

Review Article

Analyzing Effects of Naturally Occurring Missense Mutations

Zhe Zhang,^{1,2} Maria A. Miteva,² Lin Wang,¹ and Emil Alexov¹

¹ Computational Biophysics and Bioinformatics, Department of Physics and Astronomy, Clemson University, SC 29634, USA

² Université Paris Diderot, Sorbonne Paris Cité, Molécules Thérapeutiques In Silico, Inserm UMR-S 973, 35 rue Helene Brion, 75013 Paris, France

Correspondence should be addressed to Emil Alexov, ealexov@clemson.edu

Received 21 December 2011; Revised 1 February 2012; Accepted 1 February 2012

Academic Editor: Gabriela Mustata Wilson

Copyright © 2012 Zhe Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Single-point mutation in genome, for example, single-nucleotide polymorphism (SNP) or rare genetic mutation, is the change of a single nucleotide for another in the genome sequence. Some of them will produce an amino acid substitution in the corresponding protein sequence (missense mutations); others will not. This paper focuses on genetic mutations resulting in a change in the amino acid sequence of the corresponding protein and how to assess their effects on protein wild-type characteristics. The existing methods and approaches for predicting the effects of mutation on protein stability, structure, and dynamics are outlined and discussed with respect to their underlying principles. Available resources, either as stand-alone applications or web servers, are pointed out as well. It is emphasized that understanding the molecular mechanisms behind these effects due to these missense mutations is of critical importance for detecting disease-causing mutations. The paper provides several examples of the application of 3D structure-based methods to model the effects of protein stability and protein-protein interactions caused by missense mutations as well.

1. Introduction

Human DNA is not identical among individuals, and this causes natural differences among races and ethnic populations, and also among healthy individuals and individuals susceptible to disease. On the DNA sequence level, the differences could be large or small, the smallest being a difference in a single-nucleotide. If such a difference occurs in some fraction of the population, but not in a single case, the difference is termed single nucleotide polymorphism (SNP) [1, 2]. Some of the SNPs occur at the noncoding, while other SNPs happen in the coding regions [3]. The SNPs occurring at the noncoding region does not affect the gene product, that is, the protein sequence is not changed, such types of SNPs are termed silent mutations [4]. However, silent mutation could also be found in the coding region because each amino acid is coded by more than one codon. Thus, even the mutation changes the codon, though it is still possible that the protein sequence is not affected. However, a silent mutation still could affect the function of the cell by altering the gene's expression and regulation.

On the other side of the spectrum are nonsynonymous SNPs (nsSNPs), which cause changes in protein sequence. The most dramatic change is induced by the nonsense mutations, which result in a premature stop codon and produce truncated, usually nonfunctional proteins [4]. Missense mutation, on the other hand, is a change of a single amino acid into another. Such a mutation could be polymorphism if it is observed in significant fraction of the population, or it could be a rare missense mutation if found in an individual or small group of people, as, for example, in a family. In both cases, on protein level, these mutations are termed single-point mutations and they are the primary focus of this paper.

Missense mutations and, in general, nsSNPs were extensively investigated in the past to reveal their plausible effects on protein stability [5–10], protein-protein interactions [11], the characteristics of the active site [5, 6], and many others [12–20]. In parallel, significant efforts were invested to catalog naturally occurring genetic differences, those found in general population and presumably harmless as the SNP database [21–23] and those known to be disease associated as the Online Mendelian Inheritance in Man (OMIM) [24–26]. The OMIM includes the full-text description of

disease phenotypes and genes, mapping, molecular genetics, PubMed references, and many other features [26, 27]. OMIM is currently provided by the U.S. National Center for Biotechnology Information (NCBI) [28] and edited by Dr. Victor A. McKusick at John Hopkins University. By 2011, more than 21,000 entries including data and over 13,100 established gene loci and phenotypic descriptions are contained in OMIM Entrez database [28]. SNP database is used to organize and systematize the huge amount of information of gene sequencing. So far, several SNP databases have been developed such as the dbSNP database [21, 29, 30], the Human Genome Variation Database (HGVBbase) [22], the Human Gene Mutation Database (HGMD) [31], and the TopoSNP database [23, 32–35].

The existence of such databases combined with available biochemical data of the effects of single-point mutations on protein stability and interactions prompted the development of *in silico* methods to predict the effects of mutations on the wild-type characteristics of the corresponding proteins or assemblages. Currently, the approaches can be classified into several categories: first principle methods, which calculate the folding or binding free energy change based on detailed atomic models [36–46]; methods based on statistical potentials [47–55] and utilizing known protein structures in the Protein Data Bank [56]; methods using empirical potential combining both physical force fields and free parameters fitted with experimental data [57–65]; machine learning approaches, which are trained against known experimental databases, and then used to predict the effect of the newly found mutations [66–72].

2. Overview of Plausible Effects Induced by Genetic Differences

Genetic differences can potentially affect the function of the cell in a variety of ways, which can be broadly classified into several categories outlined below.

2.1. Active Sites, Reaction Kinetics, and the Reaction Parameters. If a mutation occurs in an active site, then it should be considered lethal since such substitution will affect critical components of the biological reaction, which, in turn, will alter the normal protein function [73, 74]. At the same time, the biochemical reaction is very sensitive to the precise geometry of the active sites for both of the reactants and products; therefore, any conformational change altering the active sites will also affect the biochemical reaction; however, conservative mutations are not expected to perturb protein function by much. Thus, even if the mutation does not occur at the active site, but quite close to it, the characteristics of the catalytic groups will be perturbed [5, 6, 75]. In such a case, the mutation may not completely abolish the biochemical reaction but can change the kinetics of the reaction [76]. Moreover, the biochemical reaction strongly relies on a particular (optimum) cellular environment such as pH, salt concentration, and temperature. Thus in the living cells, the proteins' behavior is controlled by these cellular environments [77, 78]. Changing the reaction rate,

the pH, or salt and temperature dependencies away from the native parameters can lead to a malfunctioning protein. The isoelectric point (pI) is a very important parameter that refers to the pH at which the net charge of the protein is zero. Recently, it was demonstrated that five missense mutations involving charged groups in the sodium iodide symporter (NIS) gene, which generates a protein called iodide transporter and is associated with iodide transport defect, can cause an obvious pI shift and influence the electrostatic interactions in the trans-membrane domains of the NIS protein. Even more, these substitutions will probably, in turn, affect the protein stability, protein trafficking, and iodide transport activity [79].

2.2. Kinetics of Protein Folding, Protein Stability, Flexibility, and Aggregation. Protein folding is the process of converting the linear unfolded polypeptide into the native 3D structure driven by the gradient of potential energy [80, 81]. The importance of kinetics of protein folding is manifested by the fact that protein miss-folding is involved in many diseases [12]. An amino acid substitution at a critical folding position can prevent the forming of the folding nucleus, which makes the remainder of the structure rapidly condense [12]. Protein stability is also a key characteristic of a functional protein [5, 6, 71, 82–86], and as such, a mutation on a native protein amino acid can considerably affect its stability [76, 77, 87] through perturbing conformational constraints (e.g., substituting a small side chain residue to a large one and vice versa, resulting in backbone strain or overpacking) or physicochemical effects (substitutions between hydrophilic residues and hydrophobic residues, burial of charged residues, the disruption of hydrogen bonds, loss of hydrogen bonds, of S–S bonds) [88]. It was shown that 80% of missense mutations associated with disease are amino acid substitutions that affect the stability of proteins by several kcal/mol [84]. In addition, the missense mutation can also alter the protein flexibility [5, 89, 90]. When a protein is carrying its function, frequently the reaction requires a small or large conformational change to occur that is specific for the particular biochemical reaction. Thus, if a mutation makes the protein more rigid or flexible compared to the native structure, then it will affect the protein's function [91, 92]. Additionally, conformational flexibility is the main mechanism affecting protein aggregation propensity [93], thus the influence on protein flexibility could cause protein aggregation and formation of fibrils [94].

2.3. Interactions between Protein-Protein, Protein-DNA, Protein-RNA, and Protein-Membrane. If a missense mutation occurs at hot-spots of the binding interface that are crucial in contributing to the interaction [95, 96], then the binding affinity would be dramatically affected due to geometrical constraints and/or energetic effects [7, 97]. For instance, when substituting a small side chain for a bulky side chain in a narrow binding pocket, the entrance of the partner group will be blocked and the binding process will be completely or partially prevented [6, 98–101]. Similarly, a mutation at the protein-DNA interface can affect DNA

regulation [13–15]. A mutation occurring at the protein-membrane interface can affect the signal processes across the membrane, protein association with the membrane, and function of various channels and pumps [16, 17].

2.4. Subcellular Localization and Protein Expression. Subcellular localization is a very important factor, which provides a specific environment for protein function, protein interactions, protein activity in signaling pathways, and many other features. Transporting a protein to the correct compartment allows it to form the necessary wild-type interactions with its biological partners and take part in the corresponding biological networks like signaling and metabolic pathways. Otherwise, mislocalizing the protein in a wrong subcellular compartment will have harmful effects on the other proteins which function there [20]. Typically, a mutation affecting the subcellular localization is a mutation that occurs at a signaling region. For example, missense mutations in Otopetrin 1 affects the subcellular location and causes nonsyndromic otoconia agenesis and a subsequent balance defect in mice [102]. Fanconi anemia is a genetic disease associated with the missense mutations in FANCA protein. These missense mutations affect the subcellular localization of the FANCA protein and make it unable to relocate to the nucleus and activate the FA/BRCA pathway [103].

Protein expression is a subcomponent of gene expression and commonly used to denote the measurement of the protein concentration in a particular cell or tissue. Missense mutations can affect DNA-transcription factors resulting in altering the expression of the corresponding protein. Altering the wild-type protein expression in the compartment where it is designed to function will disrupt the normal cell cycle and in turn may cause diseases [20]. Recently, functional analysis of pancreatitis-associated missense mutations was performed in the pancreatic secretory trypsin inhibitor (SPINK1) gene, which encodes pancreatic secretory trypsin inhibitor (PSTI). It was shown that one of the disease-causing missense mutations R65Q reduced protein expression by almost 60%, and four other pathogenic missense mutations G48E, D50E, Y54H, and R67C caused complete or almost complete loss of PSTI expression [104]. By excluding the possibility that reduced transcription or unstable mRNA can lead to reduced protein expression, it was surmised that these disease-causing missense mutations probably cause intracellular retention of their respective mutant proteins. This is suggestive of a potential unifying pathological mechanism underlying both the signal peptide and mature peptide mutations [104].

In this section, we presented the plausible effects which mutations can cause. In fact, mutations often affect the normal protein function by the combined molecular effects listed above [5, 6, 8]. For instance, in the studies of genotype-phenotype correlations of TGFBI (transforming growth factor, beta-induced) mutations, it was shown that a missense mutation V613G strongly destabilizes the wild-type protein keratoprotein by 3.1 kcal/mol; additionally, the same mutation might also result in an improper folding due to the backbone structure of the substituted gly is

not restricted by the presence of a side chain, thus can adopt any conformation and lead to a misfolded protein. At the same time, it was shown that V613G also facilitates formation of beta-sheet structure of TGFBI which is known to favor amyloid formation [105]. Similarly, another study performed *in silico* investigation on 18 missense mutations in electron transfer flavoprotein (ETF) associated with multiple acyl-coa dehydrogenase deficiency (MADD), and it was found that these 18 missense mutations can be classified into two groups by their molecular effects: altering protein folding and assembling, affecting the catalytic activity of functional sites, and disrupting interactions with their biological partner, that is, dehydrogenases in this case [106].

3. Methods and Approaches to Predict the Effects of Mutations

Current efforts in this field are aimed at predicting the deleterious mutations since such predictions can be used for diagnostics and drug design. The features used to make such predictions can be classified into three categories: (a) amino acids properties, such as size, side chain polarity, side chain flexibility, and its ability to form a hydrogen bond and other geometrical considerations; (b) 3D protein structural properties such as protein stability, affinity of receptor-ligand complex, and structural flexibility; (c) evolutionary properties like sequence conservation and phylogenetic trees. It is almost impossible to review these approaches one by one since most of the current methodology is using a combination of these features [27]. Table 1 shows several examples for application of molecular modeling methods, free of charge for academia, to study the molecular mechanisms of missense mutations affecting wild-type properties of proteins. Comparison of their performance is provided in references [65, 107]. In the following paragraphs, we explain in detail some of the available resources.

It is essential to identify the most informative features among the features mentioned above for making successful predictions. Such a necessity inspired several works among which a recent study evaluating 32 features using their mutual information together with the functional effects of the amino acid substitutions, as measured by *in vivo* assays. Sequentially, a greedy algorithm was performed to identify a subset of highly informative features [108]. Finally, it was concluded that two features describing the solvent accessibility of “wild-type” and “mutant” amino-acid residues and another feature of evolutionary properties based on superfamily-level multiple alignments produce the best accuracy [109]. Another investigation developed a formalism and a computational method based on a structural model and phylogenetic information to indicate the effects of amino acid substitution on protein functions. With such a protocol, approximately 26%–32% of naturally occurring missense mutations were predicted to affect the protein functions [110].

The amino acid properties are often considered an important characteristic, which could play a crucial role in protein folding, stability, interaction of protein-protein

TABLE 1

Methods	Short Summary	Examples (references)*	Some tools based on this method
Molecular dynamics	The trajectories of molecules are determined at atomic level by numerically solving the Newton's equation of motion	(i) Thrombosis-related R2-FV haplotype: D2194G, Coagulation Factor V, domain C2 [8] (ii) Parahemophilia, Factor V new brunswick: A221V, Coagulation Factor V, domain A [9] (iii) FPLD, R482W; Lamin A/C [159] (iv) Intellectual Disability: H101Q; CLIC2 protein [10] (v) Snyder-Robin syndrome: G56S, V132G, I150T; spermine synthase; [5]	Eris [112, 132, 133] Tinker [158] GROMACS [160]
Molecular mechanics	Using molecular mechanics force field and optimization to model molecular systems	(i) 21-Hydroxylase-Deficiency: R132C, R149C, M283V, E431K; CYP450; C21 [161] (ii) Cancer: A159V, A161V, N235I, N239Y, T256I, S269I; p53 [162] (iii) Intellectual Disability: H101Q; CLIC2 protein; [10] (iv) Mutability of human spermine synthase: all amino acids substitution at disease associated missense mutation sites G56, V132, and I150; human spermine synthase [6] (v) Studying effects of nsSNPs on protein-protein interactions: nsSNPs in OMIM and non-OMIM; 264 protein-protein complexes with known nsSNPs located at the interface; [11]	FoldX [63, 64]
Monte Carlo simulation	Applying Monte Carlo sampling to predict preferred conformational states	(i) Noonan syndrome: D61Y, Tyrosine phosphatase SHP-2 [163]	IMC [164]
Electrostatic calculation	Calculating electrostatics energy and pKa/ionized states changes/electrostatic stability upon the missense mutations	(i) Snyder-Robinson Syndrome; G56S, V132G, I150T human spermine synthase [5] (ii) Thrombosis-related R2-FV haplotype: D2194G, Coagulation Factor V, domain C2 [8] (iii) Noonan syndrome: D61Y, Tyrosine phosphatase SHP-2 [163] (iv) Studying effects of nsSNPs on protein-protein interactions: nsSNPs in OMIM and non-OMIM; 264 protein-protein complexes with known nsSNPs located at the interface; [11]	DelPhi [165] MCCE [166–168] pKD [169]
Evolutionary properties	Based on structure and sequence analysis, for example, highly conserved residues in a protein family	(i) Homocystinuria: 204 mutations; cystathionine beta synthase; [170]	SNPs3D [138] PolyPhen [86]
Machine learning	learn the behavior of a system based on training datasets	(i) Snyder-Robinson Syndrome: G56S, V132G, I150T; human spermine synthase; [5] (ii) Gastrointestinal stromal tumors: 19 mutations; KIT receptor [171]	I-Mutant 2.0/3.0 [71, 72, 134]
Graph methods	A branch of discrete mathematics. In protein science, this method is used to analyze the topological details of proteins with known structure	(i) Cancer: Y220C, R273H, R273C, R282W, and G245S; p53 protein; [172] (ii) Predicting the structural effects of nsSNPs: 506 disease-associated nsSNPs; [173]	Bongo [173]

TABLE 1: Continued.

Methods	Short Summary	Examples (references)*	Some tools based on this method
Statistical Potential	Based on the knowledge of statistical mechanics such as inverse Boltzmann law, $\Delta G = -kT \log [g_{ij}(\mathbf{r})]$	(i) Human X-linked Agammaglobulinemia (XLA): 16 missense mutations; Bruton's tyrosine kinase (Btk); [174, 175] (ii) Severe form of phenylketonuria: G46S; human phenylalanine hydroxylase (hPAH); [176]	DFIRE [55, 177, 178] PoPMuSiC-2.0 [179, 180] CUPSAT [181–183]
The BellKor collaborative filtering (CF) algorithm	Model relations of the known data points and the parameters of the model are learnt by the training database	(i) Using the known $\Delta\Delta G$ value to predict the $\Delta\Delta G$ value of other missense mutations at the same substitution site; 4803 mutants were used; [184]	Pro-Maya [184]

complexes, and protein function, although sometimes they may be misleading [11]. The side chain properties such as volume, polarity, acidity, basicity, conformational flexibility and the ability to form a hydrogen-bond and salt bridge, are distinguishable. Therefore, the compatibility of a substitution at the dominant allele could be used to make the prediction as it was done in a recent study [111], which combined amino acid properties and structural information to identify deleterious mutations by analyzing the effects on protein stability.

An alternative approach to assess the effect of mutation on protein stability is to evaluate the change of folding free energy $\Delta G(\text{folding})$. The difference between $\Delta G(\text{folding})$ of the wild-type protein and the mutant, typically described as $\Delta\Delta G(\text{folding})$, is a measure of the effect of mutation on protein stability [5, 6, 64, 65, 107, 112]. If the change in $\Delta\Delta G(\text{folding})$ is negative, then the prediction is that the mutation will destabilize the protein. In contrast, if the calculated change is positive, the mutation is expected to stabilize the protein. The same considerations are valid in the case of predicting the effect on receptor-ligand binding [113]. Numerous investigations were reported in the past to reveal the change in the stability of the native structure [71, 82–86], the macromolecular interactions [11], or altering the wild-type (WT) hydrogen-bond network, in terms of affecting the stability [5, 114, 115]. Currently, several distinctive approaches to predict the protein stability and affinity changes due to mutations have been developed and they can be classified into four categories: (a) first principle methods that using the detailed atomic models to calculate the folding/binding free energy changes caused by mutations [36–46]—these approaches are scientifically sound, but are quite computationally expensive and may not be the best choice in the cases of large sets of mutations [116]; (b) methods based on the statistical potentials [47, 48] were shown to be successful in predicting the change of protein stability upon the mutations [49–55]; (c) Methods utilizing empirical potential, combining both physical force fields, and free parameters fitted with experimental data [57–62]; (d) machine learning methods, utilizing a training database [66–70].

The 3D structure of proteins can be used not only for energy calculations, as described above, but to map mutations onto it and to use geometrical considerations

to predict the effects of mutations [117]. Recently, such an approach, the alpha-shape method from computational geometry, was used to divide all nsSNP sites into three categories: (a) Type P: nsSNPs located in a pocket or a void; (b) Type S: nsSNPs occurred on a convex region; and (c) Type I: nsSNP sites are completely buried inside the protein. It was found that 88% of pathogenic nsSNPs are of type P and rarely of type I [32]. Along the same line, 3D structures were used in combinations with machine learning (SVM) and random forest methods. It was demonstrated that these methods outperformed the SIFT algorithm developed by Ng and Henikoff [118], and was indicated that incorporating structural information is crucial to make an accurate prediction if no sufficient evolutionary information is available [119]. Based on the 3D structures, the solvent-accessibility term is also an important feature, which is often used for investigating the effects of missense mutations. It has been shown that using a solvent-accessibility term, the $C\beta$ density, and a score derived from homologous sequences will make the most accurate prediction [120]. Recent studies took into account several protein structural parameters such as solvent accessibility, location within beta strands, or active sites to predict the effects on nsSNPs. It was found that approximately 70% of the disease-associated mutations are buried and solvent inaccessible [121–124] and that such mutations have strong effects on protein structure, folding, stability, and normal function [121, 123].

Another important feature reviewed here is evolutionary properties. Among homologous proteins, the highly conserved residues are generally considered to be critical for protein stability, interaction, and function. One of the evolutionary approaches, which assumes that residues located at a highly conserved position are most likely crucial, is to extract conservation scores from a multiple sequence alignment of homologous proteins. Another widely used computational technique is named the “evolutionary trace” method [125–127]. It uses phylogenetic information based on homologous sequences to rank residues according to evolutionary importance based on their conserved residues in the protein family. After that, such evolutionary conserved residues are mapped on the representative structure. In addition, a group of conserved residues could occur at the interface of a protein-protein complex. Based on the extraction of functionally important residues, an approach

was developed utilizing the evolutionary trace method to identify active sites and functional interfaces of proteins based on their available structures. The method was tested on SH2 and SH3 modular signaling domains and the DNA binding domain of the hormone receptors. It was demonstrated this method can delineate the functional epitope and identify the essential residues for binding specificity [128].

In order to train machine learning algorithms properly and to have a benchmarking case, appropriate databases are required. A particular example is the Catalytic Site Atlas database [129], which collects 177 original hand-annotated entries and 2608 homologous entries and covers about 30% of all enzyme in the Protein Data Bank [56]. At the same time, the computational methods of predicting the functional residues were also well-developed [125, 130, 131]. A selection of structure and sequence-based features was used to indicate an amino acid polymorphism effect on protein function, and it was found that ~26%–32% of the naturally occurring nsSNPs will affect the protein's function [110].

4. Webservers for Analyzing the Effects of Mutations

In past years, several methods were implemented into webservers to predict the effects on protein stability due to mutations. The Eris webserver is based on Medusa force field [132], and it was benchmarked on 595 mutants with available experimental data resulting in RMSD 2.4 kcal/mol between the predicted $\Delta\Delta G_{cal}$ (folding) and corresponding experimental values ($\Delta\Delta G_{exp}$ (folding)) [112, 132, 133]. The FoldX is perhaps the most popular web server [63] for predicting the folding free energy changes due to the mutations, and it is based on the empirical potentials [64]. The I-Mutant 2.0/3.0 is Support Vector Machine-based (SVM: a machine learning method) webserver utilizing the 3D structural or sequential information to predict protein stability change upon single-point mutations [71, 72, 134]. Other webservers include the Site Directed Mutator (SDM) [54] and the Mupro method [135].

In parallel, there are webservers predicting the effects of mutations on protein-protein interaction. The COILCHECK is an interactive webserver, which measures the strength of interactions between two helices involved in coiled coil structures utilizing nonbonded and electrostatic interactions and the presence of hydrogen bonds and salt bridges. It can be used to assess the strength of coiled coil regions, to recognize weak and strong regions, to rationalize the phenotypic behavior of single mutations and to design mutation experiments [136]. Recently, DrugScore^{PPI} was reported, which is a fast and accurate computational approach to predict impacts on binding affinity by the change of the binding free energy upon alanine mutations at protein-protein interfaces. The primary motivation of developing this webserver is to identify hotspot residues at protein-protein interfaces, which will guide both biological experiments and the development of protein-protein interaction modulators [137].

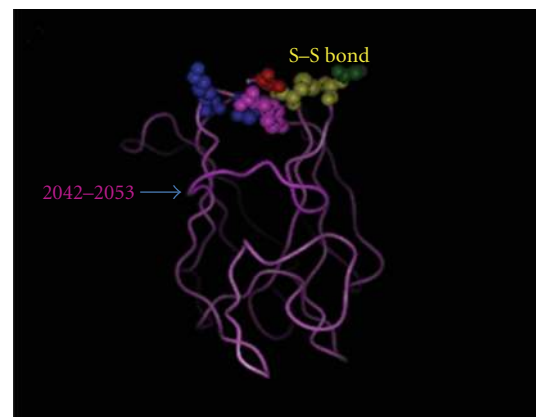
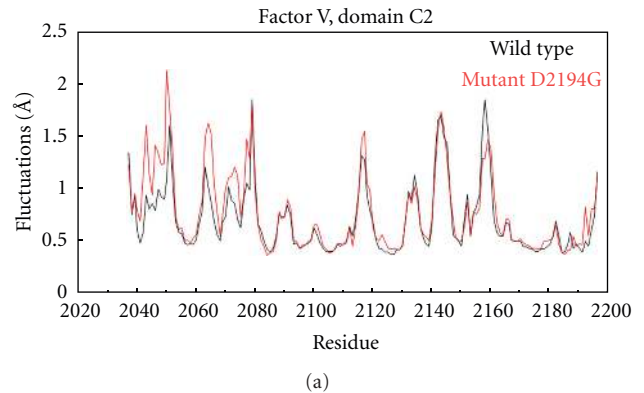
There are many webservers which are designed to predict if the mutation is pathogenic or not without providing information about the magnitude of expected energy changes. The SNPs3D [138] is a primary resource and database, which provides various disease/gene relationships at the molecular level. This server has three modules: (a) identifying the gene candidates involved in a specific disease; (b) relationships between the sets of candidate genes; and (c) analyzing the possible effects of nsSNPs on normal protein function. It is very convenient for the users to quickly obtain the available information and so develop models of gene-pathway-disease interaction. Another online predictor of molecular and structural effects of protein-coding variants was recently developed, the SNPeffect 4.0 [139]. It uses sequence- and structure-based bioinformatics tools such as aggregation prediction (TANGO) [140], amyloid prediction (WALTZ) [141], chaperone-binding prediction (LIMBO) [142], and protein stability analysis (FoldX) [63] to predict the effect of SNPs. In addition, it also contains the information of effects on catalytic sites, posttranslational modifications, and all known human protein variants from Uniprot. At the same time, SNPeffect allows users to submit custom protein variants for analyzing the SNP effects and plot correlations between phenotypic features for a user-selected set of variants [139]. The dbSNP database in NCBI lists over 9 million SNPs in the human genome but includes very limited annotation information. To fill this gap, the LS-SNP was developed to annotate the nsSNPs [83]. It can map nsSNPs onto protein sequences, functional pathways, and comparative protein structure models and predicts the positions where nsSNPs cause the effects. The results can be used to find out the functional SNP candidates within a gene, haplotype, or pathway, and also in understanding the molecular mechanisms responsible for functional effects of nsSNPs [83]. At the same time, a protocol based on Sorting Intolerant From Tolerant (SIFT) [118] was reported to predict if a missense mutation will affect the protein function. To assess the effects of a missense mutation, SIFT utilizes evolutionary properties of the protein and considers the substitutions at the conserved positions which may affect protein function. Thus, SIFT makes a prediction on effects of all possible substitutions at each position in the protein sequence by using sequence homology [143]. The Polyphen (Polymorphism Phenotype) is a tool that predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. It combines a variety of features such as sequences, evolutionary properties, and structural information to predict if an nsSNP will affect the protein function and performs optimally if the structural information is available. More than 11000 nsSNPs are annotated by this webserver [86]. A new version of Polyphen, namely, Polyphen-2, was recently released [144]. Its features include high quality multiple sequence alignment pipeline and probabilistic classifier based on a machine-learning method, and it is optimized for high-throughput analysis of the next-generation sequencing data [144]. After the development of SIFT and Polyphen, the Parepro (Predicating the amino acid replacement probability) was created, based

on two independent databases HumVar and NewHumVar. The predictions are if an nsSNP will be either deleterious or will have no effect on protein function. Compared to SIFT and Polyphen, Parepro achieved a higher Matthews correlation coefficient (MCC) and overall accuracy (Q2) when predictions were made using a 20-fold cross validation test on the HumVar dataset [145]. StSNP is a webserver referencing the data from dbSNP in NCBI, the gene and protein database from Entrez, the protein structures from the PDB, and pathway information from KEGG and makes an effort to provide combined, integrated reports about nsSNPs. Researchers can use the metabolic pathways in StSNP to examine the likely relationship between the disease-related pathways and particular nsSNPs, and link the disease with the current available molecular structure data [146]. AUTOMUTE is a knowledge-based computational mutagenesis used to predict the disease potential of human nsSNPs. In this study, 1790 neutral and disease-associated human nsSNPs on 243 diverse human protein structures were used. With a trained model, this method achieves 76% cross-validation accuracy [147].

5. Application of Structure-Based Methods to Predict the Effects of Mutations on Protein Stability and Protein-Protein Interactions

In this section, we outline several examples of utilizing structural information to predict the effects of mutations on wild-type characteristics of proteins and protein complexes.

5.1. Application of Molecular Dynamics (MD) Simulation for Predicting the Effects of Mutations. Coagulation factor V (FV) is the precursor of an essential procoagulant cofactor that accelerates FXa-catalyzed prothrombin activation in the coagulation system. It is a large glycoprotein containing several domains, A1-A2-B-A3-C1-C2 [8]. A missense mutation D2194G in its C2 domain was shown to cause low expression level and to have plausible effect on stability of the corresponding protein. To investigate the molecular mechanism of D2194G affecting the wild type of the corresponding protein, MD simulations were carried out on both of the WT and mutant structure to reveal the flexibility change upon this mutation [8]. The program CHARMM [148] was used, and the total simulation time was 900 ps. The root mean square fluctuations (RMSFs) for the α -carbon atoms of the C2 domain per residue were calculated for series of snapshots. The comparison for the WT and mutant structures is shown in Figure 1(a). It was concluded that the regions 2075–2085 and 2140–2150 in both WT structure and mutant structures are flexible. The loop 2042–2053 (Figure 1(b)) which is close to the mutation site, is more flexible in the mutant structure. At the same time, loop 2060–2067 became more flexible in the mutant as well, and this effect was attributed to the increased mobility of the loop 2042–2053. The substitution of Asp for Gly will lead to a big cavity and the nonflexible C-terminus (Tyr2196) inserts itself into the domain and attempts to fill out this cavity and to compensate for the missing negative charge of the mutant.



(b)

FIGURE 1: (a) The structural and flexibility differences between the simulated WT and mutant structures. The black line represents the RMSF of the WT structure and the red line represents the mutant protein. (b) 3D structure of the C2 domain of the WT FV. The S-S bond is marked in yellow and the loop 2042–2053 is indicated by the arrow.



FIGURE 2: 3D structure of human SMS with three missense mutation sites. Two subunits were represented by ribbon in cyan and magenta. Three mutation sites were shown with sphere representation: G56S in orange, V132G in white and I150T in green. The substrates of SPD and MTA were shown in red sticks and blue sticks, respectively.

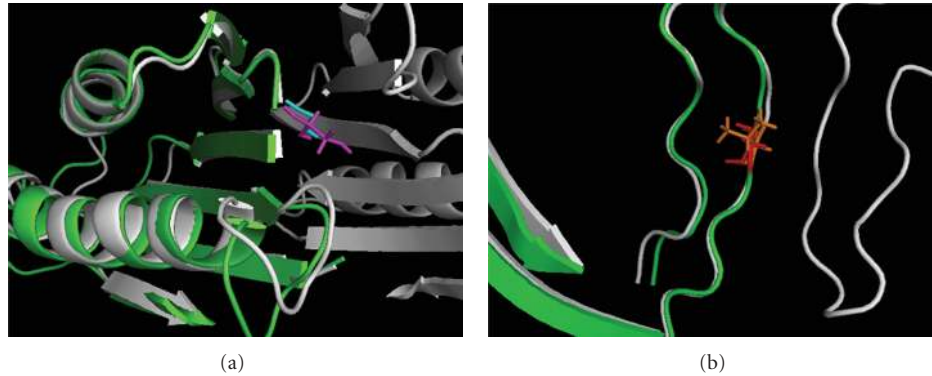


FIGURE 3: Effects on dimerization. (a) G56S: we superimposed WT structure (presented with two chains in white) and mutant structure (presented with Only one chain in green). Cyan stick represented Gly in the WT structure and magenta stick represented Ser in the mutant structure; (b) V132G: Only the region around the mutation site was shown in the figure. We superimpose the WT structure (presented with two chains in white) and mutant (presented with only one chain in green). The orange stick represented Val in the WT structure and red stick represented Gly in the mutant structure.

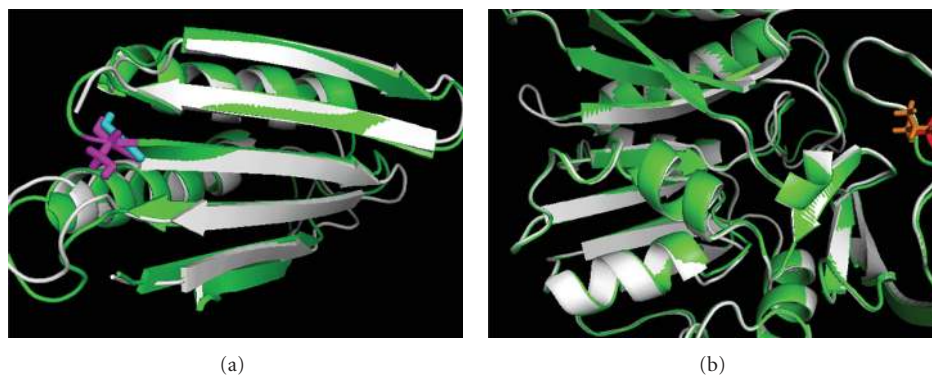


FIGURE 4: Effects on monomer stability. (a) G56S: N-terminal domain of both WT monomer (white) and mutant monomer (green) are superimposed. Cyan stick represented Gly in the WT structure and magenta represented Ser in the mutant structure; (b) V132G: C-terminal domain of both WT monomer (white) and mutant monomer (green) are superimposed. We use stick and ball representation in orange to represent Val in the WT structure and in red to represent Gly in the mutant structure.

These events could be the reason for the enhancing flexibility of the loop 2042–2053.

5.2. Application of Energy Calculation for Predicting the Effects of Mutations in Human Spermine Synthase. In this section, we describe the molecular mechanism of three missense mutations in human spermine synthase (SMS) causing Snyder-Robin Syndrome (SRS) [149–151] to demonstrate application of structure-based methods and energy calculation to predict the effects on protein stability and protein-protein interaction [5].

SMS (OMIM: 300105) is an enzyme converting spermidine (SPD) into spermine (SPM) both of which are two polyamines controlling normal mammalian cell growth and development [152–155]. The importance of SMS for the normal function is illustrated by the fact that three clinical missense mutations, c.267G > A (p.G56S) [150], c.496T > G (p.V132G) [151], and I150T [5], on SMS will cause an X-Linked mental retardation disorder named SRS (OMIM: 309583). At the same time, the 3D structures of

human SMS with either the substrates SPD or product SPM have been experimentally determined [156]. The 3D structure of SMS with the substrates SPD and product MTA (PDB ID: 3C6K) is shown in Figure 2. SMS contains two subunits forming a dimer, and each subunit includes two terminal domains: the N-terminal domain which plays a key role in dimerization and the C-terminal domain which includes the active site. The importance of dimerization for SMS function was also demonstrated by series of deletion experiments *in vitro* [156]. Additionally, two missense mutations G56S and V132G are located at the dimer interface, while the other missense mutation I150T occurred at the C-terminal domain and quite close to the active sites.

These three mutants were made *in silico* by SCAP, a program in JACKAL package [157], based on the native 3D SMS structure. Then, the TINKER package was used to perform the energy minimization and calculation [158]. It was shown that the missense mutation G56S will strongly decrease the dimer affinity by nearly 14 kcal/mol, but the other two have no impact on it. With the analysis based

on the 3D structure, it was concluded the reason that G56S strongly decreases the dimerization is because the side chain of Ser in the mutant is pointing to the dimer interface, and there is not enough room to harbor this side chain (Figure 3(a)). In contrast, while the mutation V132G is located at the dimer interface as well, the side chain of Val in the native structure does not point towards the interface and is close to large cavity (Figure 3(b)); thus this substitution can be accommodated easily without introducing any steric constraints. The third mutation, I150T, is very far away from the dimer interface (Figure 2); thus it is not supposed to affect the dimerization.

With regards to the folding energy calculation, all these missense mutations are predicted to destabilize the protein monomer by 2.8 kcal/mol (G56S), 1.1 kcal/mol (V132G), and 3.5 kcal/mol (I150T), respectively. Figure 4(a) gives the comparison of the native structure and mutant structure and is zoomed into the mutation site G56S. In the mutant structure, we can see this mutation occurs in a sharp turn, and the substitution with almost any other amino acid will introduce strain. Figure 4(b) shows the superposition of the native structure and mutant, zoomed in the mutation site V132G. It is clear that the side chain of Val points to the interior, thus the substitution with Gly will leave a big cavity inside the monomer, which in turn will affect the stability. In addition, considering the physicochemical property feature, Val and Gly have different hydrophobicity. The destabilization by I150T is mainly attributed to the totally different physicochemical properties between Ile, which is a hydrophobic residue, and Thr, which is a hydrophilic residue.

Thus, combining the 3D structure, physicochemical properties of amino acids, and energy calculations, it was shown that one can successfully predict molecular effects due to these three missense mutations. Such an analysis helps better understand how these missense mutations affect the SMS function and in turn reveal the molecular origin of SRS.

6. Conclusion

In this paper, we outlined the current state-of-the-art methods in the field of computational modeling of effects of nsSNPs and rare missense mutations. Available resources are pointed out along with short description of their functionality and accuracy. The basic concepts and major research directions are described and their advantages and disadvantages discussed.

Acknowledgments

The authors thank Nicholas Smith for proof reading the paper. The work of Z. Zhang and E. Alexov was supported in part by NIH, NLM grant, Grant no. 1R03LM009748. Z. Zhang thanks “Chateaubriand Fellowship,” which is supported by the Embassy of France in the USA.

References

- [1] P. Taillon-Miller, Z. Gu, Q. Li, L. Hillier, and P. Y. Kwok, “Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms,” *Genome Research*, vol. 8, no. 7, pp. 748–754, 1998.
- [2] S. Mooney, “Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis,” *Briefings in Bioinformatics*, vol. 6, no. 1, pp. 44–56, 2005.
- [3] M. Hagmann, “Human genome. A good SNP may be hard to find,” *Science*, vol. 285, no. 5424, pp. 21–22, 1999.
- [4] T. Strachan and A. P. Read, *Human Molecular Genetics*, 1999.
- [5] Z. Zhang, S. Teng, L. Wang, C. E. Schwartz, and E. Alexov, “Computational analysis of missense mutations causing Snyder-Robinson syndrome,” *Human Mutation*, vol. 31, no. 9, pp. 1043–1049, 2010.
- [6] Z. Zhang, J. Norris, C. Schwartz, and E. Alexov, “*In silico* and *in vitro* investigations of the mutability of disease-causing missense mutation sites in spermine synthase,” *PLoS One*, vol. 6, no. 5, Article ID e20373, 2011.
- [7] S. Akhavan, M. A. Miteva, B. O. Villoutreix et al., “A critical role for Gly25 in the B chain of human thrombin,” *Journal of Thrombosis and Haemostasis*, vol. 3, no. 1, pp. 139–145, 2005.
- [8] M. A. Miteva, J. M. Brugge, J. Rosing, G. A. F. Nicolaes, and B. O. Villoutreix, “Theoretical and experimental study of the D2194G mutation in the C2 domain of coagulation factor V,” *Biophysical Journal*, vol. 86, no. 1, part 1, pp. 488–498, 2004.
- [9] M. Steen, M. Miteva, B. O. Villoutreix, T. Yamazaki, and B. Dahlback, “Factor V new brunswick: Ala221Val associated with FV deficiency reproduced *in vitro* and functionally characterized,” *Blood*, vol. 102, no. 4, pp. 1316–1322, 2003.
- [10] S. Witham, K. Takano, C. Schwartz, and E. Alexov, “A missense mutation in *CLIC2* associated with intellectual disability is predicted by *in silico* modeling to affect protein stability and dynamics,” *Proteins*, vol. 79, no. 8, pp. 2444–2454, 2011.
- [11] S. Teng, T. Madej, A. Panchenko, and E. Alexov, “Modeling effects of human single nucleotide polymorphisms on protein-protein interactions,” *Biophysical Journal*, vol. 96, no. 6, pp. 2178–2188, 2009.
- [12] C. M. Dobson, “Protein folding and misfolding,” *Nature*, vol. 426, no. 6968, pp. 884–890, 2003.
- [13] R. N. Venkatesan, P. M. Treuting, E. D. Fuller et al., “Mutation at the polymerase active site of mouse DNA polymerase δ increases genomic instability and accelerates tumorigenesis,” *Molecular and Cellular Biology*, vol. 27, no. 21, pp. 7669–7682, 2007.
- [14] L. M. S. Elles and O. C. Uhlenbeck, “Mutation of the arginine finger in the active site of *Escherichia coli* DbpA abolishes ATPase and helicase activity and confers a dominant slow growth phenotype,” *Nucleic Acids Research*, vol. 36, no. 1, pp. 41–50, 2008.
- [15] J. D. Wright and C. Lim, “Mechanism of DNA-binding loss upon single-point mutation in p53,” *Journal of Biosciences*, vol. 32, no. 5, pp. 827–839, 2007.
- [16] L. G. Kwa, D. Wegmann, B. Brugger, F. T. Wieland, G. Wanner, and P. Braun, “Mutation of a single residue, β -glutamate-20, alters protein-lipid interactions of light harvesting complex II,” *Molecular Microbiology*, vol. 67, no. 1, pp. 63–77, 2008.
- [17] Y. Kariya, Y. Tsubota, T. Hirotsuki et al., “Differential regulation of cellular adhesion and migration by recombinant laminin-5 forms with partial deletion or mutation within the

- G3 domain of $\alpha 3$ chain,” *Journal of Cellular Biochemistry*, vol. 88, no. 3, pp. 506–520, 2003.
- [18] S. Tiede, M. Cantz, J. Spranger, and T. Bräulke, “Missense mutation in the N-acetylglucosamine-1-phosphotransferase gene (GNPTA) in a patient with mucopolisidosis II induces changes in the size and cellular distribution of GNPTG,” *Human Mutation*, vol. 27, no. 8, pp. 830–831, 2006.
- [19] M. Krumbholz, K. Koehler, and A. Huebner, “Cellular localization of 17 natural mutant variants of ALADIN protein in triple A syndrome—shedding light on an unexpected splice mutation,” *Biochemistry and Cell Biology*, vol. 84, no. 2, pp. 243–249, 2006.
- [20] C. O. Hanemann, D. D’Urso, A. A. W. M. Gabreels-Festen, and H. W. Muller, “Mutation-dependent alteration in cellular distribution of peripheral myelin protein 22 in nerve biopsies from Charcot-Marie-Tooth type 1A,” *Brain*, vol. 123, no. 5, pp. 1001–1006, 2000.
- [21] S. T. Sherry, M. H. Ward, M. Kholodov et al., “dbSNP: the NCBI database of genetic variation,” *Nucleic Acids Research*, vol. 29, no. 1, pp. 308–311, 2001.
- [22] D. Fredman, G. Munns, D. Rios et al., “HGvbase: a curated resource describing human DNA variation and phenotype relationships,” *Nucleic Acids Research*, vol. 32, pp. D516–D519, 2004.
- [23] N. O. Stitzel, T. A. Binkowski, Y. Y. Tseng, S. Kasif, and J. Liang, “topoSNP: a topographic database of non-synonymous single nucleotide polymorphisms with and without known disease association,” *Nucleic Acids Research*, vol. 32, no. Database issue, pp. D520–D522, 2004.
- [24] A. Hamosh, A. F. Scott, J. Amberger, D. Valle, and V. A. McKusick, “Online mendelian inheritance in man (OMIM),” *Human Mutation*, vol. 15, no. 1, pp. 57–61, 2000.
- [25] A. Hamosh, A. F. Scott, J. Amberger, C. Bocchini, D. Valle, and V. A. McKusick, “Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders,” *Nucleic Acids Research*, vol. 30, no. 1, pp. 52–55, 2002.
- [26] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini, and V. A. McKusick, “Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders,” *Nucleic Acids Research*, vol. 33, pp. D514–D517, 2005.
- [27] S. Teng, E. Michonova-Alexova, and E. Alexov, “Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions,” *Current Pharmaceutical Biotechnology*, vol. 9, no. 2, pp. 123–133, 2008.
- [28] E. W. Sayers, T. Barrett, D. A. Benson et al., “Database resources of the national center for biotechnology information,” *Nucleic Acids Research*, vol. 39, supplement 1, no. Database issue, pp. D38–D51, 2011.
- [29] E. M. Smigielski, K. Sirotkin, M. Ward, and S. T. Sherry, “dbSNP: a database of single nucleotide polymorphisms,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 352–355, 2000.
- [30] S. T. Sherry, M. Ward, and K. Sirotkin, “dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation,” *Genome Research*, vol. 9, no. 8, pp. 677–679, 1999.
- [31] D. N. Cooper, E. V. Ball, and M. Krawczak, “The human gene mutation database,” *Nucleic Acids Research*, vol. 26, no. 1, pp. 285–287, 1998.
- [32] N. O. Stitzel, Y. Y. Tseng, D. Pervouchine, D. Goddeau, S. Kasif, and J. Liang, “Structural location of disease-associated single-nucleotide polymorphisms,” *Journal of Molecular Biology*, vol. 327, no. 5, pp. 1021–1030, 2003.
- [33] R. B. Altman, “PharmGKB: a logical home for knowledge relating genotype to drug response phenotype,” *Nature Genetics*, vol. 39, no. 4, article 426, 2007.
- [34] B. R. Packer, M. Yeager, L. Burdett et al., “SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes,” *Nucleic acids research.*, vol. 34, pp. D617–D621, 2006.
- [35] J. Bidwell, L. Keen, G. Gallagher et al., “Cytokine gene polymorphism in human disease: on-line databases,” *Genes and Immunity*, vol. 1, no. 1, pp. 3–19, 1999.
- [36] P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman, “Free energy calculations by computer simulation,” *Science*, vol. 236, no. 4801, pp. 564–568, 1987.
- [37] Y. Duan and P. A. Kollman, “Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution,” *Science*, vol. 282, no. 5389, pp. 740–744, 1998.
- [38] S. D. Khare, M. Caplow, and N. V. Dokholyan, “FALS mutations in Cu, Zn superoxide dismutase destabilize the dimer and increase dimer dissociation propensity: a large-scale thermodynamic analysis,” *Amyloid*, vol. 13, no. 4, pp. 226–235, 2006.
- [39] B. Kuhlman and D. Baker, “Native protein sequences are close to optimal for their structures,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 19, pp. 10383–10388, 2000.
- [40] C. Lee, “Testing homology modeling on mutant proteins: predicting structural and thermodynamic effects in the Ala → Val mutants of T4 lysozyme,” *Folding and Design*, vol. 1, no. 1, pp. 1–12, 1996.
- [41] C. Lee and M. Levitt, “Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core,” *Nature*, vol. 352, no. 6334, pp. 448–451, 1991.
- [42] S. Miyazawa and R. L. Jernigan, “Protein stability for single substitution mutants and the extent of local compactness in the denatured state,” *Protein Engineering*, vol. 7, no. 10, pp. 1209–1220, 1994.
- [43] J. W. Pitera and P. A. Kollman, “Exhaustive mutagenesis *in silico*: multicoordinate free energy calculations on proteins and peptides,” *Proteins*, vol. 41, no. 3, pp. 385–397, 2000.
- [44] M. Prevost, S. J. Wodak, B. Tidor, and M. Karplus, “Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96 → Ala mutation in barnase,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 88, no. 23, pp. 10880–10884, 1991.
- [45] B. Tidor, “Simulation analysis of the stability mutant R96h of T4 lysozyme,” *Biochemistry*, vol. 30, no. 13, pp. 3217–3228, 1991.
- [46] Y. N. Vorobjev and J. Hermans, “ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model,” *Biophysical Chemistry*, vol. 78, no. 1-2, pp. 195–205, 1999.
- [47] A. Ben-Naim, “Statistical potentials extracted from protein structures: are these meaningful potentials?” *Journal of Chemical Physics*, vol. 107, no. 9, pp. 3698–3706, 1997.
- [48] P. D. Thomas and K. A. Dill, “Statistical potentials extracted from protein structures: how accurate are they?” *Journal of Molecular Biology*, vol. 257, no. 2, pp. 457–469, 1996.
- [49] D. Gilis and M. Rooman, “Stability changes upon mutation of solvent-accessible residues in proteins evaluated by

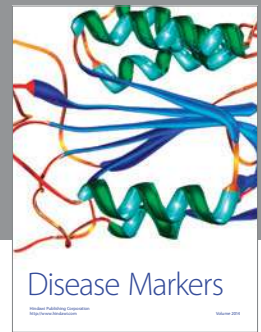
- database-derived potentials," *Journal of Molecular Biology*, vol. 257, no. 5, pp. 1112–1126, 1996.
- [50] D. Gilis and M. Rooman, "Predicting protein stability changes upon mutation using database-derived potentials: solvent accessibility determines the importance of local versus non-local interactions along the sequence," *Journal of Molecular Biology*, vol. 272, no. 2, pp. 276–290, 1997.
- [51] D. Gilis and M. Rooman, "PoPMuSiC, an algorithm for predicting protein mutant stability changes. Application to prion proteins," *Protein Engineering*, vol. 13, no. 12, pp. 849–856, 2000.
- [52] C. Hoppe and D. Schomburg, "Prediction of protein thermostability with a direction- and distance-dependent knowledge-based potential," *Protein Science*, vol. 14, no. 10, pp. 2682–2692, 2005.
- [53] M. Ota, Y. Isogai, and K. Nishikawa, "Knowledge-based potential defined for a rotamer library to design protein sequences," *Protein Engineering*, vol. 14, no. 8, pp. 557–564, 2001.
- [54] C. M. Topham, N. Srinivasan, and T. L. Blundell, "Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables," *Protein Engineering*, vol. 10, no. 1, pp. 7–21, 1997.
- [55] H. Zhou and Y. Zhou, "Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction," *Protein Science*, vol. 11, no. 11, pp. 2714–2726, 2002.
- [56] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.
- [57] A. J. Bordner and R. A. Abagyan, "Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations," *Proteins*, vol. 57, no. 2, pp. 400–413, 2004.
- [58] H. Domingues, J. Peters, K. H. Schneider et al., "Improving the refolding yield of interleukin-4 through the optimization of local interactions," *Journal of Biotechnology*, vol. 84, no. 3, pp. 217–230, 2000.
- [59] V. Munoz and L. Serrano, "Development of the multiple sequence approximation within the AGADIR Model of α -helix formation: comparison with zimm-bragg and lifson-roig formalisms," *Biopolymers*, vol. 41, no. 5, pp. 495–509, 1997.
- [60] K. Takano, M. Ota, K. Ogasahara, Y. Yamagata, K. Nishikawa, and K. Yutani, "Experimental verification of the "stability profile of mutant protein" (SPMP) data using mutant human lysozymes," *Protein Engineering*, vol. 12, no. 8, pp. 663–672, 1999.
- [61] D. Verma, D. J. Jacobs, and D. R. Livesay, "Predicting the melting point of human C-type lysozyme mutants," *Current Protein and Peptide Science*, vol. 11, no. 7, pp. 562–572, 2010.
- [62] V. Villegas, A. R. Viguera, F. X. Aviles, and L. Serrano, "Stabilization of proteins by rational design of α -helix stability using helix/coil transition theory," *Folding and Design*, vol. 1, no. 1, pp. 29–34, 1996.
- [63] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, "The FoldX web server: an online force field," *Nucleic Acids Research*, vol. 33, no. 2, pp. W382–W388, 2005.
- [64] R. Guerois, J. E. Nielsen, and L. Serrano, "Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations," *Journal of Molecular Biology*, vol. 320, no. 2, pp. 369–387, 2002.
- [65] Z. Zhang, L. Wang, Y. Gao, J. Zhang, M. Zhenirovskyy, and E. Alexov, "Predicting folding free energy changes upon single point mutations," *Bioinformatics*, vol. 28, no. 5, pp. 664–671, 2012.
- [66] E. Capriotti, P. Fariselli, and R. Casadio, "A neural-network-based method for predicting protein stability changes upon single point mutations," *Bioinformatics*, vol. 20, supplement 1, pp. i63–i68, 2004.
- [67] R. Casadio, M. Compiani, P. Fariselli, and F. Vivarelli, "Predicting free energy contributions to the conformational stability of folded proteins from the residue sequence with radial basis function networks," *Proceedings International Conference on Intelligent Systems for Molecular Biology*, vol. 3, pp. 81–88, 1995.
- [68] C. M. Frenz, "Neural network-based prediction of mutation-induced protein stability changes in staphylococcal nuclease at 20 residue positions," *Proteins*, vol. 59, no. 2, pp. 147–151, 2005.
- [69] T. Joachims, *Learning to Classify Text Using Support Vector Machines*, Springer, 2002.
- [70] M. Masso and I. I. Vaisman, "Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis," *Bioinformatics*, vol. 24, no. 18, pp. 2002–2009, 2008.
- [71] E. Capriotti, P. Fariselli, R. Calabrese, and R. Casadio, "Predicting protein stability changes from sequences using support vector machines," *Bioinformatics*, vol. 21, supplement 2, pp. ii54–ii58, 2005.
- [72] E. Capriotti, P. Fariselli, and R. Casadio, "I-mutant2.0: predicting stability changes upon mutation from the protein sequence or structure," *Nucleic Acids Research*, vol. 33, no. 2, pp. W306–W310, 2005.
- [73] Y. Yamada, Y. Banno, H. Yoshida et al., "Catalytic inactivation of human phospholipase D2 by a naturally occurring Gly901Asp mutation," *Archives of Medical Research*, vol. 37, no. 6, pp. 696–699, 2006.
- [74] G. Stevanin, V. Hahn, E. Lohmann et al., "Mutation in the catalytic domain of protein kinase C γ and extension of the phenotype associated with spinocerebellar ataxia type 14," *Archives of Neurology*, vol. 61, no. 8, pp. 1242–1248, 2004.
- [75] O. Takamiya, M. Seta, K. Tanaka, and F. Ishida, "Human factor VII deficiency caused by S339C mutation located adjacent to the specificity pocket of the catalytic domain," *Clinical and Laboratory Haematology*, vol. 24, no. 4, pp. 233–238, 2002.
- [76] S. B. Koukouritaki, M. T. Poch, M. C. Henderson et al., "Identification and functional analysis of common human flavin-containing monooxygenase 3 genetic variants," *Journal of Pharmacology and Experimental Therapeutics*, vol. 320, no. 1, pp. 266–273, 2007.
- [77] R. de Cristofaro, A. Carotti, S. Akhavan et al., "The natural mutation by deletion of Lys9 in the thrombin A-chain affects the PK_a value of catalytic residues, the overall enzyme's stability and conformational transitions linked to Na⁺ binding," *The FEBS Journal*, vol. 273, no. 1, pp. 159–169, 2006.
- [78] E. Alexov, "Numerical calculations of the pH of maximal protein stability: the effect of the sequence composition and three-dimensional structure," *European Journal of Biochemistry*, vol. 271, no. 1, pp. 173–185, 2004.
- [79] H. Fujiwara, K. I. Tatsumi, S. Tanaka, M. Kimura, O. Nose, and N. Amino, "A novel V59E missense mutation in the sodium iodide symporter gene in a family with iodide transport defect," *Thyroid*, vol. 10, no. 6, pp. 471–474, 2000.
- [80] K. A. Dill, K. M. Fiebig, and H. S. Chan, "Cooperativity in protein-folding kinetics," *Proceedings of the National*

- Academy of Sciences of the United States of America*, vol. 90, no. 5, pp. 1942–1946, 1993.
- [81] K. A. Dill, S. B. Ozkan, T. R. Weikl, J. D. Chodera, and V. A. Voelz, “The protein folding problem: when will it be solved?” *Current Opinion in Structural Biology*, vol. 17, no. 3, pp. 342–346, 2007.
- [82] Y. Ye, Z. Li, and A. Godzik, “Modeling and analyzing three-dimensional structures of human disease proteins,” *Pacific Symposium on Biocomputing*, pp. 439–450, 2006.
- [83] R. Karchin, M. Diekhans, L. Kelly et al., “LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources,” *Bioinformatics*, vol. 21, no. 12, pp. 2814–2820, 2005.
- [84] Z. Wang and J. Moulton, “SNPs, protein structure, and disease,” *Human Mutation*, vol. 17, no. 4, pp. 263–270, 2001.
- [85] Z. Wang and J. Moulton, “Three-dimensional structural location and molecular functional effects of missense SNPs in the T cell receptor V β domain,” *Proteins*, vol. 53, no. 3, pp. 748–757, 2003.
- [86] V. Ramensky, P. Bork, and S. Sunyaev, “Human non-synonymous SNPs: server and survey,” *Nucleic Acids Research*, vol. 30, no. 17, pp. 3894–3900, 2002.
- [87] H. Ode, S. Matsuyama, M. Hata et al., “Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease,” *Journal of Molecular Biology*, vol. 370, no. 3, pp. 598–607, 2007.
- [88] B. A. Shirley, P. Stanssens, U. Hahn, and C. N. Pace, “Contribution of hydrogen bonding to the conformational stability of ribonuclease T1,” *Biochemistry*, vol. 31, no. 3, pp. 725–732, 1992.
- [89] M. Karplus and J. Kuriyan, “Molecular dynamics and protein function,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 19, pp. 6679–6685, 2005.
- [90] M. A. Young, S. Gonfloni, G. Superti-Furga, B. Roux, and J. Kuriyan, “Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation,” *Cell*, vol. 105, no. 1, pp. 115–126, 2001.
- [91] K. E. S. Tang and K. A. Dill, “Native protein fluctuations: the conformational-motion temperature and the inverse correlation of protein flexibility with protein stability,” *Journal of Biomolecular Structure and Dynamics*, vol. 16, no. 2, pp. 397–411, 1998.
- [92] E. S. Song, A. Daily, M. G. Fried, M. A. Juliano, L. Juliano, and L. B. Hersch, “Mutation of active site residues of insulin-degrading enzyme alters allosteric interactions,” *The Journal of Biological Chemistry*, vol. 280, no. 18, pp. 17701–17706, 2005.
- [93] M. Valerio, A. Colosimo, F. Conti et al., “Early events in protein aggregation: molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations,” *Proteins*, vol. 58, no. 1, pp. 110–118, 2005.
- [94] P. G. Board, K. Pierce, and M. Coggan, “Expression of functional coagulation factor XIII in *Escherichia coli*,” *Thrombosis and Haemostasis*, vol. 63, no. 2, pp. 235–240, 1990.
- [95] S. E. A. Ozbabacan, A. Gursoy, O. Keskin, and R. Nussinov, “Conformational ensembles, signal transduction and residue hot spots: application to drug discovery,” *Current Opinion in Drug Discovery and Development*, vol. 13, no. 5, pp. 527–537, 2010.
- [96] A. Dixit, A. Torkamani, N. J. Schork, and G. Verkhivker, “Computational modeling of structurally conserved cancer mutations in the RET and MET kinases: the impact on protein structure, dynamics, and stability,” *Biophysical Journal*, vol. 96, no. 3, pp. 858–874, 2009.
- [97] R. Jones, M. Ruas, F. Gregory et al., “A CDKN2A mutation in familial melanoma that abrogates binding of p16 INK4a to CDK4 but not CDK6,” *Cancer Research*, vol. 67, no. 19, pp. 9134–9141, 2007.
- [98] T. R. Rignall, J. O. Baker, S. L. McCarter et al., “Effect of single active-site cleft mutation on product specificity in a thermostable bacterial cellulase,” *Applied Biochemistry and Biotechnology A*, vol. 98–100, pp. 383–394, 2002.
- [99] R. van Wijk, G. Rijksen, E. G. Huizinga, H. K. Nieuwenhuis, and W. W. van Solinge, “HK Utrecht: missense mutation in the active site of human hexokinase associated with hexokinase deficiency and severe nonspherocytic hemolytic anemia,” *Blood*, vol. 101, no. 1, pp. 345–347, 2003.
- [100] M. Hardt and R. A. Laine, “Mutation of active site residues in the chitin-binding domain ChBD_{ChIA1} from chitinase A1 of *Bacillus circulans* alters substrate specificity: use of a green fluorescent protein binding assay,” *Archives of Biochemistry and Biophysics*, vol. 426, no. 2, pp. 286–297, 2004.
- [101] M. A. Ortiz, J. Light, R. A. Maki, and N. Assa-Munt, “Mutation analysis of the pip interaction domain reveals critical residues for protein-protein interactions,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 6, pp. 2740–2745, 1999.
- [102] E. Kim, K. L. Hyrc, J. Speck et al., “Missense mutations in Otopetrin 1 affect subcellular localization and inhibition of purinergic signaling in vestibular supporting cells,” *Molecular and Cellular Neuroscience*, vol. 46, no. 3, pp. 655–661, 2011.
- [103] M. Castella, R. Pujol, E. Callen et al., “Origin, functional role, and clinical impact of fanconi anemia fanca mutations,” *Blood*, vol. 117, no. 14, pp. 3759–3769, 2011.
- [104] A. Boulling, C. le Marechal, P. Trouve, O. Ragueneas, J. M. Chen, and C. Ferec, “Functional analysis of pancreatitis-associated missense mutations in the pancreatic secretory trypsin inhibitor (*SPINK1*) gene,” *European Journal of Human Genetics*, vol. 15, no. 9, pp. 936–942, 2007.
- [105] F. Niel-Buttschi, B. Kantelip, J. Iwaszkiewicz et al., “Genotype-phenotype correlations of TGFBI p.Leu509Pro, p.Leu509Arg, p.Val613Gly, and the allelic association of p.Met502Val-p.Arg555Gln mutations,” *Molecular Vision*, vol. 17, pp. 1192–11202, 2011.
- [106] B. J. Henriques, P. Bross, and C. M. Gomes, “Mutational hotspots in electron transfer flavoprotein underlie defective folding and function in multiple acyl-CoA dehydrogenase deficiency,” *Biochimica et Biophysica Acta*, vol. 1802, no. 11, pp. 1070–1077, 2010.
- [107] S. Khan and M. Vihinen, “Performance of protein stability predictors,” *Human Mutation*, vol. 31, no. 6, pp. 675–684, 2010.
- [108] R. Battiti, “Using mutual information for selecting features in supervised neural net learning,” *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, 1994.
- [109] R. Karchin, L. Kelly, and A. Sali, “Improving functional annotation of non-synonymous SNPs with information theory,” *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pp. 397–408, 2005.
- [110] D. Chasman and R. M. Adams, “Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid

- variation,” *Journal of Molecular Biology*, vol. 307, no. 2, pp. 683–706, 2001.
- [111] P. Yue and J. Moulton, “Identification and analysis of deleterious human SNPs,” *Journal of Molecular Biology*, vol. 356, no. 5, pp. 1263–1274, 2006.
- [112] S. Yin, F. Ding, and N. V. Dokholyan, “Eris: an automated estimator of protein stability,” *Nature Methods*, vol. 4, no. 6, pp. 466–467, 2007.
- [113] J. Cai, L. Q. Cai, Y. Hong, and Y. S. Zhu, “Functional characterisation of a natural androgen receptor missense mutation (N771H) causing human androgen insensitivity syndrome,” submitted to *Andrologia*.
- [114] D. M. Hunt, J. W. Saldanha, J. F. Brennan et al., “Single nucleotide polymorphisms that cause structural changes in the cyclic AMP receptor protein transcriptional regulator of the tuberculosis vaccine strain *Mycobacterium bovis* BCG alter global gene expression without attenuating growth,” *Infection and Immunity*, vol. 76, no. 5, pp. 2227–2234, 2008.
- [115] P. Nicolao, M. Carella, B. Giometto et al., “Missense polymorphism in the human carboxypeptidase E gene alters enzymatic activity,” *Human Mutation*, vol. 18, no. 2, pp. 120–131, 2001.
- [116] P. A. Kollman, I. Massova, C. Reyes et al., “Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models,” *Accounts of Chemical Research*, vol. 33, no. 12, pp. 889–897, 2000.
- [117] A. Cavallo and A. C. R. Martin, “Mapping SNPs to protein sequence and structure data,” *Bioinformatics*, vol. 21, no. 8, pp. 1443–1450, 2005.
- [118] P. C. Ng and S. Henikoff, “SIFT: predicting amino acid changes that affect protein function,” *Nucleic Acids Research*, vol. 31, no. 13, pp. 3812–3814, 2003.
- [119] L. Bao and Y. Cui, “Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information,” *Bioinformatics*, vol. 21, no. 10, pp. 2185–2190, 2005.
- [120] C. T. Saunders and D. Baker, “Evaluation of structural and evolutionary contributions to deleterious mutation prediction,” *Journal of Molecular Biology*, vol. 322, no. 4, pp. 891–901, 2002.
- [121] S. Sunyaev, V. Ramensky, and P. Bork, “Towards a structural basis of human non-synonymous single nucleotide polymorphisms,” *Trends in Genetics*, vol. 16, no. 5, pp. 198–200, 2000.
- [122] S. R. Sunyaev, W. C. Lathe III, V. E. Ramensky, and P. Bork, “SNP frequencies in human genes: an excess of rare alleles and differing modes of selection,” *Trends in Genetics*, vol. 16, no. 8, pp. 335–337, 2000.
- [123] S. Sunyaev, V. Ramensky, I. Koch, W. Lathe III, A. S. Kondrashov, and P. Bork, “Prediction of deleterious human alleles,” *Human Molecular Genetics*, vol. 10, no. 6, pp. 591–597, 2001.
- [124] M. W. Dimmic, S. Sunyaev, and C. D. Bustamante, “Inferring SNP function using evolutionary, structural, and computational methods,” *Pacific Symposium on Biocomputing*, pp. 382–384, 2005.
- [125] R. Landgraf, I. Xenarios, and D. Eisenberg, “Three-dimensional cluster analysis identifies interfaces and functional residue clusters in proteins,” *Journal of Molecular Biology*, vol. 307, no. 5, pp. 1487–1502, 2001.
- [126] O. Lichtarge and M. E. Sowa, “Evolutionary predictions of binding surfaces and interactions,” *Current Opinion in Structural Biology*, vol. 12, no. 1, pp. 21–27, 2002.
- [127] C. A. Innis, J. Shi, and T. L. Blundell, “Evolutionary trace analysis of TGF- β and related growth factors: implications for site-directed mutagenesis,” *Protein Engineering*, vol. 13, no. 12, pp. 839–847, 2000.
- [128] O. Lichtarge, H. R. Bourne, and F. E. Cohen, “An evolutionary trace method defines binding surfaces common to protein families,” *Journal of Molecular Biology*, vol. 257, no. 2, pp. 342–358, 1996.
- [129] C. T. Porter, G. J. Bartlett, and J. M. Thornton, “The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data,” *Nucleic Acids Research*, vol. 32, pp. D129–D133, 2004.
- [130] V. Chelliah, L. Chen, T. L. Blundell, and S. C. Lovell, “Distinguishing structural and functional restraints in evolution in order to identify interaction sites,” *Journal of Molecular Biology*, vol. 342, no. 5, pp. 1487–1504, 2004.
- [131] F. Pazos and M. J. E. Sternberg, “Automated prediction of protein function and detection of functional sites from structure,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 41, pp. 14754–14759, 2004.
- [132] F. Ding and N. V. Dokholyan, “Emergence of protein fold families through rational design,” *PLoS Computational Biology*, vol. 2, no. 7, pp. 0725–0733, 2006.
- [133] S. Yin, F. Ding, and N. V. Dokholyan, “Modeling backbone flexibility improves protein stability estimation,” *Structure*, vol. 15, no. 12, pp. 1567–1576, 2007.
- [134] E. Capriotti, P. Fariselli, I. Rossi, and R. Casadio, “A three-state prediction of single point mutations on protein stability changes,” *BMC Bioinformatics*, vol. 9, supplement 2, article S6, 2008.
- [135] J. L. Cheng, A. Randall, and P. Baldi, “Prediction of protein stability changes for single-site mutations using support vector machines,” *Proteins*, vol. 62, no. 4, pp. 1125–1132, 2006.
- [136] V. Alva, D. P. Syamala Devi, and R. Sowdhagini, “COILCHECK: an interactive server for the analysis of interface regions in coiled coils,” *Protein and Peptide Letters*, vol. 15, no. 1, pp. 33–38, 2008.
- [137] D. M. Kruger and H. Gohlke, “DrugScorePPI webserver: fast and accurate *in silico* alanine scanning for scoring protein-protein interactions,” *Nucleic Acids Research*, vol. 38, supplement 2, pp. W480–W486, 2010.
- [138] P. Yue, E. Melamud, and J. Moulton, “SNPs3D: candidate gene and SNP selection for association studies,” *BMC Bioinformatics*, vol. 7, article no. 166, 2006.
- [139] G. de Baets, J. van Durme, J. Reumers et al., “Snpeffct 4.0: on-line prediction of molecular and structural effects of protein-coding variants,” *Nucleic Acids Research*, vol. 40, pp. D935–D939, 2012, submitted to *Nucleic Acids Research*.
- [140] A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano, “Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins,” *Nature Biotechnology*, vol. 22, no. 10, pp. 1302–1306, 2004.
- [141] S. Maurer-Stroh, M. Debulpaep, N. Kueemmerer et al., “Exploring the sequence determinants of amyloid structure using position-specific scoring matrices,” *Nature Methods*, vol. 7, no. 3, pp. 237–242, 2010.
- [142] J. van Durme, S. Maurer-Stroh, R. Gallardo, H. Wilkinson, F. Rousseau, and J. Schymkowitz, “Accurate prediction of DnaK-peptide binding via homology modelling and experimental data,” *PLoS Computational Biology*, vol. 5, no. 8, Article ID e1000475, 2009.
- [143] P. Kumar, S. Henikoff, and P. C. Ng, “Predicting the effects of coding non-synonymous variants on protein function using

- the SIFT algorithm,” *Nature Protocols*, vol. 4, no. 7, pp. 1073–1082, 2009.
- [144] I. A. Adzhubei, S. Schmidt, L. Peshkin et al., “A method and server for predicting damaging missense mutations,” *Nature Methods*, vol. 7, no. 4, pp. 248–249, 2010.
- [145] J. Tian, N. Wu, X. Guo, J. Guo, J. Zhang, and Y. Fan, “Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines,” *BMC Bioinformatics*, vol. 8, article 450, 2007.
- [146] A. Uzun, C. M. Leslin, A. Abyzov, and V. Ilyin, “Structure SNP (StSNP): a web server for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways,” *Nucleic Acids Research*, vol. 35, pp. W384–392, 2007.
- [147] M. Masso and I. I. Vaisman, “Knowledge-based computational mutagenesis for predicting the disease potential of human non-synonymous single nucleotide polymorphisms,” *Journal of Theoretical Biology*, vol. 266, no. 4, pp. 560–568, 2010.
- [148] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr. et al., “CHARMM: the biomolecular simulation program,” *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.
- [149] A. L. Cason, Y. Ikeguchi, C. Skinner et al., “X-linked spermine synthase gene (SMS) defect: the first polyamine deficiency syndrome,” *European Journal of Human Genetics*, vol. 11, no. 12, pp. 937–944, 2003.
- [150] G. de Alencastro, D. E. McCloskey, S. E. Kliemann et al., “New SMS mutation leads to a striking reduction in spermine synthase protein function and a severe form of Snyder-Robinson X-linked recessive mental retardation syndrome,” *Journal of Medical Genetics*, vol. 45, no. 8, pp. 539–543, 2008.
- [151] L. E. Becerra-Solano, J. Butler, G. Castaneda-Cisneros et al., “A missense mutation, p.V132G, in the X-linked spermine synthase gene (SMS) causes Snyder-Robinson syndrome,” *American Journal of Medical Genetics A*, vol. 149, no. 3, pp. 328–335, 2009.
- [152] E. W. Gerner and F. L. Meyskens Jr., “Polyamines and cancer: old molecules, new understanding,” *Nature Reviews Cancer*, vol. 4, no. 10, pp. 781–792, 2004.
- [153] Y. Ikeguchi, M. C. Bewley, and A. E. Pegg, “Aminopropyltransferases: function, structure and genetics,” *Journal of Biochemistry*, vol. 139, no. 1, pp. 1–9, 2006.
- [154] A. E. Pegg, “Mammalian polyamine metabolism and function,” *IUBMB Life*, vol. 61, no. 9, pp. 880–894, 2009.
- [155] D. Geerts, J. Koster, D. Albert et al., “The polyamine metabolism genes ornithine decarboxylase and antizyme 2 predict aggressive behavior in neuroblastomas with and without MYCN amplification,” *International Journal of Cancer*, vol. 126, no. 9, pp. 2012–2024, 2010.
- [156] H. Wu, J. Min, H. Zeng et al., “Crystal structure of human spermine synthase: implications of substrate binding and catalytic mechanism,” *The Journal of Biological Chemistry*, vol. 283, no. 23, pp. 16135–16146, 2008.
- [157] Z. Xiang and B. Honig, “Extending the accuracy limits of prediction for side-chain conformations,” *Journal of Molecular Biology*, vol. 311, no. 2, pp. 421–430, 2001.
- [158] J. W. Ponder, *Tinker-Software Tools for Molecular Design*, 1999.
- [159] V. Rajendran, R. Purohit, and R. Sethumadhavan, “*In silico* investigation of molecular mechanism of laminopathy caused by a point mutation (R482W) in lamin A/C protein,” submitted to *Amino Acids*.
- [160] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, “GROMACS: fast, flexible, and free,” *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.
- [161] C. Minutolo, A. D. Nadra, C. Fernandez et al., “Structure-based analysis of five novel disease-causing mutations in 21-hydroxylase-deficient patients,” *PLoS One*, vol. 6, no. 1, Article ID e15899, 2011.
- [162] Y. Tan and R. Luo, “Structural and functional implications of p53 missense cancer mutations,” *PMC Biophysics*, vol. 2, no. 1, article 5, 2009.
- [163] W. H. Lee, A. Raas-Rotschild, M. A. Miteva et al., “Noonan syndrome type I with PTPN11 3 bp deletion: structure-function implications,” *Proteins*, vol. 58, no. 1, pp. 7–13, 2005.
- [164] R. Abagyan and M. Totrov, “Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins,” *Journal of Molecular Biology*, vol. 235, no. 3, pp. 983–1002, 1994.
- [165] W. Rocchia, E. Alexov, and B. Honig, “Extending the applicability of the nonlinear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions,” *Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6507–6514, 2001.
- [166] E. G. Alexov and M. R. Gunner, “Calculated protein and proton motions coupled to electron transfer: electron transfer from Q_A⁻ to Q_B in bacterial photosynthetic reaction centers,” *Biochemistry*, vol. 38, no. 26, pp. 8253–8270, 1999.
- [167] R. E. Georgescu, E. G. Alexov, and M. R. Gunner, “Combining conformational flexibility and continuum electrostatics for calculating PK_as in proteins,” *Biophysical Journal*, vol. 83, no. 4, pp. 1731–1748, 2002.
- [168] Y. Song, J. Mao, and M. R. Gunner, “MCCE2: improving protein PK_a calculations with extensive side chain rotamer sampling,” *Journal of Computational Chemistry*, vol. 30, no. 14, pp. 2231–2247, 2009.
- [169] B. M. Tyman-Connolly and J. E. Nielsen, “pKD: re-designing protein PK_a values,” *Nucleic Acids Research*, vol. 34, pp. W48–W51, 2006.
- [170] Q. Wei, L. Wang, Q. Wang, W. D. Kruger, and R. L. Dunbrack Jr., “Testing computational prediction of missense mutation phenotypes: functional characterization of 204 mutations of human cystathionine beta synthase,” *Proteins*, vol. 78, no. 9, pp. 2058–2074, 2010.
- [171] R. Rajasekaran and R. Sethumadhavan, “Exploring the cause of drug resistance by the detrimental missense mutations in KIT receptor: computational approach,” *Amino Acids*, vol. 39, no. 3, pp. 651–660, 2010.
- [172] S. M. Abdur Rauf, M. Ismael, K. K. Sahu et al., “A graph theoretical approach to the effect of mutation on the flexibility of the DNA binding domain of p53 protein,” *Chemical Papers*, vol. 63, no. 6, pp. 654–661, 2009.
- [173] T. M. K. Cheng, Y. E. Lu, M. Vendruscolo, P. Lio, and T. L. Blundell, “Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single nucleotide polymorphisms,” *PLoS Computational Biology*, vol. 4, no. 7, Article ID e1000135, 2008.
- [174] S. Danielian, J. El-Hakeh, G. Basilico et al., “Bruton tyrosine kinase gene mutations in Argentina,” *Human Mutation*, vol. 21, no. 4, p. 451, 2003.
- [175] H. Iqbal, A. Mir, and R. Faryal, “Molecular modeling of mutant kinase domain of Btk, a Tec family member, for structure prediction,” *African Journal of Biotechnology*, vol. 10, no. 17, pp. 3274–3289, 2011.

- [176] J. Leandro, N. Simonsen, J. Saraste, P. Leandro, and T. Flatmark, "Phenylketonuria as a protein misfolding disease: the mutation pG46S in phenylalanine hydroxylase promotes self-association and fibril formation," *Biochimica et Biophysica Acta*, vol. 1812, no. 1, pp. 106–120, 2011.
- [177] Y. Yang and Y. Zhou, "Specific interactions for ab initio folding of protein terminal regions with secondary structures," *Proteins*, vol. 72, no. 2, pp. 793–803, 2008.
- [178] Y. Yang and Y. Zhou, "Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions," *Protein Science*, vol. 17, no. 7, pp. 1212–1219, 2008.
- [179] Y. Dehouck, A. Grosfils, B. Folch, D. Gilis, P. Bogaerts, and M. Rooman, "Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0," *Bioinformatics*, vol. 25, no. 19, pp. 2537–2543, 2009.
- [180] Y. Dehouck, J. M. Kwasigroch, D. Gilis, and M. Rooman, "PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality," *BMC Bioinformatics*, vol. 12, article 151, 2011.
- [181] V. Parthiban, M. M. Gromiha, and D. Schomburg, "CUPSAT: prediction of protein stability upon point mutations," *Nucleic Acids Research*, vol. 34, pp. W239–W242, 2006.
- [182] V. Parthiban, M. M. Gromiha, C. Hoppe, and D. Schomburg, "Structural analysis and prediction of protein mutant stability using distance and torsion potentials: role of secondary structure and solvent accessibility," *Proteins*, vol. 66, no. 1, pp. 41–52, 2007.
- [183] V. Parthiban, M. M. Gromiha, M. Abhinandan, and D. Schomburg, "Computational modeling of protein mutant stability: analysis and optimization of statistical potentials and structural features reveal insights into prediction model development," *BMC Structural Biology*, vol. 7, article 54, 2007.
- [184] G. Wainreb, L. Wolf, H. Ashkenazy, Y. Dehouck, and N. Ben-Tal, "Protein stability: a single recorded mutation aids in predicting the effects of other mutations in the same amino acid site," *Bioinformatics*, vol. 27, no. 23, pp. 3286–3292, 2011.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

