

## **In-Silico Methods for Investigating the Effect of Single Nucleotide Polymorphisms on the Structure and Function of Proteins: A Review** — [Source link](#)

Ebenezer Asiedu

**Published on:** 10 Dec 2019

**Topics:** Missense mutation, Single-nucleotide polymorphism, Docking (molecular) and In silico

Related papers:

- [In silico prediction of protein-ligand interactions](#)
- [Evaluation of influence of single nucleotide polymorphisms in cytochrome P450 2B6 on substrate recognition using computational docking and molecular dynamics simulation.](#)
- [Estimation of Binding Sites of Efavirenz with 3EO9 Receptor: In Silico Molecular Docking and Molecular Dynamics Studies](#)
- [In Silico Tertiary Structure Prediction and Molecule Docking Analysis of SNAIL Protein](#)
- [In silico Homology Modeling and Docking Study of Translationally Controlled Tumor protein of labeorohita](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/in-silico-methods-for-investigating-the-effect-of-single-3snzyif3cw>

Review

## In-silico methods for investigating the effect of single nucleotide polymorphisms on the structure and function of proteins

Ebenezer Asiedu<sup>1</sup>

<sup>1</sup> Department of Biochemistry, School of Biological Sciences, University Of Cape Coast, Ghana; ebenezer.asiedu@stu.ucc.edu.gh

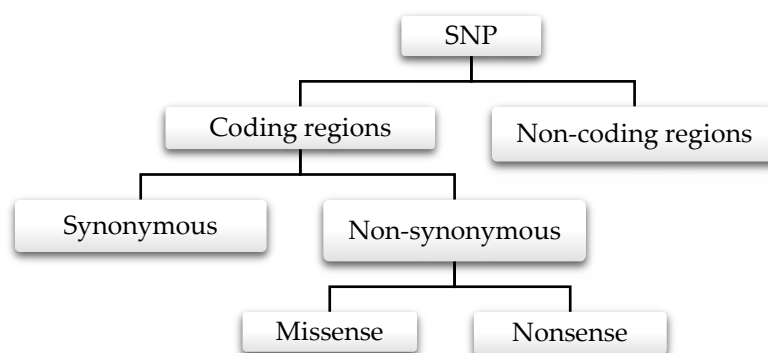
\* Correspondence: ebenezer.asiedu@stu.ucc.edu.gh; Tel +233-542753683 (E.A.)

**Abstract:** Single nucleotide polymorphisms (SNPs) are associated with diseases and drug response variabilities in humans. Elucidating the damaging and disease-associated SNPs using wet-laboratory approaches can be challenging and resource-demanding due to the large number of SNPs in the human genome. Due to the growth in the field of computational biology and bioinformatics, algorithms have been developed to help screen and filter out the most deleterious SNPs that are worth considering for wet-laboratory studies. This article reviews the existing in-silico based methods used to predict and characterize the effects of SNPs on protein structure and function. This cutting-edge approach will facilitate the search for novel therapeutics, help understand the etiology of diseases and fast-track the personalized medicine agenda.

**Keywords:** single nucleotide polymorphism; docking; molecular dynamics; in-silico studies; protein dynamics; missense; prediction algorithm; mutation

### 1. Introduction

Single nucleotide polymorphisms are the most common types of genetic variation in the human genome, occurring within coding [1] and non-coding regions [2] including intergenic stretches [3]. The wide distribution of SNPs in regions of the human genome results in different types of the mutation (Figure 1) with varying phenotypic characteristics. SNPs in non-coding regions do not alter protein sequences but may affect gene expression, non-coding RNA sequence, and gene splicing. In coding regions, SNPs may alter the amino acid sequence of proteins (non-synonymous) or not (synonymous) since the genetic code is degenerate. Synonymous mutations are not always silent, as they have been reported to alter transcription/translation processes [4] and protein conformation [5]. A nonsense mutation occurs when a single nucleotide variation results in a stop codon, thereby producing a premature protein. The missense mutation, however, results in a codon that codes for a different amino acid.



**Figure 1:** Types of single nucleotide polymorphisms in the human genome

Associating the various SNPs with their phenotypic characteristics using wet laboratory approaches can be challenging. There is growth in the use of high-throughput DNA sequencing techniques which has resulted in a plethora of genetic mutation data [6]. Filtering out the functional SNPs in this large pool of data using wet-laboratory approaches could be resource-demanding and cumbersome. Development in the field of computational biology and bioinformatics have contributed massively to alleviate this challenge. Researchers have used computational methods to characterize functional SNPs and predict the effect of the mutations on protein structure and function [7, 8, 9, 10, 11, 12]. Most of the computational algorithms make their predictions based on protein sequence and structure information [13], physicochemical properties of amino acids [14] and evolutionary conservation [15] to classify the effect as damaging or tolerable. Computational modeling studies help to elucidate the structural dynamics and biological activities of the protein molecules, providing detailed information on the atomic fluctuations and conformational changes of the protein.

## 2. SNP database

### 2.1. dbSNP (Single Nucleotide Polymorphism Database)

The National Center for Biotechnology Information (NCBI) created the dbSNP to serve as a central repository for genetic variation. In addition to single nucleotide variations, dbSNP contains other nucleotide sequence variations such as microsatellite repeats and small insertion/deletion (indel) polymorphism [16]. The mutations in dbSNP include publications related to the variations, sequence context of the polymorphism, population frequency, the experimental methods used to assay the variations, genomic and RefSeq mapping information [17]. Associating sequence variations to phenotypic variations is of prime importance to genetic research, thus, the dbSNP facilitates studies in wide range of fields including, pharmacogenomics, association studies and evolutionary studies.

### 2.2. HGMD (Human Gene Mutation Database)

HGMD is a repository of all known genetic mutations in germline responsible for inherited diseases and other disease-oriented genetic polymorphisms available in peer-reviewed scientific literature [18]. The classes of variants available at selected genetic variation database including dbSNP and HGMD are listed in Table 1.

## 3. In-silico prediction tools

### 3.1. Sorting Intolerant from Tolerant (SIFT)

SIFT web server (<https://sift.bii.a-star.edu.sg/>) uses sequence conservation and amino acid physicochemical properties to compute the potential impacts of single nucleotide polymorphisms (SNPs) and frameshifting insertions/deletions (indel) mutations in coding regions on protein function [19]. SIFT is a sequence-based prediction algorithm which assumes that highly conserved amino acids are vital for protein function and structure [20]. Based on this assumption, amino acid mutations at these conserved regions as a result of SNP may confer deleterious effect on protein activity. To ascertain evolutionary conserved sequences/regions, SIFT searches protein databases to obtain homologous sequences of the query and then performs a multiple sequence alignments (MSA) to identify conserved regions [20]. The SIFT algorithm then calculates the probability that a given substitution can be tolerated at a position based on the physicochemical properties of the native residue [14]. The substitution potentially affects protein function if its probability score falls lower than a cutoff score of 0.05 [21].

**Table 1:** Some notable genetic variation database and the type of mutation they host.

Mutation database	Mutation type	Reference
dbSNP ( <a href="https://www.ncbi.nlm.nih.gov/snp/">https://www.ncbi.nlm.nih.gov/snp/</a> )	Single nucleotide variations, small-scale deletions or insertions, retroposable element insertions, microsatellite repeat variations/short tandem repeats, multi-nucleotide variations	[16, 17]
dbVar ( <a href="https://www.ncbi.nlm.nih.gov/dbvar/">https://www.ncbi.nlm.nih.gov/dbvar/</a> )	Copy number variants (CNV), insertions, deletions, inversions, and translocations 50 base pairs (bp).	[63]
HGMD ( <a href="http://www.hgmd.cf.ac.uk/ac/index.php">http://www.hgmd.cf.ac.uk/ac/index.php</a> )	Missense substitution, Nonsense substitutions, Splicing substitutions (intronic and exonic), Regulatory substitutions (exonic, intronic, 5'- and 3'- untranslated regions), Micro-deletions $\leq 20$ bp, Micro-insertions/duplications $\leq 20$ bp, Micro-indels $\leq 20$ bp, Gross deletions $>20$ bp, Gross insertions/duplications $>20$ bp, Complex rearrangements (including inversions, translocations and complex indels), Repeat variations	[18]
OMIM ( <a href="https://omim.org/">https://omim.org/</a> )	Genetic variations associated with known genetic disorders.	[64]
1000 Genome ( <a href="https://www.internationalgenome.org/">https://www.internationalgenome.org/</a> )	A catalog of human genetic variations: single nucleotide polymorphisms, short insertions/deletions (indels), structural variants, haplotypes.	[65]
PharmKGB ( <a href="https://www.pharmkgb.org/">https://www.pharmkgb.org/</a> )	Variants (SNP, indel, haplotype, etc.) associated with drug response phenotypes	[66, 67, 68]
HGVD ( <a href="http://www.hgvd.genome.med.kyoto-u.ac.jp/">http://www.hgvd.genome.med.kyoto-u.ac.jp/</a> )	Exomic genetic variations in the Japanese population	[69]

NHLBI Exome Variant Server ( <a href="https://evs.gs.washington.edu/EVS/">https://evs.gs.washington.edu/EVS/</a> )	Exome sequence variations relating to heart, lung and blood diseases	
ClinVar ( <a href="https://www.ncbi.nlm.nih.gov/clinvar/">https://www.ncbi.nlm.nih.gov/clinvar/</a> )	The literature reviewed genetic variations significant to diseases/human health. Includes both sequence variants and structural variants.	[70]
LOVD ( <a href="https://www.lovd.nl/">https://www.lovd.nl/</a> )	A network of curated gene variant	[71]
Exome Aggregation Consortium ( <a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a> )	Exome sequence variations	[72]
COSMIC ( <a href="https://cancer.sanger.ac.uk/cosmic">https://cancer.sanger.ac.uk/cosmic</a> )	Somatic mutations related to human cancer	[73]
MitoMap ( <a href="https://www.mitomap.org/MITOMAP">https://www.mitomap.org/MITOMAP</a> )	Human mitochondrial DNA variation	[74]
Database of Genomic Variants ( <a href="http://dgv.tcag.ca/dgv/app/home">http://dgv.tcag.ca/dgv/app/home</a> )	Structural variation in the human genome: genomic alterations that involve segments of DNA that are larger than 50bp.	[75]

### 3.2. Protein Variation Effect Analyzer (PROVEAN)

PROVEAN (<http://provean.jcvi.org/index.php>) is an algorithm that predicts the impact of single or multiple amino acid substitutions, insertion and deletion mutations on the biological activity of proteins [22]. The input data for the PROVEAN algorithm are the query protein sequence and the variations. The algorithm integrates the protein blast tool (BLASTP) with a 0.1 E-value threshold and NCBI NR protein database to search and obtain homologous and distantly related sequences [22]. A BLOSSUM62 substitution matrix-based computation is performed on the supporting sequences which comprise the sequence clusters that are very similar to the query sequence [23]. The effect of the variation is quantified by the delta alignment score - the degree of impact a variation in the protein query sequence has on protein function when compared to another protein subject of high similarity [22]. The algorithm then generates a PROVEAN score (average of the delta scores), which will define

the damaging effect of the variation. The variation is predicted as deleterious if the PROVEAN score is less than or equal to a threshold of -2.5 [24].

### 3.3. Screening for Non-acceptable Polymorphisms 2 (SNAP2)

SNAP2 algorithm (<https://roslab.org/services/snap2web/>) uses a machine learning system called an artificial neural network to predict the impact of single amino acid substitutions on the function of a protein [25]. The algorithm considers evolutionarily conserved regions, protein secondary structure, biophysical properties of amino acids, residue flexibility, SWISS-PROT annotations, predicted binding residues solvent accessibility amongst others [26]. SNAP2 relies on the query protein sequence as input data and for a given amino acid substitution, computes a score depicting the potency of the mutation to affect the function of the wild type protein. The score ranges from -100 to +100 implying strong neutral prediction and strong effect prediction respectively. Using the query protein sequence, SNAP2 collects dataset from Protein Mutant Database (PMD), SWISS-PROT, Online Mendelian Inheritance in Men (OMIM), and HumVar. The algorithm fetches all amino acid variants from PMD, map the variants to the corresponding sequence and classify the impact as 'neutral' or 'effect' [26]. To obtain the enzyme dataset, SNAP2 retrieves the Enzyme Classification Commission (EC) number of the query protein from SWISS-PROT, compares it with enzymes of the same experimental annotation function and determines the residue variations among them [25]. The disease-associated variations retrieved from OMIM and Human Var are all classified as 'effect' [26]. Predictions are based on the PMD dataset, Enzyme dataset and Disease dataset.

### 3.4. Polymorphism Phenotyping 2 (PolyPhen-2)

PolyPhen-2 (<http://genetics.bwh.harvard.edu/pph2/>) predicts the impacts of non-synonymous SNPs (nsSNPs) on the stability and function of human proteins using sequence and structure-based predictive features. The structural parameters used by the Polyphen-2 algorithm include solvent accessible surface area, dihedral angles, hydrophobic propensity and B-factor [13]. Polyphen-2 requires the query protein sequence and the amino acid mutation as the input data. The algorithm automatically runs a series of programs to predict the damaging effect of missense mutations on protein function. Based on the input data, Polyphen-2 confirms the missense mutation in the gene and then characterizes the substitution site as binding or active site, a transmembrane region or metal-binding site using a sequence-based prediction feature [27]. Polyphen-2 performs a multiple sequence alignment (MSA) followed by homology sequence analysis and calculates a profile score or position-specific independent count (PSIC) [13]. PSIC represents the logarithmic ratio of the probability of a given amino acid occurring at a position to the probability of the same amino acid occurring at a different position [27, 28]. Polyphen-2 maps the amino acid variations to known 3D structures or homologous 3D structures, thus, discovering whether the mutation can potentially affect the structural and functional properties of the protein [20, 27]. Based on these derived sequences and structural parameters, Polyphen-2 predicts the impacts of the amino acid substitution as 'probably damaging', 'possibly damaging' or 'benign' [27].

### 3.5. Protein Analysis Through Evolutionary Relationships (PANTHER)

PANTHER-PSEP (position-specific evolutionary preservation) is tool for predicting the functional impact of a particular missense mutation on proteins through evolutionary analysis of coding SNPs. PANTHER-PSEP is available freely online (<http://www.pantherdb.org/tools/csnpscoreForm.jsp>) and as a command-line based tool. PANTHER-PSEP computes the period an amino acid has been evolutionary conserved in the protein of interest. The longer the preservation time of a given amino acid, the likely a substitution of the amino acid has a pathogenic or damaging effect. The input data are the query protein sequence and the amino acid variation(s). PANTHER-PSEP reconstruct the ancestral sequence using sequence alignments and phylogenetic trees retrieved from PANTHER database [29]. PANTHER-PSEP then calculates the length of time a given amino acid has been preserved.

### 3.6. Multivariate Analysis of Protein Polymorphism (MAPP)

MAPP(<http://mendel.stanford.edu/sidowlab/downloads/MAPP/index.html>) A tool is available as a command-line based software built with JAVA programming language. MAPP algorithm is a sequence-based prediction tool that evaluates the roles played by missense mutations on protein function and diseases considering amino acid physicochemical variations among homologous sequences [30]. The physicochemical properties of amino acids used in the MAPP algorithm are hydrophathy, polarity, charge, side-chain volume, free energy in  $\alpha$ -helical conformation and free energy in  $\beta$ -sheet conformation [30]. MAPP algorithm first generates multiple sequence alignment composed of the query protein sequence and its homologs, then computes the mean and variance of the physicochemical property in a particular column in the alignment matrix. The methods used by the MAPP algorithm has been outlined in detail [30].

### 3.7. PredictSNP

PredictSNP is a software tool that predicts the effect of SNPs on protein function by integrating multiple prediction tools. The performance of MAPP [30], nsSNPAnalyzer [31], PANTHER [32], PhD-SNP [33], PolyPhen-1 [27], PolyPhen-2 [13], SIFT [14] and SNAP [25]. The PredictSNP tool is a combination of the six best performing prediction tools: MAPP, PhD-SNP, PolyPhen-1, PolyPhen-2, SIFT and SNAP.

### 3.8. MutPred

MutPred2 is a machine-learning based prediction algorithm available as standalone executable software and as a web server (<http://mutpred.mutdb.org/index.html>). MutPred2 predicts the pathogenicity of amino acid substitutions and explores the molecular mechanism behind the phenotype of the mutation [34]. MutPred2 was trained with a dataset (pathogenic and unlabeled) obtained from SWISSVar, Human Gene Mutation Database (HGMD), dbSNP and inter-species pairwise alignments for accurate pathogenicity prediction [34]. The algorithm is built with 30 feed-forward neural networks that compute the probability of a substitution being pathogenic [34]. The average of the scores from each neural network represents the final prediction score, ranging from 0 (lowest probability of pathogenicity) to 1 (highest probability of pathogenicity).

### 3.9. PMut

PMut (<http://mmb.irbbarcelona.org/PMut/>) is a neural network-based prediction tool that uses residue conservation and physico-chemical properties to predict the pathological nature of amino acid variations [35]. PMut2017 prediction algorithm is written as a Python 3 module that uses popular python libraries such as NumPy, SciPy and Matplotlib [36]. Pmut2017 prediction engine was trained with a variation dataset retrieved from SWISSVar [36]. The module computes numerical data for the variations and gives a prediction score ranging from 0 to 1 [36].

### 3.10. Conservation Surface-Mapping (ConSurf) Server

ConSurf (<http://consurf.tau.ac.il>) predicts the relevance of an amino acid residue to the function and conformation of proteins by examining the evolutionary dynamics or pattern of the residues [37]. ConSurf considers the fact that residues that are relevant for optimal protein activity and perfect conformation are highly conserved, hence, evolve slowly. ConSurf computes the evolutionary rate of the residues which determine their level of importance to the protein structure and function [38]. Given an input (query sequence or structure), the ConSurf algorithm retrieves homologous sequences of the query and then build a phylogenetic tree and a multiple sequence alignment [15]. Unlike some other prediction algorithms (Table 2), ConSurf computes the evolutionary rates of the residues and represents the outcome as a color-code bar (Figure 2).





### The conservation scale:



**Figure 2:** ConSurf color-code bar. The color scheme tells the relative degree of conservation of the residues on a scale of 1 (lowly conserved) to 9 (highly conserved), It also tells whether the residue is buried ('b') or exposed ('e') and characterize the residues as functional component ('f') or structural component ('s').

**Table 2:** Other in-silico prediction tools available

Prediction tool	Algorithm	Reference
PhyreRisk	Maps genetic variants to proteins structures URL: <a href="http://phyrerisk.bc.ic.ac.uk/">http://phyrerisk.bc.ic.ac.uk/</a>	[76]
PROJECT HOPE	Project HOPE collects structural information from a range of information sources and analyzes how the mutation affects the protein structure URL: <a href="http://www.cmbi.ru.nl/hope/">http://www.cmbi.ru.nl/hope/</a>	[77]
SNPs&GO	Support Vector Machine (SVM) based classifier integrated with gene ontology (GO) predictor terms. Predict the effect of SNP on protein function based on mutation type, sequence environment information, sequence profiles URL: <a href="http://snps.biofold.org/snps-and-go/snps-and-go.html">http://snps.biofold.org/snps-and-go/snps-and-go.html</a>	[78, 79]
ParePro	An SVM-based algorithm which predicts the deleterious effect of SNP on protein function considering residue difference (RD), the status of the mutation position (SM) and the mutation sequence environment (ME) URL: <a href="http://www.mobioinfor.cn/parepro/">http://www.mobioinfor.cn/parepro/</a>	[80]



---

SNPEffect 4.0	Predicts the structural and molecular effect of SNPs on protein. Integrates aggregation prediction (TANGO), amyloid prediction (WALTZ), chaperone-binding prediction (LIMBO) and protein stability analysis (FoldX) URL: <a href="https://snpeffect.switchlab.org/">https://snpeffect.switchlab.org/</a>	[81]
LIST	Predicts the damaging effects of amino acid mutations in a protein sequence. Uses Local identity Pair-wise Sequence Alignment (LPSA). Computes the damaging effect of the mutation using shared taxa or Taxonomy distance URL: <a href="https://list-s2.msl.ubc.ca/">https://list-s2.msl.ubc.ca/</a>	[82]
SuSPect	Predicts the phenotypic impacts (especially diseases-related) of single amino acid variants on proteins. Integrates sequence conservation and network-level features URL: <a href="http://www.sbg.bio.ic.ac.uk/suspect">http://www.sbg.bio.ic.ac.uk/suspect</a>	[83]

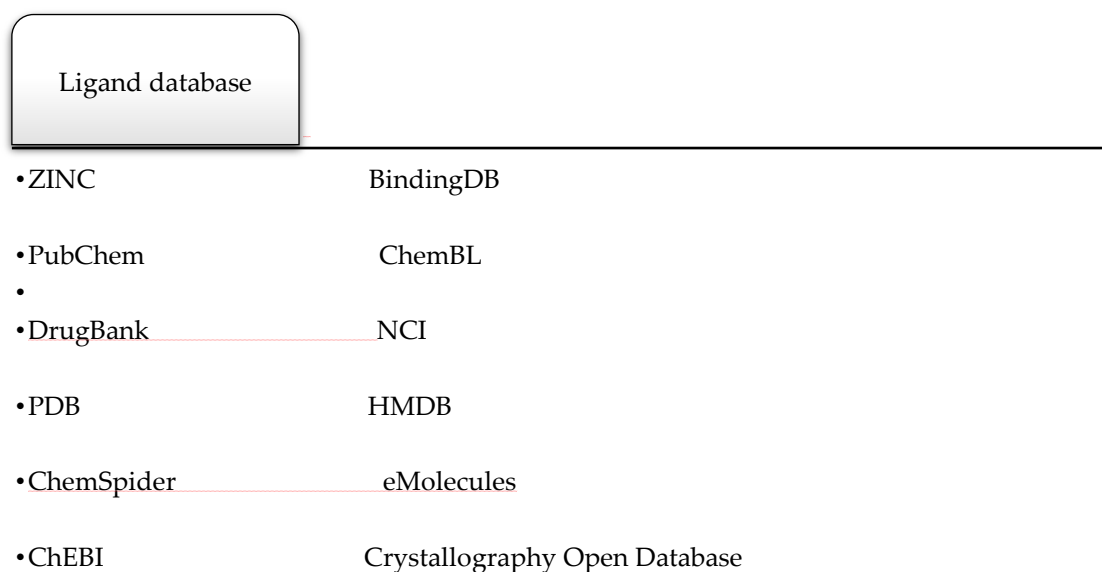
---

#### 4. Molecular modeling approaches

The structural and functional differences between the mutants and native proteins can be elucidated thoroughly using computational molecular modeling techniques. Molecular dynamics simulation and docking (molecular) studies are by far the most utilized computational modeling approaches to study protein function and structure.

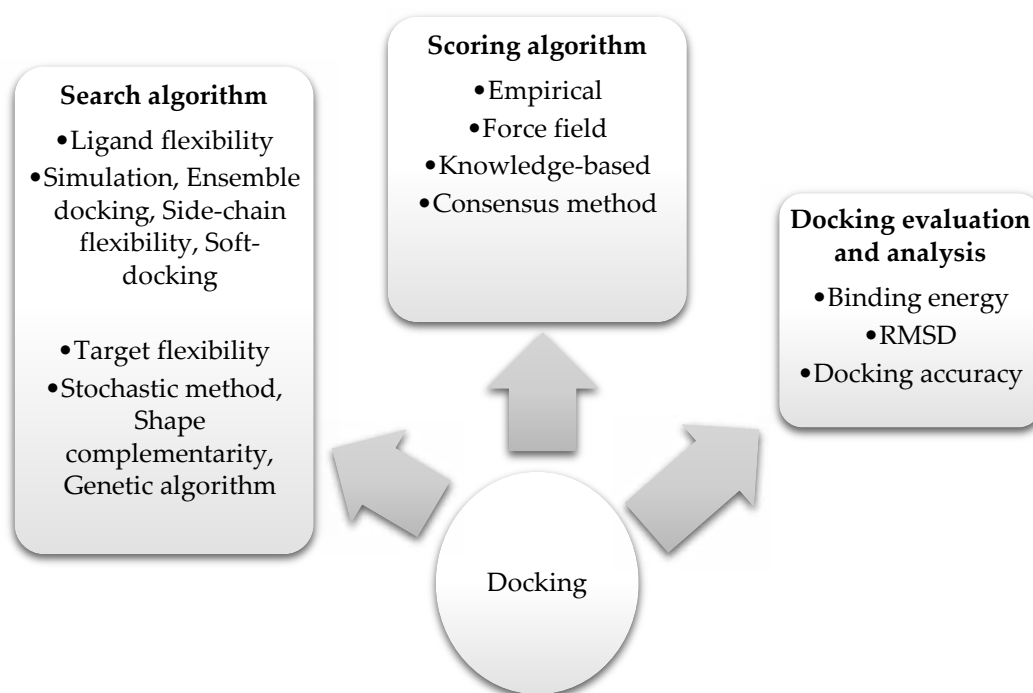
##### 4.1. Molecular Docking Studies

Molecular docking is a way of predicting the most favorable positioning of one molecule relative to another when they are brought together to form a complex. Docking tools predict the correct relative conformation and orientation of a ligand to its target (receptor) binding site [39]. The docking algorithm computationally simulates the specific interactions (hydrophobic, Van der Waals, hydrogen bonding and electrostatic forces) between the ligand and the receptor to obtain an optimized conformation that will ensure that the free energy of the system is minimized [40, 41]. The receptor is usually a protein or a nucleic acid and the ligand is often a small molecule or another protein [42]. A proper docking protocol integrates shape complementarity of ligand-receptor complex and simulation of the complex to account for ligand and receptor flexibilities [43]. Target and ligand preparation are important prerequisites for successful docking. Some experimentally determined protein models are incomplete, thus, efforts are made to obtain accurate models of the target protein. Missing hydrogen atoms and residues are expected to be present. Cofactors that are relevant to the binding interactions or activity of the protein and water molecules that are critical for optimal protein functioning are retained [44]. Preparing a library of ligands involves characterization of compounds based on physicochemical properties of interest to filter out the most prominent candidates. The library usually consists of energy minimized ligands with proper protonation, valence states, and geometry [44]. There are popular databases (Figure 3) from which users can retrieve ligand libraries for docking.



**Figure 3:** Chemical structure databases used in gathering a library of ligands for docking studies.

There are enormous possible binding modes between the ligand and the target. The search algorithm samples the various possible conformations of the ligand in the target [45], considering ligand flexibility and target flexibility. Various docking tools employ a variety of search strategies or algorithms that account for ligand and receptor flexibilities (Figure 4). Some of the freely available docking software are listed in Table 3, however, for a full list of docking tools, the reader should refer to these publications [43, 44, 46, 47, 48].



**Figure 4:** Components of a molecular docking process. The search algorithm generates poses taking into account ligand and target flexibilities. The generated poses or ligand-target alignments are then ranked using scoring algorithms. The docked ligand-target complex is evaluated using variables such as RMSD and binding energies.

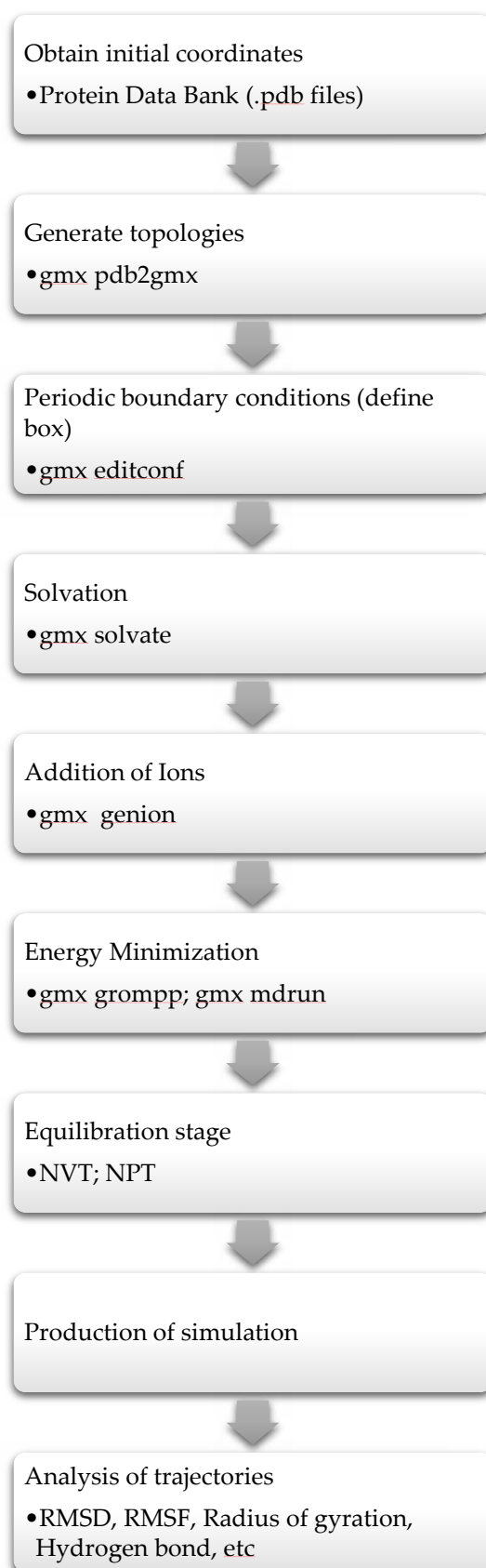
Docking tools are capable of generating many potential modes of binding thus, a method must exist to rank the binding of the ligand to the receptor. The scoring algorithm evaluates and ranks the predicted ligand conformations in the target. The various scoring functions used by scoring algorithms include a force-field based function (computes binding energy considering non-covalent interactions, torsional forces, and entropy) and empirical free energy-based function [39, 45, 49].

**Table 3:** Some freeware and open-source docking software tools available for protein-ligand and protein-protein docking.

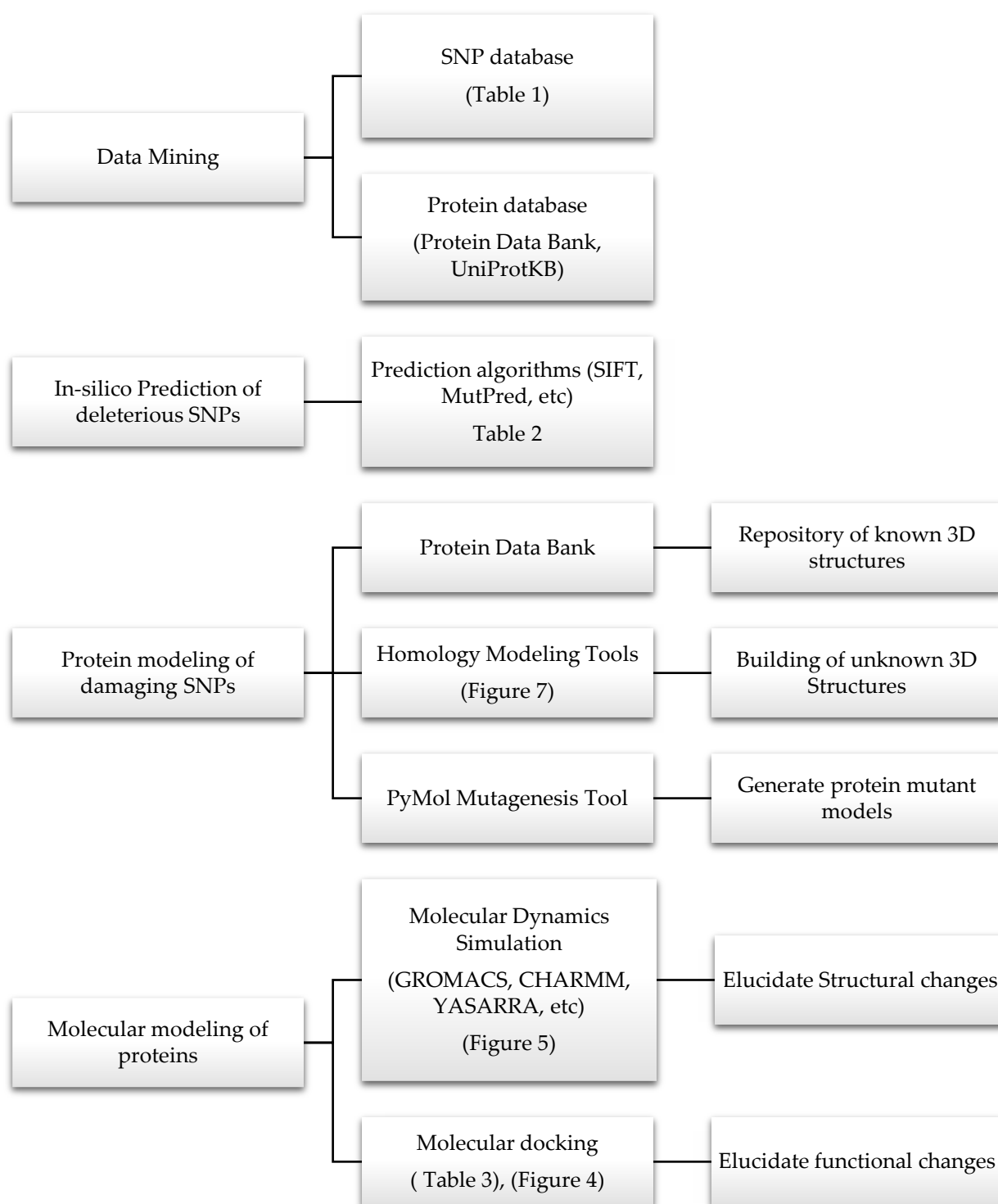
TOOL	URL	REFERENCE
<i>Protein-Ligand Docking</i>		
AutoDock 4.2	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>	[84]
SwissDock	<a href="http://www.swissdock.ch/">http://www.swissdock.ch/</a>	[85]
DOCK 6	<a href="http://dock.compbio.ucsf.edu/">http://dock.compbio.ucsf.edu/</a>	[86]
GEMDOCK	<a href="http://gemdock.life.nctu.edu.tw/dock/">http://gemdock.life.nctu.edu.tw/dock/</a>	[87]
AutoDock Vina	<a href="http://vina.scripps.edu/">http://vina.scripps.edu/</a>	[88]
FRED	<a href="https://www.eyesopen.com/oedocking">https://www.eyesopen.com/oedocking</a>	[89]
<i>Protein-Protein Docking</i>		
Clustpro 2.0	<a href="https://cluspro.bu.edu/login.php">https://cluspro.bu.edu/login.php</a>	[90]
PIPER	<a href="https://www.schrodinger.com/piper">https://www.schrodinger.com/piper</a>	[91]
PatchDock	<a href="http://bioinfo3d.cs.tau.ac.il/PatchDock/php.php">http://bioinfo3d.cs.tau.ac.il/PatchDock/php.php</a>	[92]
GRAMM-X	<a href="http://vakser.compbio.ku.edu/resources/gramm/grammx/">http://vakser.compbio.ku.edu/resources/gramm/grammx/</a>	[93]
RosettaDock 3.2	<a href="http://rosettadock.graylab.jhu">http://rosettadock.graylab.jhu</a>	[94]
pyDock and pyDockWeb	<a href="https://life.bsc.es/pid/pydock/">https://life.bsc.es/pid/pydock/</a>	[95, 96]
HADDOCK 2.2	<a href="http://haddock.science.uu.nl/services/HADDOCK2.2/">http://haddock.science.uu.nl/services/HADDOCK2.2/</a>	[97]
HDOCK	<a href="http://hdock.phys.hust.edu.cn/data/5de39cc5b25ff/">http://hdock.phys.hust.edu.cn/data/5de39cc5b25ff/</a>	[98]
SymmDock	<a href="http://bioinfo3d.cs.tau.ac.il/SymmDock/php.php">http://bioinfo3d.cs.tau.ac.il/SymmDock/php.php</a>	[99]

## 4.2. Molecular Dynamics Simulation

The main experimental technique to elucidate the molecular structures of macromolecules is the X-ray diffraction of crystallized protein [50]. This technique has led to great achievements in the field of structural biology; but, in the best cases, X-ray crystallography can only provide a static snapshot of a fully functional state. NMR has become an increasingly important technique for protein structural investigations, giving access to the flexibility of a system by revealing an ensemble of conformations [51]. Consequently, complementary tools have always been required to allow a dynamic insight into biological targets and ligand binding. The structure of macromolecules and supramolecular assemblies are not rigid and static [52]; they experience local and global movements and rearrangements at different time scales due to thermal diffusion/fluctuations, or in response to external perturbations such as pressure, concentration, pH, or temperature changes. This dynamic behavior and the dependence of the structure on external factors are key for technological applications [53]. However, it is difficult to obtain experimental structural information on individual molecules as a function of time, as well as to get fine structural information of supramolecular assemblies. Molecular dynamics simulations have offered massive progress to these shortcomings. Molecular dynamics simulations provide time-dependent microscopic properties of biomolecules, which could not be explained by experimental methods like X-ray crystallography [54]. These specifications enable MD simulations as the most widely used computational techniques for the study of dynamical properties of proteins, DNAs and other bio-macromolecules. Molecular dynamics simulation provides the methodology for detailed microscopical modeling on the atomic scale. This technique is a scheme for the study of the natural time evolution of the system that allows the prediction of the static and dynamic properties of substances directly from the underlying interactions between the molecules. In a molecular dynamics simulation, the classical equations of motion for the system of interest, say, biomolecules in solution are integrated numerically by solving Newton's equation of motion [55]. A full review of molecular dynamics simulations is available in the following publications [46, 54, 56, 57]. Popular molecular dynamics programs used in the study of biomolecules include GROMACS [58], CHARMM [59], AMBER [60], NAMD [61], YASARRA [62].



**Figure 5:** The general steps to run molecular dynamics simulation of a protein molecule in water using GROMACS 2018.2 software package.



**Figure 6:** Outline for the computational-based approach of predicting the effects of SNPs on protein structure and function. Data on the mutation and the protein of interest can be retrieved from databases such as dbSNP and UniProtKB (<https://www.uniprot.org/uniprot/>) respectively. In-silico prediction methods (Table 2) are used to filter the most deleterious variants of the protein. Proteins with known 3D structures can be fetched from protein databases like Protein data bank (PDB). However, proteins with unsolved structures are treated with homology modeling tools (Figure 7). These tools build 3D models of the query proteins using templates (solved proteins that are homologous to the query). Mutagenesis tools are then

used to obtain the mutants of the protein. Molecular dynamics simulation (Fig. 5) and docking studies are then performed to compare the wild type and mutant models on the basis of structure and function.

HOMOLOGY MODELING TOOLS	<b>SWISS-MODEL</b> <a href="https://swissmodel.expasy.org/">https://swissmodel.expasy.org/</a>
	<b>I-TASSER</b> <a href="https://zhanglab.ccmb.med.umich.edu/I-TASSER/">https://zhanglab.ccmb.med.umich.edu/I-TASSER/</a>
	<b>RaptorX</b> <a href="http://raptorx.uchicago.edu/">http://raptorx.uchicago.edu/</a>
	<b>Modeller</b> <a href="https://salilab.org/modeller/">https://salilab.org/modeller/</a>
	<b>Phyre2</b> <a href="http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index">http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index</a>
	<b>EsyPred3D</b> <a href="https://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/">https://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/</a>
	<b>Robetta</b> <a href="http://new.robetta.org/">http://new.robetta.org/</a>
	<b>PRIMO</b> <a href="http://www.immt.res.in/maxmod/">http://www.immt.res.in/maxmod/</a>
	<b>MaxMod</b> <a href="http://www.immt.res.in/maxmod/">http://www.immt.res.in/maxmod/</a>
	<b>Bhageerath-H</b> <a href="http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp">http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp</a>

**Figure 7:** Homology modeling tools available to build structures of proteins with unknown 3D models.

## 5. Summary and outlook

SNPs are ubiquitous in the human genome, occurring at both coding and non-coding regions of the DNA, thus affecting gene expression and protein dynamics. Proteins play roles in drug response and development of both metabolic and inheritable diseases. The dynamics of drug transporters, receptors and enzymes are influenced by SNPs in the human genome, hence, causing inter-individual drug response variability. Protein-protein interactions are vital for molecular and



biological activities such as cell signaling, DNA transcription, and translation. This should explain the role of SNPs in cancer, metabolic and inherited disease development. Experimentally, it is difficult to determine how SNPs affect protein dynamics due to the large number of SNPs collated. Integrating computational/bioinformatics approaches with experimental studies is the most efficient way of achieving such a task. The bioinformatics-based studies serve as a first-pass filter which predicts the most deleterious mutations for further experimental studies. Explaining the genetic basis of drug response variability in humans is the primary goal of pharmacogenomics which is a crucial element of personalized medicine and stratified medicine. Integrating computational prediction methods with computational molecular modeling studies will greatly enhance pharmacogenomics studies and revolutionize therapeutic discoveries. Nonetheless, these computational algorithms may have limitations to their performances. The researcher must therefore carefully select the most appropriate tools to handle the research problem at hand. Combining many prediction algorithms helps to complement each other, yielding more accurate results.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Våge, J.; Lingaas, F. Single nucleotide polymorphisms (SNPs) in coding regions of canine dopamine- and serotonin-related genes. *BMC genetics*, **2008**, *9*, 10. doi:10.1186/1471-2156-9-10.
2. Madelaine, R.; Notwell, J. H.; Skariah, G.; Halluin, C.; Chen, C. C.; Bejerano, G.; Mourrain, P. A screen for deeply conserved non-coding GWAS SNPs uncovers a MIR-9-2 functional mutation associated to retinal vasculature defects in human. *Nucleic acids research*, **2018**, *46*(7), 3517–3531. doi:10.1093/nar/gky166.
3. Chen, J.; Tian, W. Explaining the disease phenotype of intergenic SNP through predicted long-range regulation. *Nucleic acids research*, **2016**, *44*(18), 8641–8654. doi:10.1093/nar/gkw519.
4. Zhou, T.; Ko, E. A.; Gu, W.; Lim, I.; Bang, H.; Ko, J. H. Non-silent story on synonymous sites in voltage-gated ion channel genes. *PloS one*, **2012**, *7*(10), e48541. doi: 10.1371/journal.pone.0048541.
5. Komar, A. A. GENETICS: SNPs, Silent but Not Invisible. *Science*, **2007**, *315*(5811), 466–467. doi:10.1126/science.1138239.
6. Behjati, S.; Tarpey, P. S. What is next generation sequencing? Archives of disease in childhood. *Education and practice edition*, **2013**, *98*(6), 236–238. doi:10.1136/archdischild-2013-304340.
7. Islam, M. J.; Khan, A. M.; Parves, M. R.; Hossain, M. N.; Halim, M. A. Prediction of Deleterious Non-synonymous SNPs of Human STK11 Gene by Combining Algorithms, Molecular Docking, and Molecular Dynamics Simulation. *Scientific reports*, **2019**, *9*(1), 16426. doi:10.1038/s41598-019-52308-0.
8. Bhavaniramy, S.; Vanajothi, R.; Vishnupriya, S.; Al-Aboody, M. S.; Vijayakumar, R.; Baskaran, D. Computational characterization of deleterious SNPs in Toll-like receptor gene that potentially cause mastitis in dairy cattle. *Biocatalysis and Agricultural Biotechnology*, **2019**, *19*, 101151. doi: 10.1016/j.bcab.2019.101151.
9. Muthusamy, K.; Nagamani, S. Vitamin D receptor (VDR) non-synonymous single nucleotide polymorphisms (nsSNPs) affect the calcitriol drug response - A theoretical insight. *Journal of Molecular Graphics and Modelling*, **2018**, *81*, 14–24. doi: 10.1016/j.jmgm.2018.02.004.
10. Thirumal Kumar, D.; George Priya Doss, C.; Sneha, P.; Tayubi, I. A.; Siva, R.; Chakraborty, C.; Magesh, R. Influence of V54M mutation in giant muscle protein titin: a computational screening and molecular

dynamics approach. *Journal of Biomolecular Structure and Dynamics*, **2016**, 35(5), 917–928. doi:10.1080/07391102.2016.1166456.

11. Khan, I.; Ansari, I. A.; Singh, P.; Dass, J. F. P.; Khan, F. Identification and characterization of functional single nucleotide polymorphisms (SNPs) in Axin 1 gene: a molecular dynamics approach. *Cell Biochemistry and Biophysics*, **2017**, 76(1-2), 173–185. doi:10.1007/s12013-017-0818-1.
12. Doss, C. G.; Chakraborty, C.; Chen, L.; Zhu, H. Integrating in silico prediction methods, molecular docking, and molecular dynamics simulation to predict the impact of ALK missense mutations in structural perspective. *BioMed research international*, **2014**, 2014, 895831. doi:10.1155/2014/895831.
13. Adzhubei, I. A.; Schmidt, S.; Peshkin, L.; Ramensky, V. E.; Gerasimova, A.; Bork, P.; Kondrashov, A. S.; Sunyaev, S. R. A method and server for predicting damaging missense mutations. *Nature Methods*, **2010**, 7(4):248-249.
14. Ng, P. C.; Henikoff, S. Accounting for human polymorphisms predicted to affect protein function. *Genome Research*, **2002**, 12, 436–446.
15. Ashkenazy, H.; Abadi, S.; Martz, E.; Chay, O.; Mayrose, I.; Pupko, T.; Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic acids research*, **2016**, 44(W1), W344–W350. doi:10.1093/nar/gkw408.
16. Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic acids research*, **2001**, 29(1), 308–311. doi:10.1093/nar/29.1.308.
17. Kitts A, Sherry S. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. 2002 Oct 9 [Updated 2011 Feb 2]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda
18. Stenson, P. D.; Mort, M.; Ball, E. V.; Evans, K.; Hayden, M.; Heywood, S.; Cooper, D. N. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human genetics*, **2017**, 136(6), 665–677. doi: 10.1007/s00439-017-1779-6.
19. Sim, Ngak-Leng; Kumar, Prateek; Hu, Jing; Henikoff, Steven; Schneider, Georg; Ng, P. C. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, **2012**, 40, W452–W457. doi:10.1093/nar/gks539.
20. Mah, T. L. J.; Low, E. S. H.; Lee, Edmund. In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. *Drug Discovery Today*, **2011**, 16.
21. Ng, P. C.; Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Research*, **2001**, 11, 863–874.
22. Choi, Y.; Sims, G. E.; Murphy, S.; Miller, J. R.; Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*, **2012**, 7(10): e46688.
23. Choi, Y.; Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, **2015**, 31(16): 2745-2747.
24. Choi, Y. A Fast Computation of Pairwise Sequence Alignment Scores between a Protein and a Set of Single-Locus Variants of another Protein. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '12). ACM, New York, NY, USA, **2012**, 414-417.
25. Bromberg, Y.; Rost, B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research*, **2007**, 35(11): 3823-3835.

26. Hecht, M.; Bromberg, Y.; Rost, B. Better prediction of functional effects for sequence variant. *BMC Genomics*, **2015**, *16*(Suppl 8): S1.
27. Ramensky, V.; Bork, P.; Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research*, **2002**, *30*(17):3894-3900.
28. Sunyaev, S. R.; Eisenhaber, F.; Rodchenkov, I. V.; Eisenhaber, B.; Tumanyan, V. G.; Kuznetsov, E. N. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Engineering*, **1999**, *12*(5):387-394.
29. Tang, H.; Thomas, D. P. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics*, **2016**, *32*(14). doi: 10.1093/bioinformatics/btw222.
30. Stone, E. A.; Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Research*, **2005**, *15*: 978–986. doi:10.1101/gr.3804205.
31. Bao, L.; Zhou, M.; Cui, Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Research*, **2005**, *33*: W480–W482. doi:10.1093/nar/gki372.
32. Thomas, P. D.; Kejariwal, A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: Evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci*, **2004**, *101*, 15398–15403. doi:10.1073/pnas.0404380101.
33. 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*, 729–734. doi:10.1093/bioinformatics/btl423.
34. Pejaver, V.; Mooney, S. D.; Radivojac, P. Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *bioRxiv*, **2017**, 134981; doi: <https://doi.org/10.1101/134981>.
35. Ferrer-Costa, C.; Gelpi, J. L.; Zamakola, L.; Parraga, I.; de la Cruz, X.; Orozco, M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics (Oxford, England)*, **2005**, *21*, 3176–3178.
36. López-Ferrando, V.; Gazzo, A.; de la Cruz, X.; Orozco, M.; Gelpi, J. L. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. *Nucleic Acids Research*, **2017**, *2017*. doi: 10.1093/nar/gkx313.
37. 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 Research*, **2010**, *38*, W529-W533. doi: 10.1093/nar/gkq399; PMID: 20478830.
38. Celniker G.; Nimrod G.; Ashkenazy H.; Glaser F.; Martz E.; Mayrose I.; Pupko T.; Ben-Tal N. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal of Chemistry*, **2013**, *53*: 199-206. doi: 10.1002/ijch.201200096.
39. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews. Drug Discovery*, **2004**, *3*(11): 935–49. doi:10.1038/nrd1549.
40. Pagadala, N. S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophysical reviews*, **2017**, *9*(2), 91–102. doi:10.1007/s12551-016-0247-1.

41. Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *Journal of Molecular Biology*, **2004**, *337*(5): 1161–82. doi: 10.1016/j.jmb.2004.02.015.
42. Hernandez-Santoyo, A.; Yair, A.; Altuzar, V.; Vivanco-Cid, H.; Mendoza-Barrer, C. Protein-Protein and Protein-Ligand Docking. *Protein Engineering - Technology and Application*, **2013**. doi: 10.5772/56376.
43. Yuriev, E.; Ramsland, P. A. Latest developments in molecular docking: 2010-2011 in review. *Journal of Molecular Recognition*, **2013**, *26*(5), 215–239. doi:10.1002/jmr.2266.
44. Khanna, V.; Ranganathan, S.; Petrovsky, N. Rational Structure-Based Drug Design. *Reference Module in Life Sciences*, **2018**. doi:10.1016/b978-0-12-809633-8.20275-6.
45. Meng, X. Y.; Zhang, H. X.; Mezei, M.; Cui, M. Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, **2011**, *7*(2), 146–157. doi:10.2174/157340911795677602.
46. Naqvi, A. A. T.; Mohammad, T.; Hasan, G. M.; Hassan, M. I. Advancements in Docking and Molecular Dynamics Simulations Towards Ligand-receptor Interactions and Structure-function Relationships. *Current Topics in Medicinal Chemistry*, **2018**, *18*(20), 1755–1768.
47. Kanguane, P.; Nilofer, C. Protein-Protein Docking: Methods and Tools. *Protein-Protein and Domain-Domain Interactions*, **2018**, 161–168. doi: 10.1007/978-981-10-7347-2\_14.
48. George Priya Doss, C.; Nagasundaram, N.; Chakraborty, C.; Chen, L.; Zhu, H. Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: a molecular dynamics approach. *Human genomics*, **2013**, *7*(1), 10. doi:10.1186/1479-7364-7-10.
49. George Priya Doss, C.; Chakraborty, C.; Narayan, V.; Thirumal Kumar, D. Computational Approaches and Resources in Single Amino Acid Substitutions Analysis Toward Clinical Research. *Advances in Protein Chemistry and Structural Biology*, **2014**, *94*, 365–423. doi:10.1016/b978-0-12-800168-4.00010-x.
50. Lundstrom, K. Structural genomics for membrane proteins. *Cellular and Molecular Life Sciences*, **2006**, *63*(22), 2597–2607. doi:10.1007/s00018-006-6252-y.
51. Kovermann, M.; Rogne, P.; Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Quarterly Reviews of Biophysics*, **2016**, *49*. doi:10.1017/s0033583516000019.
52. Nussinov, R.; Papin, J. A.; Vakser, I. Computing the Dynamic Supramolecular Structural Proteome. *PLOS Computational Biology*, **2017**, *13*(1), e1005290. doi: 10.1371/journal.pcbi.1005290.
53. Yu, I.; Mori, T.; Ando, T.; Harada, R.; Jung, J.; Sugita, Y.; Feig, M. Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *eLife*, **2016**, *5*, e19274. doi:10.7554/eLife.19274.
54. Hospital, A.; Goñi, J. R.; Orozco, M.; Gelpi, J. L. Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry*, **2015**, *8*, 37–47. doi:10.2147/AABC.S70333.
55. Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, **2002**, *9*(9), 646–652. doi:10.1038/nsb0902-646.
56. Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron*, **2018**, *99*(6), 1129–1143. doi: 10.1016/j.neuron.2018.08.011.

57. Perilla, J. R.; Goh, B. C.; Cassidy, C. K.; Liu, B.; Bernardi, R. C.; Rudack, T.; Schulten, K. Molecular dynamics simulations of large macromolecular complexes. *Current Opinion in Structural Biology*, **2015**, *31*, 64–74. doi: 10.1016/j.sbi.2015.03.007.
58. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **2015**, *1-2*, 19–25. doi: 10.1016/j.softx.2015.06.001.
59. Brooks, B. R.; Brooks, C. L. 3<sup>rd</sup>; Mackerell, A. D. Jr.; et al. CHARMM: the biomolecular simulation program. *Journal of Computational Chemistry*, **2009** *30*(10), 1545–1614. doi:10.1002/jcc.21287.
60. Case, D. A.; Cheatham, T. E. 3<sup>rd</sup>; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Woods, R. J. The Amber biomolecular simulation programs. *Journal of computational chemistry*, **2005**, *26*(16), 1668–1688. doi:10.1002/jcc.20290.
61. Wang, Y.; Harrison, C. B.; Schulten, K.; & McCammon, J. A. Implementation of Accelerated Molecular Dynamics in NAMD. *Computational science & discovery*, **2011**, *4*(1), 015002. doi:10.1088/1749-4699/4/1/015002.
62. Krieger, E.; Vriend, G. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics (Oxford, England)*, **2014**, *30*(20), 2981–2982. doi:10.1093/bioinformatics/btu426.
63. Kitts, A.; Church, D.; Hefferon, T. et al. dbVar. 2014 Oct 26. In: The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK269031/>
64. Hamosh, A.; Scott, A. F.; Amberger, J. S.; Bocchini, C. A.; McKusick, V. A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*, **2005**, *33*(Database issue), D514–D517. doi:10.1093/nar/gki033.
65. Gibbs, R. A.; Boerwinkle, E.; Doddapaneni, H.; Han, Y.; Korchina, V.; Reid, J. G. A global reference for human genetic variation. *Nature*, **2015**, *526*(7571), 68–74. doi:10.1038/nature15393.
66. Barbarino, J. M.; Whirl-Carrillo, M.; Altman, R. B.; Klein, T. E. PharmGKB: A worldwide resource for pharmacogenomic information. Wiley interdisciplinary reviews. *Systems biology and medicine*, **2018**, *10*(4), e1417. doi:10.1002/wsbm.1417.
67. McDonagh, E. M.; Whirl-Carrillo, M.; Garten, Y.; Altman, R. B.; Klein, T. E. From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. *Biomarkers in Medicine*, **2011**, *5*(6):795-806.
68. Thorn, C. F.; Klein, T. E.; Altman, R. B. PharmGKB: the pharmacogenomics knowledge base. *Methods in Molecular Biology*, **2013**, doi:10.1015:311–320.
69. Higasa, K.; Miyake, N.; Yoshimura, J.; Okamura, K.; Niihori, T.; Saitsu, H.; Matsuda, F. Human genetic variation database, a reference database of genetic variations in the Japanese population. *Journal of human genetics*, **2016**, *61*(6), 547–553. doi:10.1038/jhg.2016.12.
70. Landrum, M. J.; Lee, J. M.; Benson, M.; Brown, G. R.; Chao, C.; Chitipiralla, S.; Maglott, D. R. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, **2017**, *46*(D1), D1062–D1067. doi:10.1093/nar/gkx1153.
71. Fokkema, I. F. A. C.; Taschner, P. E. M.; Schaafsma, G. C. P.; Celli, J.; Laros, J. F. J.; den Dunnen, J. T. LOVD v.2.0: the next generation in gene variant databases. *Human Mutation*, **2011**, *32*(5), 557–563. doi:10.1002/humu.21438.



72. Karczewski, K. J.; Weisburd, B.; Thomas, B.; Solomonson, M.; Ruderfer, D. M.; Kavanagh, D.; MacArthur, D. G. The ExAC browser: displaying reference data information from over 60,000 exomes. *Nucleic acids research*, **2017**, *45*(D1), D840–D845. doi:10.1093/nar/gkw971.
73. Tate, J. G.; Bamford, S.; Jubb, H. C.; Sondka, Z.; Beare, D. M.; Bindal, N.; Forbes, S. A. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Research*, **2018**. doi:10.1093/nar/gky1015.
74. Brandon, M. C. MITOMAP: a human mitochondrial genome database--2004 update. *Nucleic Acids Research*, **2004**, *33*, D611–D613. doi:10.1093/nar/gki079.
75. MacDonald, J. R.; Ziman, R.; Yuen, R. K.; Feuk, L.; Scherer, S. W. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic acids research*, **2014**, *42*, D986–D992. doi:10.1093/nar/gkt958.
76. Ofoegbu, T. C.; David, A.; Kelley, L. A.; Mezulis, S.; Islam, S. A.; Mersmann, S. F.; Sternberg, M. PhyreRisk: A Dynamic Web Application to Bridge Genomics, Proteomics and 3D Structural Data to Guide Interpretation of Human Genetic Variants. *Journal of molecular biology*, **2019**, *431*(13), 2460–2466. doi: 10.1016/j.jmb.2019.04.043.
77. Venselaar, H.; te Beek, T.A.; Kuipers, R.K. et al. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*, **2010**, *11*, 548. doi: 10.1186/1471-2105-11-548.
78. Thusberg, J.; Olatubosun, A.; Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation.*, **2011**, *32*, 358–368.
79. Capriotti, E.; Calabrese, R.; Fariselli, P.; Martelli, P. L.; Altman, R. B.; Casadio, R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*, **2013**, Suppl 3: S6.
80. Jian, T.; Ningfeng, W.; Jun, G.; Xuexia G.; Juhua Z.; Yunliu, F. Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics*, **2007**, *8*(1):450.
81. De Baets, G.; Van Durme, J.; Reumers, J.; Maurer-Stroh, S.; Vanhee, P.; Dopazo, J.; Rousseau, F. SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic acids research*, **2012**, *40*(Database issue), D935–D939. doi:10.1093/nar/gkr996.
82. Nawar, M.; Steven, J. M.; Jones, G.; Jörg, G. Improved measures for evolutionary conservation that exploit taxonomy distances. *Nature Communications*, **2019**, *10*(1):1556. doi: 10.1038/s41467-019-09583-2.
83. Yates, C. M.; Filippis, I.; Kelley, L. A.; Sternberg, M. J. E. SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. *Journal of Molecular Biology*, **2014**, *426*(14), 2692–2701. doi: 10.1016/j.jmb.2014.04.026
84. Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nature protocols*, **2016**, *11*(5), 905–919. doi:10.1038/nprot.2016.051.
85. Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic acids research*, **2011**, *39*(Web Server issue), W270–W277. doi:10.1093/nar/gkr366.
86. Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Rizzo, R. C. DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, **2015**, *36*(15), 1132–1156. doi:10.1002/jcc.23905.

87. Yang, J.-M.; Chen, C. C. GEMDOCK: A generic evolutionary method for molecular docking. *Proteins: Structure, Function, and Bioinformatics*, **2004**, *55*(2), 288–304. doi:10.1002/prot.20035.
88. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, **2010**, *31*(2), 455–461. doi:10.1002/jcc.21334.
89. McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *Journal of Chemical Information and Modeling*, **2011**, *51*(3), 578–596. doi:10.1021/ci100436p.
90. Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*, *85*(3):435-444.
91. Kozakov, D.; Brenke, R.; Comeau, S. R.; Vajda, S. PIPER: An FFT-based protein docking program with pairwise potentials. *Proteins: Structure, Function, and Bioinformatics*, **2006**, *65*(2), 392–406. doi:10.1002/prot.21117.
92. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research*, **2005**, *33*(Web Server issue), W363–W367. doi:10.1093/nar/gki481.
93. Tovchigrechko, A.; Vakser, I. A. GRAMM-X public web server for protein-protein docking. *Nucleic acids research*, **2006**, *34*(Web Server issue), W310–W314. doi:10.1093/nar/gkl206.
94. Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J. Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PloS one*, **2011**, *6*(8), e22477. doi: 10.1371/journal.pone.0022477.
95. Cheng, T. M.-K.; Blundell, T. L.; & Fernandez-Recio, J. pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins: Structure, Function, and Bioinformatics*, **2007**, *68*(2), 503–515. doi:10.1002/prot.21419.
96. Jiménez-García, B.; Pons, C.; Fernández-Recio, J. pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics*, **2013**, *29*(13), 1698–1699. doi:10.1093/bioinformatics/btt262
97. Van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastiris, P. L.; Karaca, E.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, **2016**, *428*(4), 720–725. doi: 10.1016/j.jmb.2015.09.014.
98. Yan, Y.; Zhang, D.; Zhou, P.; Li, B.; Huang, S. Y. HDock: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic acids research*, **2017**, *45*(W1), W365–W373. doi:10.1093/nar/gkx407.
99. Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research*, **2015**, *33*(Web Server issue), W363–W367. doi:10.1093/nar/gki481.