Genomics Data 2 (2014) 66–72

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Identification of deleterious nsSNPs in α, μ, π and θ class of GST family and their influence on protein structure

P. Yadav [1], A. Chatterjee [1], A. Bhattacharjee *

Department of Biotechnology & Bioinformatics, North Eastern Hill University, Shillong 793022, India

ARTICLE INFO

Article history:
Received 19 February 2014
Received in revised form 21 March 2014
Accepted 31 March 2014
Available online 9 May 2014

Keywords:
Glutathione transferase (GST)
Non-synonymous single nucleotide polymorphism (nsSNP)
Mutant modeling
Drug resistance
Xenobiotics

ABSTRACT

GST family genes have a critical role in xenobiotic metabolism and drug resistance. Among the GST family the GST-μ, GST-π, GST-α and GST-θ are the most abundant classes and have a major role in the carcinogen detoxification process. Nevertheless the activity of these enzymes may differ due to polymorphisms which ultimately results in interindividual susceptibility to cancer development. In this work, we have analyzed the potentially deleterious nsSNPs that can alter the function of these genes. As a result among the nsSNPs, 101 (42.61%) were found to be deleterious by a sequence homology-based tool, 67 (28.27%) by a structure homology based tool and a total of 59 (24.89%) by both. We propose a modeled structure of the five highly deleterious mutant proteins. Our results will provide useful information in selecting target SNPs that are likely to have an impact on GST activity and contribute to an individual's susceptibility to the disease.

Introduction

Genetic variation in the human genome is an emerging resource for studying cancer and other diseases. Single-nucleotide polymorphisms (SNPs) are the most common type of DNA sequence variation, accounting for approximately 90% of the DNA polymorphism in humans [1] and some of these have been found to be associated with some rare human diseases. As per NCBI dbSNP Build 138 statistics approx 62.67 million human SNPs have been submitted; out of that, 44.27 million SNPs are validated. Common SNPs are found, on average, every 100–300 base pairs in the 3-billion-base pair genome [2], although their density varies between regions. A non-synonymous single nucleotide polymorphism (nsSNP), which is present within the exon of a gene, is responsible for the incorporation of an alternative amino acid and known to be one of the main causes for major genetic disorders. However, tolerant nsSNPs are not deleterious and are not involved in any genetic disorders, whereas deleterious nsSNPs have a profound influence on protein structure and its interaction. Therefore, it is important to differentiate deleterious nsSNPs from tolerant nsSNPs to characterize the genetic basis of human diseases. Discovering such deleterious nsSNPs is the main task of Pharmacogenomics. However, which set of SNPs to be screened is an important issue to understand between man and diseases. A possible way to overcome this problem would be to prioritize SNPs according to their functional significance [3,4] by using Bioinformatics prediction tools, which may help to discriminate neutral SNPs from SNPs of likely functional importance and could also be useful to reveal the structural basis of disease mutations.

Glutathione transferase (GST), a class of phase II xenobiotic metabolism enzymes (EC 2.5.1.18) has received a great deal of attention owing to their importance in cellular detoxification. In fact GST, catalyzes the conjugation of toxic substrates, with glutathione (GSH) and decreases their toxic activity against cellular macromolecules (prevent adduct formation, and thus protect organisms from DNA damage or protecting chromosomes from oxidative damage) [5]. In addition to phase II metabolism, GSTs are also involved in stress response, oncogenesis, tumor progression, drug resistance, biosynthesis and metabolism of prostaglandins, steroids, and leukotrienes [6]. More recently, GST isoenzymes have also been found to modulate cell signaling pathways that control cell proliferation and cell death [6,7]. In the *GSTP*−/− knockout mouse model, the rapid development of 12-*O*-tetradecanoylphorbol-13-acetate (TPA) induced cutaneous papilloma was observed than wild-type mice, which provides evidence that the enzyme is a key determinant of the proinflammatory tumor environment [8]. The human cytosolic GST consists of GST-α (alpha), GST-μ (mu), GST-π (pi), GST-σ (sigma), GST-ω (omega), GST-θ (theta), and GST-ξ (zeta) based on their sequence similarities, substrate specificity, and immune-reactivity. Among these classes GST-μ, GST-π, GST-α and GST-θ are the most abundant and variation in GST alleles is very common in the population. This variation makes significant contribution to inter-individual differences in the metabolism of xenobiotic

* Corresponding author. Tel.: +91 364 2722406.
E-mail addresses: abhattacharjeebioinfo@gmail.com, abhattacharjee@nehu.ac.in (A. Bhattacharjee).
[1] Mobile: +919436703339.

http://dx.doi.org/10.1016/j.gdata.2014.03.004

substances and drugs. Many sequence polymorphisms in the DNA sequence of these GSTs are reported, which may affect the enzymatic activity and subsequently exert deleterious effects [9]. Many studies demonstrated that the polymorphisms of these GSTs are associated with different types of cancer [10,11]. However many studies did not show consistency which could be due to the overlapping substrate affinity of these enzymes or that these SNPs might not have an impact on enzyme structure and function. Therefore it is important to identify deleterious nsSNPs which have an impact on the structure and function of carcinogen detoxification genes. In the GST family, GST-μ, GST-π, GST-α and GST-θ are major classes involve in carcinogen detoxification process and other carcinogenesis events. Therefore the present study takes a computational approach for in silico investigation on nsSNP mutation on these GST genes. To identify and distinguish nsSNP mutation that has a functional impact on protein structure through an experimental approach is time and money consuming. Thus the computational approach can help one to select SNPs for genotyping in molecular studies by using algorithms based on the evolutionary and biochemical severity of an amino acid substitution approach.

We applied different freely available computational algorithms based on sequence homology and physicochemical properties of the amino acid residue and an in silico site directed mutagenesis tool in this work to identify the possible deleterious mutations. We proposed a modeled structure of mutant proteins and compared them with the native protein. In general, these computational methods provide a feasible, high throughput way to determine the impact of large numbers of nsSNPs on protein function.

Methods and materials

Database mining for SNPs of GST family genes

We used National Center for Biotechnology Information (NCBI) database dbSNP (http://www.ncbi.nlm.nih.gov/Projects/SNP) [12] for our computational analysis.

Functional analysis of nsSNPs by sequence and structural homology based method (SIFT and Polyphen)

Residue changes that have an impact on the biophysical and structural properties of protein are known to be pathogenic or deleterious [13]. In our study we used two complementary Bioinformatics tools for high throughput prediction of the potential function impact of the nsSNPs of GST family genes: Sorting Intolerant from Tolerent (SIFT) (http://block.fhcrc.org/sift/SIFT.html) and Polymorphism Phenotyping (PolyPhen) (http://coot.embl.de/PolyPhen/). SIFT is a sequence based homology tool which presumes that important amino acids will be conserved in protein family and so changes at well conserve protein tend to be predicted as deleterious [14]. The algorithm used a modified version of PSI-BLAST [15] and Dirichlet mixture regularization [16] to construct a multiple sequence alignment of protein that can be globally aligned to the query sequence and belong to the same clade. SIFT is a multistep procedure that, given a protein sequence: (a) searches for similar sequence; (b) chooses closely related sequences that may share similar function; (c) obtains the multiple alignment of chosen sequence; and (d) calculates the normalized probability for all possible substitutions at each position with normalized alignment. Substitution at each position from the normalize probability less than a chosen cutoff are predicted to be tolerated. SIFT scores are designated as tolerant (0.201–1.00), borderline (0.101–0.20), potentially intolerant (0.051–0.10), or intolerant (0.00–0.05) [16]. Therefore an SNP is termed as deleterious if the cutoff value in SIFT program has a tolerance index of ≤ 0.05. The value higher the tolerance index, the less functional impact a particular amino acid substitution is likely to have.

PolyPhen is a structural-homology-based tool that predicts the impact of an amino acid substitution on the structure and function of a human protein. Predictions are based on a combination of phylogenetic, structural and sequence annotation information characterizing a substitution and its position in the protein. For a given amino acid variation, PolyPhen performs several steps: (a) extraction of sequence-based features of the substitution site from the UniProt database; (b) calculation of profile scores for two amino acid variants; and (c) calculation of structural parameters and contacts of a substituted residue. It calculates the PSIC score for each of two variants and then computes the PSIC score difference between them. The higher the PSIC score difference is, the higher is the functional impact a particular amino acid likely to have and on the basis of these score polymorphisms can be classified as probably benign (0.000–0.999), borderline (1.000–1.249), potentially damaging (1.250–1.499), possibly damaging (1.500–1.999), or damaging (≥ 2.000) [17].

Modeling nsSNP locations on protein structure and their RMSD difference

Structural analysis was performed for evaluating the structural stability of native and mutant protein. A graphical program for computational aided protein engineering, TRITON has been used for modeling mutant protein [18]. TRITON uses the external program MODELLER to construct structures of mutant protein based on the wild-type structure by homology modeling method. Energy minimization for 3D structures was performed using NOMAD-Ref server [19]. This server uses Gromacs as a default force field for energy minimization based on the methods of steepest descent, conjugate gradient and L-BFGS methods [20]. A conjugate gradient method was used for optimizing the 3D structures. The deviation between the two structures was evaluated by their root mean square deviation (RMSD) values. RMSD values more than 0.15 were considered as significant structural perturbations that could have

Table 1
List of human GST genes in α, μ, π and θ family and their SNP distribution.

Gene Name	Total SNP	nsSNP	% nsSNP	sSNP	%sSNP	3′ UTR	%3′ UTR	5′ UTR	%5′ UTR	iSNP	% iSNP
GSTM1	207	13	6.28	11	5.31	7	3.38	0	0	176	85.02
GSTM2	303	13	4.29	8	2.64	7	2.31	0	0	275	90.76
GSTM3	144	12	8.33	8	5.56	33	22.92	6	4.17	85	59.03
GSTM4	210	24	11.42	12	5.71	4	1.9	1	0.48	169	80.48
GSTM5	208	30	14.42	14	6.73	4	1.92	1	0.48	159	76.44
GSTA1	281	29	10.32	17	6.05	16	5.69	9	3.2	210	74.73
GSTA2	316	26	8.22	13	4.11	6	1.9	2	0.63	269	85.13
GSTA3	291	18	6.19	8	2.75	2	0.69	5	1.72	258	88.66
GSTA4	432	9	2.08	4	0.93	9	2.08	3	0.69	407	94.21
GSTA5	382	16	4.19	1	0.26	1	0.26	2	0.52	362	94.76
GSTP1	180	17	9.44	6	3.33	3	1.67	6	3.33	148	82.22
GSTT1	87	18	20.69	9	10.34	1	1.15	5	5.75	54	62.07
GSTT2	152	12	7.89	2	1.32	2	1.32	0	0	138	90.79
Total	3193	237	7.42	113	3.54	95	2.98	40	1.25	2710	84.87

nsSNP: non synonymous SNP; sSNP: synonymous SNP; 3′UTR: 3′ Untranslated region; 5′UTR: 5′ Untranslated region; and iSNP: intronic SNP.

Table 2
List of nsSNPs that were predicted to be deleterious.

SNP ID	Gene name	Nucleotide change	Amino acid substitution	Tolerance index	PSIC ID	Validation
rs72549312	GSTM1	C–T	P179L	0.11	1.933	No
rs72549313	GSTM1	C–T	R187C	0.00	1.000	No
rs184653774	GSTM1	A–C	D9E	0.02	1.745	Yes
rs147668562	GSTM1	A–G	N85S	0.10	1.561	Yes
rs142484086	GSTM1	C–T	R145W	0.02	2.390	Yes
rs11540636	GSTM2	C–T	F148S	0.00	1.715	Yes
rs140199111	GSTM2	G–T	G12W	0.00	2.785	No
rs145910843	GSTM2	A–G	R18H	0.00	1.484	No
rs147235683	GSTM2	A–G	R78Q	0.00	2.225	No
rs146447815	GSTM2	A–G	D106G	0.00	2.034	No
rs143184866	GSTM2	A–G	G143E	0.00	1.281	No
rs141100983	GSTM2	A–G	Y161C	0.00	1.505	Yes
rs140675803	GSTM2	A–G	R96H	0.00	0.921	Yes
rs11546855	GSTM2	A–G	D42G	0.01	2.294	No
rs1803686	GSTM3	C–A	R191L	0.00	3.001	No
rs1803687	GSTM3	G–C	K128N	0.00	2.331	Yes
rs11555177	GSTM3	T–C	S48G	0.05	1.370	No
rs184721419	GSTM3	C–T	R172H	0.04	0.584	Yes
rs138797459	GSTM3	C–T	S121G	0.04	1.123	No
rs146952826	GSTM3	C–T	R86H	0.05	0.751	Yes
rs150988571	GSTM3	C–T	E33K	0.03	0.880	Yes
rs140815169	GSTM3	A–G	L23P	0.00	2.365	Yes
rs142070930	GSTM3	C–T	G10R	0.00	2.293	No
rs3211195	GSTM4	A–G	M135I	0.02	0.752	Yes
rs145606771	GSTM4	A–G	R11H	0.00	1.332	No
rs148886417	GSTM4	C–G	I17M	0.00	2.144	Yes
rs138088784	GSTM4	G–T	R18L	0.00	3.158	Yes
rs145858198	GSTM4	C–T	C78R	1.00	0.647	Yes
rs139656805	GSTM4	A–G	E91K	0.02	1.071	No
rs144284999	GSTM4	A–G	R96H	0.00	0.925	Yes
rs142265412	GSTM4	C–G	A104G	0.01	1.005	Yes
rs114328674	GSTM4	A–G	R168C	0.00	2.519	Yes
rs61734547	GSTM5	G–T	L13R	0.00	0.890	No
rs144877199	GSTM5	C–G	A16G	0.02	1.374	No
rs145616779	GSTM5	A–G	E22K	0.15	0.895	No
rs142533115	GSTM5	A–G	V29M	0.00	1.234	No
rs139457478	GSTM5	C–T	I76T	0.00	2.780	No
rs147739570	GSTM5	C–T	R78C	0.03	0.990	Yes
rs148956224	GSTM5	A–G	R96H	0.00	0.926	No
rs144915668	GSTM5	C–G	N107K	0.09	0.696	Yes
rs144530836	GSTM5	A–T	L114Q	0.00	1.750	Yes
rs140499099	GSTM5	C–T	R145W	0.02	2.147	Yes
rs150881777	GSTM5	C–T	W147R	0.00	3.864	Yes
rs150417585	GSTM5	C–T	R187C	0.00	0.917	Yes
rs137869431	GSTM5	A–G	R187H	0.03	1.000	Yes
rs144827167	GSTM5	G–T	G190V	0.04	0.500	No
rs113130058	GSTM5	A–G	S217G	0.02	0.163	Yes
rs2234953	GSTT1	G–A	E173K	0.04	2.240	Yes
rs2266637	GSTT1	G–A	V169I	0.04	0.722	Yes
rs17856199	GSTT1	T–G	F45C	0.00	2.993	Yes
rs2266635	GSTT1	G–A	A21T	0.00	0.998	Yes
rs185499198	GSTT1	A–G	R240W	0.01	2.270	Yes
rs77300908	GSTT1	C–T	E204K	0.00	0.147	Yes
rs112867476	GSTT1	C–T	R197H	0.00	2.429	No
rs139881998	GSTT1	A–T	H162L	0.03	2.480	Yes
rs150601402	GSTT1	A–G	R112W	0.02	2.037	Yes
rs141759372	GSTT1	A–G	W101R	0.00	4.138	Yes
rs149896285	GSTT1	A–G	M1T	0.00	3.563	No
rs1126752	GSTT2	C–T	S68L	0.05	0.778	No
rs146675046	GSTT2	A–G	E147K	0.01	2.092	Yes
rs1804666	GSTP1	G–A	G78E	0.03	1.794	No
rs4986949	GSTP1	G–T	D147Y	0.01	2.439	Yes
rs71534294	GSTP1	G–C	D158H	0.00	2.234	Yes
rs11553892	GSTP1	C–A	L176M	0.05	1.417	Yes
rs45549733	GSTP1	C–T	R187W	0.00	2.696	No
rs188653023	GSTP1	A–G	R183H	0.01	1.198	Yes
rs191595383	GSTP1	C–G	P197A	0.12	1.904	Yes
rs78507509	GSTP1	C–G	P124A	0.13	1.918	Yes
rs1051983	GSTA1	G–T	A216S	0.04	0.327	Yes
rs17414159	GSTA1	C–T	C112R	0.53	2.081	No
rs73740645	GSTA1	A–T	K64M	0.03	0.708	Yes
rs148795539	GSTA1	C–T	E168K	0.00	1.938	No
rs1051778	GSTA1	A–T	I128K	0.00	1.395	Yes
rs138688572	GSTA1	A–G	I75T	0.00	1.529	Yes
rs140333826	GSTA1	A–G	L72F	0.02	0.756	No
rs145721561	GSTA1	G–T	A70D	0.00	2.332	*

Table 2 (*continued*)

SNP ID	Gene name	Nucleotide change	Amino acid substitution	Tolerance index	PSIC ID	Validation
rs61734623	GSTA1	A–T	K64M	0.03	0.708	No
rs11552000	GSTA1	C–T	M57T	0.01	2.011	No
rs138678278	GSTA1	A–C	G48V	0.00	2.365	No
rs1803682	GSTA2	G–T	K196N	0.03	1.202	Yes
rs2266631	GSTA2	C–T	V149A	0.00	2.189	Yes
rs75013911	GSTA2	C–T	E32K	0.00	2.044	Yes
rs151112301	GSTA2	A–G	R131C	0.04	1.679	Yes
rs147776857	GSTA2	C–T	G83R	0.00	2.520	Yes
rs138041732	GSTA2	A–C	L180R	0.23	1.566	Yes
rs146304331	GSTA2	A–G	R155W	0.20	1.183	Yes
rs142063997	GSTA2	C–T	R20Q	0.00	2.382	Yes
rs139552194	GSTA2	G–T	H8N	0.12	2.699	Yes
rs143619808	GSTA2	A–C	K64N	0.01	0.586	No
rs61734623	GSTA2	A–T	K64M	0.03	0.722	No
rs183168307	GSTA2	A–C	M57I	0.06	1.560	Yes
rs11552000	GSTA2	C–T	M57I	0.01	2.825	No
rs41273858	GSTA3	C–T	N73D	0.10	1.974	Yes
rs1052661	GSTA3	A–C	I71L	0.00	1.629	Yes
rs17851798	GSTA3	A–C	M63I	0.04	2.057	Yes
rs59410661	GSTA3	A–G	R13W	0.00	3.035	Yes
rs149910347	GSTA3	C–G	S202T	0.01	1.281	Yes
rs143944137	GSTA3	A–G	P200L	0.00	2.841	No
rs143163780	GSTA3	A–G	T193M	0.00	1.840	Yes
rs144126679	GSTA3	A–C	Y147D	0.00	3.028	Yes
rs141590731	GSTA3	C–T	E97G	0.03	1.570	No
rs143379014	GSTA3	A–G	F52L	0.00	2.732	No
rs148359991	GSTA3	C–G	D47H	0.01	2.197	No
rs186026850	GSTA3	A–C	E32D	0.00	0.503	Yes
rs141510758	GSTA3	C–T	R20Q	0.00	2.494	No
rs45551133	GSTA4	C–T	L100P	0.25	1.208	Yes
rs141595669	GSTA4	A–T	F197I	0.00	2.569	No
rs139066992	GSTA4	C–T	G144R	0.03	0.723	No
rs151284340	GSTA4	C–T	K84R	0.01	0.890	Yes
rs140367015	GSTA5	A–T	S142C	0.02	1.959	*
rs145445113	GSTA5	A–C	K141N	0.02	1.048	*
rs146408369	GSTA5	C–T	Y74C	0.00	3.242	Yes
rs185015376	GSTA5	C–T	M1I	0.00	2.924	Yes
rs145528403	GSTA5	A–G	T193M	0.00	0.530	Yes
rs150669459	GSTA5	C–T	R20Q	0.00	2.349	No

* SNPs were omitted from dbSNP database because its subsnp_id was deleted.

functional implications for the protein [21]. Molecular graphics images were produced using the UCSF Chimera package [22].

Result & discussion

SNP dataset

A total of 13 genes of four major classes of Cytosolic GST family viz. GSTA1, GSTA2, GSTA3, GSTA4, GSTA5, GSTM1, GSTM2, GSTM3, GSTM4, GSTM5, GSTP1, GSTT1 and GSTT2 investigated in this work were retrieved from the dbSNP database. These genes contained 3193 SNPs; out of that 237 were found to be nsSNPs, and 113 to be coding synonymous SNPs (sSNPs). The noncoding SNPs consisted of 40 SNPs in the 5′ Untranslated region (UTR), 95 SNPs in the 3′ UTR region and 2708 intronic SNPs (iSNP). The number and percentage for every SNP type of individual genes are given in Table 1. The coding nonsynonymous SNPs were selected for our investigation.

Prediction of deleterious nsSNPs by SIFT and Polyphen Program

A sequence homology based tool, SIFT was used to determine the conservation level of a particular single amino acid substitution in a protein based on the alignment of orthologous and/or paralogous protein sequences. Among 237 nsSNPs, 101 nsSNPs (42.61%) were found to be deleterious by a SIFT algorithm which showed a deleterious tolerance index score between 0.00 and 0.05. Out of these deleterious nsSNPs, 55 nsSNPs (52.88%) found to be exhibited highly deleterious tolerance index score of 0.00 which could affect the protein function in these genes.

Using Insilico tool Polyphen, 67 nsSNPs (28.27%) were found to be deleterious having a PSIC score difference ≥ 1.5. Out of which 19 nsSNPs lie between a PSIC score difference ≥ 1.500 and ≥ 1.999 and were predicted to be possibly deleterious. 40 nsSNPs lie between a PSIC score difference ≥ 2.000 and ≤ 2.999 and were predicted to be deleterious. 8 nsSNPs having a PSIC score difference ≥ 3.000 were predicted as highly deleterious. It was also observed that 59 (24.89%) nsSNPs were deleterious by SIFT as well as Polyphen tools. Deleterious nsSNPs predicted by SIFT and PolyPhen for GST genes are listed in Table 2.

Table 3
RMSD and total energy of native and their respective mutant modeled structures 1TDI R13W, 1TDI Y147D, 4GTU R18L, 3GTU R191L and 2C3N W101R.

Substitution	Energy (KJ/mol) after 25,000 step minimization		RMSD
	Native	Mutant	
R13W GSTA3	−27,029.410	−25,893.779	1.535
Y147D GSTA3	−27,029.410	−26,319.830	1.368
R18L GSTM4	−26,365.328	−25,996.713	0.924
R191L GSTM3	−57,996.078	−57,176.730	1.321
W101R GSTT1	−28,807.537	−28,944.705	0.937

Mutant structure modeling and their RMSD

Highly deleterious nsSNPs having a tolerance index of 0.00 and a PSIC score difference ≥3.00 have been selected for modeling on their respective native structure. A total of 8 nsSNPs fall in that criterion. Out of these eight nsSNPs, two (rs146408369 of GSTA5 and rs150881777 of GSTM5 genes) were excluded from modeling as the native protein structure was not available. Further one more nsSNPs (rs149896285 of GSTT1 gene) was excluded as a single amino acid polymorphism (SAP) occurred at the initiation codon. Therefore finally, a total of 5 nsSNPs (rs59410661, rs144126679 of GSTA3; rs1803686 of GSTM3; rs138088784 of GSTM4 and rs141759372 of GSTT1 gene) has been selected for modeling and analysis of the mutant structure. The amino acid residue substitutions were performed by the TRITON software to get mutant modeled structures (1TDI_R13W, 1TDI_Y147D, 3GTU_R191L, 4GTU_R18L and 2C3N_W101R). Then energy minimizations were performed by the NOMAD-Ref server for native structure and their respective mutant modeled structures. It was found that the total energy of the mutant proteins 1TDI_R13W and 1TDI_Y147D were −25893.779 and −26319.830 Kcal/mol, respectively and that of the native protein (1TDI) was −27029.410 Kcal/mol, for 3GTU and mutant 3GTU_R191L total energy was −57996.078 and −57176.730 Kcal/mol, respectively; for 4GTU and mutant 4GTU_R18L, the total energy was −26365.328 and −25996.713 Kcal/mol, respectively and for 2C3N and mutant 2C3N_W101R, the total energy was found to be −28807.537 and −28944.705, respectively (Table 3). The RMSD values for the modeled mutants were significant for the pathogenicity for all missense mutations. The RMSD value between the native type (1TDI)

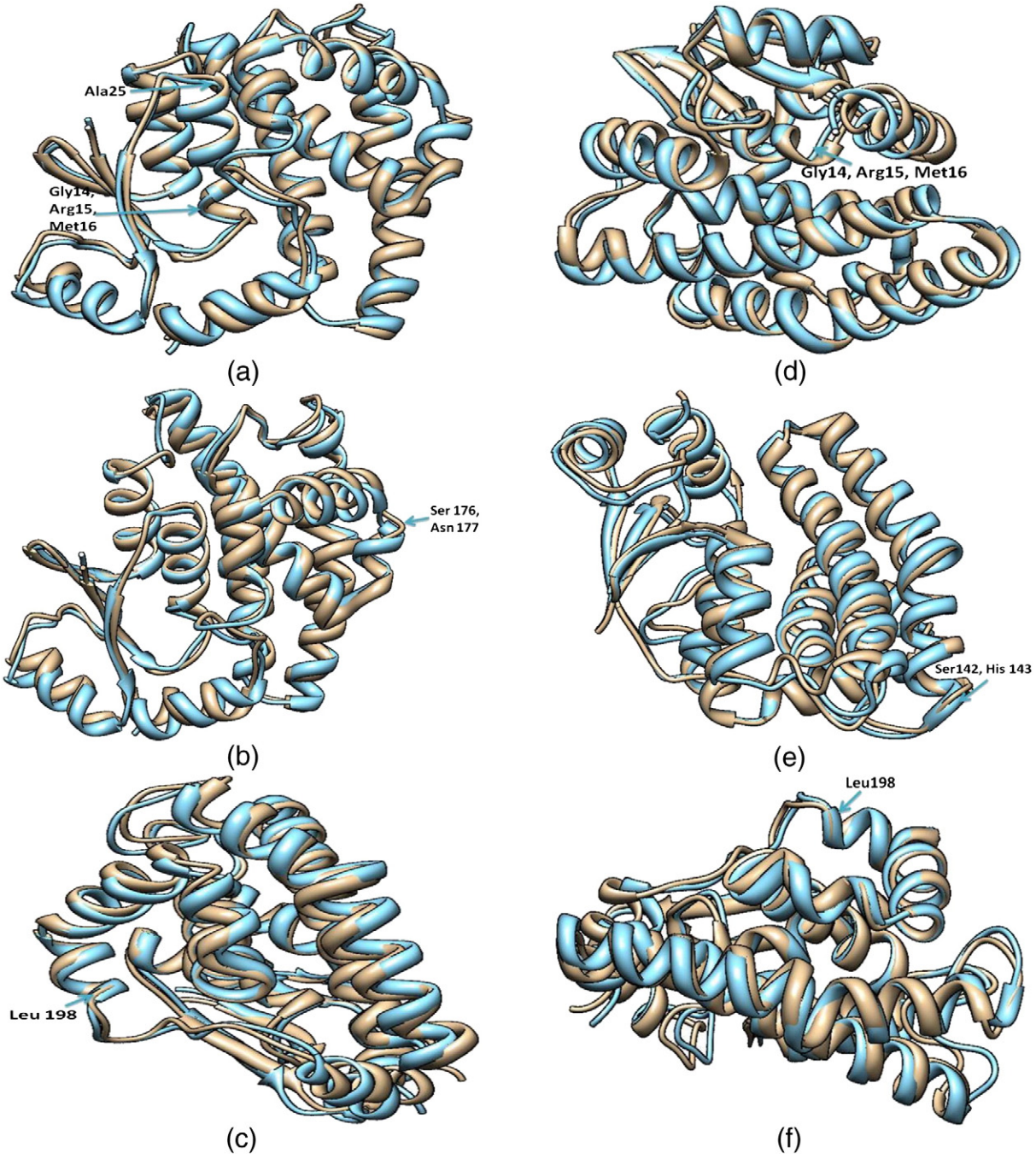

Fig. 1. Superimposed structure of native protein 1TDI (camel) with mutant protein 1TDI_R13W (Carolina blue) showing changes in secondary structure (a, b and c). Superimposed structure of native protein 1TDI (camel) with mutant protein 1TDI_Y147D (Carolina blue) showing changes in secondary structure (d, e and f).

and the mutant 1TDI_R13W and 1TDI_Y147D is 1.535 Å and 1.368 Å respectively; between the native type (3GTU) and the mutant 3GTU_R191L is 1.321 Å, between the native type (4GTU) and the mutant 4GTU_R18L is 0.924 Å and between the native type (2C3N) and the mutant 2C3N_W101R is 0.937 Å. The higher the RMSD value the more will be the deviation between the native and mutant type structures, which in turn changes their functional activity. Comparative structure analysis of wild and mutant proteins revealed the occurrence of a secondary structure and protein folding alteration due to SAP. Ser176, Asn177, and Leu198 have mutated from a loop to helix and Gly14, Arg15, Met 16, and Ala25 from a helix to loop in mutant protein 1TDI_R13W due to the SAP of arginine to tryptophan at position 13 (Fig. 1a, b & c). Likewise Ser142, His143, Ser176, Asn177, and Leu198 in mutant protein 1TDI_Y147D changed to a helix from loop and Gly14, Arg15, and Met16 to a loop from helix as tyrosine changed to aspartic acid at position 147 (Fig. 1d, e & f). In mutant 4GTU_R18L, Arg11 changed from a loop to helix and Glu171, Pro172 from a helix to loop due to the SAP of Arginine to leucine at position 18 (Fig. 2a, b). However one amino acid of mutant 3GTU_R191L, Glu195 (Fig. 2c) and four amino acids of mutant 2C3N_W101R, Gln39, His40, Leu41, and Gln102 (Fig. 2d & e) changed from a loop to helix due to the SAP of arginine to leucine at position 191 in 3GTU and tryptophan to arginine at position 101 in 2C3N,

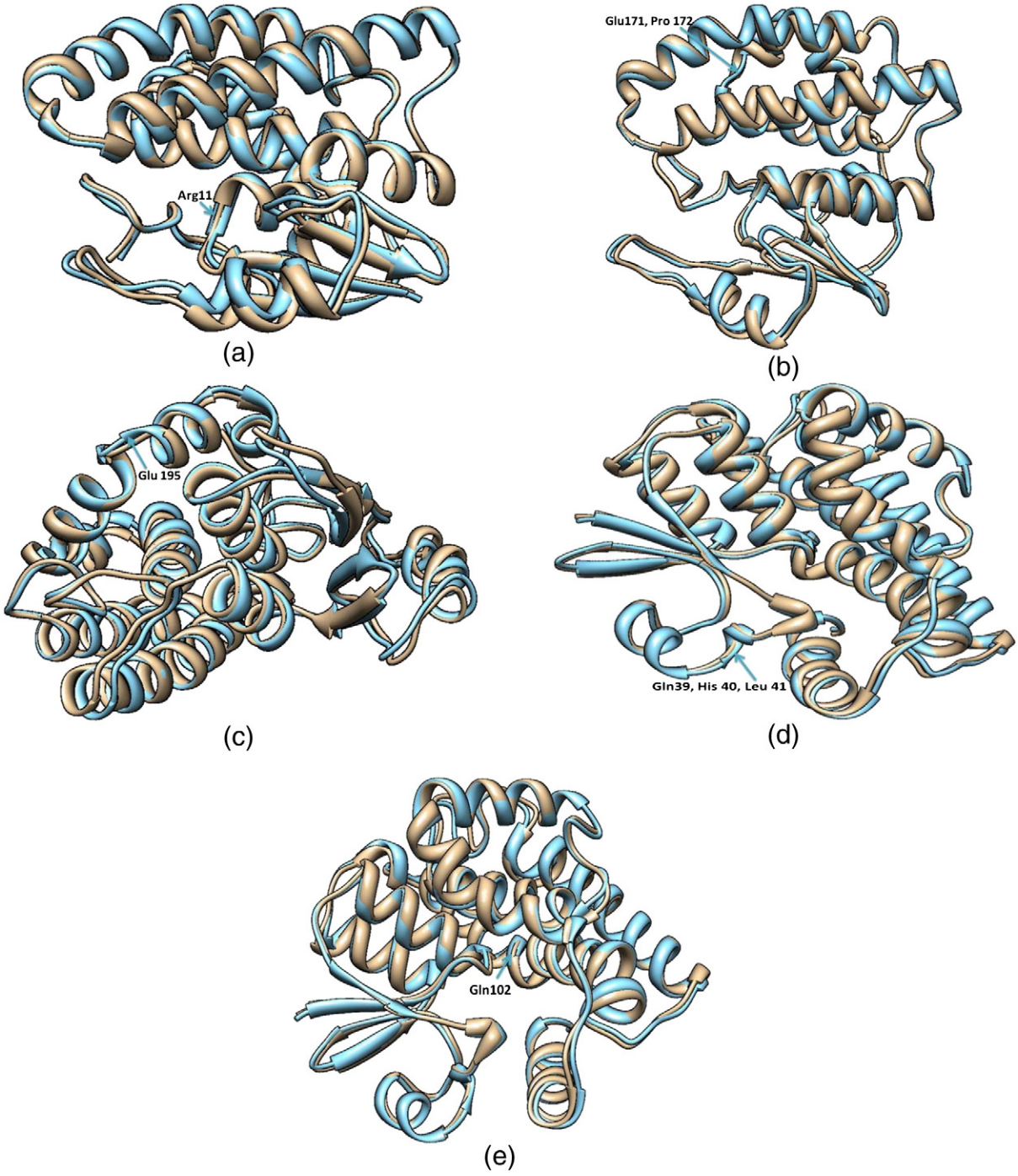

Fig. 2. Superimposed structure of native protein 4GTU (camel) with mutant protein 4GTU_R18L (Carolina blue) showing changes in secondary structure (a and b). Superimposed structure of native protein 3GTU (camel) with mutant protein 3GTU_R191L (Carolina blue) showing changes in secondary structure (c). Superimposed structure of native protein 2C3N (camel) with mutant protein 2C3N_W101R (Carolina blue) showing changes in secondary structure (d and e).

respectively. These structural changes of mutant proteins indicate that there might be alterations in the binding affinity of these proteins with glutathione and other substrates which ultimately leads to aberrant carcinogens, drugs and xenobiotic metabolism.

Conclusions

Among all cytosolic GSTs, α, μ, π and θ classes are mainly found to be involved in the carcinogen detoxification process in addition to kinase regulation and drug resistance [10,23]. However, the alteration of enzymes present in these classes by polymorphism might explain individual differences in susceptibility to cancer that exposed to the same type of carcinogens and can be influenced the pharmacokinetics of clinically-important drugs, but this is still limited to a small fraction of nsSNPs identified. In our study, we investigated deleterious nsSNPs which have functional influences on four major cytosolic classes of Glutathione transferases through computational method. As a result four genes were found to have five highly deleterious mutations viz. GSTA3 (R13W and Y147D), GSTM4 (R18L), GSTM3 (R191L) andGSTT1(W101R) with a PSIC difference score ≥3.00 and TI 0.00 and the mutant protein structure showed alteration in their structure, energy and high RMSD, which indicates their high divergence from one another. Finally we conclude that these five deleterious polymorphisms could be the prime target mutation for the altered detoxification process of their respective enzymes which ultimately leads to carcinogenesis event. Therefore our analysis will provide useful information in selecting amino acid substitutions which are supposed to increase susceptibility to certain diseases including cancer by altering xenobiotic, carcinogens and drug metabolism for further genotype–phenotype studies using molecular approaches.

References

[1] F.S. Collins, L.S. Brooks, A. Chakravarti, A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 8 (1998) 1229–1231.
[2] J.E. Lee, J.H. Choi, J.H. Lee, M.G. Lee, Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. Mutat. Res. 573 (1–2) (2005) 195–204.
[3] T. Emahazion, L. Feuk, M. Jobs, S.L. Sawyer, D. Fredman, D. St Clair, J.A. Prince, A.J. Brookes, SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. Trends Genet. 17 (7) (2001) 407–413.
[4] N.J. Schork, D. Fallin, J.S. Lanchbury, Single nucleotide polymorphisms and the future of genetic epidemiology. Clin. Genet. 58 (4) (2000) 250–264.
[5] R.N. Armstrong, Structure, catalytic mechanism, and evolution of the glutathione transferases. Chem. Res. Toxicol. 10 (1997) 2–18.
[6] E. Laborde, Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death. Cell Death Differ. 17 (2010) 1373–1380.
[7] S.G. Cho, Y.H. Lee, H.S. Park, K. Ryoo, K.W. Kang, J. Park, S.J. Eom, M.J. Kim, T.S. Chang, S.Y. Choi, J. Shim, Y. Kim, M.S. Dong, M.J. Lee, S.G. Kim, H. Ichijo, E.J. Choi, Glutathione S-transferase mu modulates the stress-activated signals by suppressing apoptosis signal-regulating kinase 1. J. Biol. Chem. 276 (16) (2001) 12749–12755.
[8] C.J. Henderson, K.J. Ritchie, A. McLaren, P. Chakravarty, C.R. Wolf, Increased skin papilloma formation in mice lacking glutathione transferase GSTP. Cancer Res. 71 (22) (2011) 7048–7060.
[9] R.J. Hung, O. van der Hel, S.V. Tavtigian, P. Brennan, P. Boffetta, M. Hashibe, Perspectives on the molecular epidemiology of aerodigestive tract cancers. Mutat. Res. 592 (2005) 102–118.
[10] C.C. McIlwain, D.M. Townsend, K.D. Tew, Glutathione S-transferase polymorphisms: cancer incidence and therapy. Oncogene 25 (2006) 1639–1648.
[11] J.D. Hayes, R.C. Strange, Glutathione S-transferase polymorphism and their biological consequences. Pharmacology 61 (3) (2000) 154–166.
[12] S.T. Sherry, M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29 (1) (2001) 308–311.
[13] C. Ferrer-Costa, M. Orozco, X. de la Cruz, Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J. Mol. Biol. 315 (2002) 771–786.
[14] P.C. Ng, S. Henikoff, SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 31 (2003) 3812–3814.
[15] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (1997) 3389–3402.
[16] K. Sjolander, K. Karplus, M. Brown, R. Hughey, A. Krogh, I.S. Mian, D. Haussler, Dirichlet mixtures: a method for improved detection of weak but significant protein sequence homology. Comput. Appl. Biosci. 12 (1996) 327–345.
[17] T. Xi, I.M. Jones, H.W. Mohrenweiser, Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics 83 (2004) 970–979.
[18] M. Prokop, J. Damborsky, J. Koca, TRITON: in silico construction of protein mutants and prediction of their activities. Bioinformatics 16 (2000) 845–846.
[19] E. Lindahl, C. Azuara, P. Koehl, M. Delarue, NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res. 34 (2006) 52–56.
[20] M. Delarue, P. Dumas, On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proc. Natl. Acad. Sci. 101 (18) (2004) 6957–6962.
[21] M. Mistri, P.M. Tamhankar, F. Sheth, D. Sanghavi, P. Kondurkar, et al., Identification of novel mutations in HEXA gene in children affected with Tay Sachs disease from India. PLoS ONE 7 (2012) e39122, http://dx.doi.org/10.1371/journal.pone.0039122 (PubMed: 22723944).
[22] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera — a visualization system for exploratory research and analysis. J. Comput. Chem. 25 (13) (2004) 1605–1612.
[23] K.D. Tew, Glutathione-associated enzymes in anticancer drug resistance. Cancer Res. 54 (1994) 4313–4320.