



Published in final edited form as:

J Mol Biol. 2013 November 1; 425(21): 3919–3936. doi:10.1016/j.jmb.2013.07.014.

Molecular mechanisms of disease-causing missense mutations

Shannon Stefl^{a,§}, Hafumi Nishi^{b,c,§}, Marharyta Petukh^a, Anna R. Panchenko^{b,*}, and Emil Alexov^{a,*}

^aComputational Biophysics and Bioinformatics, Department of Physics, Clemson University, Clemson, SC, USA, (864)-656-5307

^bNational Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA

Abstract

Genetic variations resulting in a change of amino acid sequence can have a dramatic effect on stability, hydrogen bond network, conformational dynamics, activity and many other physiologically important properties of proteins. The substitutions of only one residue in a protein sequence, so-called missense mutations, can be related to many pathological conditions, and may influence susceptibility to disease and drug treatment. The plausible effects of missense mutations range from affecting the macromolecular stability to perturbing macromolecular interactions and cellular localization. Here we review the individual cases and genome-wide studies which illustrate the association between missense mutations and diseases. In addition we emphasize that the molecular mechanisms of effects of mutations should be revealed in order to understand the disease origin. Finally we report the current state-of-the-art methodologies which predict the effects of mutations on protein stability, the hydrogen bond network, pH-dependence, conformational dynamics and protein function.

Keywords

Genetic variation; single nucleotide polymorphism; SNP; rare mutations; diseases

Introduction

The differences in human DNA sequences contribute to phenotypic variations, influence an individual's susceptibility to disease, response to the environment and drug treatment.^{1; 2} At the same time, the DNA differences lead to phenotypic differences between populations, for example differences in eye color.³ The genetic variations may involve several nucleotides or only one, the latter is called "single nucleotide polymorphism" or SNP. Technically a polymorphism represents a DNA variation found in more than one percent of the population.^{4; 5} SNPs are the most common types of genetic variations in humans^{5; 6; 7} and occur approximately every 1200 bases in the overall human population.^{6; 8} Most often, SNPs occur in the non-coding region of the genome¹ whereas SNPs in coding DNA regions

*corresponding authors: ealexov@clemson.edu, panch@ncbi.nlm.nih.gov.

^cCurrent address: Yokohama City University, Yokohama, Japan

[§]contributed equally

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

may result in changes of amino acid sequences either through amino acid substitutions (nsSNPs) or the introduction of nonsense/truncation mutations.⁵

At the opposite end of the frequency spectrum are the rare mutations (often referred to as rare variants in contrast to common variants), which are usually defined as mutations with minor allele frequency of less than 0.5 – 1%.⁹ Historically, effects of mutations were discussed in the context of a “Common Disease, Common Variant (CDCV)” or “Common Disease, Rare Variant (CDRV)” debate.¹⁰ The CDCV hypothesis argues that common disease mutations with low penetrance (the percent of individuals with disease mutations who exhibit disease phenotype) are the major contributors to genetic susceptibility to diseases. On the other hand, the CDRV hypothesis asserts that rare mutations with relatively high penetrance are the major contributors. The effects of common and rare mutations were extensively studied by the theoretical population genetics¹¹ and it was found that the frequencies of susceptibility alleles for any disease ranged from rare to high and depended on the mutation rates and the strength of purifying selection.¹¹ With the advancement of DNA sequencing technologies, many examples of rare mutations contributing to common diseases have been reported and here we list several of them. Cohen and colleagues, for example, showed that multiple rare mutations contributed significantly to the plasma levels of high-density lipoprotein (HDL)¹² and low-density lipoprotein (LDL)¹³. In the HDL study, three HDL related genes were studied and several dozen rare mutations were identified in total. Almost all mutations were found on the HDL related genes, ABC transporter A1, and two of these mutations, N1800H and W590S, drastically changed the physicochemical properties of amino acids and affected its interaction with a partner protein.

The functional effect of rare mutations has been a topic of many recent studies which tried to understand the role of rare mutations in complex traits.^{14; 15} Several studies concluded that the excess of deleterious rare mutations in the human genome was due to recent fast population growth and weak purifying selection. One study, focusing on a couple of hundreds of drug target genes, reported that rare mutations were more enriched for damaging variants than common mutations.¹⁴ Deep sequencing of human exomes also confirmed the abundance and deleterious effects of rare mutations, namely that rare mutations accounted for 86% of identified single nucleotide variants and also accounted for about 96% of variants which were predicted as functionally important.¹⁵ The same study mapped rare mutations on several known protein structures and found that they were enriched in the following categories: ligand binding, active sites, and sites participating in hydrogen bonding.¹⁵

In an effort to identify a link between mutations and chronic disease susceptibility, much data was collected from the literature for disease-causing rare and common mutations. It was shown that when the frequency of variants of adenoma patients and control groups was compared, the frequency of variants was higher in patients than for controls, which suggested that disease susceptibility could be due in part to the effects of many rare variants.¹⁶ In contrast to rare mutations contributing to Mendelian diseases, some rare mutations occur in single individuals, where the patient does not necessarily inherit the mutation from a parent. These *de novo* mutations are considered as the most extreme cases of rare mutations.¹⁷ Since *de novo* mutations have not experienced strong evolutionary selection, they tend to be more deleterious than inheritable rare mutations. Indeed, recent whole-exome sequencing studies showed that *de novo* mutations in a single gene contributed to many rare diseases such as Kabuki syndrome and others.¹⁸ It was also suggested that the scarcity of diseases caused by *de novo* mutations was related to the number of mutational target genes. Diseases caused by *de novo* mutations in a single target gene occur at very low population frequencies (<1/10000) whereas diseases caused by *de*

*nov*o mutations in more than 100 candidate genes such as intellectual disability could be relatively common ($>1/100$).¹⁷

The necessity to understand the effects of missense mutations becomes clear when one considers genetic diseases. It is known that the substitution of only one residue in a protein sequence can be related to a number of pathological conditions such as Alzheimer's, Parkinson's and Creutzfeldt-Jakob's diseases.⁶ It is also well known that accumulation of autosomal mutations can lead to cancers, and that hereditary diseases are caused by one or more germline mutations.¹⁹ The analysis of the impacts of missense mutations advances our understanding of the relationships between protein structure and function, and allows us to decipher the mechanisms of the effects of disease mutations and thereby pathogenesis.²⁰ With personalized medicine on the not-so-distant horizon, it is swiftly becoming essential to understand the process by which genetic mutations lead to disease.

The plausible effects of missense mutations^{21; 22} range from affecting the macromolecular stability to perturbing macromolecular interactions and cellular localization. Here we will review the current state-of-the-art methods in the area of predicting the effects of mutations on protein stability, the hydrogen bond network and pH-dependence, conformational dynamics, and activity.

Effects of mutations on macromolecular stability

Association between protein stability changes and human diseases

It is well established that protein function is closely related to stability of monomers and complexes;^{23;24;25} therefore in order to assess computationally the functional consequences of mutations, it is essential to identify the effect of mutations on stability and folding free energy (energy difference between folded and unfolded states). Essentially, in order for the macromolecule to carry out its function, the macromolecule, in most cases, must adopt a particular three-dimensional (3D) fold and make specific interactions with its partners. A missense mutation that affects the 3D structure and alters the stability or binding affinity of a protein complex may cause significant perturbations or complete abolishment of the function of this particular protein.

Typically the change in the folding free energy (ΔG) is used to quantify the magnitude of a mutation's effect on stability. Methods using physical potential energy functions (such as those used in molecular mechanics approaches or Monte Carlo simulations) are probably the most insightful methods for predicting the details of the effects of mutations on protein stability.²⁶ They usually are time-consuming and are frequently used for small-scale investigations.²⁷ On the other side of the spectrum are methods utilizing machine learning techniques, which are very fast and can deliver predictions on large datasets. A typical scheme of assessing the effect of a mutation on macromolecular folding free energy is illustrated in Fig. 1. Two different approaches can be used to estimate ΔG : (a) in one case the folding free energy can be calculated for the wild type and mutant proteins and then the difference can be found; (b) in the second approach, the change in free energy upon mutation can be calculated for unfolded and folded states separately and then the value of the energy of the unfolded state can be subtracted from the value of the energy of the folded state. Such detailed schemes to predict ΔG are typically employed by methods utilizing 3D structure. However, frequently structural and sequence information are used together within the same methodological framework.

Before outlining the existing methods for predicting effects of missense mutations on macromolecular stability we first review recent studies showing the connection between human diseases and protein stability changes. In general the effects on protein stability can

be grouped into two distinctive categories: (a) destabilizing and (b) stabilizing effects (see Table 1 for examples). Most frequently, missense mutations are found to destabilize the corresponding protein. Such cases include mutations in *LMNA* gene which are associated with muscular diseases,²⁸ in the VWF A2 domain causing von Willebrand disease type,²⁹ in retinal proteins causing retinal diseases,³⁰ in the perforin protein resulting in hemophagocytic lymphohistiocytosis,^{31; 32} and mutations in prion proteins associated with prion diseases.^{33; 34; 35; 36} Many neurodegenerative diseases, such as Parkinson's disease, are also associated with destabilization of the corresponding proteins.^{37; 38; 39} Sometimes the impact is local, affecting particular secondary structure element (SSE). For example, mutations of residues E22 and D23 in the Amyloid- β protein are associated with familial Alzheimer's disease and are shown to destabilize an important beta-turn⁴⁰ whereas destabilizing mutations in the core of the protein lead to an inactivation of many tumor suppressors in cancer.⁴¹ Most frequently, if the effect of mutations is related to protein stability, it usually significantly destabilizes the corresponding protein.⁴²

There are examples where missense mutations improve or enhance stability of the corresponding protein while still being deleterious. Recently a mutation in the CLIC2 protein, H101Q, was associated with a mental disorder and was predicted to increase the CLIC2 protein stability therefore obstructing its transport to the cell membrane.⁴³ Later these predictions were confirmed experimentally and it was demonstrated that, indeed, this mutation makes the CLIC2 protein thermodynamically more stable, the residence time in the membrane shorter, and the mutant interacts more strongly with the ryanodine receptor.⁴⁴ Similarly, the stability of a small 41-residue helical protein, the peripheral subunit-binding domain, was found to increase upon a replacement of a surface charge with a hydrophobic residue.⁴⁵ *In silico* modeling of the effects of mutations on stability of spermine synthase (causing Snyder-Robinson Syndrome) showed that sites harboring disease-causing mutations may or may not tolerate other (different from disease-causing) amino-acid substitutions.⁴⁶ This indicates that disease-causing effects on the stability may be both site and amino-acid type dependent.

Algorithms for predicting the effect of mutations on protein stability

The previous section listed several cases where missense mutations produce a very prominent effect on macromolecular stability and many efforts were invested to develop approaches and algorithms to predict the change of the folding free energy upon missense mutations. While mutations might have large effects on protein binding affinity and lead to many diseases,^{47; 48} here we only focus on existing approaches for predicting the effect of mutations on stability of protein monomers. In general, the existing methods can be classified into several distinct categories based on the strategy used in the calculations (see for example reviews^{49; 21}). Here we do not try to classify these methods because many of them utilize a mixture of different approaches. Table 2 includes several widely used methods, provides links to corresponding databases and servers and a short methods' description. Below we summarize several methods and resources.

The first group of methods represents machine learning approaches which are trained on different types of data relevant to protein stability and in some cases take into account experimental conditions such as temperature, salt concentration and pH values. Taking into account such parameters is important for assessing the free energy changes upon mutations at near physiological conditions.⁵⁰ Majority of these methods (MuStab⁵¹, I-Mutant and others) incorporate different physicochemical properties of amino acids and structural preferences of different sites and are trained on the experimental differences of folding free energy caused by mutations. Some methods, like I-Mutant2.0, use Support Vector Machines (SVMs), make predictions based on either structure or sequence alone,^{52; 19} and predict actual values of ΔG . Similarly, the MuPro uses SVMs leveraging both sequence and

structural information.⁵³ Other methods, such as MuStab, predict only the deleterious effects of mutations and increase or decrease (the sign) of ΔG values. Another recently introduced method, PoPMuSiC-2.0, uses a combination of statistical potential and neural networks to estimate the changes in stability; it exploits statistical potentials which take into account the coupling between four protein sequence and structure descriptors, and the amino acid volume variation upon mutation.⁵⁴

The second group of methods exploits the evolutionary conservation data under an assumption that changes at conserved positions in the multiple sequence alignments tend to be deleterious. Although such approaches do not predict the effect of mutations on protein stability directly, they are typically used in conjunction with the above mentioned methods to achieve consensus predictions. For example, a prediction of large ΔG caused by a mutation will have a higher accuracy if supported by sequence based analysis. There are many sequence based methods. For example, Sorting Intolerant From Tolerant (SIFT⁵⁵) method scores the normalized probabilities for all possible substitutions for a site and calculates the conditional probability that an amino acid is tolerated compared to the most frequent tolerated amino acid. PolyPhen on the other hand, predicts damaging amino acid substitutions using sequence-based as well as structure-based features such as sequence conservation, structure, and a position-specific independent counts matrix derived from the multiple sequence alignment.^{4; 56} Additionally, PolyPhen-2 uses eight sequence-based and three structure-based predictive features and a Naïve Bayes classifier to predict the functional significance of the mutation.⁵⁷

The third group of methods relies on structural information, assuming that the ability of a protein to function properly depends on the fundamental physicochemical properties which can be derived only from structures.²³ One research group, for example, investigated the relevance of combining coarse-grained structure-based stability predictions with a simple comparative modeling procedure.⁵⁸ CUPSAT (Cologne University Protein Stability Analysis Tool) uses structural environment specific atom potentials and torsion angle potentials to predict ΔG ⁵⁹ whereas Site Directed Mutator (SDM) uses a statistical potential energy function;⁶⁰ Some methods use empirically derived energy functions which are quite accurate in part because they are trained on the experimental data. The most prominent example is the empirical force field of FoldX which has been optimized for point mutations,¹⁹ and implemented into a computer algorithm and webserver, FOLDEF, to predict folding free energy change.^{61; 56} Another method which utilizes structural information is ERIS which applies the Medusa force field while including backbone flexibility to make predictions.⁶² Recently an interesting approach was introduced (MutPred)⁶³ which uses a broad range of different attributes based on protein sequence, structure and dynamics. This method models the changes of sequence and structural features between wild-type and mutant sequences where changes are expressed as probabilities of gain or loss of an attribute.⁶³ Another approach for predicting changes of the folding free energy upon mutations is implemented in a method called CC/PBSA.⁶⁴ It utilizes Concoord⁶⁵ to generate a structural ensemble of the target protein and calculates the folding free energy with Poisson-Boltzmann (Delphi⁶⁶) Surface Area (PBSA) method. Structural information is utilized in HOPE as well, either using experimentally available structures or structures built by homology in conjunction with energy calculations done with WhatIf.⁶⁷

Several papers reported comparisons of the performance between different methods.^{68; 69; 70; 71} In order to assess the performance, a systematic analysis is necessary. Although there is no single measure that accurately gauges the performance, some helpful measures include ROC, AUC, sensitivity, specificity, positive and negative predictive values.⁷² Different factors may influence the prediction accuracy including the type of substituted amino acid, protein structural class and flexibility, structural environment of an

affected site. All these factors should be considered when judging the effectiveness of a given approach.⁷³ It is outside of the scope of this review to rank the performance of the existing tools (including Dmutant,⁷⁴ MultiMutate,⁷⁵ SCide,⁷⁶ Scpred,⁷⁷ SRide,⁷⁸ nsSNPAnalyzer,⁷⁹ Panther,⁸⁰ PhD-SNP,⁶ SNAP,⁸¹ and SNPs&GO,⁸² among other well-known tools), because different benchmarking papers (see reviews^{83;68;73;84}) report conflicting results. The main conclusion is that various groups of methods complement each other and are suited for different types of tasks.

Effect of mutations on hydrogen bond network and ionization states

Protein stability is determined by many different factors and the formation of hydrogen bonds is among the most important ones. Hydrogen atoms are an essential component of the atomic structure of biological macromolecules. Among all hydrogens, those carrying significant positive partial charge, the so-called polar hydrogens, are particularly important because of their ability to form hydrogen bonds. In structures of biological macromolecules the hydrogen bond is formed between a polar hydrogen and a negatively charged hydrogen acceptor, typically an oxygen atom. In water phase, the water hydrogens and oxygens form a complex network of interactions resulting in water clusters. The arrangement of hydrogen bonds at the macromolecular surface is even more complicated involving interactions between macromolecular and water atoms. The groups of hydrogen bonds often form a cluster or a web of interactions resulting in so-called hydrogen bond network. Hydrogen bonds and hydrogen bond networks participate in biological functions of macromolecules and provide the pH-dependence associated with many biological reactions. They contribute to protein structural integrity, provide “proton wires” for proton translocation, the source of proton uptake/release and, finally, participate in many catalytic reactions. Below we briefly outline the mechanisms of how missense mutations may affect the hydrogen bond networks.

Hydrogen bond networks and macromolecular structure

As was mentioned above, the hydrogen bonds are key constituents of biomolecular structures,⁸⁵ participating in the formation of SSEs,⁸⁶ tertiary and quaternary structure.⁸⁷ A mutation resulting in the removal or addition of a hydrogen donor or acceptor is expected to have a significant impact on the structural integrity. Even a conservative mutation which results in different donor-acceptor positions, may still be quite deleterious for the structure and may disrupt the entire hydrogen bond network (Table 3).

Figure 2 shows one example of mutation N550K in Isoform 3 Lactosylceramide alpha-2,3-sialyltransferase, which is associated with Autosomal Recessive Neurocutaneous Condition. As can be seen from this figure, the wild type residue, N550, is involved in several hydrogen bonds, which are deleted in the mutant. In another example of the Aldosterone synthase deficiency, it was demonstrated that molecular mechanism of this disease involved mutation R374W in CYP11B2 protein, which changed the properties of the hydrogen bond network of the wild type Arg residue.⁸⁸ Similarly, in patients suffering from mitochondrial fatty acid oxidation disorder and bearing a R595W mutation in the CPT1 protein, the disease was attributed to a disruption of the wild type hydrogen bond network formed by the wild type R374 residue.⁸⁹ In another case the rearrangement of the hydrogen bond network due to a missense mutation (I150T) in spermine synthase was predicted to be the main cause of the Snyder Robinson syndrome.⁹⁰ Further mutational analyses of the same protein indicated that almost any mutation affecting this residue involved in a wild type hydrogen bond network will have drastic effects on the wild type properties of the hydrogen bond network and macromolecular stability.⁴⁶

The effect of mutations on hydrogen bond network is especially pronounced if the network involves active site residues. For example, a recent study showed that the low catalytic

efficiency of a mutant, T560M, found in human lipoxygenase is caused by the alterations of a hydrogen bond network interconnecting this residue with active sites; this in turn may lead to a predisposal to cardiovascular diseases.⁹¹ The wild type hydrogen bond network and stability of tertiary structure was also reported to be altered by missense mutations causing congenital hereditary cataract disease.⁹² These limited examples confirm the importance of the hydrogen bond network for structural integrity of macromolecules and point to the necessity of predicting such effects in the context of missense mutations.

Hydrogen bond network in protein aggregation and flexibility

The aggregation, or formation of amyloid fibrils is one of the primary sources of many diseases.^{93; 94; 95; 96} Missense mutations disrupting the native hydrogen bond(s) can destabilize the local structure and thus expose the hydrophobic core of the protein to the water phase. Such an event frequently triggers aggregation and fibril formation. Several examples are listed in Table 3. Biological reactions frequently are accompanied by small or large protein conformational changes.⁹⁷ Consequently, the ability of macromolecules to retain their conformational flexibility is crucial to their function. Altering the hydrogen bond network by missense mutation can affect the conformational flexibility needed for allosteric regulation and conformational gating. Several examples of an altered protein flexibility causing diseases are listed in Table 3. At the same time, altering the wild type flexibility through mutations may not necessarily result in the pathogenic effect.⁹⁸ A detailed analysis is usually required in order to attribute the change of flexibility to a particular molecular function.

Hydrogen bond networks and pH-dependence

Practically all biological processes are pH-dependent and pH is an important regulator of cellular function. Different compartments of the cell have different characteristic pH (see reviews^{99,100, 24, 101}). Macromolecules that shuttle between different cellular compartments, as for example, prolactin receptor,^{102; 103} have their pH-dependent characteristics precisely tuned to the local pH. Thus, in order to function properly and interact with its biological partners, the protein needs maintain specific pH dependent characteristics²² which can be very sensitive to single point mutations.¹⁰⁴ In addition, single point mutations may affect a protein's cellular location.¹⁰⁵ More examples are provided in Table 3.

Hydrogen bond networks and catalytic reaction

Hydrogen bonds participate in catalytic reactions in various ways: they coordinate the substrate and can be formed or disrupted during the reaction process (donating or accepting a hydrogen via general acid/base residues) (Table 3). Typically the catalytic reaction is mostly affected by a substitution of an amino acid directly involved in the catalysis. However, it should be mentioned that active site residues are rarely mutated and mutations often occur in the sites connected to the active site region via hydrogen bond networks. The most drastic change of the hydrogen bond network is caused by protonation/deprotonation of titratable groups. Thus, the pKa shifts of catalytic residues induced by a mutation will disrupt the general acid/basic reaction and will dramatically affect the catalysis.

Approaches to model effects of mutations on hydrogen bond networks

Assessing the effect of mutations on hydrogen bond networks is not a straightforward task since the positions of protons (hydrogens) are typically not resolved experimentally, but rather should be generated *in silico*. Even more, the protonation states of titratable groups, which are responsible for pH-dependence of biological processes including general acid/basic catalysis, are typically unknown and have to be predicted as well.

In the simplest case scenario when mutation does not involve the protonation change, the analysis of the hydrogen bond network begins with placing the missing hydrogens onto the wild type and mutant 3D structures. By doing so, one typically assumes default protonation states of titratable groups at pH equal to 7. However, many biological reactions occur at pH different from neutral and careful analysis would require obtaining the biochemical data on the optimal/characteristic pH.^{106; 100;24} At the next step, protons' positions are generated with the standard Molecular Dynamics (MD) packages such as NAMD,¹⁰⁷ Charmm,¹⁰⁸ Amber,¹⁰⁹ Gromos,¹¹⁰ Gromacs,¹¹¹ or other stand-alone programs such as REDUCE¹¹² and PDB2PQR¹¹³. Then one would compare hydrogen bonds in the minimized structures of wild type and mutant proteins.⁹⁰ A more sophisticated analysis would involve the comparison of the hydrogen bonds in the snapshot structures obtained in MD simulations. All MD packages have tools for the analysis of hydrogen bond networks. A recently developed standalone program (HBonanza,¹¹⁴ hydrogen-bond analyzer; <http://www.nbc.net/hbonanza>), allows the analysis and visualization of hydrogen-bond networks. HBonanza can be used to analyze single structures or many structures of a molecular dynamics trajectory.

Cases where mutation induces changes in the protonation state of titratable groups are much more complex. It should be mentioned that such ionization changes may occur even if mutation does not involve titratable groups.⁴⁶ Predictions of protonation states can be done by calculating the pKa's of titratable groups and then by assigning the appropriate charge states depending on the characteristic pH for a given protein. There are many approaches for computing pKa's which are reviewed in¹¹⁵. Some of them are standalone programs such as MCCE,¹¹⁶ ProPKA,¹¹⁷ while others are implemented into web servers such as H++¹¹⁸.

Effect of mutations on conformational dynamics

Biological macromolecules may adopt different conformations along the pathway of the corresponding biochemical reaction^{119; 120} and their intrinsic flexibility, the ability to sample alternative conformations is crucial for protein function.^{121; 122} Furthermore, a significant fraction of macromolecules is either disordered or has disordered segments at a particular stage of the biological reaction.^{123; 124} Missense mutations can affect the flexibility of the entire molecule or just a small region, can shift the equilibrium between different conformations or can affect the entire conformational dynamics of the molecule. In a most recent study of the NFAT5 transcription factor, different mutations from the same DNA-binding loop were analyzed.¹²⁵ It was shown that even though these mutations are located very close to each other in sequence and space, their effect on protein dynamics and DNA binding is drastically different. Typically the changes of conformational dynamics are assessed computationally via monitoring the RMSD of the wild type and mutant structures and recorded via the snapshots obtained in Molecular Dynamics (MD).

Below we outline different aspects of effects of mutations on protein dynamics and provide examples of recent finding with this regard. The most commonly used approach to study protein dynamics is MD simulations, although Monte Carlo methods and Normal Mode Analysis can be utilized as well. Most widely used packages are NAMD,¹⁰⁷ Charmm,¹⁰⁸ Amber,¹⁰⁹ Gromos,¹¹⁰ Gromacs,¹¹¹ TINKER,¹²⁶ and others.

Effect on protein flexibility

Molecular flexibility is reflected in the ability of a macromolecule to sample alternative conformations, as for example to open/close the gate of a channel, to mediate the recognition of the receptor or to facilitate the allosteric reactions. Typical examples are centrosomes, which are central regulators of mitosis often amplified in cancer cells. Specifically, the centrosomal protein CEP63 is associated with an aneuploidy and solid

tumors in humans. When genetic alterations such as L61P occur, they increase flexibility of the protein, as seen in MD simulations, which was suggested to be the cause of the disease.¹²⁷ Although mutations may cause different diseases, the common trend is the same, namely, disease can be associated with a change in the conformational or dynamical properties of the corresponding protein (Table 4).

Another case is phosphatase and tensin homolog (PTEN) protein which plays essential roles in cellular processes including survival, proliferation, energy metabolism, and cellular architecture. Mutations in PTEN are implicated in diabetes and cancer. A particular mutation (H61D) was studied using MD simulations and it was shown that it increases the flexibility, radius of gyration, and solvent accessibility of PTEN.¹²⁸ The list of examples would be incomplete without mentioning the “conformational diseases”, where native protein conformers convert to pathological intermediates that can polymerize. Recently a *forme fruste* deficiency variant of (1)-antitrypsin (K154N) was investigated by the Nuclear Magnetic Resonance spectroscopy and it was found that this mutation alters the wild type interaction of the side chain of K154 with the backbone carbonyl oxygen of K174, affecting protein flexibility and resulting in polymerization.¹²⁹

Disease mutations might not only lead to increased flexibility, on the contrary, they can restrict the conformational transitions. One example includes mutations in the human protoporphyrinogen oxidase (hPPO) gene which are responsible for the dominantly inherited disorder variegate porphyria (VP). Two missense mutations (R59Q and R59G) were investigated in a recent work¹³⁰ where MD modeling revealed that these mutations affect the catalytic activity of hPPO by changing its ability to sample different conformations. Analysis of mutation H101Q in the CLIC2 protein demonstrated that this mutation restricted the mobility of the N-terminal domain and prevented the conformational change presumed to be required for entering of CLIC2 into the membrane.^{43; 44} In this regard, channels and pores are very interesting objects to study since their selectivity is regulated by structural fragments. For instance, aquaporins play physiological roles in several organs and tissues, and their alteration is associated with disorders of water regulation. A mutant, D184E, was shown to affect the mobility of the aquaporin D-loop, which acquires a higher propensity to equilibrate in a “closed conformation”, thus affecting the rate of water flux¹³¹.

Protein intrinsic disorder and disease mutations

Disordered proteins, existing in dynamic equilibrium between various conformers, may provide a way to tolerate many mutations due to the loosely packed cores. However, the relationship between disorder, stability, and function is not well understood. According to some studies, about 10% of inherited disease mutations from HGMD database affect disordered regions,⁶³ and cancer-associated proteins are especially enriched in intrinsically disordered regions.¹³² In particular, disease-related genes encoding disordered proteins or proteins with the extended disordered regions, include α -synuclein (Parkinson’s disease), BRCA1 (breast and/or ovarian cancer), p53 (cancer), huntingtin (Huntington’s disease).^{133; 134}

There are many ways in which mutations might impact disordered regions and lead to dysfunctional proteins. If a residue that promotes disorder is mutated into a residue that favors structural regions, such a mutation can have a drastic effect on disorder-order transition, post-translational modifications, binding, and other functions relevant to intrinsic disorder. Examples of such substitutions are abundant among substitutions of disorder-promoting arginine into other residues types. Overall, about a quarter of disease mutations in disordered regions were predicted to disrupt disorder-promoting properties. Therefore the prediction and prioritization of disease mutations should account for the disorder-promoting

tendency of the region where mutations occur.^{135; 136} Several such examples are provided in Table 4.

Protein aggregation, misfolding and dynamics

Protein aggregation is driven by the exposure of the hydrophobic core of the macromolecule to the water phase due to the changes in native structure and dynamics. It was shown that mutations in peripheral myelin protein 22 resulted in the common peripheral neuropathy Charcot-Marie-Tooth disease. One of them, L16P, caused misfolding which led to a loss of function and toxic accumulation of aggregates.¹³⁷ Another recent work studied the effects of several mutations (V180I, F198S, V203I and V210I) known to cause prion disease and showed that these mutations induced the misfolding and aggregation of the prion protein.¹³⁸ Another case involved mutations or deletions in the FMRP protein, which participates in the regulation of mRNA metabolism in brain, leading to the Fragile X syndrome. A severe manifestation of the disease has been associated with the I304N mutation, located on the KH2 domain of the protein. This mutation was found to destabilize the hydrophobic core causing a partial unfolding and a displacement of alpha-helices.¹³⁹

Similarly, mutation E342K in human serine protease inhibitor (serpin) α_1 -antitrypsin causes polymerization in the endoplasmic reticulum of hepatocytes and is associated with a lack of secretion into the circulation. It was shown that this mutation increases local flexibility, favors polymerization and promotes aggregation.¹⁴⁰ Another case involves mutations in the human prion protein resulting in Creutzfeldt-Jakob or Gerstmann-Straussler-Scheinker diseases. The molecular dynamics modeling suggested that these mutations promote amyloid formation.¹⁴¹ Other examples of disease causing mutations in relation to aggregation and misfolding are shown in Table 4.

Effect of mutations on protein activity

Challenges of modeling of mutation effects on protein function

Protein structure-function relationships are complex and crucial for understanding and predicting the effects of benign or disease-causing mutations on the fitness. Proteins largely evolve through the acquisition of new mutations, the majority of them are destabilizing but neutral. At the same time some mutations can be deleterious or damaging or may result in advantageous novel functions. Interestingly, mutations which modify and produce novel binding specificities were shown to have larger destabilizing effects compared to mutations occurring on protein surfaces.¹⁴² On the other hand, since protein functional regions can be energetically unfavorable, mutations of functionally important residues, especially of polar and charged residues, may often result in more stable structures.^{143; 144; 145} Protein stability is necessary but not sufficient for protein functioning and proteins are not necessarily optimized to maximize their stability^{146; 147}. Therefore in order to assess the damaging effect of mutations, it would be crucial to understand how they will impact functionally important sites. Functional site prediction methods can be subdivided into several categories: those that use evolutionary conservation of binding site motifs, those that use information about a structure of a complex and docking methods. Some of them include PHUNCTIONER,¹⁴⁸ FIRESTAR,¹⁴⁹ IBIS,¹⁵⁰ ConCavity,¹⁵¹ and others.

The effects of missense mutations on proteins and their function can be understood and modeled within the framework of the energy landscape theory which describes the potential energy of a protein as a function of conformational coordinates¹²⁷. Different conformational states might be characterized by different functional specificities and structural differences. The equilibrium between conformational states can be shifted by binding of different ligands, post-translational modifications, by changing the environmental conditions, and, finally, by mutations. Recently, human spermine synthase activity was

enhanced by engineered novel mutations¹⁵² and it was demonstrated that the activity depended on electrostatic field distribution, the intrinsic flexibility of protein domains, and the overall protein stability. This indicates the complexity of biochemical functions involving various factors which in turn may be affecting each other. Table 5 lists exemplary mutations and their effect on protein activity.

Case study: Receptor Tyrosine Kinases

Here we analyze the mechanisms of cancer mutations on protein stability, dynamics, and activity using an example of the well-studied human receptor tyrosine kinase (RTK) family that is frequently mutated in cancer. The connection between cancer and kinase activation was found fairly recently¹⁵³ and an increased RTK activity in tumor tissues was attributed to gene amplifications, enhanced transcription, translation, and to mutations. Several cancer mutation hotspots were identified in several RTKs, most of them were located in the activation loop, P-loop, and DFG loops. Moreover, driver mutations were found to be more often associated with the functionally important regions in kinases than passenger mutations.¹⁵⁴ The activation of certain RTKs is tightly linked with their dimerization and high RTK activity is sometimes achieved by promoting dimerization. According to long-timescale molecular dynamics simulations, it was suggested that cancer mutations may suppress the intrinsic disorder on the dimer interface and stimulate the EGFR dimerization.¹⁵⁵

Structural studies of kinases revealed different structural perturbations in response to cancer mutations. In particular, the mechanisms of kinase activation in cancer is probably linked to transitions between the active and inactive states.^{156; 157} Indeed, the thorough analysis of the effect of cancer mutations on kinase structure and activity showed that some mutations disturbed autoinhibitory interactions and considerably accelerated the catalysis.¹⁵⁶ At the same time the crystal structure of the *EGFR* L858R mutant revealed that this mutation prevented the activation loop from adopting the inactive conformation¹⁵⁷ thereby activating the kinase. On the other hand, the secondary *EGFR* T790M mutation facilitated the transition between inactive and active conformations and increased the stability of the active conformation.¹⁵⁸ Enhanced mobility was observed near cancer mutations sites and modeling of autoinhibited conformations revealed that these mutations disrupted the local stabilizing interactions and destabilized the inactive form.¹⁵⁹ Recently the effect of cancer mutations was analyzed using the dataset of all available structural pairs of active in inactive states. It was shown that cancer mutations destabilized active states, but to a lesser degree than the random mutations; moreover, to a lesser degree than mutations destabilized inactive states.¹⁶⁰ This led to kinase activation. The same study found a relationship between the statistics-based estimate of oncogenic potential of mutation and its activation effect calculated based on thermodynamics principles. Namely, more frequent mutations had a higher activating effect.

Multiple mutations

A more comprehensive understanding of molecular mechanisms of disease mutations may come from the analysis of the effects of multiple mutations occurring in the same gene. It was demonstrated previously that the number of cases where multiple mutations may occur simultaneously in one gene is higher than the number predicted from the random mutation distribution.¹⁶¹ Such multiple mutations may be the result of an accumulation of single mutations in sequential cell replications or in the same cell cycle. The latter synchronous mutations are usually characterized by non-random proximal spacing in higher eukaryotes.¹⁶² At the same time, a recent study showed that the extent of the clustering of cancer mutations might differ between oncogenes and tumor suppressors and the former are found to be clustered while latter are not.⁴¹ A significant fraction of cancer-associated

mutations comes in doublets or triplets where overall about 6–8% of all cancer mutations represent double cancer mutations.^{160; 162; 163} Similarly to single mutations, the majority of cancer multiple mutations occur in non-synonymous codon sites implying that doublets are under positive selection.^{160; 164} The observed and simulated spectra of double mutations generated from the spectra of single mutations were found to be significantly different for receptor tyrosine kinase (RTK) genes, which pointed to different mechanisms underlying single and double mutation spectra.¹⁶⁰

Moreover, multiple disease mutations might have a synergistic phenotypic effect manifested in enhanced or decreased protein stability or activity. For example, a double mutation was found in β -amyloid precursor protein which resulted in an increased production and secretion of amyloid- β -peptide causing early-onset Alzheimer's disease.¹⁶⁵ On the other hand double mutations in the β -galactosidase encoding gene, which cause Fabry disease, resulted in reduced or non-detectable activity compared to single mutants.¹⁶⁶ A positive epistasis was found for many double cancer mutants in RTK genes,¹⁶⁰ namely, the effect of multiple mutations on kinase activity was higher than a total of individual mutations. This trend was especially pronounced for double mutations observed in more than one tumor sample.

Concluding remarks

The effects of human missense mutations on various biophysical characteristics, stability, hydrogen bond network, dynamics and activity, were reviewed in this work. It was demonstrated that mutations can frequently affect several biophysical characteristics simultaneously and may or may not cause diseases (see also ¹⁶⁷). There is no clear threshold of how large the change of the wild type characteristics should be in order to alter protein function and result in disease. Predictions of damaging effects of mutations are further complicated by the observations that enhanced activity¹⁵², higher stability or binding affinity,^{43; 44} are not necessarily advantageous for the cell and protein and can be disease causing.

Acknowledgments

S.S., M.P. and E.A. acknowledge the support from NIH, grant number R01GM093937. H.N. and A.R.P. were supported by the Intramural Research Program of the National Library of Medicine at the U.S. National Institutes of Health.

References

1. Taillon-Miller P, Gu Z, Li Q, Hillier L, Kwok PY. Overlapping genomic sequences: a treasure trove of single-nucleotide polymorphisms. *Genome Res.* 1998; 8:748–54. [PubMed: 9685323]
2. De Baets G, Van Durme J, Reumers J, Maurer-Stroh S, Vanhee P, Dopazo J, Schymkowitz J, Rousseau F. SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Research.* 2012; 40:D935–D939. [PubMed: 22075996]
3. Sturm RA, Frudakis TN. Eye colour: portals into pigmentation genes and ancestry. *Trends Genet.* 2004; 20:327–32. [PubMed: 15262401]
4. Fernald GH, Capriotti E, Daneshjou R, Karczewski KJ, Altman RB. Bioinformatics challenges for personalized medicine. *Bioinformatics.* 2011; 27:1741–8. [PubMed: 21596790]
5. Mooney S. Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. *Brief Bioinform.* 2005; 6:44–56. [PubMed: 15826356]
6. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics.* 2006; 22:2729–34. [PubMed: 16895930]

7. Collins FS, Guyer MS, Chakravarti A. Variations on a Theme: Cataloging Human DNA Sequence Variation. *Science*. 1997; 278:1580–1581. [PubMed: 9411782]
8. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001; 29:308–11. [PubMed: 11125122]
9. Raychaudhuri S. Mapping rare and common causal alleles for complex human diseases. *Cell*. 2011; 147:57–69. [PubMed: 21962507]
10. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*. 2009; 19:212–9. [PubMed: 19481926]
11. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001; 69:124–37. [PubMed: 11404818]
12. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*. 2004; 305:869–72. [PubMed: 15297675]
13. Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, Hobbs HH. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A*. 2006; 103:1810–5. [PubMed: 16449388]
14. Nelson MR, Wegmann D, Ehm MG, Kessner D, St Jean P, Verzilli C, Shen J, Tang Z, Bacanu SA, Fraser D, Warren L, Aponte J, Zawistowski M, Liu X, Zhang H, Zhang Y, Li J, Li Y, Li L, Woollard P, Topp S, Hall MD, Nangle K, Wang J, Abecasis G, Cardon LR, Zollner S, Whittaker JC, Chisoe SL, Novembre J, Mooser V. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*. 2012; 337:100–4. [PubMed: 22604722]
15. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM, Broad GO, Seattle GO, Project NES. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*. 2012; 337:64–9. [PubMed: 22604720]
16. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*. 2008; 40:695–701. [PubMed: 18509313]
17. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet*. 2012; 13:565–75. [PubMed: 22805709]
18. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C, Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010; 42:790–3. [PubMed: 20711175]
19. Shen B, Bai J, Vihinen M. Physicochemical feature-based classification of amino acid mutations. *Protein Engineering Design and Selection*. 2008; 21:37–44.
20. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C. Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proceedings of the National Academy of Sciences*. 2003; 100:8424–8429.
21. Teng S, Michonova-Alexova E, Alexov E. Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. *Curr Pharm Biotechnol*. 2008; 9:123–33. [PubMed: 18393868]
22. Zhang Z, Miteva MA, Wang L, Alexov E. Analyzing Effects of Naturally Occurring Missense Mutations. *Computational and Mathematical Methods in Medicine*. 2012; 2012:15.
23. Araya CL, Fowler DM, Chen W, Muniez I, Kelly JW, Fields S. A fundamental protein property, thermodynamic stability, revealed solely from large-scale measurements of protein function. *Proceedings of the National Academy of Sciences*. 2012; 109:16858–16863.
24. Talley K, Alexov E. On the pH-optimum of activity and stability of proteins. *Proteins*. 2010; 78:2699–706. [PubMed: 20589630]
25. Zhang Z, Wang L, Gao Y, Zhang J, Zhenirovskyy M, Alexov E. Predicting folding free energy changes upon single point mutations. *Bioinformatics*. 2012; 28:664–71. [PubMed: 22238268]

26. Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Research*. 2011; 39:W215–W222. [PubMed: 21593128]
27. Kang S, Chen G, Xiao G. Robust prediction of mutation-induced protein stability change by property encoding of amino acids. *Protein Engineering Design and Selection*. 2009; 22:75–83.
28. Zwerger M, Jaalouk DE, Lombardi ML, Isermann P, Mauermann M, Dialynas G, Herrmann H, Wallrath LL, Lammerding J. Myopathic lamin mutations impair nuclear stability in cells and tissue and disrupt nucleo-cytoskeletal coupling. *Human Molecular Genetics*. 2013
29. Xu AJ, Springer TA. Mechanisms by which von Willebrand disease mutations destabilize the A2 domain. *J Biol Chem*. 2013; 288:6317–24. [PubMed: 23322777]
30. Rakoczy EP, Kiel C, McKeone R, Stricher F, Serrano L. Analysis of disease-linked rhodopsin mutations based on structure, function, and protein stability calculations. *J Mol Biol*. 2011; 405:584–606. [PubMed: 21094163]
31. An O, Gursoy A, Gurgey A, Keskin O. Structural and functional analysis of perforin mutations in association with clinical data of familial hemophagocytic lymphohistiocytosis type 2 (FHL2) patients. *Protein Sci*. 2013; 22:823–839. [PubMed: 23592409]
32. Goransdotter Ericson K, Fadeel B, Nilsson-Ardnor S, Soderhall C, Samuelsson A, Janka G, Schneider M, Gurgey A, Yalman N, Revesz T, Egeler R, Jahnukainen K, Storm-Mathiesen I, Haraldsson A, Poole J, de Saint Basile G, Nordenskjold M, Henter J. Spectrum of perforin gene mutations in familial hemophagocytic lymphohistiocytosis. *Am J Hum Genet*. 2001; 68:590–7. [PubMed: 11179007]
33. Prusiner S. Molecular biology of prion diseases. *Science*. 1991; 252:1515–1522. [PubMed: 1675487]
34. Collinge J. Prion diseases of humans and animals: their causes and molecular basis. *Annu Rev Neurosci*. 2001; 24:519–50. [PubMed: 11283320]
35. Kretzschmar H, Tatzelt J. Prion disease: a tale of folds and strains. *Brain Pathol*. 2013; 23:321–32. [PubMed: 23587138]
36. Jetha NN, Semenchenko V, Wishart DS, Cashman NR, Marziali A. Nanopore analysis of wild-type and mutant prion protein (PrP(C)): single molecule discrimination and PrP(C) kinetics. *PLoS One*. 2013; 8:e54982. [PubMed: 23393562]
37. Lin W, Kang UJ. Characterization of PINK1 processing, stability, and subcellular localization. *J Neurochem*. 2008; 106:464–74. [PubMed: 18397367]
38. Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Hum Mutat*. 2010; 31:763–80. [PubMed: 20506312]
39. Morais VA, Verstreken P, Roethig A, Smet J, Snellinx A, Vanbrabant M, Haddad D, Frezza C, Mandemakers W, Vogt-Weisenhorn D, Van Coster R, Wurst W, Scorrano L, De Strooper B. Parkinson's disease mutations in PINK1 result in decreased Complex I activity and deficient synaptic function. *EMBO Mol Med*. 2009; 1:99–111. [PubMed: 20049710]
40. Grant MA, Lazo ND, Lomakin A, Condrion MM, Arai H, Yamin G, Rigby AC, Teplow DB. Familial Alzheimer's disease mutations alter the stability of the amyloid beta-protein monomer folding nucleus. *Proc Natl Acad Sci U S A*. 2007; 104:16522–7. [PubMed: 17940047]
41. Stehr H, Jang SH, Duarte JM, Wierling C, Lehrach H, Lappe M, Lange BM. The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors. *Mol Cancer*. 2011; 10:54. [PubMed: 21575214]
42. Shi Z, Moulton J. Structural and functional impact of cancer-related missense somatic mutations. *J Mol Biol*. 2011; 413:495–512. [PubMed: 21763698]
43. Witham S, Takano K, Schwartz C, Alexov E. A missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins*. 2011; 79:2444–54. [PubMed: 21630357]
44. Takano K, Liu D, Tarpey P, Gallant E, Lam A, Witham S, Alexov E, Chaubey A, Stevenson RE, Schwartz CE, Board PG, Dulhunty AF. An X-linked channelopathy with cardiomegaly due to a CLIC2 mutation enhancing ryanodine receptor channel activity. *Hum Mol Genet*. 2012; 21:4497–507. [PubMed: 22814392]

45. Spector S, Wang M, Carp SA, Robblee J, Hendsch ZS, Fairman R, Tidor B, Raleigh DP. Rational Modification of Protein Stability by the Mutation of Charged Surface Residues†. *Biochemistry*. 2000; 39:872–879. [PubMed: 10653630]
46. Zhang Z, Norris J, Schwartz C, Alexov E. In silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase. *PLoS One*. 2011; 6:e20373. [PubMed: 21647366]
47. Teng S, Madej T, Panchenko A, Alexov E. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys J*. 2009; 96:2178–88. [PubMed: 19289044]
48. Nishi H, Tyagi M, Teng S, Shoemaker BA, Hashimoto K, Alexov E, Wuchty S, Panchenko AR. Cancer missense mutations alter binding properties of proteins and their interaction networks. *PLoS One*. 2013; 8:e66273. [PubMed: 23799087]
49. Zhang Z, Miteva MA, Wang L, Alexov E. Analyzing effects of naturally occurring missense mutations. *Comput Math Methods Med*. 2012; 2012:805827. [PubMed: 22577471]
50. Kang S, Chen G, Xiao G. Robust prediction of mutation-induced protein stability change by property encoding of amino acids. *Protein Eng Des Sel*. 2009; 22:75–83. [PubMed: 19054789]
51. Teng S, Srivastava AK, Wang L. Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BMC Genomics*. 2010; 11(Suppl 2):S5. [PubMed: 21047386]
52. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*. 2005; 33:W306–W310. [PubMed: 15980478]
53. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics*. 2006; 62:1125–1132.
54. Dehouck Y, Grosfils A, Folch B, Gilis D, Bogaerts P, Rooman M. Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics*. 2009; 25:2537–2543. [PubMed: 19654118]
55. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*. 2003; 31:3812–3814. [PubMed: 12824425]
56. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*. 2002; 30:3894–3900. [PubMed: 12202775]
57. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7:248–9. [PubMed: 20354512]
58. Gonnelli G, Rooman M, Dehouck Y. Structure-based mutant stability predictions on proteins of unknown structure. *J Biotechnol*. 2012; 161:287–93. [PubMed: 22782143]
59. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Research*. 2006; 34:W239–W242. [PubMed: 16845001]
60. Worth CL, Preissner R, Blundell TL. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res*. 2011; 39:W215–22. [PubMed: 21593128]
61. Guerois R, Nielsen JE, Serrano L. Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. *Journal of Molecular Biology*. 2002; 320:369–387. [PubMed: 12079393]
62. Ding F, Dokholyan NV. Emergence of Protein Fold Families through Rational Design. *PLoS Comput Biol*. 2006; 2:e85. [PubMed: 16839198]
63. Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009; 25:2744–2750. [PubMed: 19734154]
64. Benedix A, Becker CM, de Groot BL, Caflisch A, Bockmann RA. Predicting free energy changes using structural ensembles. *Nat Methods*. 2009; 6:3–4. [PubMed: 19116609]
65. de Groot BL, van Aalten DM, Scheek RM, Amadei A, Vriend G, Berendsen HJ. Prediction of protein conformational freedom from distance constraints. *Proteins*. 1997; 29:240–51. [PubMed: 9329088]

66. Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M, Alexov E. DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* 2012; 5:9. [PubMed: 22583952]
67. Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics.* 2010; 11:548. [PubMed: 21059217]
68. Khan S, Vihinen M. Performance of protein stability predictors. *Hum Mutat.* 2010; 31:675–84. [PubMed: 20232415]
69. Masso M, Vaisman II. Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics.* 2008; 24:2002–2009. [PubMed: 18632749]
70. Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Engineering Design and Selection.* 2009; 22:553–560.
71. Thiltgen G, Goldstein RA. Assessing predictors of changes in protein stability upon mutation using self-consistency. *PLoS One.* 2012; 7:e46084. [PubMed: 23144695]
72. Vihinen M. How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC Genomics.* 2012; 13(Suppl 4):S2. [PubMed: 22759650]
73. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011; 32:358–68. [PubMed: 21412949]
74. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci.* 2002; 11:2714–26. [PubMed: 12381853]
75. Deutsch C, Krishnamoorthy B. Four-body scoring function for mutagenesis. *Bioinformatics.* 2007; 23:3009–15. [PubMed: 17921497]
76. Dosztanyi Z, Magyar C, Tusnady G, Simon I. SCide: identification of stabilization centers in proteins. *Bioinformatics.* 2003; 19:899–900. [PubMed: 12724305]
77. Kurgan L, Cios K, Chen K. SCPRED: accurate prediction of protein structural class for sequences of twilight-zone similarity with predicting sequences. *BMC Bioinformatics.* 2008; 9:226. [PubMed: 18452616]
78. Magyar C, Gromiha MM, Pujadas G, Tusnady GE, Simon I. SRide: a server for identifying stabilizing residues in proteins. *Nucleic Acids Res.* 2005; 33:W303–5. [PubMed: 15980477]
79. Bao L, Zhou M, Cui Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* 2005; 33:W480–2. [PubMed: 15980516]
80. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.* 2003; 13:2129–41. [PubMed: 12952881]
81. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* 2007; 35:3823–35. [PubMed: 17526529]
82. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat.* 2009; 30:1237–44. [PubMed: 19514061]
83. Gromiha MM, Huang LT. Machine learning algorithms for predicting protein folding rates and stability of mutant proteins: comparison with statistical methods. *Curr Protein Pept Sci.* 2011; 12:490–502. [PubMed: 21787301]
84. Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipinski AJ. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. *Genome Res.* 2009; 19:1562–9. [PubMed: 19546171]
85. Nisius L, Grzesiek S. Key stabilizing elements of protein structure identified through pressure and temperature perturbation of its hydrogen bond network. *Nat Chem.* 2012; 4:711–7. [PubMed: 22914191]
86. Gong S, Worth CL, Bickerton GR, Lee S, Tanramluk D, Blundell TL. Structural and functional restraints in the evolution of protein families and superfamilies. *Biochem Soc Trans.* 2009; 37:727–33. [PubMed: 19614584]

87. Horowitz S, Trievel RC. Carbon-oxygen hydrogen bonding in biological structure and function. *J Biol Chem.* 2012; 287:41576–82. [PubMed: 23048026]
88. Nguyen HH, Hannemann F, Hartmann MF, Malunowicz EM, Wudy SA, Bernhardt R. Five novel mutations in CYP11B2 gene detected in patients with aldosterone synthase deficiency type I: Functional characterization and structural analyses. *Mol Genet Metab.* 2010; 100:357–64. [PubMed: 20494601]
89. Fontaine M, Dessein AF, Douillard C, Dobbelaere D, Brivet M, Boutron A, Zater M, Mention-Mulliez K, Martin-Ponthieu A, Vianey-Saban C, Briand G, Porchet N, Vamecq J. A Novel Mutation in CPT1A Resulting in Hepatic CPT Deficiency. *JIMD Rep.* 2012; 6:7–14. [PubMed: 23430932]
90. Zhang Z, Teng S, Wang L, Schwartz CE, Alexov E. Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Hum Mutat.* 2010; 31:1043–9. [PubMed: 20556796]
91. Schurmann K, Anton M, Ivanov I, Richter C, Kuhn H, Walther M. Molecular basis for the reduced catalytic activity of the naturally occurring T560M mutant of human 12/15-lipoxygenase that has been implicated in coronary artery disease. *J Biol Chem.* 2011; 286:23920–7. [PubMed: 21558275]
92. Wang S, Zhao WJ, Liu H, Gong H, Yan YB. Increasing betaB1-crystallin sensitivity to proteolysis caused by the congenital cataract-microcornea syndrome mutation S129R. *Biochim Biophys Acta.* 2013; 1832:302–11. [PubMed: 23159606]
93. Gorbenko G, Trusova V. Protein aggregation in a membrane environment. *Adv Protein Chem Struct Biol.* 2011; 84:113–42. [PubMed: 21846564]
94. Goto Y, Yagi H, Yamaguchi K, Chatani E, Ban T. Structure, formation and propagation of amyloid fibrils. *Curr Pharm Des.* 2008; 14:3205–18. [PubMed: 19075701]
95. Khare SD, Dokholyan NV. Molecular mechanisms of polypeptide aggregation in human diseases. *Curr Protein Pept Sci.* 2007; 8:573–9. [PubMed: 18220844]
96. Naeem A, Fazili NA. Defective protein folding and aggregation as the basis of neurodegenerative diseases: the darker aspect of proteins. *Cell Biochem Biophys.* 2011; 61:237–50. [PubMed: 21573992]
97. Wilbur JD, Hwang PK, Brodsky FM, Fletterick RJ. Accommodation of structural rearrangements in the huntingtin-interacting protein 1 coiled-coil domain. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:314–8. [PubMed: 20179344]
98. Mason AB, Halbrooks PJ, James NG, Byrne SL, Grady JK, Chasteen ND, Bobst CE, Kaltashov IA, Smith VC, MacGillivray RT, Everse SJ. Structural and functional consequences of the substitution of glycine 65 with arginine in the N-lobe of human transferrin. *Biochemistry.* 2009; 48:1945–53. [PubMed: 19219998]
99. Garcia-Moreno B. Adaptations of proteins to cellular and subcellular pH. *J Biol.* 2009; 8:98. [PubMed: 20017887]
100. Mitra RC, Zhang Z, Alexov E. In silico modeling of pH-optimum of protein-protein binding. *Proteins.* 2011; 79:925–36. [PubMed: 21287623]
101. Chan P, Warwicker J. Evidence for the adaptation of protein pH-dependence to subcellular pH. *BMC Biol.* 2009; 7:69. [PubMed: 19849832]
102. Kulkarni MV, Tettamanzi MC, Murphy JW, Keeler C, Myszka DG, Chayen NE, Lolis EJ, Hodsdon ME. Two independent histidines, one in human prolactin and one in its receptor, are critical for pH-dependent receptor recognition and activation. *J Biol Chem.* 2010; 285:38524–33. [PubMed: 20889499]
103. Wang L, Witham S, Zhang Z, Li L, Hodsdon M, Alexov E. In Silico Investigation of pH-Dependence of Prolactin and Human Growth Hormone Binding to Human Prolactin Receptor. 2013
104. Reichold M, Zdebek AA, Lieberer E, Rapedius M, Schmidt K, Bandulik S, Sterner C, Tegtmeyer I, Penton D, Baukrowitz T, Hulton SA, Witzgall R, Ben-Zeev B, Howie AJ, Kleta R, Bockenbauer D, Warth R. KCNJ10 gene mutations causing EAST syndrome (epilepsy, ataxia, sensorineural deafness, and tubulopathy) disrupt channel function. *Proc Natl Acad Sci U S A.* 2010; 107:14490–5. [PubMed: 20651251]

105. Hanemann CO, D'Urso D, Gabreels-Festen AA, Muller HW. Mutation-dependent alteration in cellular distribution of peripheral myelin protein 22 in nerve biopsies from Charcot-Marie-Tooth type 1A. *Brain*. 2000; 123 (Pt 5):1001–6. [PubMed: 10775544]
106. Petukh M, Stefl S, Alexov E. The Role of Protonation States in Ligand-Receptor Recognition and Binding. *Curr Pharm Des*. 2012
107. Nelson MT, Humphrey W, Gursoy A, Dalke A, Kalé LV, Skeel RD, Schulten K. NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *International Journal of High Performance Computing Applications*. 1996; 10:251–268.
108. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*. 1983; 4:187–217.
109. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem*. 2005; 26:1668–88. [PubMed: 16200636]
110. Christen M, Hunenberger PH, Bakowies D, Baron R, Burgi R, Geerke DP, Heinz TN, Kastenholz MA, Krautler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF. The GROMOS software for biomolecular simulation: GROMOS05. *J Comput Chem*. 2005; 26:1719–51. [PubMed: 16211540]
111. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*. 1995; 91:43–56.
112. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *Journal of Molecular Biology*. 1999; 285:1735–1747. [PubMed: 9917408]
113. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*. 2007; 35:W522–W525. [PubMed: 17488841]
114. Durrant JD, McCammon JA. HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. *J Mol Graph Model*. 2011; 31:5–9. [PubMed: 21880522]
115. Murray AE. A routine method for the quantification of physical change in melanocytic naevi using digital image processing. *J Audiov Media Med*. 1988; 11:52–7. [PubMed: 3243943]
116. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pK(a)s in proteins. *Biophys J*. 2002; 83:1731–48. [PubMed: 12324397]
117. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins*. 2005; 61:704–21. [PubMed: 16231289]
118. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res*. 2005; 33:W368–71. [PubMed: 15980491]
119. Maier T, Leibundgut M, Boehringer D, Ban N. Structure and function of eukaryotic fatty acid synthases. *Q Rev Biophys*. 2010; 43:373–422. [PubMed: 20731893]
120. Svetlov V, Nudler E. Macromolecular micromovements: how RNA polymerase translocates. *Curr Opin Struct Biol*. 2009; 19:701–7. [PubMed: 19889534]
121. Kokkinidis M, Glykos NM, Fadouloglou VE. Protein flexibility and enzymatic catalysis. *Adv Protein Chem Struct Biol*. 2012; 87:181–218. [PubMed: 22607756]
122. Chow MK, Lomas DA, Bottomley SP. Promiscuous beta-strand interactions and the conformational diseases. *Curr Med Chem*. 2004; 11:491–9. [PubMed: 14965229]
123. Oldfield CJ, Xue B, Van YY, Ulrich EL, Markley JL, Dunker AK, Uversky VN. Utilization of protein intrinsic disorder knowledge in structural proteomics. *Biochim Biophys Acta*. 2013; 1834:487–98. [PubMed: 23232152]
124. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn*. 2012; 30:137–49. [PubMed: 22702725]

125. Li M, Shoemaker BA, Thangudu RR, Ferraris JD, Burg MB, Panchenko AR. Mutations in DNA-Binding Loop of NFAT5 Transcription Factor Produce Unique Outcomes on Protein-DNA Binding and Dynamics. *J Phys Chem B*.
126. Ponder J, Richards F. TINKER molecular modeling package. *J Comput Chem*. 1987; 8:1016–1024.
127. Kumar A, Purohit R. Computational investigation of pathogenic nsSNPs in CEP63 protein. *Gene*. 2012; 503:75–82. [PubMed: 22555018]
128. George Priya Doss C, Rajith B. A New Insight into Structural and Functional Impact of Single-Nucleotide Polymorphisms in PTEN Gene. *Cell Biochem Biophys*. 2012
129. Nyon MP, Segu L, Cabrita LD, Levy GR, Kirkpatrick J, Roussel BD, Patschull AO, Barrett TE, Ekeowa UI, Kerr R, Waudby CA, Kalsheker N, Hill M, Thalassinou K, Lomas DA, Christodoulou J, Gooptu B. Structural dynamics associated with intermediate formation in an archetypal conformational disease. *Structure*. 2012; 20:504–12. [PubMed: 22405009]
130. Wang B, Wen X, Qin X, Wang Z, Tan Y, Shen Y, Xi Z. Quantitative Structural Insight into Human Variegate Porphyria Disease. *J Biol Chem*. 2013; 17:11731–40. [PubMed: 23467411]
131. Nicchia GP, Ficarella R, Rossi A, Giangreco I, Nicolotti O, Carotti A, Pisani F, Estivill X, Gasparini P, Svelto M, Frigeri A. D184E mutation in aquaporin-4 gene impairs water permeability and links to deafness. *Neuroscience*. 2011; 197:80–8. [PubMed: 21952128]
132. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J Mol Biol*. 2002; 323:573–84. [PubMed: 12381310]
133. Uversky VN, Oldfield CJ, Midic U, Xie H, Xue B, Vucetic S, Iakoucheva LM, Obradovic Z, Dunker AK. Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics*. 2009; 10(Suppl 1):S7. [PubMed: 19594884]
134. Bugg CW, Isas JM, Fischer T, Patterson PH, Langen R. Structural features and domain organization of huntingtin fibrils. *J Biol Chem*. 2012; 287:31739–46. [PubMed: 22801429]
135. Hu Y, Liu Y, Jung J, Dunker AK, Wang Y. Changes in predicted protein disorder tendency may contribute to disease risk. *BMC Genomics*. 2011; 12(Suppl 5):S2.
136. Vacic V, Iakoucheva LM. Disease mutations in disordered regions--exception to the rule? *Mol Biosyst*. 2012; 8:27–32. [PubMed: 22080206]
137. Sakakura M, Hadziselimovic A, Wang Z, Schey KL, Sanders CR. Structural basis for the Trembler-J phenotype of Charcot-Marie-Tooth disease. *Structure*. 2011; 19:1160–9. [PubMed: 21827951]
138. van der Kamp MW, Daggett V. Pathogenic mutations in the hydrophobic core of the human prion protein can promote structural instability and misfolding. *J Mol Biol*. 2010; 404:732–48. [PubMed: 20932979]
139. Di Marino D, Achsel T, Lacoux C, Falconi M, Bagni C. Molecular dynamics simulations show how the FMRP Ile304Asn mutation destabilizes the KH2 domain structure and affects its function. *J Biomol Struct Dyn*. 2013
140. Kass I, Knaupp AS, Bottomley SP, Buckle AM. Conformational properties of the disease-causing Z variant of alpha1-antitrypsin revealed by theory and experiment. *Biophys J*. 2012; 102:2856–65. [PubMed: 22735536]
141. Jahandideh S, Zhi D. Systematic investigation of predicted effect of nonsynonymous SNPs in human prion protein gene: a molecular modeling and molecular dynamics study. *J Biomol Struct Dyn*. 2013
142. Tokuriki N, Stricher F, Serrano L, Tawfik DS. How protein stability and new functions trade off. *PLoS Comput Biol*. 2008; 4:e1000002. [PubMed: 18463696]
143. Dessailly BH, Lensink MF, Wodak SJ. Relating destabilizing regions to known functional sites in proteins. *BMC Bioinformatics*. 2007; 8:141. [PubMed: 17470296]
144. Elcock AH. Prediction of functionally important residues based solely on the computed energetics of protein structure. *J Mol Biol*. 2001; 312:885–96. [PubMed: 11575940]
145. Studer RA, Dessailly BH, Orengo CA. Residue mutations and their impact on protein structure and function: detecting beneficial and pathogenic changes. *Biochem J*. 2013; 449:581–94. [PubMed: 23301657]

146. Meiering EM, Serrano L, Fersht AR. Effect of active site residues in barnase on activity and stability. *J Mol Biol.* 1992; 225:585–9. [PubMed: 1602471]
147. Shoichet BK, Baase WA, Kuroki R, Matthews BW. A relationship between protein stability and protein function. *Proc Natl Acad Sci U S A.* 1995; 92:452–6. [PubMed: 7831309]
148. Pazos F, Sternberg MJ. Automated prediction of protein function and detection of functional sites from structure. *Proc Natl Acad Sci U S A.* 2004; 101:14754–9. [PubMed: 15456910]
149. Lopez G, Valencia A, Tress ML. firestar--prediction of functionally important residues using structural templates and alignment reliability. *Nucleic Acids Res.* 2007; 35:W573–7. [PubMed: 17584799]
150. Shoemaker BA, Zhang D, Tyagi M, Thangudu RR, Fong JH, Marchler-Bauer A, Bryant SH, Madej T, Panchenko AR. IBIS (Inferred Biomolecular Interaction Server) reports, predicts and integrates multiple types of conserved interactions for proteins. *Nucleic Acids Res.* 2012; 40:D834–40. [PubMed: 22102591]
151. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. *PLoS Comput Biol.* 2009; 5:e1000585. [PubMed: 19997483]
152. Zhang Z, Zheng Y, Petukh M, Pegg A, Ikeguchi Y, Alexov E. Enhancing human spermine synthase activity by engineered mutations. *PLoS Comput Biol.* 2013; 9:e1002924. [PubMed: 23468611]
153. Martin GS. The road to Src. *Oncogene.* 2004; 23:7910–7. [PubMed: 15489909]
154. Izarzugaza JM, Redfern OC, Orengo CA, Valencia A. Cancer-associated mutations are preferentially distributed in protein kinase functional sites. *Proteins.* 2009; 77:892–903. [PubMed: 19626714]
155. Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE. Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. *Cell.* 2012; 149:860–70. [PubMed: 22579287]
156. Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, Eck MJ. Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell.* 2007; 11:217–27. [PubMed: 17349580]
157. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell.* 2006; 125:1137–49. [PubMed: 16777603]
158. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M, Eck MJ. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A.* 2008; 105:2070–5. [PubMed: 18227510]
159. Dixit A, Verkhivker GM. Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. *PLoS Comput Biol.* 2009; 5:e1000487. [PubMed: 19714203]
160. Hashimoto K, Rogozin IB, Panchenko AR. Oncogenic potential is related to activating effect of cancer single and double somatic mutations in receptor tyrosine kinases. *Hum Mutat.* 2012; 33:1566–75. [PubMed: 22753356]
161. Colgin LM, Hackmann AF, Emond MJ, Monnat RJ Jr. The unexpected landscape of in vivo somatic mutation in a human epithelial cell lineage. *Proc Natl Acad Sci U S A.* 2002; 99:1437–42. [PubMed: 11818556]
162. Chen JM, Ferec C, Cooper DN. Closely spaced multiple mutations as potential signatures of transient hypermutability in human genes. *Hum Mutat.* 2009; 30:1435–48. [PubMed: 19685533]
163. Chen Z, Feng J, Buzin CH, Sommer SS. Epidemiology of doublet/multiplier mutations in lung cancers: evidence that a subset arises by chronocoordinate events. *PLoS One.* 2008; 3:e3714. [PubMed: 19005564]
164. Meng L, Lin L, Zhang H, Nassiri M, Morales AR, Nadji M. Multiple mutations of the p53 gene in human mammary carcinoma. *Mutat Res.* 1999; 435:263–9. [PubMed: 10606817]
165. Haass C, Lemere CA, Capell A, Citron M, Seubert P, Schenk D, Lannfelt L, Selkoe DJ. The Swedish mutation causes early-onset Alzheimer's disease by beta-secretase cleavage within the secretory pathway. *Nat Med.* 1995; 1:1291–6. [PubMed: 7489411]

166. Yasuda M, Shabbeer J, Benson SD, Maire I, Burnett RM, Desnick RJ. Fabry disease: characterization of alpha-galactosidase A double mutations and the D313Y plasma enzyme pseudodeficiency allele. *Hum Mutat.* 2003; 22:486–92. [PubMed: 14635108]
167. Gong S, Blundell TL. Structural and functional restraints on the occurrence of single amino acid variations in human proteins. *PLoS One.* 2010; 5:e9186. [PubMed: 20169194]
168. Zhang Z, Norris J, Kalscheuer V, Wood T, Wang L, Schwartz C, Alexov E, Van Esch H. A Y328C missense mutation in spermine synthase causes a mild form of Snyder-Robinson syndrome. *Hum Mol Genet.* 2013
169. Pandit RA, Chen CJ, Butt TA, Islam N. Identification and analysis of a novel mutation in PEPD gene in two Kashmiri sibs with prolidase enzyme deficiency. *Gene.* 2013; 516:316–9. [PubMed: 23287645]
170. Attanasio F, Convertino M, Magno A, Caflisch A, Corazza A, Haridas H, Esposito G, Cataldo S, Pignataro B, Milardi D, Rizzarelli E. Carnosine Inhibits Abeta42 Aggregation by Perturbing the H-Bond Network in and around the Central Hydrophobic Cluster. *Chembiochem.* 2013; 14:583–92. [PubMed: 23440928]
171. Murray MM, Krone MG, Bernstein SL, Baumketner A, Condron MM, Lazo ND, Teplow DB, Wyttenbach T, Shea JE, Bowers MT. Amyloid beta-protein: experiment and theory on the 21–30 fragment. *J Phys Chem B.* 2009; 113:6041–6. [PubMed: 19341254]
172. Mizuguchi M, Yokoyama T, Nabeshima Y, Kawano K, Tanaka I, Niimura N. Quaternary structure, aggregation and cytotoxicity of transthyretin. *Amyloid.* 2012; 19(Suppl 1):5–7. [PubMed: 22439750]
173. Liou B, Kazimierczuk A, Zhang M, Scott CR, Hegde RS, Grabowski GA. Analyses of Variant Acid -Glucosidases: EFFECTS OF GAUCHER DISEASE MUTATIONS. *Journal of Biological Chemistry.* 2006; 281:4242–4253. [PubMed: 16293621]
174. Bal NC, Sharon A, Gupta SC, Jena N, Shaikh S, Gyorke S, Periasamy M. The catecholaminergic polymorphic ventricular tachycardia mutation R33Q disrupts the N-terminal structural motif that regulates reversible calsequestrin polymerization. *J Biol Chem.* 2010; 285:17188–96. [PubMed: 20353949]
175. Kumar A, Purohit R. Computational screening and molecular dynamics simulation of disease associated nsSNPs in CENP-E. *Mutat Res.* 2012; 738–739:28–37.
176. Vieyra D, Senger DL, Toyama T, Muzik H, Brasher PM, Johnston RN, Riabowol K, Forsyth PA. Altered subcellular localization and low frequency of mutations of ING1 in human brain tumors. *Clin Cancer Res.* 2003; 9:5952–61. [PubMed: 14676120]
177. Kunishima S, Matsushita T, Kojima T, Sako M, Kimura F, Jo EK, Inoue C, Kamiya T, Saito H. Immunofluorescence analysis of neutrophil nonmuscle myosin heavy chain-A in MYH9 disorders: association of subcellular localization with MYH9 mutations. *Lab Invest.* 2003; 83:115–22. [PubMed: 12533692]
178. Yokoyama T, Mizuguchi M, Nabeshima Y, Kusaka K, Yamada T, Hosoya T, Ohhara T, Kurihara K, Tomoyori K, Tanaka I, Niimura N. Hydrogen-bond network and pH sensitivity in transthyretin: Neutron crystal structure of human transthyretin. *J Struct Biol.* 2012; 177:283–90. [PubMed: 22248451]
179. Garai K, Baban B, Frieden C. Self-association and stability of the ApoE isoforms at low pH: implications for ApoE-lipid interactions. *Biochemistry.* 2011; 50:6356–64. [PubMed: 21699199]
180. Imbard A, Boutron A, Vequaud C, Zater M, de Lonlay P, de Baulny HO, Barnerias C, Mine M, Marsac C, Saudubray JM, Brivet M. Molecular characterization of 82 patients with pyruvate dehydrogenase complex deficiency. Structural implications of novel amino acid substitutions in E1 protein. *Mol Genet Metab.* 2011; 104:507–16. [PubMed: 21914562]
181. Smith AT, Su Y, Stevens DJ, Majtan T, Kraus JP, Burstyn JN. Effect of the disease-causing R266K mutation on the heme and PLP environments of human cystathionine beta-synthase. *Biochemistry.* 2012; 51:6360–70. [PubMed: 22738154]
182. Padhi AK, Kumar H, Vasaikar SV, Jayaram B, Gomes J. Mechanisms of loss of functions of human angiogenin variants implicated in amyotrophic lateral sclerosis. *PLoS One.* 2012; 7:e32479. [PubMed: 22384259]

183. Musayev FN, Di Salvo ML, Saavedra MA, Contestabile R, Ghatge MS, Haynes A, Schirch V, Safo MK. Molecular basis of reduced pyridoxine 5 -phosphate oxidase catalytic activity in neonatal epileptic encephalopathy disorder. *J Biol Chem.* 2009; 284:30949–56. [PubMed: 19759001]
184. Wang D, Robertson IM, Li MX, McCully ME, Crane ML, Luo Z, Tu AY, Daggett V, Sykes BD, Regnier M. Structural and functional consequences of the cardiac troponin C L48Q Ca(2+)-sensitizing mutation. *Biochemistry.* 2012; 51:4473–87. [PubMed: 22591429]
185. Kumar A, Rajendran V, Sethumadhavan R, Purohit R. Evidence of Colorectal Cancer-Associated Mutation in MCAK: A Computational Report. *Cell Biochem Biophys.* 2013
186. Tang M, Facchiano A, Rachamadugu R, Calderon F, Mao R, Milanese L, Marabotti A, Lai K. Correlation assessment among clinical phenotypes, expression analysis and molecular modeling of 14 novel variations in the human galactose-1-phosphate uridylyltransferase gene. *Hum Mutat.* 2012; 33:1107–15. [PubMed: 22461411]
187. Acsadi G, Moore SA, Cheron A, Delalande O, Bennett L, Kupsy W, El-Baba M, Le Rumeur E, Hubert JF. Novel mutation in spectrin-like repeat 1 of dystrophin central domain causes protein misfolding and mild Becker muscular dystrophy. *J Biol Chem.* 2012; 287:18153–62. [PubMed: 22453924]
188. Uversky, VN.; Iakoucheva, LM.; Dunker, AK. eLS. John Wiley & Sons, Ltd; 2001. *Protein Disorder and Human Genetic Disease.*
189. Doring G. Serine proteinase inhibitor therapy in alpha(1)-antitrypsin inhibitor deficiency and cystic fibrosis. *Pediatr Pulmonol.* 1999; 28:363–75. [PubMed: 10536068]
190. Zhang Z, Liu M, Li B, Wang Y, Yue J, Liang L, Sun J. Exploring the mechanism of a regulatory SNP of KLK3 by molecular dynamics simulation. *J Biomol Struct Dyn.* 2013; 31:426–40. [PubMed: 22877366]
191. Kote-Jarai Z, Amin Al Olama A, Leongamornlert D, Tymrakiewicz M, Saunders E, Guy M, Giles GG, Severi G, Southey M, Hopper JL, Sit KC, Harris JM, Batra J, Spurdle AB, Clements JA, Hamdy F, Neal D, Donovan J, Muir K, Pharoah PD, Chanock SJ, Brown N, Benlloch S, Castro E, Mahmud N, O'Brien L, Hall A, Sawyer E, Wilkinson R, Easton DF, Eeles RA. Identification of a novel prostate cancer susceptibility variant in the KLK3 gene transcript. *Hum Genet.* 2011; 129:687–94. [PubMed: 21465221]
192. O'Mara TA, Nagle CM, Batra J, Kedda MA, Clements JA, Spurdle AB. Kallikrein-related peptidase 3 (KLK3/PSA) single nucleotide polymorphisms and ovarian cancer survival. *Twin Res Hum Genet.* 2011; 14:323–7. [PubMed: 21787114]
193. Singaraja RR, Visscher H, James ER, Chroni A, Coutinho JM, Brunham LR, Kang MH, Zannis VI, Chimini G, Hayden MR. Specific mutations in ABCA1 have discrete effects on ABCA1 function and lipid phenotypes both in vivo and in vitro. *Circ Res.* 2006; 99:389–97. [PubMed: 16873719]
194. Nagao K, Zhao Y, Takahashi K, Kimura Y, Ueda K. Sodium taurocholate-dependent lipid efflux by ABCA1: effects of W590S mutation on lipid translocation and apolipoprotein A-I dissociation. *J Lipid Res.* 2009; 50:1165–72. [PubMed: 19202195]
195. Wang LJ, Wang J, Li N, Ge L, Li BL, Song BL. Molecular characterization of the NPC1L1 variants identified from cholesterol low absorbers. *J Biol Chem.* 2011; 286:7397–408. [PubMed: 21189420]
196. De Jong JC, Van Der Vliet WA, Van Den Heuvel LP, Willems PH, Knoers NV, Bindels RJ. Functional expression of mutations in the human NaCl cotransporter: evidence for impaired routing mechanisms in Gitelman's syndrome. *J Am Soc Nephrol.* 2002; 13:1442–8. [PubMed: 12039972]

- Missense mutations affect protein stability, hydrogen bonds, dynamics and activity and cause diseases
- Any deviation away from wild type characteristics can be deleterious
- Changes in stability, dynamics, hydrogen bonds and activity upon mutations are interconnected
- Rare mutations are typically more deleterious than common variants

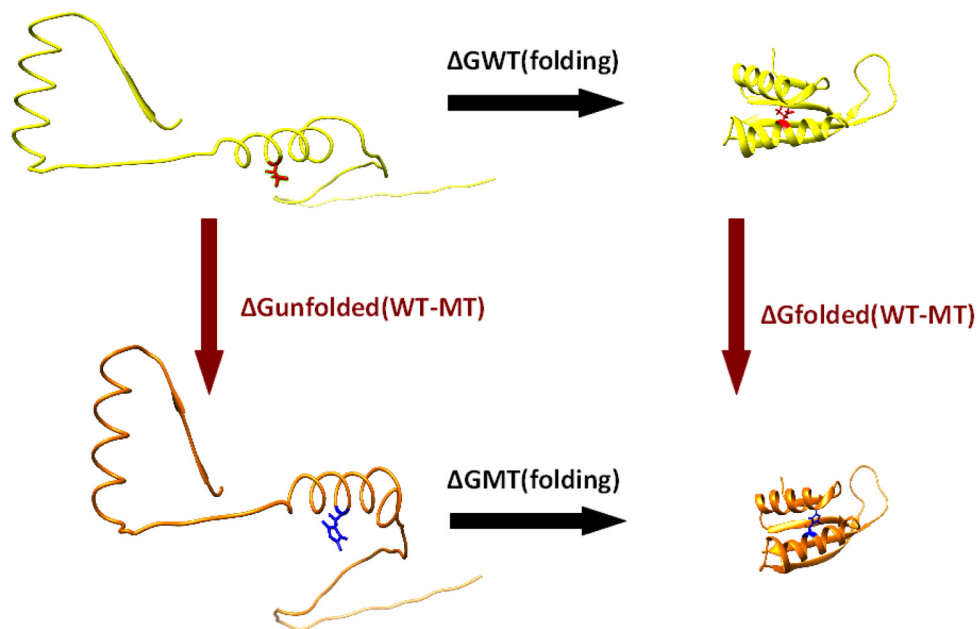


Figure 1.

The change in the folding free energy (ΔG) may be evaluated using two different methods: (1) using the difference in the folding free energy values calculated from the transition from an unfolded to a folded state of the wild type ($\Delta G^{WT}(\text{folding})$) minus the mutant type ($\Delta G^{MT}(\text{folding})$) shown with black arrows; or (2) using the difference in the folding free energy between the folded state of the wild type and mutant type ($\Delta G^{\text{folded}}(\text{WT-MT})$) minus the folding free energy between the unfolded state of the wild type and mutant type ($\Delta G^{\text{unfolded}}(\text{WT-MT})$) shown with red arrows.

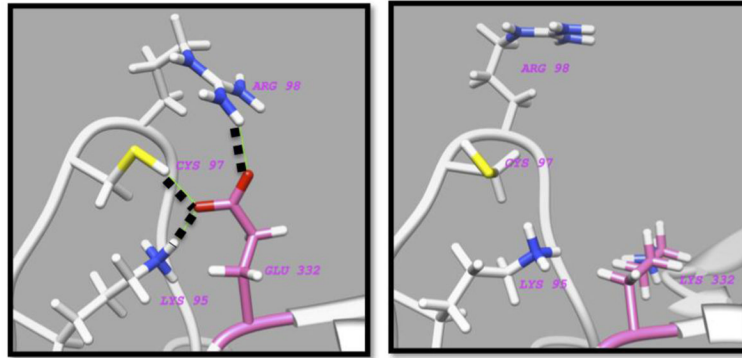


Figure 2. The region at the mutation site N550 of the wild type 3D structure of Isoform 3 Lactosylceramide alpha-2,3-sialyltransferase and mutant model which was built using 2wmlA.pdb as a template: (a) the wild type with N550 in the left panel and (b) the mutant N550K in the right panel, both centered at the mutation site. The hydrogen bonds between N550 and its neighbors in the wild type are shown with dashed black lines.

Table 1
The list of servers and programs to predict the effect of mutations on protein molecular characteristics. (

See ref 63 and 69 for other methods and comparison between the methods.)

Name	Input	Description	URL
MuStab	Sequence	Predicts decrease/increase of stability upon mutation using support vector machine	http://bioinfo.ggc.org/mustab/
MUPro	Sequence	Predicts decrease/increase of stability upon mutation using support vector machine	http://mupro.proteomics.ics.uci.edu/
SIFT	Sequence	Estimates deleterious effect of mutation using sequence homology and site conservation	http://sift-dna.org
MutPred	Sequence	Estimates deleterious effect of mutation using SIFT and gain/loss of structural or functional properties predicted from sequences	http://mutpred.mutdb.org/
SNPdbe	Sequence	Pre-calculated database of effects of known mutations calculated based on the neural network method and SIFT	http://www.roslab.org/services/snpdbe/
Polyphen(2)	Sequence/structure	Estimates deleterious effect of mutation using sequence homology, site conservation and structural features	http://genetics.bwh.harvard.edu/pph2/
I - Mutant 2.0	Sequence/structure	Estimates G upon mutation using support vector machine	http://gpcr2.biocomp.unibo.it/%7Eemidio/I-Mutant/I-Mutant.htm
Site Directed Mutator (SDM)	Structure	G upon mutation using statistical potential energy function	http://www-crysl.bioc.cam.ac.uk/~sdm/sdm.php
CUPSAT	Structure	Estimates G upon mutation using mean force atom pair and torsion angle potentials	http://cupsat.tu-bs.de/
FOLD-X	Structure	Estimates G upon mutation using empirical force field	http://fold-x.embl-heidelberg.de
AUTO-MUTE	Structure	Estimates G upon mutation using knowledge-based statistical contact potential and machine learning methods	http://proteins.gmu.edu/automute
ERIS	Structure	Estimates G upon mutation using physical force field with atomic modeling	http://dokhlab.unc.edu/tools/eris/
PopMuSIC 2.0	Structure	Estimates G upon mutation using a combination of statistical potential and neural networks	http://babylone.ulb.ac.be/popmusic
CC/PBSA	Structure	Poisson-Boltzmann Calculations combined with surface accessibility term to calculate the energy on an ensemble of structures generated with Concoord	http://ccpbdsa.bioinformatik.unisaarland.de/
HOPE	Structure/sequence	Uses experimentally determined or homology-based structures with conjunction with Whatif to make predictions	http://www.cmbi.ru.nl/hope/

Table 2

Diseases associated with changes in stability as a result of mutations

Change In Stability	Disease	Gene/Mutation	Reference
Destabilizing	Muscular Diseases	LMNA; multiple sites	28
	Von Willebrand Disease	VWF A2; R1597W, M1528V	29
	Retinal Diseases	Rhodopsin; multiple sites	30
	Hemophagocytic Lymphohistiocytosis	Perforin; multiple sites	31; 32
	Neurodegenerative/Prion Diseases	Prion; D178N	36
	Autosomal recessive Parkinson's Disease	PINK1; multiple sites	37; 39
	Familial Alzheimer's	Amyloid protein; Glu-22, Asp-23	40
	Inactivation of tumor suppressants	Multiple genes; multiple sites	41
	Snyder-Robinson Syndrome	Spermine Synthase; G56S, V132G, I150T, Y328C	46; 168
Stabilizing	Mental Disorder	CLIC2; H101Q	43

Table 3

Diseases resulting from changes in the hydrogen bond network as a consequence of mutations

Effect on	Disease	Gene/Mutation	Reference
Structure	Prolidase Deficiency	PEPD; 304insA	169
	Aldosterone Synthase Deficiency	CYP11B; R374W	88
	Mitochondrial Fatty Acid Oxidation Disorder	CPT1; R595W	89
	Snyder-Robinson Syndrome	Spermine synthase; I150T	46
	Cardiovascular Disease Predisposition	Human lipoxygenase; T560M	91
	Congenital Hereditary Cataract Disease	B1-crystallin; S129R	92
Aggregation/Folding	Alzheimer's Disease	amyloid- peptide aggregation; multiple sites	170
	Familial Alzheimer's Disease	Fragment of amyloid beta- protein; multiple sites	171
	Amyloidosis (with severe cardiomyopathy)	Aggregation/cytotoxicity of transthyretin; S112I	172
Flexibility	Aneuploidy and Solid Tumors	CEP63; L61P	127
	Gaucher disease	acid- glucosidase; N370S	173
	Lethal Catecholaminergic Polymorphic Ventricular Tachycardia	Calsequestrin; R33Q	174
	Promotion of Aneuploidy and Tumorigenesis	Centromere-associated protein-E; Y63H	175
Cellular Localization	Brain Tumors	ING1B; p33	176
	May-Hegglin anomaly, Sebastian syndrome, and Fechtner syndrome	MYH9; multiple sites	177
pH	Human Amyloidosis	TTR; several residues	178
	Mental Retardation	CLIC2; H101Q	44
	Alzheimer's Disease	Apolipoprotein E, apoE4 isoform	179
Activity	Pyruvate Dehydrogenase Complex	pyruvate dehydrogenase	180
	Deficiency	E1 subunit protein; multiple sites	
	Classical Homocystinuria	cystathionine -synthase; R266K	181
	Amyotrophic Lateral Sclerosis Disease	angiogenin protein; K17I	182
	Neonatal Epileptic Encephalopathy	pyridoxine 5 -phosphate oxidase; R229W	183
	Snyder Robinson Syndrome	spermine synthase; I150T	90

Table 4

Effects of disease mutations on conformational dynamics

Effect	Disease Associated	Gene/Mutation	Reference
Conformational Flexibility	Diabetes and Cancer	PTEN; H61D	128
	Multiple Diseases	cNtnC; L48Q	184
	Conformational Diseases	(1)-antitrypsin variant; K154N	129
	Variagate Porphyria Disorder	hPPO; R59Q, R59G	130
	Colorectal Cancer	mitotic centromere-associated kinesin protein; E403K	185
	Classic Galactosemia	GALT; multiple sites	186
	Becker Muscular Dystrophy	dystrophin protein; L427P	187
Disordered Regions	Parkinson disease	-synuclein; multiple sites	133
	Ovarian Cancer	HPV proteins; multiple sites	133
	Cancer	p53 protein; multiple sites	133
	Huntington's disease	huntingtin; multiple sites	133
Order-Disorder Transitions	Stargardt Disease (type 1)	Multiple mutations	188
	Adrenoleukodystrophy X-linked	Multiple mutations	188
	Androgen Insensitivity Syndrome	Multiple mutations	188
	Citrullinemia (type 1)	Multiple mutations	188
Aggregation	Alpha(1)-antitrypsin Deficiency and Cystic Fibrosis	human serine protease inhibitor (serpin) -1 antitrypsin; E342K	140; 189
	Creutzfeldt-Jakob disease and Gerstmann-Straussler-Scheinker Disease	prion protein; multiple sites	141
Local Motion	Cancers	KLK3; multiple sites	190; 191; 192

Table 5

Examples of disease-related mutations that affect protein activity and function.

Gene/Protein	Disease	Mutation	Effect	Reference
ATP-binding cassette transporter (ABCA1)	Tangier disease(reduction of HDL Cholesterol in plasma)	N1800H	Changes the localization of protein, resulting in intracellular accumulation of ABCA1 and loss of interaction with ApoA-I.	193
		W590S	Impairs dissociation of apoA-I from ABCA1, which may affect ATP-dependent lipid translocation.	194
Niemann–Pick C1-Like protein 1 (NPC1L1)	Hypocholesterolemia(reduction of LDL Cholesterol in the plasma)	T61M, S881L	Severely dysfunctional mutations; affect cholesterol uptake function, cholesterol-regulated recycling, glycosylation and stability of the protein.	195
		G402S, R1268H	Partially dysfunctional mutations; affect cholesterol-regulated recycling.	195
Thiazide-sensitive sodium-chloride cotransporter (SLC12A3)	Gitelman syndrome (salt wasting and low blood pressure)	G439S, G741R	Abolish sodium uptake function due to loss of localization at plasma membrane, which may be caused by incomplete glycosylation.	196
Chloride intracellular channel protein 2(CLIC2)	X-linked intellectual disability	H101Q	Stabilizes the protein, which in turn may increase ryanodine receptor activity.	44
Histone-lysine N-methyltransferase (MLL2)	Kabuki syndrome	W5065X, R5179H	May affect the epigenetic control of active chromatin states via methylation of histones.	18