Published in final edited form as:

Methods Mol Biol. 2017 ; 1550: 235–260. doi:10.1007/978-1-4939-6747-6_17.

Annotating mutational effects on proteins and protein interactions: designing novel and revisiting existing protocols

Minghui Li[#], Alexander Goncarenko[#], and Anna R. Panchenko^{*}

National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894

Abstract

In this review we describe a protocol to annotate the effects of missense mutations on proteins, their functions, stability, and binding. For this purpose we present a collection of the most comprehensive databases which store different types of sequencing data on missense mutations, we discuss their relationships, possible intersections, and unique features. Next, we suggest an annotation workflow using the state-of-the-art methods and highlight their usability, advantages, and limitations for different cases. Finally, we address a particularly difficult problem of deciphering the molecular mechanisms of mutations on proteins and protein complexes to understand the origins and mechanisms of diseases.

Keywords

Protein–protein interactions; Databases; Mutations

1 Introduction

The era of genome sequencing has unraveled a large number of human genetic variations, as illustrated by the milestone 1000 Genomes project [1]. Mutations and genetic recombinations may occur naturally during the cell division and at the same time may be caused by extrinsic factors. A single nucleotide substitution that results in a codon change encoding a different amino acid is called “missense point mutation” (called “mutation” thereafter). Germline missense mutations are passed to progeny, whereas somatic mutations are not inherited. Due to predominantly neutral effects of many genetic variations, they have accumulated in human population and can be responsible for many individual phenotypic traits in humans and may be used for genetic fingerprinting. Generally, a variant frequently occurring in a population is termed a polymorphisms and single nucleotide polymorphisms (SNP) are one of the most common types of genetic variations.

Missense mutations can render proteins nonfunctional and may be responsible for many diseases. From the clinical perspective, these non-neutral mutations affecting human health represent the main interest. For some diseases and genes, particularly following the Mendelian inheritance patterns, the causal genotype–phenotype relationship has been

^{*}corresponding author, panch@ncbi.nlm.nih.gov.

[#]authors contributed equally to this work

already established, while for complex polygenic diseases involving multiple factors it is still unknown. Moreover, genetic variants with low penetrance, weakly associated with disease phenotypes, can only be annotated for large samples and for many diseases their genetic determinants have to be discovered.

2 Materials

2.1 Software

1. Molecular dynamics packages: NAMD (<http://www.ks.uiuc.edu/Research/namd/>) [2] and CHARMM (<http://www.charmm.org/>) [3].
2. Structural visualization packages: VMD (<http://www.ks.uiuc.edu/Research/vmd/>) [4], Chimera (<http://www.cgl.ucsf.edu/chimera/>) [5] and CN3D (<http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>) [6].

2.2 Online Resources

1. FTP site at NCBI (ftp://ftp.ncbi.nih.gov/pub/panch/Mutation_binding/) includes examples of configuration and runfiles for VMD, NAMD, and CHARMM used in protocols of Subheadings 3.5 and 3.6.
2. Databases collecting human genetic variations, mutations, and data on their clinical relevance (Table 1).
3. Web servers for characterization of structural features of mutations (Table 2).
4. Web servers for predicting the phenotypic effects of mutations (Table 3)
5. Web servers and programs for predicting the effects of mutations on protein stability (Table 4).
6. Web servers and programs for predicting the effects of mutations on protein–protein binding affinity (Table 5).

3 Methods

Recent advances in experimental methods reduced the cost of the DNA sequencing and equipped many labs and hospitals with genotyping and sequencing assays, so that these data along with the clinical profiles of patients can be deposited into the central archival facilities. Some of them include the databases of Genotypes and Phenotypes (dbGaP) [7] and Genetic Testing Registry (GTR) [8] developed at NCBI, NIH. The genotypes are mapped to a reference genome assembly and typically include two major categories of variations: (1) single nucleotide polymorphisms (SNP) deposited in the dbSNP database [9] and (2) larger structural genetic variations in genomes, including long indels, inversions, and copy number variants (CNVs), deposited in the dbVar database [10]. The phenotypes in these databases are mainly organized by disease names or clinical conditions, and diseases are classified according to the Disease Ontology (DO) [11] and Medical Subject Headings (MeSH) terms. Additionally, the NCBI phenotype–genotype integrator (PheGenI) [12] merges genetic variations identified by genome-wide association studies (GWAS) with dbVar, OMIM, GTR, and dbSNP databases.

The clinical implications of genetic variations are recorded in several other databases. The human gene mutation (HGMD) [13] and Online Mendelian inheritance in Man (OMIM) [14] are the main databases that integrate the information on genetic Mendelian disorders with genes and mutations reported to be causative. ClinVar [15] is another archive, which collects the data on genetic variations from dbSNP and dbVar and integrates them with the available clinical evidence of these variants obtained from multiple studies. In ClinVar each variant is assigned a score which shows the consistency of the reported clinical effect in different studies. ClinVar annotates variants as: benign, likely benign, likely pathogenic, pathogenic, of uncertain significance, or variants with conflicting interpretations. Importantly, it also annotates the variants associated with individual drug effects.

Germline mutations constitute the bulk of mutations in OMIM, ClinVar, and HGMD, while somatic missense mutations are found predominantly in cancer cells and are not inherited. The information on location, tissue type, and frequency of somatic mutations in tumor samples together with the cancer type can be obtained from COSMIC[16] and TCGA databases [17], while other resources, for instance, MutDB [18] and SwissVar [19], integrate the data on germline and somatic mutations for specific genes and diseases.

Figure 1 describes a computational pipeline for exploring mutations and assessing their effects on protein structure, function, and interactions. Different sections of this chapter follow this pipeline and suggest the protocols to solve each specific problem.

3.1 Collecting and Integrating the Data on Human Polymorphisms and Mutations

First, we will describe a protocol for extracting clinically relevant mutations from ClinVar, COSMIC, and TCGA databases to further analyze them with respect to their clinical and functional impacts. Here we use human monoclonal Casitas B-lineage lymphoma c-CBL protein (CBL) as an example [20]. The links and references to web resources used in this protocol are listed in Table 1.

1. NCBI variation viewer (<http://www.ncbi.nlm.nih.gov/variation/view/>) [21] can be searched by gene name (CBL), Refseq accession (NP_005179), or Uniprot ID (P22681.2). A snapshot of variation viewer shows the genetic variation from ClinVar in the locus of CBL gene (Fig. 2a). Different variants in this viewer are depicted as separate tracks below the CBL gene locus. The ClinVar track displays multiple variants as boxes with the number of variants listed within each box. Variants with benign effect are shown using green color, whereas the purple boxes show pathogenic variants.
2. Additionally, the user can download all mutations from the COSMIC ftp server in VCF format (CosmicCodingMuts.vcf.gz) and display these mutations in a separate track in the variation viewer (choose menu “your data” and select “add track” option). Note that it is necessary to select the same genome assembly in both variation viewer and in COSMIC (e.g., GRCh37).
3. Each variant with a valid ClinVar annotation is linked to a corresponding dbVar or dbSNP record. Here we will focus on a single nucleotide variant of CBL gene with the dbSNP accession rs267606706. As shown in Fig. 2b, it is a missense

mutation where nucleotide substitution T->C results in p.Y371H amino acid substitution. There is clinical evidence that this variant can cause a Noonan syndrome-like disorder and/or juvenile myelomonocytic leukemia. The GTR studies cited in ClinVar show [22] that p.Y371H is a heterozygous germline substitution.

4. The dbSNP page for accession rs267606706 contains a link to “3D structure mapping” (found under the “NCBI resources” section) which points to the SNP3D page where several synonymous and missense variants are mapped onto the CBL protein structure. By default, this variant is selected, but it is possible to select other variants for display by clicking on the “Cn3D selected” button. Additionally, the SNP3D page contains a link “CD” that shows conserved domains from the CDD database mapped onto this protein [23]. Figure 2c depicts the structure of CBL using Cn3D with mutated Tyr residue side chain colored in yellow.
5. As was shown previously using ClinVar, the germline p.Y371H mutation may be linked to leukemia, however, many cancer mutations are somatic and therefore are not present in ClinVar. In order to explore somatic mutations we switch to cBioPortal, which allows exploring mutations from the TCGA and COSMIC databases. Open cBioPortal web page and submit a query “CBL” as the user-defined gene set. A summary for different types of cancers will be shown for the CBL gene. Open the second tab “Mutations”, which display mutations on CBL protein sequence. Figure 2d depicts missense CBL mutations mapped by the cBioPortal onto the corresponding CDD domain context. Additionally, the cBioPortal provides locations of mutations on protein structures. The blue footprints in Fig. 2d show that several structures cover CBL protein sequence and could be explored interactively. All missense mutations from COSMIC and TCGA databases are mapped on the representative structure (in green) and are displayed in the web browser (Fig. 2e).
6. Each mutation in cBioPortal “Mutations” tab is shown as a pin indicating its position in CBL protein sequence. The height of the pin corresponds to the number of known mutations. Place the mouse cursor over the Znf domain (zinc finger, shown in yellow) and over the first pin from the left in Znf domain. As a result, a window with a list of mutations and cancer types will pop up. For zinc finger domain the first two missense mutations, C384R and C384W, are associated with glioblastoma and melanoma, respectively. By searching for Cys384 residue in the table below and pressing the “3D” button, the structural location of this mutation is displayed. We will explain how this mutation could be interpreted in the context of molecular interactions in the next section.

3.2 Finding Protein Domains and Functional Sites Affected by Mutations

Now we consider the domain context of CBL mutations described in the previous section (p.Y371H and p.C384R) and annotate their impacts on protein interactions and signaling pathways. The web resources used in this protocol are listed in Table 2.

1. Evolutionarily conserved sites in a multiple sequence alignment usually correspond to functionally important sites and mutations in these sites can be harmful to protein function. If a protein of interest has a known PDB structure, conservation profiles can be downloaded from the PDBsum resource otherwise ConSurf server [24] can be used. In addition, the CDD server can offer functional annotations of sites in conserved protein domains, whereas IBIS server provides locations of binding sites for different types of binding partners (protein, small molecule, nucleic acid, ion, and peptides), and facilitates the mapping of a comprehensive biomolecular interaction network for a given protein query (with or without structure) [25–27]. Similar binding sites in IBIS are clustered together based on their sequence and structure conservation.
2. Open IBIS web page and search for 1FBV structure, chain A. Go to “protein–protein” tab and click on the balloon with the annotation “RING” domain to display interactions of the CBL RING domain with other domains/proteins (Fig. 3a). Binding sites are shown on CBL sequence as triangles and highly conserved binding sites are shown in red color. In the list of interaction partners below, the first conserved binding site cluster is formed between RING domain of CBL and ubiquitin conjugating enzyme from UBCc family. By clicking on the plus sign next to “UBCc”, one can see the corresponding binding site, the alignment of similar binding sites found in different CBL-UBCc complexes. By opening the link to Cn3D viewer, one can explore the interfaces and binding sites in these protein complexes (Fig. 3c).
3. The interaction graph in Fig. 3b shows the observed (black lines) and predicted interaction partners of CBL. Next, we will focus on interactions of CBL with zinc ions and UBCc ubiquitin ligase. Note that self-links indicate interactions between domains in CBL protein within or between CBL chains.
4. The structure in Fig. 3c shows that conserved cysteine residues in the binding site coordinate two zinc ions, apparently playing an important structural role. A substitution of cysteine by arginine disrupts the coordination of Zn, which affects the structure and stability of the zinc finger RING domain and may also affect CBL function.
5. Structural and biochemical analyses [28] show that CBL inactive state adopts an autoinhibited interaction. Substrate binding and Tyr371 phosphorylation activates CBL by producing a large conformational change in order to place the RING domain and UBCc in close proximity to the substrate necessary for effective catalysis. Importantly, mutation p.Y371H may prohibit activation of CBL by phosphorylation and may also affect the interaction with UBCc.
6. The impact of mutations on signaling pathways can be explored using recently developed PathiVar server [29]. Alternatively, it is possible to explore all pathways in which CBL interacts with UBCc using KEGG, Reactome, or NCBI Biosystems databases.

7. Search for 'CBL AND Y731 and "Homo sapiens"[Organism]' in the NCBI Biosystems portal. The first record will point to the "Regulation of signaling by CBL" pathway and the role of Y731 phosphorylation will be explained in the pathway description.

In addition to the NCBI IBIS server that allows analyzing the domain context and structural determinants of interactions, several other servers (DMDM, PolyDoms, and muPIT) provide the mapping of mutations on protein domains and protein structures.

3.3 Assessing If Mutations Have Damaging or Benign Effects on Proteins

Many methods have been proposed to predict the effects of missense mutations on proteins, classifying them as damaging or benign. These methods differ in terms of the properties of mutations or proteins used during the training procedure as well as in terms of the algorithms applied for decision-making. For example, machine learning algorithms train their models to distinguish known disease-associated from neutral mutations. Other methods do not explicitly train their models but almost all methods described in this section exploit the evolutionary conservation assuming that changes at conserved positions tend to be more deleterious. Besides sequence conservation various other sequence and structural features are used, which may include: changes in physicochemical properties between wild-type and substituted amino acid, structural features (mostly solvent accessibility), site mutability in DNA, and sequence context of the site.

An unbiased testing and comparison of machine learning methods is obviously an issue since they are usually trained on all available datasets of mutations and it is difficult to obtain a test set which would not overlap with training set. There are several experimental studies on variants in P53, LacI, and ABCA1 proteins which can be regarded as unbiased test cases. Comparisons of different methods on these experimental sets reported up to 70 % TPR (True positive rate) at 10 % FPR (False positive rate) [30]. Models trained to distinguish Mendelian variants with pronounced deleterious effects are more appropriate and accurate for predicting the effects of Mendelian mutations. The accuracy of these models is much higher than of those models that aim to assess the effects of mutations from complex diseases including cancer. This is evident from the evaluation of PolyPhen-2 performance which yields 0.70–0.77 TPR at 10 % FPR when trained on the HumDiv dataset (Mendelian disease mutations) and drops to 0.50–0.52 TPR for HumVar (all human disease causing mutations) trained models [31]. It should be mentioned that there are several methods that are trained to distinguish cancer mutations from neutral polymorphisms (Table 3); however, no existing method can accurately identify driver and passenger mutations within the pool of cancer mutations.

One of the most comprehensive comparisons of different methods to predict phenotypic effects of mutations was performed recently [32]. To avoid any bias in evaluation of these methods, most of which were trained on all available sets of disease mutations and neutral polymorphisms, the authors of this study tested different methods on an independent set of experimental studies. They concluded that there was a variation between various methods in terms of their accuracy and applicability, with SNP&GO and MuPred being the most reliable predictors. Interestingly, despite the fact that different methods use similar sets of features,

only half of their correctly predicted cases overlap [32]. Since this study was published, several new methods have been introduced (see Table 3). For example, in contrast to many methods, that assess amino acid frequency distribution in a given site of interest, a recently developed method PROVEAN accounts for the sequence context around the site of interest and poorly aligned regions/sites are assigned very low scores. Overall, the effect of alignment quality on the performance of all methods is largely undetermined but suspected to be very large. Therefore, a user-based construction of accurate alignments of homologous proteins would be very advantageous for accurate annotations of the effects of mutations.

3.4 Predicting the Impact of Mutations on Protein Stability

While methods, which provide a classification of damaging effects of mutations, are widely used by the genomics community, a new level of annotation is needed to offer an explanation of why and how these mutations damage proteins. Algorithms and servers described in the next several sections address these tasks. Proteins may evolve through the acquisition of new mutations, most of which are destabilizing but phenotypically neutral. Stability of a protein may be directly related to its functional activity and incorrect folding and decreased stability can be the major consequences of pathogenic missense mutations [33, 34]. However, protein stability is necessary but not sufficient for protein function, and proteins do not evolve to maximize their stability. Typically, the magnitude of effects of mutations on stability can be quantified by changes in unfolding free energy (ΔG_{fold}) (Fig. 4).

$$\Delta\Delta G_{\text{fold}} = \Delta G_{\text{fold}}^{\text{mut}} - \Delta G_{\text{fold}}^{\text{WT}} \quad (1)$$

Table 4 lists several state-of-the art methods for predicting the quantitative changes in unfolding free energy upon mutations and provides short descriptions and links to corresponding programs/servers. Methods described in this section differ in terms of energy functions, procedures used for optimization and sampling, and algorithms used for training, if applicable. Energy functions may vary from physics-based force fields, which describe fundamental physical forces between atoms, to knowledge-based potentials, which are based on statistical analyses of protein structures and residue properties. The majority of these methods require the coordinates of protein structures while methods like MuStab or iPTREE-STAB do not use structural data but their performance is also limited. The performance of different methods was evaluated in several studies [35–37]. In the first study [35] the following performance ranking was reported: EGAD>CC/PBSA>I-Mutant2.0>FoldX>Hunter>Rosetta with correlation coefficients between experimental and predicted ΔG values in the range of 0.59 and 0.26 and standard deviation in the range of 0.95 and 2.32 kcal mol⁻¹. However, the servers of the top performing methods EGAD and CC/PBSA are no longer available. In the second study I-Mutant3.0, Dmutant, and FoldX were found to be the most reliable predictors [36].

There are several servers to assess the effects of mutations on stability that are straightforward and easy to use. Here we present a protocol on how to use FoldX software.

1. Run RepairPDB module of FoldX to correct errors in the structure produced during the refinement (nonstandard angles or distances) (Runfile can be obtained from <http://foldxsuite.crg.eu/command/RepairPDB>).
2. Run BuildModel module of FoldX to introduce a mutation on the optimized wild-type structure <http://foldxsuite.crg.eu/command/BuildModel>. BuildModel module optimizes the configurations of the side chains of amino acids in the vicinity of the mutated site and calculates the difference in unfolding free energy (ΔG_{fold}) between mutant and repaired wild-type structure. The total unfolding free energy and each energy term can be obtained from the “Dif_BuildModel_*.fxout” output file.

3.5 Predicting the Effects of Mutations on Protein–Protein Binding Affinity

Crucial prerequisite for proper biological function is a protein’s ability to establish highly selective interactions with macromolecular partners. A missense mutation that affects protein interactions [38–40] may cause significant perturbations or complete abolishment of protein function, potentially leading to diseases. Typically, the change in binding free energy (ΔG_{bind}) is used to quantify the magnitude of mutational effects on protein–protein interactions (Fig. 4).

$$\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind}}^{\text{mut}} - \Delta G_{\text{bind}}^{\text{WT}} \quad (2)$$

The binding energy is calculated as a difference between the free energies of a complex AB and unbound proteins A and B :

$$\Delta G_{\text{bind}} = G^{\text{AB}} - G^{\text{A}} - G^{\text{B}} \quad (3)$$

There are very few methods that estimate actual G_{bind} values and these methods require all-atom or at least protein backbone atom coordinates of a wild-type and/or mutated protein. Some of the methods use coarse-grained predictors based on statistical or empirical potentials, others apply molecular mechanics force fields with different solvation models. For example, the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method and its derivatives have been shown to yield very good agreement between predicted and experimental values with correlation coefficients up to 0.69 [41]. For all methods, the right choices of minimization protocols, energy functions, and solvation models are crucial for achieving reasonable prediction accuracy. In addition, prediction accuracy strongly depends on the type of mutation and its location in a protein complex. For example, if a residue is located on the protein–protein interface, its mutation might have a larger effect on protein–protein interaction and binding affinity compared to a non-interfacial mutation [41]. A location of mutated sites can be mapped by SPPIDER (<http://sppider.cchmc.org/>) [42] or Meta-PPISP (<http://pipe.scs.fsu.edu/meta-ppisp.html>) [43] servers. These servers are recommended by two assessments of computational methods for predicting protein–protein

interaction sites [44, 45]. Users can also analyze structures and locations of mutations using software Chimera or VMD.

Below is a step-by-step protocol reported in our previous paper [41] to predict the impact of mutations on binding affinity. This protocol combines molecular mechanics force fields with statistical (BeAtMuSiC) and empirical (FoldX) energy functions. All files are provided via the ftp site ftp://ftp.ncbi.nih.gov/pub/panch/Mutation_binding. The improved version can be available from our MutaBind server <https://www.ncbi.nlm.nih.gov/projects/mutabind/>.

1. Install software VMD, NAMD, and CHARMM.
2. Download a structure for your protein of interest from the Protein Data Bank (PDB).
3. Add hydrogen atoms, a rectangular box (10 Å) of water molecules, and Na⁺ and Cl⁻ ions (ionic concentration of 150 mM) to the structure using VMD (“vmd.pgn”).
4. Carry out 5000-step energy minimization with harmonic restraints (with the force constant of 5 kcal mol⁻¹ Å⁻²) applied on the backbone atoms of all residues (“minimization1.conf”), followed by a 35,000-step energy minimization on the whole system (“minimization2.conf”) with NAMD program using the CHARMM27 force field.
5. Introduce a mutation using “mutator” plugin of VMD software on the final minimized model from **step 4**.
6. Run an additional 300-step minimization for the whole mutant structure (“minimization2.conf”).
7. Run CHARMM program using the last frame from **step 4** (for wild-type structure) and **step 6** (for mutant structure) to obtain van der Waals interaction energy (E_{vdw}), polar solvation energy of solute in water (G_{solv}) for the wild-type and mutant, and interface area (SA_{mut}) for mutant (Runfile is “binding_energy.str”).
8. Submit your structure from **step 2** and a mutation to BeAtMuSiC webserver (Table 5) to obtain the binding affinity change G_{BM} .
9. Run AnalyseComplex module of FoldX to obtain the binding affinity change G_{FD} . (<http://foldxsuite.crg.eu/command/AnalyseComplex>).
10. Obtain the binding affinity changes using the following combination of energy terms from [41]:

$$\Delta\Delta G_{bind} = \alpha\Delta\Delta E_{vdw} + \beta\Delta\Delta G_{solv} + \gamma\Delta SA_{mut} + \varepsilon\Delta\Delta G_{BM} + \lambda\Delta\Delta G_{FD} + \delta$$

$$\left(\begin{array}{l} \alpha=0.122, \beta=0.101, \gamma=0.043, \\ \varepsilon=0.446, \lambda=0.168 \text{ and } \delta=1.326 \end{array} \right).$$

Recently a new computational method MutaBind [46] was developed to evaluate the effects of mutations on protein–protein interactions (<http://www.ncbi.nlm.nih.gov/projects/>

[mutabind/](#)). The MutaBind method uses molecular mechanics force fields, statistical potentials, and fast side-chain optimization algorithms. It maps mutations on a protein complex structure, calculates the associated changes in binding affinity, determines the deleterious effect of a mutation, estimates the confidence of this prediction, and produces a mutant structural model for download.

3.6 Assessing the Changes in Protein Conformations and Hydrogen Bond Networks Induced by Mutations

Proteins may adopt different conformations along the pathway of a biochemical reaction and their intrinsic flexibility and ability to sample alternative conformations are crucial for protein function. Mutations might shift the equilibrium between different conformations (Fig. 4) and as a result, the most populated conformation of a mutated protein can be different in structure, stability, and functional activity from the wild-type conformation. It is extremely difficult to model structural changes in a protein backbone produced by mutations and large conformational shifts can be predicted correctly only in a few cases. In fact, most algorithms discussed in the previous sections do not account for the backbone flexibility. If several conformations of the same protein are available in the structural databank, all of them ideally should be used to provide a complete picture of dynamical and energetic effects of mutations [20].

Mutations can either change the global conformation of an entire molecule or have a localized effect in a small region. In a recent study of the NFAT5 transcription factor [47], different mutations from the same DNA-binding loop were analyzed and it was shown that effects of these mutations on protein dynamics and DNA binding were drastically different although they were located very close to each other in sequence and structure. Protein dynamics can be studied through performing molecular dynamics (MD) simulations using NAMD [2], CHARMM [3] and Amber [48] MD packages. NAMD, for example, is fast and easy to use; it can be applied with CHARMM or Amber force fields, whereas VMD or CHARMM packages can be used to analyze the MD trajectories produced by NAMD.

Changes in structure may also be assessed through the analyses of hydrogen bond networks and their differences between mutant and wild-type proteins since hydrogen bonds are important in determining protein stability. A mutation disrupting hydrogen bonds might have a significant impact on protein conformation, stability, and dynamics (reviewed in [49], Fig. 4). Hydrogen bonds can be calculated using HBOND (<http://caps.ncbs.res.in/iws/hbond.html>) [50] or PIC servers (<http://pic.mbu.iisc.ernet.in/>) [51] and visualized by Chimera.

Below is a step-by-step protocol to assess the conformational changes induced by mutations:

1. Download the structure and introduce a mutation using VMD (refer to **steps 2** and **5** from Subheading 3.5).
2. Build the model systems with VMD (Refer to **step 3** from Subheading 3.5).
3. Carry out the energy minimizations (Refer to **step 4** from Subheading 3.5). The number of minimization steps can be chosen based on the size of system.

4. Heat the system to 300 K over 300 ps with harmonic constraints applied to protein backbone atoms using NAMD (“heat.conf”).
5. Perform unconstrained MD simulation on the system with NAMD (“md.conf”).
6. Load MD trajectories into the VMD software to monitor the conformational changes and calculate the root-mean-square deviation (RMSD) between the wild-type and mutant structures.

Assessing the effects of mutations on hydrogen bond networks using Chimera.
7. Load your protein structure of interest into Chimera.
8. Select residues of interest and input “findhbond selRestrict both distSlop 0.35 angleSlop 60.0 saveFile filename” in the command line. The hydrogen bonds will be shown on the screen and details will be saved in the file “filename”. One can adjust the distance (distSlop) and angle (angleSlop) parameters in the definition of hydrogen bonds.
9. (Optional) Go to Tools -> Structural Analysis -> FindHBond to find the hydrogen bonds.

Acknowledgments

This work was supported by the Intramural Research Program of the National Library of Medicine.

References

1. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. Genomes Project C. A map of human genome variation from population-scale sequencing. *Nature*. 2010; 467(7319):1061–1073. [PubMed: 20981092]
2. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005; 26(16):1781–1802. [PubMed: 16222654]
3. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. Charmm – a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem*. 1983; 4(2):187–217.
4. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996; 14(1): 33–38. 27–38. [PubMed: 8744570]
5. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 25(13):1605–1612. [PubMed: 15264254]
6. Wang Y, Geer LY, Chappey C, Kans JA, Bryant SH. Cn3D: sequence and structure views for Entrez. *Trends Biochem Sci*. 2000; 25(6):300–302. [PubMed: 10838572]
7. Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. *Nat Genet*. 2007; 39(10):1181–1186. [PubMed: 17898773]
8. Rubinstein WS, Maglott DR, Lee JM, Kattman BL, Malheiro AJ, Ovetsky M, Hem V, Gorelenkov V, Song G, Wallin C, Husain N, Chitipiralla S, Katz KS, Hoffman D, Jang W, Johnson M, Karmanov F, Ukrainchik A, Denisenko M, Fomous C, Hudson K, Ostell JM. The NIH genetic testing registry: a new, centralized database of genetic tests to enable access to comprehensive

information and improve transparency. *Nucleic Acids Res.* 2013; 41(Database issue):D925–D935. [PubMed: 23193275]

9. Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* 1999; 9(8):677–679. [PubMed: 10447503]
10. Lappalainen I, Lopez J, Skipper L, Hefferon T, Spalding JD, Garner J, Chen C, Maguire M, Corbett M, Zhou G, Paschall J, Ananiev V, Flícek P, Church DM. DbVar and DGVa: public archives for genomic structural variation. *Nucleic Acids Res.* 2013; 41(Database issue):D936–D941. [PubMed: 23193291]
11. Kibbe WA, Arze C, Felix V, Mitraka E, Bolton E, Fu G, Mungall CJ, Binder JX, Malone J, Vasant D, Parkinson H, Schriml LM. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic Acids Res.* 2015; 43(Database issue):D1071–D1078. [PubMed: 25348409]
12. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, Feolo M, Hindorff LA. Phenotype-Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. *Eur J Hum Genet.* 2014; 22(1):144–147. [PubMed: 23695286]
13. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeyasinghe S, Krawczak M, Cooper DN. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat.* 2003; 21(6): 577–581. [PubMed: 12754702]
14. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.* 2015; 43(Database issue):D789–D798. [PubMed: 25428349]
15. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* 2014; 42(Database issue):D980–D985. [PubMed: 24234437]
16. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* 2011; 39(suppl 1):D945–D950. [PubMed: 20952405]
17. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. Cancer Genome Atlas Research N. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; 45(10):1113–1120. [PubMed: 24071849]
18. Singh A, Olowoyeye A, Baenziger PH, Dantzer J, Kann MG, Radivojac P, Heiland R, Mooney SD. MutDB: update on development of tools for the biochemical analysis of genetic variation. *Nucleic Acids Res.* 2008; 36(Database issue):D815–D819. [PubMed: 17827212]
19. Mottaz A, David FP, Veuthey AL, Yip YL. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics.* 2010; 26(6):851–852. [PubMed: 20106818]
20. Li M, Kales SC, Ma K, Shoemaker BA, Crespo-Barreto J, Cangelosi AL, Lipkowitz S, Panchenko AR. Balancing protein stability and activity in cancer: a new approach for identifying driver mutations affecting CBL ubiquitin ligase activation. *Cancer Res.* 2015; 76(3):561–571. [PubMed: 26676746]
21. NR Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2014; 42(Database issue):D7–D17. [PubMed: 24259429]
22. Perez B, Mechinaud F, Galambrun C, Ben Romdhane N, Isidor B, Philip N, Derain-Court J, Cassinat B, Lachenaud J, Kaltenbach S, Salmon A, Desiree C, Pereira S, Menot ML, Royer N, Fenneteau O, Baruchel A, Chomienne C, Verloes A, Cave H. Germline mutations of the CBL gene define a new genetic syndrome with predisposition to juvenile myelomonocytic leukaemia. *J Med Genet.* 2010; 47(10):686–691. [PubMed: 20543203]
23. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 2015; 43(D1):D222–D226. [PubMed: 25414356]

24. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* 2010; 38(suppl 2):W529–W533. [PubMed: 20478830]
25. Shoemaker BA, Zhang D, Thangudu RR, Tyagi M, Fong JH, Marchler-Bauer A, Bryant SH, Madej T, Panchenko AR. Inferred Biomolecular Interaction Server – a web server to analyze and predict protein interacting partners and binding sites. *Nucleic Acids Res.* 2010; 38(Database issue):D518–D524. [PubMed: 19843613]
26. Goncarenco A, Shaytan AK, Shoemaker BA, Panchenko AR. Structural perspectives on the evolutionary expansion of unique protein-protein binding sites. *Biophys J.* 2015; 109(6):1295–1306. [PubMed: 26213149]
27. 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(Database issue):D834–D840. [PubMed: 22102591]
28. Dou H, Buetow L, Hock A, Sibbet GJ, Vousden KH, Huang DT. Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. *Nat Struct Mol Biol.* 2012; 19(2):184–192. [PubMed: 22266821]
29. Hernansaiz-Ballesteros RD, Salavert F, Sebastian-Leon P, Aleman A, Medina I, Dopazo J. Assessing the impact of mutations found in next generation sequencing data over human signaling pathways. *Nucleic Acids Res.* 2015; 43(W1):W270–W275. [PubMed: 25883139]
30. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012; 7(10):e46688. [PubMed: 23056405]
31. 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(4):248–249. [PubMed: 20354512]
32. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011; 32(4):358–368. [PubMed: 21412949]
33. 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(11):1566–1575. [PubMed: 22753356]
34. Schleich JP, Narayan M, Alford C, Mittendorf KF, Carter BD, Li J, Sanders CR. Conformational stability and pathogenic mis-folding of the integral membrane protein PMP22. *J Am Chem Soc.* 2015; 137(27):8758–8768. [PubMed: 26102530]
35. Potapov V, Cohen M, Schreiber G. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel.* 2009; 22(9):553–560. [PubMed: 19561092]
36. Khan S, Vihinen M. Performance of protein stability predictors. *Hum Mutat.* 2010; 31(6):675–684. [PubMed: 20232415]
37. Zhang Z, Wang L, Gao Y, Zhang J, Zhenirovskyy M, Alexov E. Predicting folding free energy changes upon single point mutations. *Bioinformatics.* 2012; 28(5):664–671. [PubMed: 22238268]
38. 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(6):e66273. [PubMed: 23799087]
39. Teng S, Madej T, Panchenko A, Alexov E. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys J.* 2009; 96(6):2178–2188. [PubMed: 19289044]
40. Ghersi D, Singh M. Interaction-based discovery of functionally important genes in cancers. *Nucleic Acids Res.* 2014; 42(3):e18. [PubMed: 24362839]
41. Li M, Petukh M, Alexov E, Panchenko AR. Predicting the impact of missense mutations on protein-protein binding affinity. *J Chem Theory Comput.* 2014; 10(4):1770–1780. [PubMed: 24803870]
42. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins.* 2007; 66(3):630–645. [PubMed: 17152079]

43. Qin S, Zhou H-X. meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*. 2007; 23(24):3386–3387. [PubMed: 17895276]
44. Zhou H-X, Qin S. Interaction-site prediction for protein complexes: a critical assessment. *Bioinformatics*. 2007; 23(17):2203–2209. [PubMed: 17586545]
45. Porollo A, Meller J. Computational methods for prediction of protein-protein interaction sites. *Protein-Protein Interactions – Computational and Experimental Tools*. 2012; 472:3–26.
46. Li M, Simonetti FL, Goncarenco A, Panchenko AR. MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. *Nucleic Acids Res*. 2016 Jul 8; 44(W1):W494–501. [PubMed: 27150810]
47. 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*. 2013; 117(42):13226–13234. [PubMed: 23734591]
48. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem*. 2005; 26(16): 1668–1688. [PubMed: 16200636]
49. Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E. Molecular mechanisms of disease-causing missense mutations. *J Mol Biol*. 2013; 425(21):3919–3936. [PubMed: 23871686]
50. Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP. JOY: protein sequence-structure representation and analysis. *Bioinformatics*. 1998; 14(7):617–623. [PubMed: 9730927]
51. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. *Nucleic Acids Res*. 2007; 35(suppl 2):W473–W476. [PubMed: 17584791]
52. Stenson P, Mort M, Ball E, Shaw K, Phillips A, Cooper D. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014; 133(1):1–9. [PubMed: 24077912]
53. Thorn CF, Klein TE, Altman RB. Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics*. 2010; 11(4):501–505. [PubMed: 20350130]
54. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2(5):401–404. [PubMed: 22588877]
55. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjana V, Muthusamy B, Gandhi TK, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Blobe GC, Dang CV, Garcia JG, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res*. 2003; 13(10):2363–2371. [PubMed: 14525934]
56. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, Chen C, Duesbury M, Dumousseau M, Feuermann M, Hinz U, Jandrasits C, Jimenez RC, Khadake J, Mahadevan U, Masson P, Pedruzzi I, Pfeifferberger E, Porras P, Raghunath A, Roechert B, Orchard S, Hermjakob H. The IntAct molecular interaction database in 2012. *Nucleic Acids Res*. 2012; 40(Database issue):D841–D846. [PubMed: 22121220]
57. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, Jassal B, Gopinath GR, Wu GR, Matthews L, Lewis S, Birney E, Stein L. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res*. 2005; 33(Database issue):D428–D432. [PubMed: 15608231]
58. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res*. 2006; 34(Database issue):D354–D357. [PubMed: 16381885]
59. Niknafs N, Kim D, Kim R, Diekhans M, Ryan M, Stenson PD, Cooper DN, Karchin R. MuPIT interactive: webserver for mapping variant positions to annotated, interactive 3D structures. *Hum Genet*. 2013; 132(11):1235–1243. [PubMed: 23793516]

60. Peterson TA, Adadey A, Santana-Cruz I, Sun Y, Winder A, Kann MG. DMDM: domain mapping of disease mutations. *Bioinformatics*. 2010; 26(19):2458–2459. [PubMed: 20685956]
61. Jegga AG, Gowrisankar S, Chen J, Aronow BJ. PolyDoms: a whole genome database for the identification of non-synonymous coding SNPs with the potential to impact disease. *Nucleic Acids Res*. 2007; 35(Database issue):D700–D706. [PubMed: 17142238]
62. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003; 31(13):3812–3814. [PubMed: 12824425]
63. Yue P, Melamud E, Moul J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics*. 2006; 7:166. [PubMed: 16551372]
64. Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*. 2005; 21(14):3176–3178. [PubMed: 15879453]
65. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. 2007; 35(11):3823–3835. [PubMed: 17526529]
66. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, Day IN, Gaunt TR. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013; 34(1):57–65. [PubMed: 23033316]
67. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011; 39(17):e118. [PubMed: 21727090]
68. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, Vogelstein B, Karchin R. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res*. 2009; 69(16):6660–6667. [PubMed: 19654296]
69. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*. 2013; 14(Suppl 3):S6.
70. Wang M, Zhao XM, Takemoto K, Xu H, Li Y, Akutsu T, Song J. FunSAV: predicting the functional effect of single amino acid variants using a two-stage random forest model. *PLoS One*. 2012; 7(8):e43847. [PubMed: 22937107]
71. Bao L, Zhou M, Cui Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res*. 2005; 33(Web Server Issue):W480–482. [PubMed: 15980516]
72. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*. 2013; 41(Database issue):D377–D386. [PubMed: 23193289]
73. 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(22):2729–2734. [PubMed: 16895930]
74. Al-Numair NS, Martin AC. The SAAP pipeline and database: tools to analyze the impact and predict the pathogenicity of mutations. *BMC Genomics*. 2013; 14(Suppl 3):S4.
75. Yates CM, Filippis I, Kelley LA, Sternberg MJ. SuSPect: enhanced prediction of single amino acid variant (SAV) phenotype using network features. *J Mol Biol*. 2014; 426(14):2692–2701. [PubMed: 24810707]
76. Simonetti FL, Tornador C, Nabau-Moreto N, Molina-Vila MA, Marino-Buslje C. Kin-Driver: a database of driver mutations in protein kinases. *Database*. 2014; 2014:bau104. [PubMed: 25414382]
77. McSkimming DI, Dastgheib S, Talevich E, Narayanan A, Katiyar S, Taylor SS, Kochut K, Kannan N. ProKinO: a unified resource for mining the cancer kinome. *Hum Mutat*. 2015; 36(2):175–186. [PubMed: 25382819]
78. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*. 2002; 320(2):369–387. [PubMed: 12079393]
79. 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(19):2537–2543. [PubMed: 19654118]

80. Yin S, Ding F, Dokholyan NV. Eris: an automated estimator of protein stability. *Nat Methods*. 2007; 4(6):466–467. [PubMed: 17538626]
81. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res*. 2006; 34(Web Server Issue):W239–242. [PubMed: 16845001]
82. Potapov V, Cohen M, Inbar Y, Schreiber G. Protein structure modelling and evaluation based on a 4-distance description of side-chain interactions. *BMC Bioinformatics*. 2010; 11:374–374. [PubMed: 20624289]
83. Deutsch C, Krishnamoorthy B. Four-body scoring function for mutagenesis. *Bioinformatics*. 2007; 23(22):3009–3015. [PubMed: 17921497]
84. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res*. 2003; 31(13):3316–3319. [PubMed: 12824316]
85. Giollo M, Martin AJ, Walsh I, Ferrari C, Tosatto SC. NeEMO: a method using residue interaction networks to improve prediction of protein stability upon mutation. *BMC Genomics*. 2014; 15(Suppl 4):S7.
86. Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic Acids Res*. 2014; 42(Web Server Issue):W314–319. [PubMed: 24829462]
87. Laimer J, Hofer H, Fritz M, Wegenkittl S, Lackner P. MAESTRO – multi agent stability prediction upon point mutations. *BMC Bioinformatics*. 2015; 16(1):116. [PubMed: 25885774]
88. Capriotti E, Fariselli P, Rossi I, Casadio R. A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics*. 2008; 9(Suppl 2):S6.
89. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*. 2006; 62(4):1125–1132. [PubMed: 16372356]
90. Chen CW, Lin J, Chu YW. iStable: off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics*. 2013; 14(Suppl 2):S5.
91. Teng S, Srivastava A, Wang L. Sequence feature-based prediction of protein stability changes upon amino acid substitutions. *BMC Genomics*. 2010; 11(Suppl 2):1–8. [PubMed: 20044946]
92. Huang L-T, Gromiha MM, Ho S-Y. iPTREE-STAB: interpretable decision tree based method for predicting protein stability changes upon mutations. *Bioinformatics*. 2007; 23(10):1292–1293. [PubMed: 17379687]
93. Dehouck Y, Kwasigroch JM, Rooman M, Gilis D. BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations. *Nucleic Acids Res*. 2013; 41(W1):W333–W339. [PubMed: 23723246]
94. Berliner N, Teyra J, Çolak R, Garcia Lopez S, Kim PM. Combining structural modeling with ensemble machine learning to accurately predict protein fold stability and binding affinity effects upon mutation. *PLoS One*. 2014; 9(9):e107353. [PubMed: 25243403]
95. Kruger DM, Gohlke H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic Acids Res*. 2010; 38(Web Server Issue):W480–486. [PubMed: 20511591]
96. Zhao N, Han JG, Shyu CR, Korkin D. Determining effects of non-synonymous SNPs on protein-protein interactions using supervised and semi-supervised learning. *PLoS Comput Biol*. 2014; 10(5):e1003592. [PubMed: 24784581]

Fig. 1.
Mutation assessment workflow

Fig. 2. Identification of clinically relevant mutations in ClinVar, COSMIC, and TCGA. **(a)** NCBI Variation Viewer showing the CBL gene locus on chromosome 11; ClinVar and dbSNP data shown as tracks below the gene, pathogenic mutations are presented as *purple squares* and closely-located mutations are grouped together in ClinVar track; **(b)** one of the pathogenic mutations p.Y371H shown in dbSNP with the corresponding disease annotation in GTR; **(c)** a representative structure of CBL protein visualized in SNP3D with Cn3D software, mutated p.Y371H is shown with the *yellow side chain*; **(d)** cBioPortal view of CBL protein with missense mutations mapped onto the corresponding domains; **(e)** a representative structure visualized with Jmol directly in cBioPortal with all mapped missense mutations shown in *green*

Fig. 3. Analysis of conserved functional and binding sites in mutated proteins using IBIS method. (a) conserved protein binding site in the RING domain of CBL; (b) interaction graph of CBL protein (represented by 1FBV PDB structure); (c) visualization of protein interface between CBL and UBCc in Cn3D software. Position 384 in 1FBV corresponds to position 338 in the full-length PDB sequence

Fig. 4.
Annotation of the effects of mutations on proteins with available structures

Table 1

A summary of databases integrating the data on human genetic variations, mutations, and their clinical relevance

Database	Description	URL	Reference
COSMIC	Somatic mutations in cancer	http://cancer.sanger.ac.uk/cosmic	[16]
HGMD	Published gene lesions responsible for human inherited disease	http://www.hgmd.cf.ac.uk/ac/	[52]
TCGA	Cancer Genome Atlas	http://cancergenome.nih.gov/	[17]
OMIM	Human genes, inherited genetic disorders, and germline mutations	http://www.omim.org/	[14]
dbGaP	Clinical studies linking genotypes to disease phenotypes	http://www.ncbi.nlm.nih.gov/gap	[7]
GTR	Genetic Testing Registry	http://www.ncbi.nlm.nih.gov/gtr/	[8]
PheGenI	Phenotype–genotype integrator	http://www.ncbi.nlm.nih.gov/gap/phegeni	[12]
ClinVar	Genomic variations and their relationship to human health	http://www.ncbi.nlm.nih.gov/clinvar/	[15]
dbSNP	Single nucleotide polymorphisms	http://www.ncbi.nlm.nih.gov/SNP/	[9]
dbVar	Large genetic variations	http://www.ncbi.nlm.nih.gov/dbvar	[10]
SwissVar	Disease-related variants in Uniprot, provides structural mapping	http://swissvar.expasy.org	[19]
PharmGKB	Associates genes with drugs. Catalogs genetic variations known to impact drug response	https://www.pharmgkb.org/	[53]
MutDB	Integrates human variations and COSMIC mutations, maps to structural data, KEGG pathways, and includes predictions of effects of variations on phenotype	http://www.mutdb.org/	[18]
CBioPortal	Visualization and analysis of large cancer studies. It is based on TCGA and incorporates the overlapping data from COSMIC	http://cbioportal.org/	[54]

Table 2

Online resources for exploring the structural, cellular, and genomic context of mutated proteins

Database	Description	URL	Reference
HPRD	Integrates information pertaining to domain architecture, posttranslational modifications, interaction networks, and disease association	http://www.hprd.org	[55]
EBI IntAct	Molecular interaction data	http://www.ebi.ac.uk/intact/	[56]
BioSystems	Provides integrated access to genes, proteins, small molecules, and pathways	http://www.ncbi.nlm.nih.gov/biosystems/	[21]
Reactome	Curated and peer-reviewed pathways	http://www.reactome.org	[57]
KEGG	Manually curated pathways	http://www.genome.jp/kegg/pathway.html	[58]
CDD	Annotates functional and binding sites in protein domain families	http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml	[23]
IBIS	Annotates protein–protein and protein–DNA/RNA/ion/small molecule interactions and binding sites. Identifies conserved binding sites in homologous protein complexes	http://www.ncbi.nlm.nih.gov/Structure/ibis/ibis.cgi	[27]
muPIT	Interactive exploration of mutations and their structural context	http://mupit.icm.jhu.edu	[59]
DMDM	Domain mapping of mutations	http://bioinf.umbc.edu/dmdm/	[60]
PolyDoms	Mapping of mutations to protein domains and prediction of structural and functional impact of mutations	https://polydoms.cchmc.org/polydoms/	[61]

Table 3

A summary of online resources for predicting the phenotypic effects of mutations

Name	Features and methods	URL	Reference
SIFT	Sequence homology and physical properties of amino acids	http://sift.bii.a-star.edu.sg	[62]
PolyPhen-2	Bayesian models based on substitution scores, phylogenetic conservation, and structural features	http://genetics.bwh.harvard.edu/pph2/	[31]
SNPs3D	Sequence conservation and changes in physical properties of amino acids affecting protein stability	http://snps3d.org	[63]
PMut	Neural network classifier	http://mmb2.pcb.ub.es/PMut/	[64]
SNAP	Neural network classifier based on protein structural properties predicted from protein sequence	https://www.rostlab.org/services/snap/	[65]
Fathmm	Cancer-specific and other disease-specific Hidden Markov Models for predicting the functional consequences of coding and noncoding variants	http://fathmm.biocompute.org.uk	[66]
MutationAssessor	Predicts the effects of mutations on subfamily specific sites	http://mutationassessor.org/	[67]
CHASM	Random Forest classifier for cancer somatic mutations	http://karchinlab.org/apps/appChasm.html/	[68]
MutPred	Uses sequence conservation and structural features, posttranslational modifications	http://mutpred.mutdb.org/	[18]
SNPs&GO	SVM classifier of disease-related variations based on protein functional annotation (GO)	http://snps.biofold.org/snps-and-go/	[69]
PROVEAN	Predicts the effects of amino acid substitutions, insertions, and deletions on protein function, allows scanning multiple mutations	http://provean.jcvi.org/	[30]
FunSAV	Random Forest-based classifier, uses structural features and network properties of mutated proteins	http://sunflower.kuicr.kyoto-u.ac.jp/sjn/FunSA/	[70]
nsSNPAnalyzer	Random Forest-based classifier, uses structural and evolutionary information	http://snpanalyzer.utmem.edu/	[71]
PANTHER	Uses phylogenetic and evolutionary information	http://www.pantherdb.org/tools/csnpscoreForm.jsp/	[72]
PhD-SNP	SVM classifier, uses sequence profiles	http://gpcr.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi/	[73]
SAAP	Precalculated database of predicted effects of known variants, considers structural properties and clashes resulting from amino acid substitutions	http://www.bioinf.org.uk/saap/	[74]
SusPect	Incorporates sequence conservation and network-based features	www.sbg.bio.ic.ac.uk/suspect/	[75]
KinDriver	Annotates driver mutations in protein kinases with experimental evidence demonstrating their functional role	http://kin-driver.leloir.org.ar/	[76]

Name	Features and methods	URL	Reference
ProKinO	Unified resource for mining the cancer kinome	http://vulcan.cs.uga.edu/prokino	[77]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

A summary of online software resources for predicting the effects of mutations on protein stability. The second column indicates the kind of data a method requires as an input: protein sequence, structure, or any of the two. Here “G” refers to G_{fold} .

Resource	Input	Type of output, method, and energy function	URL	Reference
FoldX	Structure	G using empirical force fields	http://foldxsuite.org.eu/	[78]
PoPMuSiC-2.0	Structure	G using a combination of statistical potentials and neural networks	http://dezyme.com/	[79]
ERIS	Structure	G using physical force fields with atomic modeling	http://dokhlab.unc.edu/tools/eris/	[80]
CUPSAT	Structure	G using mean force atom pair and torsion angle potentials	http://cupsat.tu-bs.de/	[81]
Hunter	Structure	G using knowledge-based potentials	http://bioinfo41.weizmann.ac.il/hunter/	[82]
MultiMutate	Structure	G using four-body scoring functions	http://www.math.wsu.edu/math/faculty/bkrishna/DT/Mutate/	[83]
AUTO-MUTE	Structure	G using knowledge-based potentials	http://proteins.gmu.edu/automute	[84]
NeEMO	Structure	G using residue interaction networks	http://protein.bio.unipd.it/neemo/	[85]
DUET	Structure	G using an integrated computational approach of mCSM and SDM	http://bleoberis.bioc.cam.ac.uk/duet/stability	[86]
MAESTRO	Structure	G using multi agent stability prediction	http://biwww.che.sbg.ac.at/MAESTRO	[87]
I-Mutant3.0	Structure/Sequence	G using SVMs	http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi	[88]
MUPro	Structure/Sequence	Predicts qualitative decrease/increase of stability using SVM	http://mupro.proteomics.ics.uci.edu/	[89]
iStable	Structure/Sequence	G using SVM	http://predictor.nchu.edu.tw/iStable	[90]
MuStab	Sequence	Predicts qualitative decrease/increase of stability using SVM	http://bioinfo.ggc.org/mustab/	[91]
iPTREE-STAB	Sequence	G using decision tree methods	http://bioinformatics.myweb.hinet.net/iptree.htm	[92]

Table 5

A summary of online and software resources for predicting the effects of mutations on protein–protein binding affinity. Here “ G ” refers to G_{bind} . All resources need structure as an input

Resource	Type of output, method, and energy function	URL	Reference
MutaBind	G using molecular mechanics force fields, statistical potentials, and fast side-chain optimization algorithms. Produces a model of mutant.	https://www.ncbi.nlm.nih.gov/projects/mutabind/	[46]
BeAtMuSiC	G using a set of statistical potentials, does not produce a model of mutant	http://babylone.ulb.ac.be/beatmusic	[93]
ELASPIC	G for mutations located on interface using a combination of semi-empirical energy terms and molecular features; does not produce a model of mutant	http://www.kimlab.org/software/elaspic	[94]
DrugScore ^{PPI}	G for alanine-scanning mutations located on interface using knowledge-based scoring functions; does not produce a model of mutant	http://cpclab.uni-duesseldorf.de/dsppi/	[95]
SNP-IN	classifies effects of mutations on protein–protein interactions using supervised and semi-supervised learning; does not produce a model of mutant	http://andromeda.rnet.missouri.edu/snptool/	[96]
FoldX	G using empirical force field, Produces a model of mutant.	http://foldxsuite.crg.eu/	[78]