. This clearly indicates that all 22 mutant structures were less stable than the native structure. We further evaluated the effect of these detrimental missense mutations by performing binding analysis between ASPA and NAA using docking studies.

2.8 Binding efficiency of native and mutant ASPA with its substrate

To determine the binding efficiency of ASPA with its substrate we selected the PDB ID 2O4H model structure. The LCT program was used to calculate contacts between the binding residues of ASPA and N-phosphonomethyl-L-aspartate. Thirteen amino acids (Arg (63), Asp (68), Asn (70), Arg (71), Lys (103), Asp (104), Asn (117), Ile (127), Tyr (164), Arg (168), Glu (178), Lys (228) and Tyr (288)) act as binding residues in aspartoacylase with N-phosphonomethyl-L-aspartate (Table 3). To determine the binding efficiency of NAA with both the native and mutants of ASPA, we unbound N-phosphonomethyl-L-aspartate from ASPA in PDB ID 2O4H. We again performed *in silico*

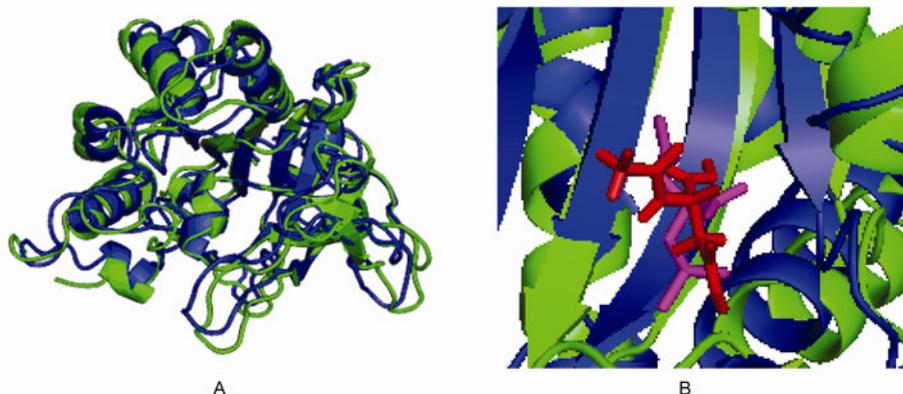


Figure 1 A, Superimposed structure of the native protein (green) with mutant P280L (blue). B, Superimposed structure of the binding region of the native complex (native aspartoacylase (green) and NAA (pink)) with the mutant I143T complex (mutant I143T (blue) and NAA (red)).

Table 2 RMSD, total energy, stabilizing residues and ACE for the native protein and mutants^{a)}

Variants	RMSD (Å)	Total energy (kcal mol ⁻¹)	No. of SR	Stabilizing residues	ACE (kcal mol ⁻¹)
Native	0	-18595.142	20	Val14, Ala15, Ile16, Gly18, Gly19, Gly22, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Pro280, Val281, Ala296, Thr298	-117.39
I16T	1.61	-17567.387	18	Val14, Ala15, Ile16, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Phe262, Pro280, Val281, Phe295, Ala296, Thr298	-52.24
H21P	1.71	-16806.029	18	Val14, Ala15, Ile16, Gly22, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Phe262, Pro280, Val281, Ala296, Thr298	-142.78
E24G	1.46	-16841.162	17	Val14, Ala15, Ile16, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Phe262, Pro280, Val281, Ala296, Thr298	-38.61
G27R	1.7	-17043.043	19	Val14, Ala15, Ile16, Gly22, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-76.28
A57T	1.49	-16774.873	19	Val14, Ala15, Ile16, Gly22, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-141.41
D68A	1.84	-16824.373	19	Val14, Ala15, Ile16, Gly22, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-52.47
D114Y	1.59	-16710.324	15	Val14, Ala15, Asp110, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Phe262, Pro280, Val281, Ala296, Thr298	-36.2
G123E	1.92	-16968.52	18	Val14, Ala15, Ile16, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-48.2
I143T	1.67	-16930.9	19	Val14, Ala15, Ile16, Gly22, Glu80, Asp110, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-47.32
C152R	1.55	-17020.063	17	Val14, Ala15, Ile16, Gly22, Glu89, Asp110, Phe113, Gly193, Thr125, Leu126, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-130.68
C152W	1.74	-16922.334	17	Val14, Ala15, Ile16, Glu89, Asp110, Phe113, Ile127, Leu156, Leu156, Gly176, Val222, Tyr223, Phe262, Pro280, Val281, Ala296, Thr298	-19.06
C152Y	0.24	-16922.334	17	Val14, Ala15, Ile16, Glu89, Asp110, Phe113, Ile127, Leu156, Leu156, Gly176, Val222, Tyr223, Phe262, Pro280, Val281, Ala296, Thr298	-68.28
R168H	1.67	-16333.039	18	Val14, Ala15, Ile16, Gly22, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Gly212, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-120.25
P181T	0.24	-12294.522	18	Val14, Ala15, Ile16, Gly22, Thr125, Leu126, Ile127, Leu156, Gly176, Gly212, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-62.08
H244R	1.53	-16938.777	17	Val14, Ala15, Ile16, Asp110, Phe113, Thr125, Leu156, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-26.92
D249V	2.03	-16709.76	19	Val14, Ala15, Ile16, Gly22, Glu89, Asp110, Phe113, Thr125, Leu126, Gly176, Gly212, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-19.89
G274R	1.55	-17226.846	16	Val14, Ala15, Asp110, Thr125, Leu126, Ile127, Leu156, Gly176, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-83.8
P280L	2.13	-16871.318	17	Val14, Ala15, Ile16, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly212, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-141.78
P280S	1.48	-16926.385	19	Val14, Ala15, Ile16, Gly22, Glu89, Asp110, Phe113, Ile127, Leu156, Gly176, Gly212, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-128.88
E285A	0.25	-16713.943	19	Val14, Ala15, Ile16, Gly22, Glu89, Asp110, Phe113, Thr125, Leu126, Ile127, Gly212, Val222, Tyr223, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-44.58
F295S	0.23	-12187.107	19	Val14, Ala15, Ile16, Gly22, Glu89, Asp110, Phe113, Thr125, Leu126, Ile127, Leu156, Gly176, Gly212, Asp251, Phe262, Pro280, Val281, Ala296, Thr298	-27.53
A305E	1.55	-16949.135	14	Val14, Ala15, Ile16, Asp110, Thr125, Leu126, Ile127, Val222, Tyr223, Phe262, Pro280, Val281, Ala296, Thr298	-24.5

a) RMSD, root mean square deviation; SR, stabilizing residues; the common stabilizing residues are shown in bold; ACE, atomic contact energy.

Table 3 Comparison of normalized mean square displacement of substrate binding amino acids of the native protein and mutants^{a)}

Binding residues (active site)	Normalized mean square displacement (R^2)															
	Native	I16T	E24G	G27R	D68A	D114Y	G123E	I143T	C152W	P181T	H244R	D249V	G274R	E285A	F295S	A305E
Arg63	0.0158	*0.0119	*0.012	*0.0111	*0.0114	*0.0124	*0.0104	*0.0112	*0.0116	*0.0129	*0.0119	*0.0119	*0.0109	*0.0124	*0.0131	*0.0124
Asp68	0.0149	*0.0118	*0.012	*0.0111	*0.0112	*0.012	*0.0109	*0.0111	*0.0115	*0.0122	*0.0117	*0.0116	*0.0107	*0.0123	0.0163	*0.0121
Asn70	0.0114	*0.009	*0.0092	*0.0082	*0.0088	*0.0094	*0.0083	*0.0084	*0.0092	*0.0092	*0.009	*0.0088	*0.0084	*0.0094	*0.0093	*0.0092
Arg71	0.0135	*0.0098	*0.0098	*0.0092	*0.0095	*0.0102	*0.009	*0.0092	*0.0098	*0.0104	*0.0098	*0.0097	*0.0093	*0.0102	*0.0105	*0.0092
Lys103	0.0143	*0.0102	*0.011	*0.0088	*0.0105	*0.0091	*0.0086	*0.0091	*0.0103	*0.0118	*0.0095	*0.0097	*0.0086	*0.0103	*0.0118	*0.0109
Asp104	0.0117	0.0121	0.013	*0.0104	0.0124	*0.0102	*0.01	*0.0112	0.0138	0.0156	*0.0116	*0.0116	*0.0102	0.0122	0.0156	0.0144
Asn117	0.0091	*0.0088	*0.0088	*0.0077	*0.0088	*0.0086	*0.0078	*0.0083	*0.0086	*0.008	*0.0086	0.0091	*0.0081	*0.009	*0.0082	*0.009
Ile127	0.0080	0.0082	0.0080	*0.0067	*0.0078	*0.0077	*0.0066	*0.0075	*0.007	*0.0069	*0.0076	*0.0079	*0.007	0.0082	*0.0007	*0.0078
Tyr164	0.0146	*0.011	*0.0116	*0.0094	*0.0109	*0.011	*0.0097	*0.0101	*0.0108	*0.0121	*0.0107	*0.0107	*0.0094	*0.011	0.0152	*0.0109
Arg168	0.0066	*0.0051	*0.0053	*0.004	*0.0052	*0.0051	*0.0043	*0.0044	*0.0051	*0.0053	*0.0048	*0.0047	*0.0043	*0.0049	0.0072	*0.0047
Glu178	0.0083	*0.0079	*0.0078	*0.007	*0.0078	*0.0074	*0.0068	*0.0073	*0.0077	*0.0071	*0.0075	*0.0077	*0.0072	0.0083	*0.0072	*0.0078
Lys228	0.0127	*0.0108	*0.0112	*0.0093	*0.0108	*0.0094	*0.0088	*0.0102	*0.0099	0.0131	*0.0102	*0.0108	*0.0089	*0.0108	0.0129	*0.0106
Tyr288	0.0078	*0.0039	*0.0042	*0.0043	*0.0048	*0.0038	*0.0052	*0.0039	*0.0047	*0.0052	*0.0032	*0.0035	*0.0051	*0.004	*0.0050	*0.0042

a) Bold numbers indicate amino acids with increased flexibility in the mutant compared with the native protein; * indicates amino acids with decreased flexibility in the mutants compared with the native protein.

point mutation using the SWISSPDB viewer for the 22 variants. Energy minimization was performed by NOMAD-Ref [30], followed by stimulated annealing by ifold [32] server to obtain optimized structures for both the native and mutant proteins.

Docking was performed using the PatchDock server between NAA and the native and mutant modeled structures of ASPA to determine the binding efficiency in the form of the ACE. The ACE between NAA and native ASPA was -117.39 kcal mol⁻¹, whereas the ACE values of the mutants were between -19.06 and -142.78 kcal mol⁻¹. Six mutants (H21P, A57T, C152R, R168H, P280L and P280S) had higher binding affinity with NAA than the native ASPA. The remaining 15 mutants were found less binding affinity with NAA than native Aspartoacylase in terms of ACE. These data indicate that the increased binding effect of NAA with the six mutants (H21P, A57T, C152R, R168H, P280L and P280S) might be a result of the 3D conformation of NAA, which had a comfortable fit into the 3D space of the binding residues of these mutants as compared with the native. Table 2 shows that the most detrimental 15 point mutants, namely, I16T, E24G, G27R, D68A, D114Y, G123E, I143T, C152W, P181T, H244R, D249V, G274R, E285A, F295S and A305E had less binding affinity with NAA in terms of ACE and also had RMSD between ≥ 0.23 Å and ≤ 2.13 Å. The mutant C152Y has a very low RMSD value and lower binding affinity compared to other mutants. Only mutants with higher binding affinity based on the ACE score was chosen for the further work. Thus this mutant was not taken for further studies. Moreover, they were also commonly found to be less stable, deleterious and damaging by the I-mutant 2.0, SIFT and PolyPhen servers, respectively. Figure 1B shows the docked complex of both

the native protein and the I143T mutant modeled structure with the substrate NAA as an illustrative example. These 15 mutants were also confirmed as detrimental by experimental and clinical observations performed elsewhere. These studies did not use structural and binding analysis but rather used molecular genetic analysis, population prevalence and epidemiology survey. [6,7,51,52]. By contrast, we have reported 15 mutants were shown to be deleterious by structural analysis (total energy calculation, RMSD and Patchdock score). Hence, we submitted these 15 potential detrimental point mutations for normal mode analysis to understand the flexibility of the active site region for the native and mutant structures.

2.9 The majority of amino acids in active site show loss of flexibility

To understand the cause of the lower substrate binding efficiency of the 15 detrimental missense mutations, we used the program ElNémo [49] to compare the flexibility of amino acids that are involved in binding with NAA of both the native protein and the mutants. Table 3 depicts the flexibility of the amino acids in the substrate binding pocket (active site) of both the native and mutant proteins by means of the normalized mean square displacement (R^2). These data were further sorted into three different categories of flexibility. One is where the (R^2) of the amino acids in the substrate binding pocket of the mutant was the same as that of the native protein (termed identical flexibility). The second category was where the (R^2) of the amino acids in the substrate binding pocket of the mutant was higher than that of the native protein (termed increased flexibility). The last category was where the (R^2) of the amino acids in the sub

Table 4 Substrate binding amino acids of mutants with different ranges of flexibility based on $\langle R^2 \rangle^a$

Mutants	A	B	C
	$\langle R^2 \rangle$		
I16T	11	2	0
E24G	11	1	1
G27R	13	0	0
D68A	12	1	0
D114Y	13	0	0
G123E	13	0	0
I143T	13	0	0
C152W	12	1	0
P181T	11	2	0
H244R	13	0	0
D249V	12	0	1
G274R	13	0	0
E285A	10	2	1
F295S	8	5	0
A305E	12	1	0

a) 'A' denotes amino acids with decreased flexibility in the mutants compared with the native protein; 'B' denotes amino acids with increased flexibility in both the native protein and mutants; 'C' denotes amino acids with identical flexibility in both the native protein and the mutants.

strate binding pocket of a mutant was lower than that of the native protein (termed decreased flexibility). From this analysis, we found that fewer substrate binding amino acids of these 15 mutants have identical flexibility than have increased and decreased flexibility (Table 4). However, fewer substrate binding amino acids have increased flexibility than have decreased flexibility (Table 4). Thus the majority of the amino acids participating in substrate binding of these mutants lost their flexibility, leading to a loss of binding efficiency with the substrate.

3 Conclusion

Of the 30 variants that were retrieved from Swissprot, 28 variants were found less stable by I-Mutant2.0, 27 variants were found to be deleterious by SIFT and 25 variants were considered damaging by PolyPhen. Twenty-two variants were selected as potentially detrimental point mutations because they were commonly found to be less stable, deleterious and damaging by the I-Mutant 2.0, SIFT and PolyPhen servers, respectively. The structures of these 22 variants were modeled and the RMSD between the mutants and native structures ranged from 0.23Å to 2.13Å. Docking analysis between NAA and the native and mutant modeled structures generated ACE scores between -142.78 and -19.06. Finally, we concluded that the lower binding affinity of 15 mutants (I16T, E24G, G27R, D68A, D114Y, G123E, I143T, C152W, P181T, H244R, D249V, G274R, E285A, F295S and A305E) with NAA compared with ASPA in terms of their ACE and RMSD scores identified them as deleterious mutations. Normalized mean square

displacement $\langle R^2 \rangle$ by normal mode analysis allows us to conclude that the majority of amino acids in the mutants bind to NAA (i.e., are in the active site) had decreased flexibility which could be the cause for their decreased substrate binding affinity. Thus the results indicate that our approach successfully allowed us to (i) consider computationally a suitable protocol for missense mutation (point mutation/single amino acid polymorphism) analysis before wet lab experimentation and (ii) provided an optimal path for further clinical and experimental studies to characterize ASPA mutants in depth.

The authors thank the management of Vellore Institute of Technology for providing the facilities to carry out this work.

- 1 Matalon R, Michals-Matalon K, Sebesta M, et al. Aspartoacylase deficiency and N-acetylaspartic aciduria in patient with Canavan disease. *Am J Med Genet*, 1988, 29: 463-471
- 2 Surendran S, Michals-Matalon K, Quast M J, et al. Canavan disease: a monogenic trait with complex genomic interaction. *Mol Genet Metab*, 2003, 80: 74-80
- 3 Zeng B J, Wang Z H, Ribeiro L A, et al. Identification and characterization of novel mutations of the aspartoacylase gene in non-Jewish patients with Canavan disease. *J Inher Metab Dis*, 2002, 25: 557-570
- 4 Matalon R, Michals-Matalon K. Spongy degeneration of the brain, Canavan disease: biochemical and molecular findings. *Front Biosci D*, 2000, 5: 307-311
- 5 Shaag A, Anikster Y, Christensen E, et al. The molecular basis of Canavan (aspartoacylase deficiency) disease in European non-Jewish patients. *Am J Hum Genet*, 1995, 57: 572-580
- 6 Kaul R, Gao G P, Michals K, et al. Novel (cys 125 arg) missense mutation in an Arab patient with Canavan disease. *Hum Mutat*, 1995, 5: 269-271

- 7 Kaul R, Gao G P, Matalon R, et al. Identification and expression of eight novel mutations among non-Jewish patients with Canavan disease. *Am J Hum Genet*, 1996, 59: 95–102
- 8 Kobayashi K, Tsujino S, Ezoe T, et al. A missense mutation I143T in a Japanese patient with Canavan disease. *Hum Mutat*, 1998, 1: S308–S309
- 9 Adornato B T, O'Brien J S, Lampert P W. Cerebral spongy degeneration of infancy: a biochemical and ultrastructural study of affected twins. *Neurology*, 1972, 22: 202–210
- 10 Baslow M H. Molecular water pumps and the aetiology of Canavan disease: a case of the sorcerer's apprentice. *J Inher Metab Dis*, 1999, 22: 99–101
- 11 Adachi M, Torii J, Schneck L, et al. Electron microscopic and enzyme histochemical studies of the cerebellum in spongy degeneration (van Bogaert and Bertrand type). *Acta Neuropathol*, 1972, 20: 22–31
- 12 Luo Y, Huang K. Spongy degeneration of the CNS in infancy. *Arch Neurol*, 1984, 41: 164–170
- 13 Barash V, Flhor D, Morag B, et al. A radiometric assay for aspartoacylase activity in human fibroblasts: application for the diagnosis of Canavan's disease. *Clin Chim Acta*, 1991, 201: 175–181
- 14 Bitto E, Bingman C A, Wesenberg G E, et al. Structure of aspartoacylase, the brain enzyme impaired in Canavan disease. *Proc Natl Acad Sci USA*, 2007, 104: 456–61
- 15 Hershfield J R, Pattabiraman N, Madhavarao C N, et al. Mutational analysis of aspartoacylase: implications for Canavan disease. *Brain Res*, 2008, 1148: 1–14
- 16 Matalon R. Canavan disease: diagnosis and molecular analysis. *Genet Test*, 1997, 1: 21–25
- 17 Rajasekaran R, Sudandiradoss C, Doss C G, et al. Identification and *in silico* analysis of functional SNPs of the BRCA1 gene. *Genomics*, 2007, 90: 447–452
- 18 Rajasekaran R, Priya Doss C G, Sudandiradoss C, et al. *In silico* analysis of structural and functional consequences in p16INK4A by deleterious nsSNPs associated CDKN2A gene in malignant melanoma. *Biochimie*, 2008, 90: 1523–1529
- 19 Rajasekaran R, Sethumadhavan R. Application of molecular mechanics and molecular dynamic for investigating the detrimental missense mutations in tumour suppressor protein SMAD4. *J Bionanosci*, 2009, 3: 80–87
- 20 Yip Y L, Scheib H, Diemand A V, et al. The Swiss-Prot variant page and the ModSNP database: a resource for sequence and structure information on human protein variants. *Hum Mutat*, 2004, 23: 464–470
- 21 Yip Y L, Famiglietti M, Gos A, et al. Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase. *Hum Mutat*, 2008, 29: 361–366
- 22 Boeckmann B, Bairoch A, Apweiler R, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res*, 2003, 31: 365–370
- 23 Berman H M, Westbrook J, Feng Z, et al. The Protein Data Bank. *Nucleic Acids Res*, 2000, 28: 235–242
- 24 Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*, 2005, 33: 306–310
- 25 Bava K A, Gromiha M M, Uedaira H, et al. ProTherm, version 4.0: thermodynamic database for proteins and mutants. *Nucleic Acids Res*, 2004, 32: 120–121
- 26 Ng P C, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 2003, 31: 3812–3814
- 27 Ng P C, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res*, 2001, 11: 863–874
- 28 Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 2002, 30: 3894–3900
- 29 Cavallo A, Martin A C. Mapping SNPs to protein sequence and structure data. *Bioinformatics*, 2005, 21: 1443–1450
- 30 Lindahl E, Azuara C, Koehl P, et al. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic acids Res*, 2006, 34: 52–56
- 31 Delarue M, Dumas P. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc Natl Acad Sci USA*, 2004, 101: 6957–6962
- 32 Sharma S, Ding F, Nie H, et al. iFold: a platform for interactive folding simulation of proteins. *Bioinformatics*, 2006, 22: 2693–2694
- 33 Han J H, Kerrison N, Chothia C, et al. Divergence of interdomain geometry in two-domain proteins. *Structure*, 2006, 14: 935–945
- 34 Varfolomeev S D, Uporov I V, Fedorov E V. Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases. *Biochemistry (Mosc)*, 2002, 67: 1099–1108
- 35 Leach A R. *Molecular Modeling: Principles and Applications*. 2nd ed. Sussex: Pearson Education EMA, 2001
- 36 Chou K C, Carlacci L. Simulated annealing approach to the study of protein structures. *Protein Eng*, 1991, 4: 661–667
- 37 Magyar C, Gromiha M M, Pujadas G, et al. SRide: a server for identifying stabilizing residues in proteins. *Nucleic Acids Res*, 2005, 33: W303–W305
- 38 Lo'pez G, Valencia A, Tress M L. Firestar—prediction of functionally important residues using structural templates and alignment reliability. *Nucleic Acids Res*, 2007, 35: 573–577
- 39 Feldman H J, Synder K A, Ticoll A, et al. A complete small molecule dataset from the protein data bank. *FEBS Lett*, 2006, 580: 1649–1653
- 40 Duhovny D, Nussinov R, Wolfson H J. Efficient unbound docking of rigid molecules. In: *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI) Lecture Notes in Computer Science*, Rome, Italy, 2002. 2452: 185–200
- 41 Connolly M L. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 1983, 221: 709–713
- 42 Schneidman-Duhovny D, Inbar Y, Nussinov R, et al. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res*, 2005, 33: 363–367
- 43 Zhang C, Vasmatzis G, Cornette J L, et al. Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol*, 1997, 267: 707–726
- 44 Yuan Z, Bailey T L, Teasdale R D. Prediction of protein B-factor profiles. *Proteins*, 2005, 58: 905–912
- 45 Ringe D, Petsko G A. Study of protein dynamics by X-ray diffraction. *Methods Enzymol*, 1986, 131: 389–433
- 46 Parthasarathy S, Murthy M R. Protein thermal stability: insights from atomic displacement parameters (B values). *Protein Eng*, 2000, 13: 9–13
- 47 Carlson H A, McCammon J A. Accommodating protein flexibility in computational drug design. *Mol Pharmacol*, 2000, 57: 213–218
- 48 Hinkle A, Tobacman L S. Folding and function of the troponin tail domain, effects of cardiomyopathic troponin T mutations. *J Biol Chem*, 2003, 278: 506–513
- 49 Suhre K, Sanejouand Y H. ElNe'mo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*, 2004, 32: 610–614
- 50 Teng S, Madej T, Panchenko A, et al. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys J*, 2009, 96: 2178–2188
- 51 Elepeleg O N, Shaag A. The spectrum of mutations of the aspartoacylase gene in Canavan disease in Non-Jewish patients. *J Inher Metab Dis*, 1999, 22: 531–534
- 52 Hussain R, Daud S, Kakar N. A missense mutation (p.G274R) in gene ASPA causes Canavan disease in a Pakistani family. *Mol Bio Rep*, 2012, 39: 6197–6201
- 53 Sisternans E A, de Coe R F, van Beerendonk H M, et al. Mutation detection in the aspartoacylase gene in 17 patients with Canavan disease: four new mutations in the non-Jewish population. *J Hum Genet*, 2000, 8: 557–560
- 54 Olsen T R, Tranebjaerg L, Kvittingen E A, et al. Two novel aspartoacylase gene (ASPA) missense mutations specific to Norwegian and Swedish patients with Canavan disease. *J Med Genet*, 2002, 39:

E55

- 55 Kaul R, Gao G P, Balamurugan K, *et al.* Cloning of the human aspartoacylase cDNA and a common missense mutation in Canavan disease. *Nat Genet*, 1993, 5: 118–123
- 56 Moore R A, Le Coq J, Faehnle C R. Purification and preliminary characterization of brain aspartoacylase. *Arch Biochem Biophys*, 2003, 413: 1–8
- 57 Kaul R, Gao G P, Aloya M. Canavan disease: mutations among Jewish and non-Jewish patients. *Am J Hum Genet*, 1994, 55: 34–41
- 58 Kumar P, Henikoff S, Ng P C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, 2009, 4: 1073–1081
- 59 Capriotti E, Fariselli P, Rossi I, *et al.* A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics*, 2008, 9: S6
- 60 Adzhubei I A, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods*, 2010, 7: 248–249

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.