Front. Biol. 2014, 9(6): 504–518
DOI 10.1007/s11515-014-1320-4

RESEARCH ARTICLE

Influence of the SNPs on the structural stability of CBS protein: Insight from molecular dynamics simulations

C. GEORGE PRIYA DOSS (✉)[1], B. RAJITH[1], R. MAGESH[2], A. ASHISH KUMAR[3]

[1] *Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore-14, TamilNadu, India*
[2] *Department of Biotechnology, Faculty of Biomedical Sciences, Technology & Research, Sri Ramachandra University, Chennai-600116, TamilNadu, India*
[3] *Bioinformatics Division, School of Biosciences and Technology, VIT University, Vellore-14, TamilNadu, India*

Abstract **Cystathionine β-synthase is an essential enzyme of the trans-sulfuration pathway that condenses serine with homocysteine to form cystathionine. Missense mutations in CBS are the major cause of inherited homocystinuria, and the detailed effect of disease associated amino acid substitutions on the structure and stability of human CBS is yet unknown. Here, we apply a unique approach in combining *in silico* tools and molecular dynamics simulation to provide structural and functional insight into the effect of SNP on the stability and activity of mutant CBS. In addition, principal component analysis and free energy landscape were used to predict the collective motions, thermodynamic stabilities and essential subspace relevant to CBS function. The obtained results indicate that C109R, E176K and D376N mutations have the diverse effect on dynamic behavior of CBS protein. We found that highly conserved D376N mutation, which is present in the active pocket, affects the protein folding mechanism. Our strategy may provide a way in near future to understand and study effects of functional nsSNPs and their role in causing homocystinuria.**

Keywords CBS, *in silico*, molecular dynamics simulation, SNPs

Introduction

Cystathionine β synthase (CBS) is an essential cytosolic enzyme that catalyzes the condensation of homocysteine and serine to cystathionine with the release of water or hydrogen sulfide, an irreversible step in the trans-sulfuration pathway (Afman et al., 2003). CBS displays a complex structural architecture consisting of an N-terminal heme binding domain, central catalytic domain containing a pyridoxal-5′-dependent (PLP) molecule and C-terminal regulatory domain containing a tandem of CBS domain. Mutagenesis studies have demonstrated the involvement of Single Nucleotide Polymorphisms (SNPs) in modulating the enzymatic activity and loss of function of CBS. Deficiency in the CBS activity is one of the most common causes of classical homocystinuria (HCU), an inherited human genetic disorder of sulfur amino acid metabolism (Mudd et al., 2001). It was proposed that a deficiency in *CBS* gene is mainly caused by the occurrence of single amino acid substitution which induces CBS misfolding (Janosík et al., 2001). Furthermore, a couple of mutations of arginine residue at 266 positions were found to be pathogenic which leads to protein misfolding (Kim et al., 1997; Katsushima et al., 2006). Most of the amino acid substitution in coding the region was found to be originated from deamination of methylcytosine in CpG dinucleotides (Kraus et al., 1999). Previous studies have reported the association of SNP rs5742905, which results in an amino acid substitution of Ile to Thr at 844 positions, is significantly associated to myelomeningoceles risk (Martinez et al., 2009). Analysis of the crystal structure of the dimeric truncated human CBS (residues 1–413) inferred that disease-causing mutations are distributed in various areas: the dimer interface, the active site, the heme binding site, and the predicted interface region between the catalytic domain and the missing regulatory domain (Meier et al., 2001; Meier et al., 2003). Owing to the importance of *CBS* gene in human disorders, the functional analysis of mutations could hold a significant advantage in diagnosis and treatment. Over the past few years *in silico* studies have improved significantly in screening the functional SNPs and predicting the conformational changes upon single amino acid substitution in proteins (Rabbani et al., 2012; Lino Cardenas et al., 2011). Therefore, to investigate possible associations between genetic mutation and pheno-

Received April 7, 2014; accepted June 3, 2014
Correspondence: C. GEORGE PRIYA DOSS
E-mail: georgecp77@yahoo.co.in

typic variations, different algorithms like Sorting Intolerant from Tolerant (SIFT) (Ng and Henikoff 2003) and Phenotype Polymorphism (PolyPhen 2.0) (Adzhubei et al., 2010), I Mutant 3.0 (Capriotti et al., 2008) were used for the prioritization of high-risk non-synonymous (ns) mutations in the coding region. Above mentioned methods vary in the properties of the variant they take into account, also the nature and classification methods. SIFT is based on evolutionary information whereas, PolyPhen 2 combine both protein structural/functional parameters and sequence analysis derived information. Similarly, I Mutant 3.0 is used to analyze the change in structural stability. Many landmark expression quantitative trait loci studies in humans have been conducted which showed the involvement of polymorphisms in *cis* or *trans* regulatory regions in affecting the gene expression (Cheung et al., 2010) or exonic variants that alter transcript stability or splicing. In this study, conducted *in silico* functional analysis of SNPs by using F-SNP and is-rSNP to explore the potential associated regulatory mechanisms that may be involved in disease (Lee and Shatkay 2008; Macintyre et al., 2010). Even though the information and location about the mutation of *CBS* gene are available in CBS mutation database (http://cbs.lf1.cuni.cz/cbsdata/cbs_02.htm), the structural and functional impact of these mutations in *CBS* gene is still poorly understood. To answer this, in the absence of further investigations, we mapped the deleterious nsSNPs on to 3D structure (PDB ID: 1JBQ). Then, we subjected MD simulation analysis in native and mutant protein complexes of CBS to analyze the structural level changes in time scale level. An atomic level look at the protein behavior using molecular dynamics simulations helped in better understanding the impact of these mutations on the protein structure, which in turn helped us in investigating how an amino acid variation can create a ripple effect throughout the protein structure and ultimately affect function. Moreover, the folding dynamics of native and mutant protein are studied here by constructing free energy landscape along the principal components obtained from a principal component analysis (PCA) (Jolliffe, 2002) which typically captures most of the total displacement from the average protein structure with the first few PCs during a simulation. Mapping the conformational free energies of native and mutant CBSs onto a principal component space enables identification of population changes in various conformations on mutation. The main purpose of this proposed work is to find out the most functional mutations which affect the folding mechanism of CBS.

Materials and methods

Data set used for SNP annotation

Human *CBS* gene information data was collected from Online Mendelian Inheritance in Man (OMIM) (Amberger et al., 2009) and Entrez Gene on National Centre for Biological Information (NCBI). The SNP information (Protein accession number (NP), mRNA accession number (NM) and SNP ID) of CBS was retrieved from the NCBI dbSNP (Sherry et al., 2001), and SWISS-Prot databases (Amos and Rolf 1996). Protein 3D structure was obtained from Protein Data Bank (PDB) (Berman et al., 2000).

Predicting functional context of missense mutation

The functional context of nsSNPs was predicted using SIFT, PolyPhen 2.0 and I Mutant 3.0. SIFT is a sequence homology based tool that predicts variants as neutral or functionally significant using normalized probability score. Variants at position with normalized probability score less than 0.05 are predicted to be functionally significant and score greater than 0.05 are predicted to be neutral (Ng and Henikoff 2003). PolyPhen 2.0 is based on the combination of sequence and structure based attributes and uses naive Bayesian classifier for the identification of amino acid substitution and the impact of mutation. The output levels of probably damaging (0.85–1.0) and possibly damaging (0.15–0.84) were classified as functionally significant and the benign level being (0–0.14) classified as tolerated (Adzhubei et al., 2010). I Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) is a support vector machine (SVM) based tool. We used the sequence based version of I Mutant 3.0 that classifies the prediction into three classes: neutral mutation ($-0.5 \leqslant \mathrm{DDG} \geqslant 0.5$kcal/mol), the large decrease (< -0.5 kcal/mol) and a large increase (> 0.5kcal/mol). The free energy change (DDG) predicted by I Mutant 3.0 is based on the difference between unfolding Gibbs free energy change of mutant and native protein (kcal/mol) (Capriotti et al., 2008).

Analysis of SNPs in regulatory region

The F-SNP database (Lee and Shatkay, 2008) was used to predict functional effects on protein coding, splicing regulation, transcriptional regulation and post translation based on 16 different tools. Functional significance scores (FS) are defined by F-SNP, which ranges between 0 and 1. An FS of 0 means none of the tools predict a deleterious effect; whereas an FS of 1 suggests all tools predict a deleterious effect. FS were deemed significant if $\mathrm{FS} \geqslant 0.5$, although it has to be taken into account that still 45% of disease-related SNPs were previously found to be < 0.5 (Lee and Shatkay, 2009). As a cross-reference for the transcription factor results within F-SNP, we used is-rSNP (Macintyre et al., 2010) to predict whether any of the SNPs would map to a potential transcription factor binding site. This tool utilizes the non-redundant human TF database JASPAR (Sandelin et al., 2004) to determine if any of the two SNP alleles are significantly predicted to be localized in a potential transcription factor binding site and also alter the binding score. The is-rSNP tool uses a standard cut-off of $P < 0.05$ for

Benjamini-Hochberg (Benjamini and Hochberg, 1995) corrected *P*-values of the observed difference between the alleles.

Model generation

The crystal structure of human CBS was obtained from the PDB (1JBQ) to generate the starting models for the simulation (Meier et al., 2001). Three different point mutations were introduced into the wild-type crystal structure that corresponds to C109R, E176K and D376N respectively. After mutation, the structure was subjected to optimization and energy minimized using GROMACS force field. Four different simulations were carried out that include the wild type and the three mutated models and each system was simulated for 6 ns.

Molecular dynamics simulation

All the molecular dynamics simulations were carried out using the program package GROMACS 4.5.3 (Hess et al., 2008) along with GROMOS9643a1 force field. Initially, all models were solvated with the 0.9 nm simple point charge (SPC) water embedded in the simulation boxes. To neutralize the systems, one chlorine ion was added to replace one SPC water molecule (Jorgensen, 1983). Subsequently, all the systems investigated were subjected to a steepest descent energy minimization until reaching a tolerance of 100 kJ/mol. After the solvent molecules were equilibrated with the fixed protein at 300 K for a while, the entire system was gradually relaxed and heated up to 300 K. Finally, 6ns MD simulations were performed under the normal temperature and pressure with the coupling time constant 1.0 ps. The particle mesh Ewald method (Essmann et al., 1995) was used to treat long-range Coulombic interactions and the simulations performed using the SANDER module. The SHAKE algorithm was used to constrain bond lengths involving hydrogen, permitting a time step of 2 fs. Van der Waals force was maintained at 1.4 nm, and coulomb interactions were truncated at 0.9 nm.

Principal component analysis and free energy landscape

The Principal component analysis is a technique that reduces the complexity of the data and extracts the concerted motion in simulations that are essentially correlated and presumably meaningful for biologic function (Amadei et al., 1993). In the Principal component analysis, a variance/covariance matrix was constructed from the trajectories after removal of the rotational and translational movements. The calculation of the eigenvectors and eigenvalues, and their projection along the first two principal components was carried out according to protocol within the GROMACS software package. A set of eigenvectors and eigenvalues was identified by diagonalizing the matrix. The eigenvalues represented the amplitude of the eigenvectors along the multidimensional space, and the displacement of atoms along each eigenvector showed the concerted motions of protein along each direction. The movements of protein in the essential subspace were identified by projecting the Cartesian trajectory coordinates along the most important eigenvectors from the analysis. For the simulation of both native and mutant protein backbone were included in the definition of the covariance matrices for the protein.

Analysis of molecular dynamics trajectory

The trajectory files were analyzed by using g_rms, g_rmsf, g_sas GROMACS utilities in order to obtain the root-mean-square deviation (RMSD), root-mean square fluctuation (RMSF), solvent accessibility surface area (SASA). Number of distinct intermolecular hydrogen bonds formed between during the simulation was calculated using g_hbond utility. Number of hydrogen bond is prominent, when donor-acceptor distance is smaller than 3.9 nm and donor–hydrogen-acceptor angle is larger than 90 nm. The trajectory files of PCA were analyzed through the use of g_covar and g_anaeig of GROMACS utilities in order to perform PCA. The free-energy landscape of the protein was obtained from the conformational sampling by using the g_sham module implemented in GROMACS.

Results

SNP annotation

CBS gene polymorphism data investigated in this work was retrieved from NCBI dbSNP and Swiss-Prot database. A total of 131 nsSNPs were selected for our analysis, out of which 8 nsSNPs were mapped on to the heme binding domain; 101 nsSNPs were mapped on to the central catalytic core unit and remaining 22 nsSNPs were mapped on the C-terminal domain.

Prediction of functional mutations

SIFT predicts whether an single amino acid substitution affects the protein function based on the physical properties of amino acid and sequence homology. SIFT program focuses more on sequence conservation over evolutionary time and the nature of amino acids in predicting the effect of residue substitutions on function. About, 69 nsSNPs were predicted as highly functional, exhibited a SIFT score of 0.00, and 28 nsSNPs exhibited a score ranging from 0.01 to 0.05 were predicted as functionally significant. Thus, 97 nsSNPs were predicted to be intolerant, that could not bring about a change in protein function. PolyPhen 2.0 evaluates the location of the amino acid replacement within identified functional domains and 3D structures. Unlike SIFT, PolyPhen does not solely depend on sequence homology alone to make SNP functional

Table 1 Analysis of functional SNPs in *CBS* gene using SIFT, PolyPhen 2 and I Mutant 3.0 tools

Variants	Amino Acid position	SIFT	Polyphen 2.0	I Mutant 3
VAR_046921	R18C	0.18	0.453	–
rs201827340	R18C	0.18	0.453	–
rs201372812	R45W	0.01	0.998	−1.27
rs148865119	P49L	0.01	0.977	−0.08
VAR_008049	P49L	0.01	0.977	−0.08
VAR_008050	R58W	0.01	0.982	0.16
rs199507134	E62K	0.98	0.012	−1.04
VAR_021790	H65R	0	0.982	0.05
rs17849313	A69P	0.94	0.00	0.32
rs185581633	P70L	0.68	0	0.6
rs192232907	K72I	0.16	0.001	0.12
VAR_002171	P78R	0.05	0.968	−0.21
VAR_008051	G85R	0	1	−0.3
VAR_002172	P88S	0	1	−1.37
rs71322503	R91K	0.37	0.008	−1.14
rs112029370	F99Y	1	0	−0.81
VAR_021791	L101P	0	1	−0.12
rs34040148	K102Q	0.07	0.301	−1.36
VAR_002173	K102N	0	0.994	−1.61
VAR_021792	**C109R**	**0**	**1**	**−2.32**
rs121964964	A114V	0.05	0.988	−1.55
VAR_002174	A114V	0.05	0.988	−1.55
VAR_008053	G116R	0	1	0.03
VAR_008054	**R121C**	**0**	**1**	**−1.95**
VAR_008055	**R121H**	**0**	**1**	**−2.18**
VAR_008056	R121L	0	1	−1.11
VAR_046923	**R125P**	**0**	**1**	**−2.2**
VAR_002175	R125Q	0.02	1	−2.19
VAR_008057	R125W	0	1	−1.64
VAR_008058	M126V	0	0.997	−0.11
VAR_008059	E128D	0.04	0.396	−1.11
VAR_002176	E131D	0	0.964	−0.5
rs140002610	R132C	0.01	0.998	0.6
rs147474549	G134R	0	0.999	−0.08
rs144832032	T135M	0.23	0.057	−0.48
VAR_008060	G139R	0	1	−1.44
VAR_021793	I143M	0	1	0.12
VAR_002177	E144K	0	1	−0.75
VAR_002178	P145L	0	1	−0.64
VAR_008061	G148R	0	1	−0.83
VAR_008062	G151R	0	1	−0.41
VAR_008064	I152M	0	0.985	0.78
VAR_046924	L154Q	0	1	−0.82
VAR_008065	A155T	0	0.999	−2.06
VAR_046925	A155V	0	1	−0.7
rs199817801	A157T	0.12	0.106	−2.28
VAR_002179	C165Y	0.07	1	−0.63
VAR_046926	V168A	0.01	0.947	−0.75
rs121964970	V168M	0.01	1	−0.21
VAR_002180	V168M	0.01	1	−0.21
VAR_046927	M173V	0	0.651	−1.1
VAR_008066	**E176K**	**0**	**1**	**−2.43**

(*Continued*)

Variants	Amino Acid position	SIFT	Polyphen 2.0	I Mutant 3
VAR_008067	V180A	0.25	0.014	−0.49
rs149649130	R182W	0	1	−0.02
rs138314784	R182Q	0.01	0.919	−0.55
VAR_008068	T191M	0	1	−1.72
VAR_008069	D198V	0	0.997	−0.27
VAR_066099	P200L	0	0.993	−1.31
rs201118737	K211R	0.2	0	−1.22
rs139456571	R224C	0.02	0.872	−0.79
VAR_002181	R224H	0.03	0.782	−0.72
VAR_008070	A226T	0.11	0.353	−0.94
VAR_021794	N228K	0	1	−0.55
VAR_046928	N228S	0	1	−0.66
VAR_046929	A231P	0	0.97	0.07
VAR_008071	D234N	0.01	0.998	−0.99
VAR_002182	E239K	0	1	−0.7
rs148257986	M250I	0.01	0.667	0.54
VAR_002183	T257M	0	1	0.21
rs143124288	G259S	0	1	−1.16
VAR_008072	T262M	0	1	−0.31
VAR_021795	T262R	0	1	−0.51
rs121964969	R266K	0.09	0.59	−0.88
VAR_008073	R266G	0	0.995	–
VAR_021796	C275Y	0	0.998	−0.2
VAR_066100	I278S	0	0.99	−0.64
rs117019516	I278T	0	0.967	−0.12
VAR_066101	D281N	0.02	0.998	−1.9
rs147040567	I286V	0.16	0.002	0.19
VAR_046932	A288P	0	0.98	−0.3
VAR_046933	A288T	0	0.986	−0.28
rs141502207	A288S	0.03	0.593	–
VAR_002185	P290L	0	0.997	–
rs201155833	E291D	0.3	0	−0.67
VAR_008076	E302K	0.17	0.616	−0.15
VAR_008077	G305R	0	1	0.27
VAR_002186	G307S	0	1	−0.22
VAR_008078	V320A	0.07	0.977	−1.12
VAR_066102	D321V	0	1	−0.9
VAR_008079	A331E	0	0.711	−0.33
VAR_002187	A331V	0.01	0.845	−0.16
VAR_002188	R336C	0	1	−0.62
VAR_008080	R336H	0	1	−1.32
VAR_021797	L338P	0	1	−0.5
VAR_021798	G347S	0	0.999	−1.93
VAR_021799	S349N	0	0.986	−0.72
VAR_008081	S352N	0.01	0.566	−0.21
VAR_008082	T353M	0	0.102	−0.17
rs121964972	T353M	0	0.102	−0.17
VAR_008083	V354M	1	0.016	−0.42
VAR_021800	A355P	0.19	0.66	−0.03
rs148589243	V358M	0.09	0.412	−0.16
VAR_046934	A361T	0.01	0.482	−0.83

(*Continued*)

Variants	Amino Acid position	SIFT	Polyphen 2.0	I Mutant 3
VAR_008084	R369C	0	1	−0.44
rs11700812	R369P	0	1	−0.09
rs11700812	R369H	0	1	−0.72
VAR_008085	C370Y	0	1	−0.54
VAR_002190	V371M	0	1	0.06
VAR_046935	**D376N**	**0**	**1**	**−2.07**
VAR_021801	R379Q	0.01	0.847	−0.99
VAR_046936	R379W	0	1	−0.31
VAR_002191	K384E	0.01	0.996	−1.2
rs121964967	K384E	0.01	0.996	−1.2
VAR_008086	K384N	0	0.999	−1.69
VAR_008087	M391I	0.01	0.414	0.13
rs28934892	P422L	0.11	0.9	–
rs138211175	V425M	0.01	0.983	−0.8
VAR_008088	T434N	0	0.86	−0.54
VAR_008089	I435T	0.07	0.528	−0.64
VAR_008090	R439Q	0.19	0.212	−0.25
rs28934891	D444N	0.05	0.122	−0.46
VAR_066103	A446S	0	0.336	−1.79
rs201585750	A452V	0.07	0.012	–
VAR_002193	V454E	1	0	–
VAR_021803	L456P	0.04	0.999	−0.56
rs141428279	M464T	0	0.999	−2.21
rs121964971	S466L	0.01	0.376	0.2
VAR_008091	S466L	0.01	0.376	0.2
rs201098477	G471R	0.21	0.752	−1.62
VAR_008092	R491C	0.08	0.002	−0.45
rs200613751	M505I	0.43	0	–
rs145228319	E514K	0.5	0.012	−1.28
rs201916339	G522R	0.02	0.027	0.1
VAR_046937	Q526K	0.98	0	–
VAR_008093	V534D	0	0.996	−1.37
VAR_002194	L539S	0	0.996	−0.93
rs121964968	L539S	0	0.996	−0.93
rs139651937	A545S	0.65	0.001	−1.91
rs150828989	R548Q	0.7	0.007	−0.67
VAR_046938	R548Q	0.7	0.007	−0.67

SNP highlighted in bold are predicated to be highly functional based on SIFT, PolyPhen 2.0 and I Mutant 3.0.

prediction, but also on structural information. PolyPhen predicted 50 of the nsSNPs to be "Probably damaging"; 47 nsSNPs as "Possibly damaging" and the remaining 34 nsSNPs were characterized as benign. Most of the mutations predicted to be functionally significant were also predicted to be damaging by PolyPhen (Table 1). Based on the difference in Gibbs free energy value of native and mutant proteins, 74 of nsSNPs are found to destabilize the protein. The results obtained from these *in silico* tools were compared with experimentally proved information for the validation of our predicted results. To narrow down the screening of most functional mutations, total energy change of native and mutant protein were computed before and after energy minimization.

Concordance analysis of predicted results using *in silico* tools

The accuracy of functional SNPs predicted can be increased by combining different computational methods. Out of 131 nsSNPs, 74% nsSNPs were predicted to be functionally significant by all three tools. To prioritize the most potent nsSNPs associated with *CBS* gene, the result obtained above were integrated into a single coherent framework. Thus, by

comparing the results obtained from all three tools, 5 nsSNPs (3%) in the coding region were predicted to have maximum functional effect on protein function and stability. We found that 3 SNPs at positions C109R, E176K, D356N were showing drastic shift in total free energy change (Table 2). Hence these three mutations were selected for further structural analysis. Difference *P*-value: significance of the change in binding score between the two SNP alleles, calculated by the is-rSNP tool. Adjusted difference *P*-value (BH): The Benjamini-Hochberg corrected *P*-value of the observed change in binding score between the two SNP alleles, calculated by the is-rSNP tool (shown are elements with BH-corrected $P < 0.05$). functional significance score calculated by F-SNP tool. SNP id highlighted in bold were found to be experimentally validated. Abbreviations: SR- splicing regulator; TF- transcription factor.

Table 2 Potential *cis*-acting regulatory elements affected by nsSNPs identified by expression quantitative trait loci analysis

SNP ID	Allele (A/a)	Regulatory element	Type	Difference *P*-value[1]	Adjusted difference *P*-value [2]	FS score[3]
rs706209	C/T	LM226	TF	1.25E-06	0.012	0.176
rs4987122	C/T	TGCGCANK	TF	1.95E-05	0.189	0.189
rs1051316	C/T	—	TF, SR	—	—	1
rs3788050	G/T	hlh-2::hlh-3	TF	3.20E-06	0.032	0.158
rs8127973	C/T	—	TF	—	—	0.158
rs2124458	C/T	CCCNNNAWT	TF	1.93E-05	0.188	0.176
rs2124459	C/T	—	TF	—	—	0.176
rs2124460	C/T	YGTCCTTGR	TF	1.49E-05	0.144	0.176
rs2124461	C/T	LM176	TF	2.04E-05	0.199	0.199
rs8132811	C/T	usp	TF	1.47E-05	0.143	0.176
rs760124	A/G	Egr1	TF	1.57E-05	0.153	0.05
rs234701	A/G	LM141	TF	1.87E-05	0.182	0.176
rs11700992	A/G	PUT3	TF	9.61E-06	0.093	0.176
rs234702	C/G	LM145	TF	1.35E-05	0.138	0.176
rs6586282	C/T	Tcfcp2l1	TF	9.08E-06	0.088	0.176
rs2895956	C/T	opa	TF	3.71E-06	0.036	0.176
rs9325622	A/G	Tcfe2a_2	TF	1.25E-06	0.012	0.176
rs11203172	C/T	TGCGCANK	TF	1.95E-05	0.188	0.176
rs234704	A/G	SOX2	TF	2.88E-05	0.035	0.176
rs1801181	A/G	Pou5f1	TF, SR	1.06E-06	0.189	0.289
rs2014564	A/G	REL	TF,	1.49E-05	0.006	0.208
rs4920037	C/G	Spz1	TF	9.00E-06	0.088	0.176
rs7276378	C/T	TEAD1	TF	3.21E-06	0.046	0.176
rs1789953	C/T	Ik-2	TF	2.25E-06	0.112	0.05
rs2228298	A/G	LM176	SR	2.95E-05	0.198	0.103
rs5742905	C/T	LM226	SR	1.99E-05	0.044	0.648
rs9978861	A/G	NIT2	TF	2.14E-05	0.899	0.178
rs9978863	A/G	SC35	TF	1.77E-05	0.142	0.208
rs2849727	A/G	RUNX1	TF	1.57E-05	0.153	0.208
rs234705	C/G	HSF	TF	1.87E-05	0.182	0.176
rs1788466	C/T	Spz	TF	9.61E-06	0.093	0.208
rs234706	C/T	Egr1	TF	2.04E-05	0.199	0.103
rs2298758	A/G	Ik-2	TF	1.47E-05	0.143	0.269
rs2298759	C/T	hlh-2::hlh-3	TF	1.57E-05	0.153	0.208
rs2298760	A/G	NIT-2	TF	1.87E-05	0.182	0.242
rs2298761	A/G	Sp1	TF	9.61E-06	0.093	0.242
rs234707	A/G	HSF	TF	2.04E-05	0.199	0.242
rs234708	C/G	Egr1	TF	1.47E-05	0.143	0.242

[1] Difference *P*-value: significance of the change in binding score between the two SNP alleles, calculated by the is-rSNP tool. [2] Adjusted difference *P*-value (BH): The Benjamini-Hochberg corrected *P*-value of the observed change in binding score between the two SNP alleles, calculated by the is-rSNP tool (shown are elements with BH-corrected $P < 0.05$). [3] functional significance score calculated by F-SNP tool. SNP id highlighted in bold were found to be experimentally validated. Abbreviations: SR- splicing regulator; TF- transcription factor.

Screening for functional expression quantitative trait loci SNPs

Using an *in silico* prediction tools F-SNP and is-rSNP, we identified 38 putative regulatory expression quantitative trait loci SNPs with altered CBS expression level out of which 36 SNPs were having transcription factors regulatory mechanisms and 2 SNPs were having splicing regulator mechanisms. Two SNPs rs1051316 and rs1801181 were found have the functional role in both transcription factors and splicing regulators (Table 3). To test if the functional predictions were likely to be biologically relevant, we searched the literature for prior evidence that the predicted regulatory elements were previously associated with any human disorder. Out of the 38 regulatory elements that may be influenced by the expression quantitative trait loci SNPs, 11 have previously been associated with hematologic cancers or, supporting the biologic plausibility of our findings (Boyles et al., 2006; Fan et al., 2008; Boyles et al., 2008; Paré et al., 2009; Martinez et al., 2009; Steinmaus et al., 2010; Metayer et al., 2011; Wernimont et al., 2011; Tilley et al., 2012).

Table 3 Total energy of native and mutant structures before and after energy minimization.

Amino acid change	Total energy (kJ/mol)	
	before minimization	After minimization
Native	– 11063.460	– 15831.080
C109R	**– 10127.572**	**– 15343.102**
R121H	– 10711.939	– 15217.965
R125P	– 10153.237	– 15604.006
E176K	**– 10078.544**	**– 15217.565**
D376N	**– 10009.343**	**– 15118.850**

Molecular dynamics simulation

To understand the structural consequences of the prioritized functional mutants, we conducted molecular dynamics simulation for the native and mutant CBS protein. Three factors, namely, tolerance index, PSIC score and change in free energy score correspond to the conformational changes in protein residues due to the mutation, which in turn affect the functional behavior of the protein molecule. The result obtained from the above analysis provoked us to study the dynamic behavior of native and mutant structures. We analyzed RMSD, RMSF, Rg, SASA and NH bond variations between the native and mutant structures C109R, E176K and D376N. The RMSD for all the Cα atoms from the initial protein structure was considered to measure the protein system. In Fig. 1A, native and e mutant models showed a similar fashion of deviation till ~1400 ps from their starting structure, resulting in backbone RMSD ~0.05 to 0.248 nm during the simulation. After 1500 ps, mutant D376N showed a different deviation pattern till the end of the simulation resulting in backbone RMSD of ~0.248 to 0.375 nm, whereas mutant C109R did not stay too far from the native protein. Mutant model E176K remained distinguished throughout the simulation resulting in backbone RMSD of ~0.03 to 0.34 nm. Despite the fact that mutant D376N and E176K deviated most from its initial conformation, toward the end of the simulation time, both mutants were observed to have similar RMSD values (a difference of less than 0.05 Å). This magnitude of fluctuation together with a small difference in average RMSD value after the relaxation period (approx 1500 ps) led to the conclusion that the simulation produced stable trajectory, thus providing a suitable basis for further study. With the aim of determining whether, mutation affected the dynamic behavior of residues; the RMSF values of native and mutant model were monitored. It was observed that D376N mutation affects neighboring residues at the maximum of around 0.246 nm fluctuation indicating a gain of flexibility due to mutation (Fig. 1B). Further, the RMSF values of native and mutant models C109R and E176K observed a similar fluctuation throughout the process. The radius of gyration (Rg) is defined as the mass-weighted root mean square distance of a collection of atoms from their common center of mass. The compactness, shape, and folding of the overall CBS structure at different time points during the trajectory can be seen in the plot of Rg (Fig. 1C). In the first 3000 ps, Rg value of mutant model D376N and E176K experienced a similar Rg pattern, whereas mutant C109R and native protein exhibited the similar pattern of simulation throughout the simulation period. Toward the end of the simulation, native protein exhibited lowest Rg value of around 1.91 nm, whereas mutant model D376N showed a maximum deviation of about 1.97 nm at around 5000 ps. Despite the fact that D376N mutant deviate the most from its initial conformation, toward the end of the simulation, native along with all mutant model plateaued around 6000 ps. We tried to understand the impact of mutation of protein activity by examining the change in SASA. The change of SASA of the native and mutant protein with time is shown in Fig. 1D. Native protein and mutant model C109R, E176K, indicated a similar pattern of deviation till 4500 ps, after which native protein exhibited lower surface accessibility value of around 81 $nm^{2.}$ After a relaxation period of 500 ps, the hydrophobic SASA of the native protein was found to be 81–96 nm^2 whereas the hydrophobic SASA of mutant D376N was found to be 8–101 nm^2. Mutant model D376N indicated a clear increase in hydrophobic SASA after 3000 ps when compared to the native protein. After 5000 ps, native along with three mutant models were equilibrated.

Intermolecular NH bond is calculated for native and mutant structure during the simulation time. Notable differences in protein–solvent interactions are evident in native and mutant, and it is shown in Fig. 1E. Mutant D376E shown least number of intermolecular NH bond with a range of ~218 to ~280 and in native protein maximum number of NH bond was observed within the range of ~220 to ~285. The average number of NH bond in the native protein was found to be ~255 and in mutant model D376N it was found to be ~240. Similar pattern of NH bond was observed in the case of both the mutant model

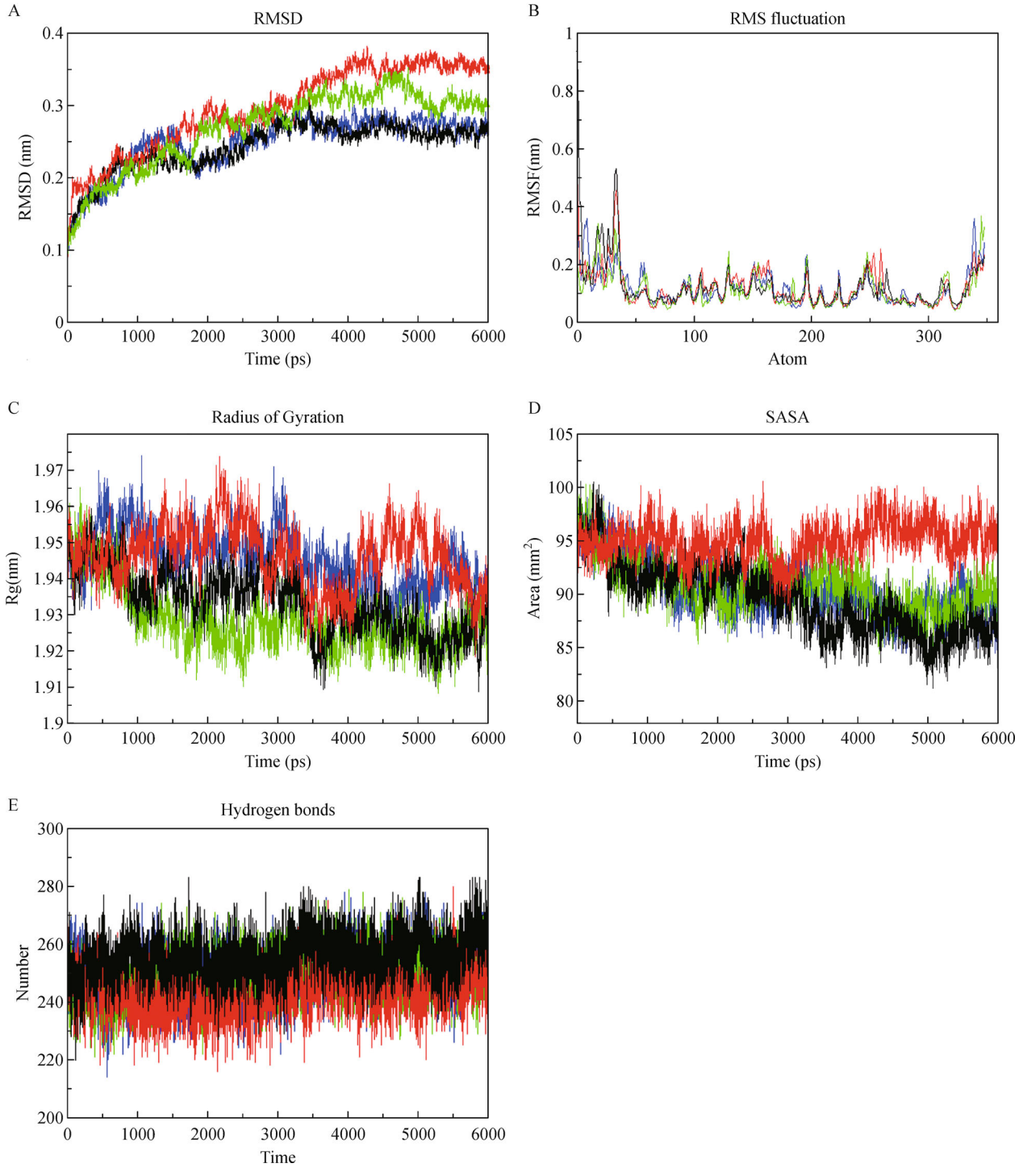

Figure 1 (A) Time evolution of backbone RMSDs are shown as a function of time of the wild and mutant structures at 6000 ps. The symbol coding scheme is as follows: wild (black color), mutant C109R (blue color), E176K (green color) and D376N (red color). (B) RMSF of the backbone carbon alpha over the entire simulation. The ordinate is RMSF (nm), and the abscissa is atom. The symbol coding scheme is as follows: wild (black color), mutant C109R (blue color), E176K (green color) and D376N (red color). (C) Rg of the protein backbone over the entire simulation. The ordinate is Rg (nm), and the abscissa is residue. The symbol coding scheme is as follows: wild (black color), mutant C109R (Blue color), E176K (green color) and D376N (red color). (D) Solvent accessible surface of protein over the entire simulation. The ordinate is SASA (nm^2), and the abscissa is atom. The symbol coding scheme is as follows: wild (black color), mutant C109R (blue color), E176K (green color) and D376N (red color). (E) Analysis of Intermolecular NH bond of native and mutant model protein at 6000ps. Average number of intermolecular hydrogen bond in native and mutant versus time. The symbol coding scheme is as follows: wild (black color), mutant C109R (blue color), E176K (green color) and D376N (red color).

C109R and E176K with an average number of 245 intermolecular NH bond. This might help to maintain its rigidity while less tendency of the mutant to involve in participating in hydrogen bonding with solvent makes it more flexible.

Principal component analysis

The calculated eigenvalues and cumulative contribution in the collective motion were calculated for the first 50 eigenvectors. More than 80% of the motion of the system was described by the first two eigenvectors, which explained the overall positional fluctuations contributing to the largest motions. The eigenvalues associated with each vector represents the variance of the molecular motion along that vector. In the case of native protein (Fig. 2A), the projection of first two PC components shows the populated clustered motions with trace of the covariance matrix after diagonalizing of 17.302 nm^2. Further analysis of Gibbs free energy landscape shows 5 global minima conformations (Fig. 2B). The minima are designated by blue dots that signify the stable state of protein in the folding free energy landscape. Mutations introduce large differences in the movement along the PCs, as evidenced by eigen value distributions (Fig. 2C-2H). All the three mutants showed random walk and less collective motions in essential subspace. Trace of the covariance matrix after diagonalizing obtained for the mutant model D376N exhibited less clustered motion with increased in covariance matrix value after diagonalizing of 18.414 nm^2 when compared to the native protein and other two mutant models (18.054 nm^2 and 18.348 nm^2) followed by Gibbs free energy landscape of 1 global minima conformations. Our observation thus signifies that; mutant model D376N covers the larger region of space and lesser thermodynamics stability than native at 300K.

Discussion

The functional relevance of the entire 131 mutations found in the *CBS* gene was assayed based on three different *in silico* approaches. To further elucidate the functional importance of nsSNPs on gene transcription and splicing mechanisms, we applied *in silico* methods like F SNP and is-rSNP. In addition, the stability affecting the function of protein was analyzed based on molecular dynamics simulation study followed by principal component analysis. The results obtained suggest that, among all, D376N mutation could affect the stability and in turn can be pathogenic which was supported by experimental evidence (Kruger et al., 2003). The use of different bioinformatics tools to predict the pathogenicity of genetic variations is always debatable because of the often discrepancy between some of them, even when analyzing the same variation. Hence there is a need to optimize the protocol into more integrated platforms, preferentially with the full processing in one single virtual environment which is not available yet. We recently applied such analysis to better comprehend a missense substitution in *HGD, G6PD,* and *ATM* gene (Magesh and George Priya Doss, 2012; Rajith and George Priya Doss, 2012; George Priya Doss et al., 2014). Our findings revealed that the incorporation of different algorithms often serves as powerful tools for prioritizing candidate functional nsSNPs. This was also supported mounting studies which utilized different set of *in silico* methods with diverse principles (Chan et al., 2007; Chun and Fay, 2009; Wei et al., 2010; Schwarz et al., 2010; Thusberg et al., 2011; Hicks et al., 2011; Hao et al., 2011). Thusberg and Vihinen (2009) compared different *in silico* tools, out of which SIFT and PolyPhen were reported to have better performance in identifying functional nsSNPs among other *in silico* tools. The accuracy of SIFT and PolyPhen 2.0 was further validated by Hicks et al. (2011), which makes these tools more suitable for the prediction. Khan and Vihinen (2010) suggested I Mutant 3.0 as one of the most reliable predictor for identifying the structural stability upon mutation. Based on this, we used a combination of prediction methods based on evolutionary information and protein structure and/or functional parameters were used in order to increase the prediction accuracy. By comparing the results obtained from the above methods, the following nsSNPs C109R, R121H, R125P, E176K and D376N were found to be highly significant. Three nsSNPs with mutation position C109R, E176K and D376N demonstrated significant change total in total energy difference before and after minimization. A literature search revealed that these three mutations were already implicated in human disorders (Steck et al., 1997; Kruger et al., 2003; Doniger et al., 2008).

In CBS-C109R, the mutation introduces a less hydrophobic residue which is important for multimerisation, and therefore this mutation could affect the multimer contacts (Aly et al., 2006). Furthermore, the mutant residue introduces a charge in a buried residue which can lead to protein folding problems. Gaustadnes et al. (2002) in his work validated the association on C109R mutation leading to loss of catalytic activity of protein which is the most common cause of homocystinuria. Our result suggests that, in E176K, the charge of the buried wild-type residue is reversed which can cause repulsion between residues in the protein core. The substitution of glutamic acid by lysine leads to disruption of hydrogen bond formed by the wild type residue which was found to be involved in multimer contact. In addition previous study has reported that E176K mutation leads to protein misfolding thereby reduce the catalytic activity of protein (Hnízda et al., 2012). In D376N mutation, aspartic acid residue which was mapped very near to active site pocket, forms hydrogen bond with asparagine 380, tyrosine 381 and asparagines 149 residues and hence the mutation would likely disrupt access to the active site of the enzyme (Singh et al., 2007). Moreover, D376N mutation which was present in the highly conserved region of the protein was

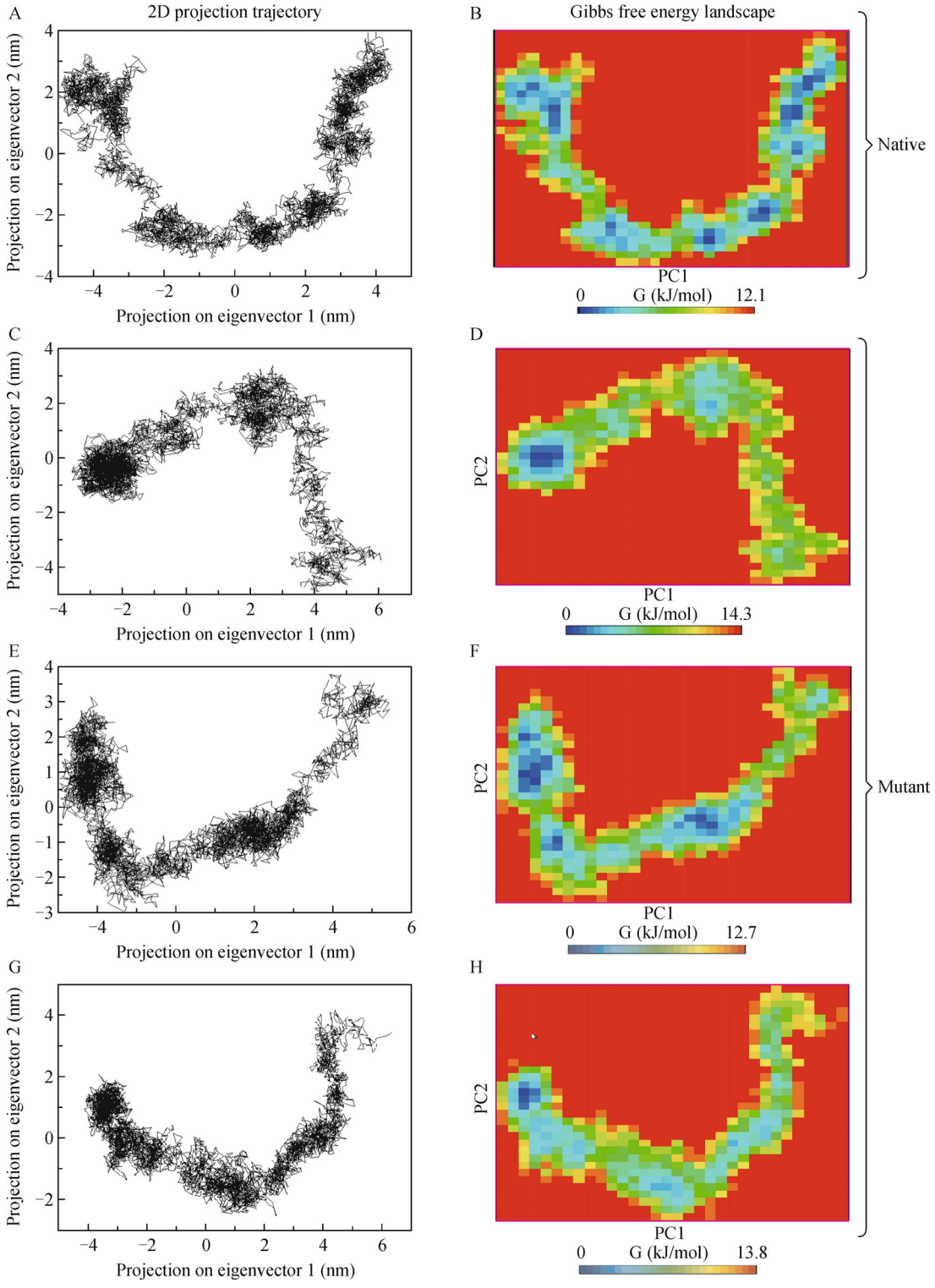

Figure 2 2D projection of first two principal eigenvector and comparison of free energy landscape of native and mutant CBS protein at 300K. (A) 2D Projection of native CBS at 300K. (B) Free energy landscape of native CBS. (C) 2D Projection of mutant C109R CBS at 300K. (D) Free energy landscape of Mutant CBS. (E) 2D Projection of mutant E176K CBS at 300K. (F) 2D Free energy landscape of mutant E176K CBS. (G) 2D Projection of mutant D376N CBS at 300K. (H) Free energy landscape of mutant D376N CBS.

predicted to have a significant impact on protein function (Fig. 3). Few studies have validated significant role of SNPs within evolutionary conserved regions (Aly et al., 2006; Doniger et al., 2008). Currently, there has been more focus on

the functional SNPs affecting regulatory regions. It has been validated that polymorphisms in 3′ UTR region affect the gene expression pattern during translation of mRNA whereas the polymorphisms in 5′ UTR region affect the RNA half-life by altering the polyadenylation (Wang et al., 2005; Sandberg et al., 2008). From our analysis, we identified 28 different regulatory elements that may influence the expression level of *CBS* gene. In spite of the above mentioned findings, assessment of how these residue changes propagate into the protein structure leading to a functionally disruptive effect can be obtained by comparing native and mutant protein model to MD simulations. Although it would be speculative to examine the detailed solvent behavior of protein models in order to investigate the difference in stability and dynamics behavior of native and C109R, K176K and D376N mutant models. A clear insight of stability loss was observed in the RMSF, RMSD, SASA, Rg and NH bond for D376N when compared to the native protein and other mutant models (R47G and V343E). Less intermolecular NH bond in D376N mutant structure might help to lose its rigidity and makes it more flexible. Further, PCA analysis was used to obtain the information about essential subspaces. The collective (correlated) motions of the atom in the protein play a key role in the protein function (van Aalten et al., 1995; Beer et al., 1996). PCA helped us to understand the motion of protein in free space. The increase in the deviation of mutant D376N might

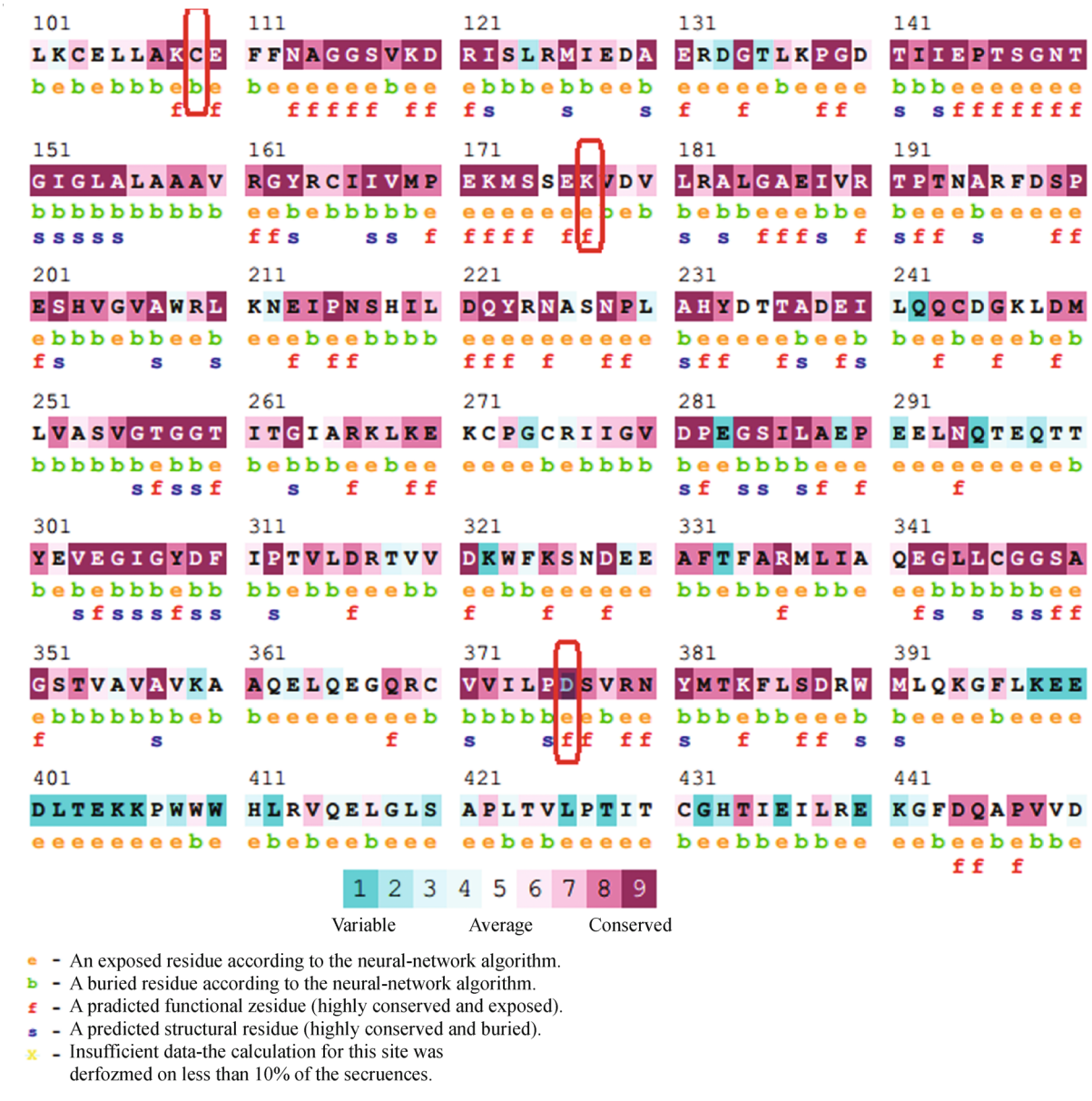

Figure 3 The conservation pattern of amino acid sequence in CBS. The location of amino acid residues in CBS based on the evolutionary conservation pattern. Red color box indicate the position of native amino acid position which are predicted to be most functionally significant by various *in silico* tools. Color intensity increases with the degree of conservation.

be due to disruption of secondary structure, which in turn affects the protein folding thereby decreasing the stability of protein.

This study also indicates that the identification of representative subspaces from the PCA is useful for elucidating the structure-function relationship for CBS. Hegger et al. (2007) defined the dimension of the free energy landscape by number of PCs that contain the most important molecular conformations. From our analysis, the D376N had only one global basin with local minima. Therefore, we would suggest that D376N mutation would have a great impact on protein folding which was in good concordance with the experimental results (Kruger et al., 2003; Singh et al., 2007).

Conclusions

In conclusion, this study comprises a comprehensive effort in genetic screening of functional SNPs in *CBS* gene associated with disease. One striking observation was the identification of D376N mutation that could impart maximum functional effect on CBS function. The prediction of nsSNPs in human *CBS* gene would be useful for further genotype-phenotype studies on the individual variation in drug metabolism and clinical response. In clinical laboratories, many new mutations in different genes are diagnosed daily. However, any standard protocol for evaluating the effects of new mutations on proteins structure and function has been not produced till date. To distinguish polymorphisms from harmful missense mutations, this method which incorporated different *in silico* tools in combination with molecular dynamics approach does remarkably well in a number of different experiments.

Compliance with ethics guidelines

All authors have not any potential conflict of interest. This article does not contain any studies with human or animal subjects performed by any of the authors.

Acknowledgements

The authors take this opportunity to thank the management of Vellore Institute of Technology and Sri Ramachandra University for providing the facilities and encouragement to carry out this work.

References

Adzhubei I A, Schmidt S, Peshkin L, Ramensky V E, Gerasimova A, Bork P, Kondrashov A S, Sunyaev S R (2010). A method and server for predicting damaging missense mutations. Nat Methods, 7(4): 248–249

Afman L A, Lievers K J A, Kluijtmans L A J (2003). Gene-gene interaction between the cystathionine b-synthase 31 base pair variable number of tandem repeats and the methylenetetrahydrofolate reductase 677C > T polymorphism on homocysteine levels and risk for neural tube defects. Mol Genet Metab, 78(3): 211–215

Aly T A, Eller E, Ide A, Gowan K, Babu S R, Erlich H A, Rewers M J, Eisenbarth G S, Fain P R (2006). Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1–B8-DR3 haplotype. Diabetes, 55(5): 1265–1269

Amadei A, Linssen A B M, Berendsen H J C (1993). Essential dynamics of proteins. Proteins, 17(4): 412–425

Amberger J, Bocchini C A, Scott A F, Hamosh A (2009). Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res, 37 (Database): D793–D796

Amos B, Rolf A (1996). The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res, 24(1): 21–25

Beer H D, Wohlfahrt G, McCarthy J E, Schomburg D, Schmid R D (1996). Analysis of the catalytic mechanism of a fungal lipase using computer-aided design and structural mutants. Protein Eng, 9(6): 507–517

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser A Stat Soc, 57: 289–300

Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N, Bourne P E (2000). The protein data bank. Nucleic Acids Res, 28(1): 235–242

Boyles A L, Billups A V, Deak K L, Siegel D G, Mehltretter L, Slifer S H, Bassuk A G, Kessler J A, Reed M C, Nijhout H F, George T M, Enterline D S, Gilbert J R, Speer M C (2006). NTD Collaborative Group. Neural tube defects and folate pathway genes: family-based association tests of gene-gene and gene-environment interactions. Environ Health Perspect, 114(10): 1547–1552

Boyles A L, Wilcox A J, Taylor J A, Meyer K, Fredriksen A, Ueland P M, Drevon C A, Vollset S E, Lie R T (2008). Folate and one-carbon metabolism gene polymorphisms and their associations with oral facial clefts. Am J Med Genet A, 146A(4): 440–449

Capriotti E, Fariselli P, Rossi I, Casadio R (2008). A three-state prediction of single point mutations on protein stability changes. BMC Bioinformatics, 2(Suppl 2): S6

Chan P A, Duraisamy S, Miller P J, Newell J A, McBride C, Bond J P, Raevaara T, Ollila S, Nyström M, Grimm A J, Christodoulou J, Oetting W S, Greenblatt M S (2007). Interpreting missense variants: comparing computational methods in human disease genes CDKN2A, MLH1, MSH2, MECP2, and tyrosinase (TYR). Hum Mutat, 28(7): 683–693

Cheung V G, Nayak R R, Wang I X, Elwyn S, Cousins S M, Morley M, Spielman R S (2010). Polymorphic Cis- and Trans-Regulation of Human Gene Expression. PLoS Biol, 8(9): e1000480

Chun S, Fay J C (2009). Identification of deleterious mutations within three human genomes. Genome Res, 19(9): 1553–1561

Doniger S W, Kim H S, Swain D, Corcuera D, Williams M, Yang S P, Fay J C (2008). Catalog of neutral and deleterious polymorphism in yeast. PLoS Genet, 4(8): e1000183

Essmann U, Perera L, Berkowitz M L, Darden T, Lee H, Pedersen L G (1995). A smooth particle meshes Ewald method. J Chem Phys, 103 (19): 8577–8593

Fan B J, Chen T, Grosskreutz C, Pasquale L, Rhee D, DelBono E, Haines J L, Wiggs J L (2008). Lack of association of polymorphisms in homocysteine metabolism genes with pseudoexfoliation syndrome and glaucoma. Mol Vis, 14: 2484–2491

Gaustadnes M, Wilcken B, Oliveriusova J, McGill J, Fletcher J, Kraus J

P, Wilcken D E (2002). The molecular basis of cystathionine beta-synthase deficiency in Australian patients: genotype-phenotype correlations and response to treatment. Hum Mutat, 20(2): 117–126

George Priya Doss C (2012). In Silico Profiling of deleterious Amino Acid Substitutions of Potential Pathological Importance in Hemophilia A and Hemophilia B. BMC J Biomed Sci, 19: 30

George Priya Doss C, Chakraborty C, Syed Haneef S A, NagaSundaram N, Chen, Zhu H (2014). Evolution and structure-based computational design to reveal the impact of deleterious missense mutations in type 2 maturity-onset diabetes of the young. Theranostics, 4: 366–385

Hao D C, Feng Y, Xiao R, Xiao P G (2011). Non-neutral nonsynonymous single nucleotide polymorphisms in human ABC transporters: the first comparison of six prediction methods. Pharmacol Rep, 63(4): 924–934

Hegger R, Altis A, Nguyen P H, Stock G (2007). How complex is the dynamics of peptide folding? Phys Rev Lett, 98(2): 028102

Hess B, Kutzner D, Spoel D (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput, 4(3): 435–447

Hicks S, Wheeler D A, Plon S E, Kimmel M (2011). Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Hum Mutat, 32(6): 661–668

Hnízda A, Majtan T, Liu L, Pey A L, Carpenter J F, Kodíček M, Kožich V, Kraus J P (2012). Conformational properties of nine purified cystathionine β-synthase mutants. Biochemistry, 51(23): 4755–4763

Janosík M, Oliveriusová J, Janosíková B, Sokolová J, Kraus E, Kraus J P, Kozich V (2001). Impaired heme binding and aggregation of mutant cystathionine betasynthase subunits in homocystinuria. Am J Hum Genet, 51(6): 1506–1513

Jolliffe I T (2002). Principal Component Analysis. New York: Springer

Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W, Klein M L (1983). Comparison of simple potential functions for simulating liquid water. J Chem Phys, 79(2): 926

Katsushima F, Oliveriusova J, Sakamoto O, Ohura T, Kondo Y, Iinuma K, Kraus E, Stouracova R, Kraus J P (2006). Expression study of mutant cystathionine beta-synthase found in Japanese patients with homocystinuria. Mol Genet Metab, 87(4): 323–328

Khan S, Vihinen M (2010). Performance of protein stability predictors. Hum Mutat, 31(6): 675–678

Kim C E, Gallagher P M, Guttormsen A B, Refsum H, Ueland P M, Ose L, Folling I, Whitehead A S, Tsai M Y, Kruger W (1997). Functional modeling of vitamin responsiveness in yeast: a common pyridoxine-responsive cystathionine b synthase mutation in homocystinuria. Hum Mol Genet, 6(13): 2213–2221

Kraus J P, Janosík M, Kozich V, Mandell R, Shih V, Sperandeo M P, Sebastio G, de Franchis R, Andria G, Kluijtmans L A, Blom H, Boers G H, Gordon R B, Kamoun P, Tsai M Y, Kruger W D, Koch H G, Ohura T, Gaustadnes M(1999). Cystathionine beta-synthase mutations in homocystinuria. Hum Mutat, 13: 362–375

Kruger W D, Wang L, Jhee K H, Singh R H, Elsas L J2nd (2003). Cystathionine beta-synthase deficiency in Georgia (USA): correlation of clinical and biochemical phenotype with genotype. Hum Mutat, 22(6): 434–441

Lee P H, Shatkay H (2008). F-SNP: computationally predicted functional SNPs for disease association studies. Nucleic Acids Res, 36(Database): D820–D824

Lee P H, Shatkay H (2009). An integrative scoring system for ranking SNPs by their potential deleterious effects. Bioinformatics, 25(8): 1048–1055

Lino Cardenas C L, Renault N, Farce A, Cauffiez C, Allorge D, Lo-Guidice J M, Lhermitte M, Chavatte P, Broly F, Chevalier D (2011). Genetic polymorphism of CYP4A11 and CYP4A22 genes and in silico insights from comparative 3D modelling in a French population. Gene, 487(1): 10–20

Macintyre G, Bailey J, Haviv I, Kowalczyk A (2010). is-rSNP: a novel technique for in silico regulatory SNP detection. Bioinformatics, 26 (18): i524–i530

Magesh R, George Priya Doss C (2012). Computational methods to work as first-pass filter in deleterious SNP analysis of alkaptonuria. ScientificWorldJournal, 2012: 738423

Martinez CA, Northrup H, Lin J I, Morrison A C, Fletcher J M, Tyerman G H, Au K S (2009). Genetic association study of putative functional single nucleotide polymorphisms of genes in folate metabolism and spina bifida. Am J Obstet Gynecol 201: 394e1–11

Meier M, Janosik M, Kery V, Kraus J P, Burkhard P (2001). Structure of human cystathionine beta-synthase: a unique pyridoxal 5′-phosphate-dependent heme protein. EMBO J, 20(15): 3910–3916

Meier M, Oliveriusova J, Kraus J P, Burkhard P (2003). Structural insights into mutations of cystathionine beta-synthase. Biochim Biophys Acta, 1647(1–2): 206–213

Metayer C, Scélo G, Chokkalingam A P, Barcellos L F, Aldrich M C, Chang J S, Guha N, Urayama K Y, Hansen H M, Block G, Kiley V, Wiencke J K, Wiemels J L, Buffler P A (2011). Genetic variants in the folate pathway and risk of childhood acute lymphoblastic leukemia. Cancer Causes Control, 22(9): 1243–1258

Mudd S H, Levy H, Kraus J P (2001). Disorders in transsulfuration. In: The Metabolic and Molecular Bases of Inherited Disease. McGraw-Hill, NY, pp. 2007–2056

Ng P C, Henikoff S (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res, 31(13): 3812–3814

Paré G, Chasman D I, Parker A N, Zee R R, Mälarstig A, Seedorf U, Collins R, Watkins H, Hamsten A, Miletich J P, Ridker P M (2009). Novel associations of CPS1, MUT, NOX4, and DPEP1 with plasma homocysteine in a healthy population: a genome-wide evaluation of 13 974 participants in the Women's Genome Health Study. Circ Cardiovasc Genet, 2(2): 142–150

Rabbani B, Mahdieh N, Haghi Ashtiani M T, Setoodeh A, Rabbani A (2012). *In silico* structural, functional and pathogenicity evaluation of a novel mutation: an overview of HSD3B2 gene mutations. Gene, 503(2): 215–221

Sandberg R, Neilson J R, Sarma A, Sharp P A, Burge C B (2008). Proliferating cells express mrnas with shortened 39 untranslated regions and fewer microRNA target sites. Science, 320(5883): 1643–1647

Sandelin A, Alkema W, Engström P, Wasserman W W, Lenhard B (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res, 32(90001): D91–D94

Schwarz J M, Rödelsperger C, Schuelke M, Seelow D (2010). MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods, 7(8): 575–576

Sherry S T, Ward M, Sirotkin K (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res, 29(1): 308–311

Singh L R, Chen X, Kozich V, Kruger W D (2007). Chemical chaperone rescue of mutant human cystathionine beta-synthase. Mol Genet Metab, 91(4): 335–342

Steck P A, Pershouse M A, Jasser S A, Yung W K, Lin H, Ligon A H, Langford L A, Baumgard M L, Hattier T, Davis T, Frye C, Hu R, Swedlund B, Teng D H, Tavtigian S V (1997). Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet, 15(4): 356–362

Steinmaus C, Yuan Y, Kalman D, Rey O A, Skibola C F, Dauphine D, Basu A, Porter K E, Hubbard A, Bates M N, Smith M T, Smith A H (2010). Individual differences in arsenic metabolism and lung cancer in a case-control study in Cordoba, Argentina. Toxicol Appl Pharmacol, 247(2): 138–145

Thusberg J, Olatubosun A, Vihinen M (2011). Performance of mutation pathogenicity prediction methods on missense variants. Hum Mutat, 32(4): 358–368

Thusberg J, Vihinen M (2009). Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. Hum Mutat, 30(5): 703–714

Tilley M M, Northrup H, Au K S (2012). Genetic studies of the cystathionine beta-synthase gene and myelomeningocele. Birth Defects Res A Clin Mol Teratol, 94(1): 52–56

van Aalten D M, Amadei A, Linssen A B, Eijsink V G, Vriend G, Berendsen H J (1995). The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulation in vacuum and water. Proteins, 22(1): 45–54

Wang G, Guo X, Floros J (2005). Differences in the translation efficiency and mRNA stability mediated by 59-UTR splice variants of human *SP-A1* and *SPA2* genes. AJP- Lung Physiol, 289: L497–L508

Wei Q, Wang L, Wang Q, Kruger W D, Dunbrack R L (2010). Testing computational prediction of missense mutation phenotypes: functional characterization of 204 mutations of human cystathionine beta synthase. Proteins, 78: 2058–2074

Wernimont S M, Clark A G, Stover P J, Wells M T, Litonjua A A, Weiss S T, Gaziano J M, Tucker K L, Baccarelli A, Schwartz J, Bollati V, Cassano P A (2011). Folate network genetic variation, plasma homocysteine, and global genomic methylation content: a genetic association study. BMC Med Genet, 12(1): 150

Supporting information

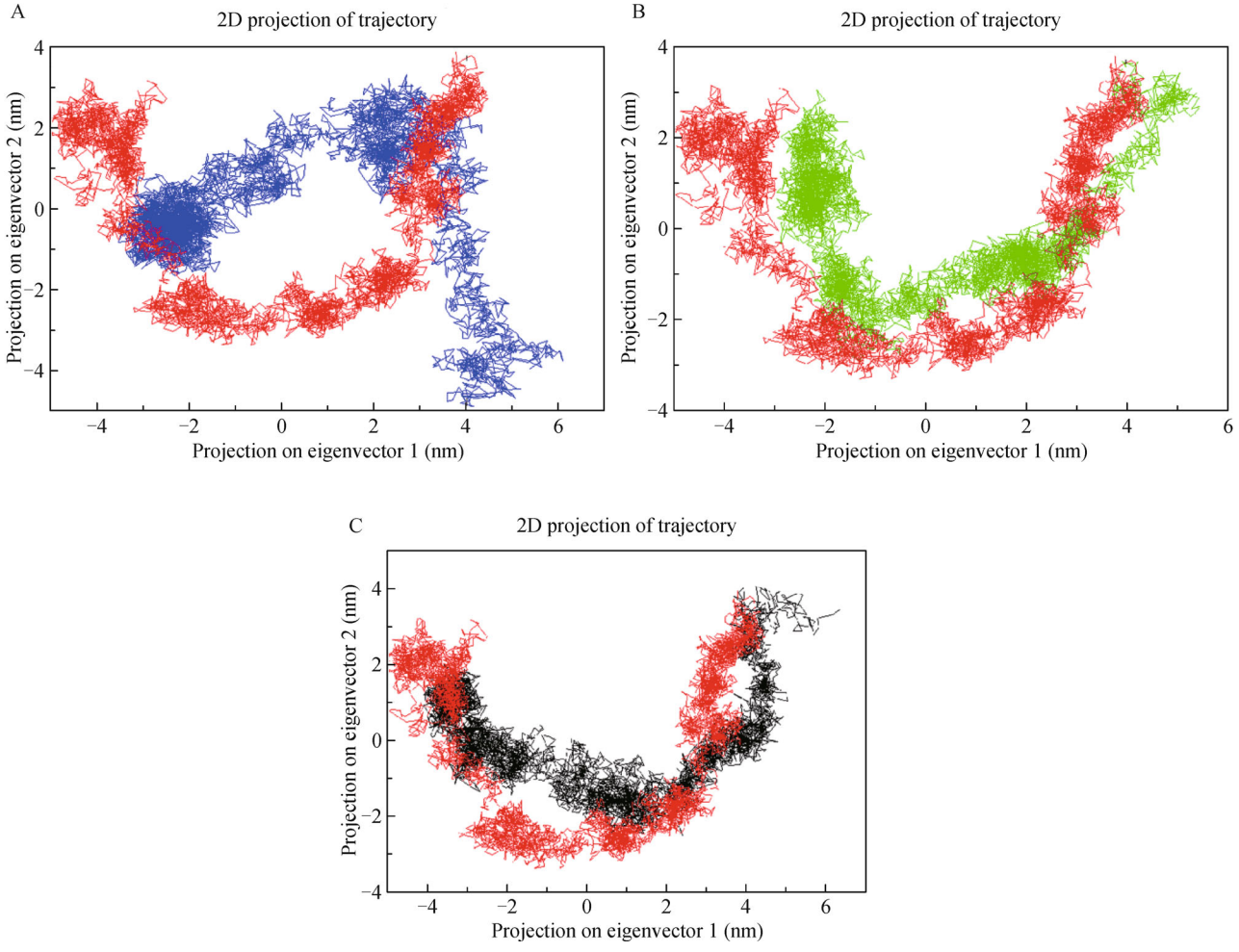

Figure S1 Projection of the motion of the protein in phase space along the first two principal eigenvectors and Gibbs free energy landscape at 300 K. (A) Wild type (Red color) vs. C109R (Blue color), (B) Wild type (Red color) vs. K176E (Green color). (C) Wild type (Red color) vs. D376N (black color).