CHAPTER TEN

Computational Approaches and Resources in Single Amino Acid Substitutions Analysis Toward Clinical Research

C. George Priya Doss*,[1], Chiranjib Chakraborty†, Vaishnavi Narayan‡, D. Thirumal Kumar*

*Medical Biotechnology Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu, India
†Department of Bio-Informatics, School of Computer and Information Sciences, Galgotias University, Greater Noida, Uttar Pradesh, India
‡BioMolecules & Genetics Division, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu, India
[1]Corresponding author: e-mail address: georgecp77@yahoo.co.in; georgepriyadoss@vit.ac.in

Contents

Abstract

Single amino acid substitutions (SAPs) belong to a class of SNPs in the coding region, which alter the protein function during the translation process. Storage of more information regarding SAPs in public databases will soon become a major hurdle in

Advances in Protein Chemistry and Structural Biology, Volume 94
ISSN 1876-1623
http://dx.doi.org/10.1016/B978-0-12-800168-4.00010-X

characterizing the functional SAPs. In such a demanding era, biology has to rely on bioinformatics, which can work its way through to solve the problems at hand by cutting huge amount of time and resources that are otherwise wasted. Here, we describe an overview of the existing repositories of variant databases and computational methods in predicting the effects of functional SAPs on protein stability, structure, function, drug response, and protein dynamics. This chapter will inspire many biologists with a greater promise in identifying the functional SAPs at the structural level, thereby understanding the molecular effects that are critical for personalized medicine diagnosis, prognosis, and treatment for diseases.

1. INTRODUCTION

Technological advances in high-throughput research have modernized the whole field of biology and medicine with the introduction of terms like genomics, proteomics, pharmacogenomics, and epigenomics. The completion of the Human Genome Project in 2003 (International Human Genome Sequencing Consortium, 2004) and HapMap project in 2007 (Frazer et al., 2007; The International HapMap Consortium, 2003) followed by initiation of 1000 Genomes Project (The 1000 Genomes Project Consortium, 2010) and the Exome Sequencing Project has led to deposition of large volume of genetic variation information in public databases. Numerous efforts were underway in understanding the effects of genetic variation between the individuals and consequences in phenotypic variation and disease susceptibility (De Baets et al., 2012; Taillon-Miller, Gu, Li, Hillier, & Kwok, 1998). The DNA variation consists of insertions, deletions, copy number variations, and single nucleotide substitutions (SNPs). Change in a single nucleotide base from any one of the four nucleotides (A, T, G, and C) to another one is termed as single nucleotide substitution and is found to be the most common DNA variation (ENCODE Project Consortium, 2012). SNP alleles are created either by transition (C/T or G/A) or transversion (C/T, A/G, C/A, or T/G) substitutions. All of these transition and transversion events appear to be more or less similar in occurrence, except for the extreme overabundance of the C to T transition. Over 70% of all the SNPs found in the human genome involve a C to T transition (Kimura, 1980). So far 11 million SNPs have been cataloged; among them 7 million SNPs are designated as common variants occurring with a minor allele frequency above 5%, while the remaining SNPs with minor allele frequency below 5% are designated as rare variants

(Frazer, Murray, Schork, & Topol, 2009; Raychaudhuri, 2011). SNPs are not only designated as markers for constructing genetic maps but also have the potential to direct functional polymorphic variants that are involved in monogenic and complex disorders such as diabetes, cardiovascular diseases, and cancers. Understanding the involvement of functional SNPs might shed some light in disease susceptibility to monogenic and complex disorders and also help in designing more effective treatments to individuals by monitoring adverse drug effects. SNPs are classified based on the location within coding sequences of genes, in noncoding regions of genes, or in the intergenic regions between genes (Risch, 2000). Maximum numbers of SNPs are found in noncoding regions without any biological function of a protein (silent), though they may affect gene expression or splicing. An SNP in which nucleotide substitution leads to no change in amino acid sequence is termed as synonymous (silent mutation), whereas substitution of nucleotide that leads to alteration in the amino acid sequence is defined as nonsynonymous (missense or "nonsense mutations") also called as single amino acid substitutions (SAPs) (Mooney, 2005; Stenson et al., 2003). SNPs within coding or regulatory region of a gene are of biological significance (Pastinen, Ge, & Hudson, 2006). Tennessen et al. (2012) estimated the occurrence of 13,000 exonic SNPs per person, of which 58% are nonsynonymous. Among the class of SNPs, SAPs are of broad research interest due to their accountability in causing half of the known gene lesions responsible for human inherited diseases (Krawczak et al., 2000; Stenson et al., 2003). Therefore, these SAPs are classified as deleterious ones, which have an impact on protein function, thereby leading to dramatic phenotypic change (Sunyaev et al., 2001; Wang & Moult, 2001; Yue & Moult, 2006). However, majority of the SAPs were hypothesized to be neutral or tolerant SAPs, which do not contribute to any phenotype (Masso & Vaisman, 2010; Ng & Henikoff, 2006; Shastry, 2006a, 2006b). Differentiation of deleterious from neutral or tolerant SAPs is very essential in characterizing the genetic basis and pathogenesis of human disease in medical genetics, thereby able to access individual susceptibility to disease (Dimmic, Sunyaev, & Bustamante, 2005). SAPs affect the functional roles of proteins in signal transduction of visual, hormonal, and other stimulants (Dryja et al., 1990; Smith et al., 1994) in gene regulation by altering DNA and transcription factor binding (Barroso et al., 1999), and in maintaining the structural integrity of cells and tissues (Thomas et al., 1999). SAPs inactivate functional sites of enzymes or alter splice sites and thereby form defective gene products (Jaruzelska et al., 1995; Yoshida, Huang, & Ikawa, 1984).

SAPs may affect drug–receptor or drug–enzyme interactions by inducing structural change in receptors or active target-enzyme sites (Bonnardeaux et al., 1994; Erdin, Ward, Venner, & Lichtarge, 2010; Rignall et al., 2002; Ung, Lu, & McCammon, 2006; Vatsis, Martell, & Weber, 1991), ion channels (Wang et al., 1996), and proteins involved in the detoxification pathways (Hassett, Aicher, Sidhu, & Omiecinski, 1994). Furthermore, SAPs may destabilize proteins, or reduce protein solubility (Proia & Neufeld, 1982), and also have functional effects on transcriptional regulation, by affecting transcription factor binding sites in the promoter or intronic enhancer regions (Prokunina & Alarcon-Riquelme, 2004), or alternatively splicing regulation by disrupting exonic splicing enhancers or silencers (Cartegni & Krainer, 2002).

To understand the mechanism of phenotypic variations due to SAPs, it is important to measure the structural consequences due to change in amino acid residue. A well-known classical example is sickle-cell anemia, studied by Sir John Kendrew 55 years ago, which results from the substitution of V instead of E in sixth position of the beta chain of hemoglobin reducing the solubility of the deoxygenated form of hemoglobin markedly (Stryer, 1995). Several studies have illustrated the importance of SAPs in affecting cellular function in the variety of ways. It includes occurrence of SAPs in the active sites (Stevanin et al., 2004; Yamada et al., 2006) or surrounding amino acid residue involved in ligand binding or amino acid residue involved in contact with surrounding proteins will alter the function of the protein. When an SAP occurs near the active site, it might alter the characteristic of the catalytic groups (Koukouritaki et al., 2007; Takamiya, Seta, Tanaka, & Ishida, 2002; Zhang, Norris, Schwartz, & Alexov, 2011; Zhang, Wang, et al., 2010). This will alter the kinetic properties (optimum cellular environment) such as pH, temperature, and salt concentration (Alexov, 2004; Fujiwara et al., 2000). Furthermore, these SAPs can affect the protein stability (Dobson, 2003; Gromiha, Oobatake, Kono, Uedaira, & Sarai, 1999; Ode et al., 2007; Shirley, Stanssens, Hahn, & Pace, 1992; Wang & Moult, 2001), protein flexibility (Karplus & Kuriyan, 2005; Song et al., 2005; Tang & Dill, 1998; Young, Gonfloni, Superti-Furga, Roux, & Kuriyan, 2001), protein folding (Dobson, 2003; Thomas, Qu, & Pedersen, 1995), solvent accessibility (Gromiha et al., 1999; Karchin, Diekhans, et al., 2005; Karchin, Kelly, & Sali, 2005; Kleina & Miller, 1990; Rennell, Bouvier, Hardy, & Poteete, 1991; Rose & Wolfenden, 1993; Stitziel et al., 2003), secondary structure elements (Chasman & Adams, 2001; Ferrer-Costa, Orozco, & de la Cruz, 2002; Gromiha & Ponnuswamy, 1993;

Saunders & Baker, 2002), protein aggregation (Board, Pierce, & Coggan, 1990; Keage et al., 2009; Valerio et al., 2005; Wong, Fritz, & Frishman, 2005), protein–protein interaction (Akhavan et al., 2005; Dixit, Torkamani, Schork, & Verkhivker, 2009; Hardt & Laine, 2004; Jones et al., 2007; Ma, Elkayam, Wolfson, & Nussinov, 2003; Ortiz, Light, Maki, & Assa-Munt, 1999; Ozbabacan, Gursoy, Keskin, & Nussinov, 2010; Rignall et al., 2002; Teng, Madej, Panchenko, & Alexov, 2009; van Wijk, Rijksen, Huizinga, Nieuwenhuis, & van Solinge, 2003; Zhang et al., 2011), protein–DNA interaction (Elles & Uhlenbeck, 2008; Venkatesan et al., 2007; Wright & Lim, 2007), subcellular localization (Boulling et al., 2007; Castella et al., 2011; Hanemann, D'Urso, Gabreëls-Festen, & Müller, 2000; Kim, Hyrc, et al., 2011; Kim, Kim, et al., 2011; Laurila & Vihinen, 2009; Moosawi & Mohabatkar, 2009), protein expression (Boulling et al., 2007; Hanemann et al., 2000), and posttranslational modifications (Grasbon-Frodl et al., 2004; Radivojac et al., 2008; Ryu et al., 2009; Thomas et al., 2004; Tolkacheva et al., 2001; Vazquez, 2000; Vogt et al., 2007). These mounting studies imply the varying functional role of SAPs, which can have a large effect on an organism or species. It is assumed that SAPs in the protein sequences that are observed among living organisms have survived natural selection. Population genetic studies describe that a significant fraction of functional SAPs was present in the highly conserved regions. Residues that evolve under strong selective pressure are found to be significantly associated with human diseases (Arbiza et al., 2006). Disease-causing or deleterious mutations are most likely to correspond to evolutionarily conserved positions in protein sequence due to their functional importance (Tavtigian et al., 2006; Thusberg & Vihinen, 2009). Generally, functional consequences of SAPs fall into two types, namely, disease-associated (deleterious) and benign (no observable phenotypic effect) (Bao & Cui, 2006). The researches of structural and evolutionary features that discriminate the two classes of SAPs have many important applications. First, such features will help to identify disease-associated SAPs from the majority of benign SAPs and to reveal the molecular background of genetic diseases (Karchin, Kelly, et al., 2005). Second, such features will help to determine crucial residues and to elucidate the sequence–structure–function paradigms for individual proteins (Murphy, Barrantes-Reynolds, Kocherlakota, Bond, & Greenblatt, 2004; Wang & Moult, 2003). Finally, such features can be used to guide the selection of target sites in artificial mutagenesis experiments (Dambosky, Prokop, & Koca, 2001). Importantly, SAPs result in altered protein products, which might lead to change in

drug–target phenotypes and thereby cause dysfunction of drugs. Moreover, SAPs may produce altered effects in drug transporters, drug-metabolizing enzymes, and drug–target proteins (Ingelman-Sundberg, Sim, Gomez, & Rodriguez-Antona, 2007; Tomalik-Scharte, Lazar, Fuhr, & Kirchheiner, 2008; Zhou et al., 2009), which results in variability of patient–drug responses. To address this, gaining a detailed understanding of the effect of genetic variants on patient–drug response and underlying mechanism is a key part in the establishment of personalized medicine (Fernald, Capriotti, Daneshjou, Karczewski, & Altman, 2011; Rodriguez-Casado, 2012).

Knowledge of a protein's three-dimensional (3D) structure is not only used for energy calculations but also necessary for a full understanding of a mutational effect on its functionality (Capriotti & Altman, 2011), drug–target interaction, and the relationships between mutations and drug response (Lahti, Tang, Capriotti, Liu, & Altman, 2012; Rodriguez-Casado, 2012; Weigelt, 2010). To understand the structure–function relationship, it is necessary to map a mutation onto known protein 3D structure, which acts as a powerful tool in revealing the mechanistic explanation of their effects on function. Proteins with mutations do not always have 3D structures that are solved and deposited in Protein Data Bank (PDB). Therefore, it is necessary to construct 3D models using homology modeling by locating the variation in 3D. This acts as a powerful tool to reveal what kind of adverse effects a mutation can have on protein. This process of detailed structural analysis of protein–drug interactions was not always feasible in the past, but advances in structural genomics have resulted in an explosion of high-resolution structures of known and potential drug–target proteins.

The last decade has witnessed a drastic increase in genomic information, especially SNPs in public databases and lends itself to an informatics approach. Bioinformatics, especially computational molecular biology, is playing a vital role in extracting knowledge from the vast amount of genomic information generated by different genomics technologies. As a result, various computational resources were developed to aid identification and characterization of the functional SAPs and to study the impact of SAPs in protein, patient–drug response, and current therapeutic targets. In this chapter, we intend to provide an overview of the existing computational methods in SAP analysis available on the World Wide Web. Summing up, the techniques presented in this chapter will build a bridge between computational methods and clinicians toward personalized medicine in tailoring new treatment strategies.

2. COMPUTATIONAL METHODS IN SAP ANALYSIS

With the ever-increasing influx of high-throughput technologies and the pronounced ascendency of the Human Genome Project (HGP), the scientific communities of disease research and drug design have witnessed a paradigm shift toward single amino acid substitutions (SAP). With a successfully completed genome, genome-wide association studies were initiated using SAPs as markers to study disease–gene associations. The commonality of SAPs in a genome makes them suitable detection points for studying disease susceptibility (Curtis, North, & Sham, 2001). The last decade has witnessed extraordinary advances in experimental and computational technologies to identify, characterize, and differentiate pathogenic SAPs from neutral ones. Traditional methods find a unique place in disease diagnosis and treatment of an individual. Employing traditional methods for SAPs analysis will consume precious time, require increased labor, and thus often turn detrimental to the study itself. Computational techniques with cutting edge software and innovative algorithms are fast improving the arduous task of analyzing genetic polymorphism data. This will serve as a powerful screening process to single out potentially deleterious SAPs from the whole stack and examine it as a possible drug–target through stability studies (Mah & Chia, 2007). Increased deliberation in this regard led to the generation of numerous algorithms and methods to provide an edge to the study of SAPs. Many computational methods were developed to predict whether an SAP is deleterious to the structure or the function of the gene and will, therefore, lead to disease. These predictions of SAP are of three classes: (1) predict the effect on protein function, (2) predict the effect based on the pathogenicity, and (3) predict the effect on protein structure stability. Researchers have taken many input features such as sequence-based properties, physical properties of the wild-type and mutant amino acids, protein structural properties (solvent accessibility, location within beta strands or active sites, and participation in disulphide bridges), and evolutionary properties derived from a phylogeny or sequence alignment to predict an SAP as deleterious/disease/pathogenic/intolerable/nonneutral or neutral/tolerable. To classify whether an SAP will be tolerable, a training set is usually constructed of mutations known to be deleterious. For example, these training sets can be derived from saturation mutagenesis experiments, where the mutation severity is determined in activity assays (Cai et al., 2004;

Chasman & Adams, 2001; Krishnan & Westhead, 2003; Ng & Henikoff, 2001; Saunders & Baker, 2002), multiple sequence alignments where tolerance to mutation is derived from evolutionary analyses of sequence positions (Sunyaev et al., 2001), or known deleterious human mutations. In this chapter, we have provided a brief classification and description of some well-known reliable SAP prediction methods available to this day. For the accurate description of the process of SAP analysis, we have provided a pictorial representation of the entire process in Fig. 10.1.

The computational prediction methods can be divided into four most important families of algorithms: machine-learning methods, empirical rule-based methods, physics-based models, and evolutionary theory-based models to classify SAPs. Machine-learning approaches develop classification models that automatically learn from the training data, extract patterns from complex data, and make predictions of new cases. Random forests (Bao, Zhou, & Cui, 2005; Li et al., 2009; Mathe et al., 2006), neural networks (Bromberg & Rost, 2007; Ferrer-Costa et al., 2005; Linding, Russell, Neduva, & Gibson, 2003; Yang, Thomson, McNeil, & Esnouf, 2005), Decision Trees (Yuan et al., 2006), support vector machines (Calabrese, Capriotti, Fariselli, Martelli, & Casadio, 2009; Capriotti & Altman, 2011; Capriotti et al., 2013; Capriotti, Fariselli, & Casadio, 2005; Capriotti, Fariselli, Rossi, & Casadio, 2008; Karchin, Kelly, et al., 2005; Mathe et al., 2006; Parthiban, Gromiha, Hoppe, & Schomburg, 2007; Tian et al., 2007; Yue & Moult, 2006), naive Bayes approach (Adzhubei et al., 2010; Schwarz, Rodelsperger, Schuelke, & Seelow, 2010), hidden Markov models (Mi, Muruganujan, & Thomas, 2013; Shihab et al., 2013), and rule-based methods (Kumar, Henikoff, & Ng, 2009; Ramensky, Bork, & Sunyaev, 2002; Reva, Antipin, & Sander, 2011; Tavtigian et al., 2006; Zhou & Zhou, 2002) are among the most widely used machine-learning methods for SAP analysis. Few meta-analysis suite tools are also available, which combine the prediction information from the abovementioned prediction methods (Jegga, Gowrisankar, Chen, & Aronow, 2007; Lee & Shatkay, 2008; Olatubosun, Valiaho, Harkonen, Thusberg, & Vihinen, 2012; Schaefer, Meier, Rost, & Bromberg, 2012; Wang, Ronaghi, Chong, & Lee, 2011) and SAP prioritization tools (Conde et al., 2004; Freimuth, Stormo, & McLeod, 2005; Lee & Shatkay, 2008; Wjst, 2004; Xu et al., 2005). The abovementioned methods use several different inputs such as NCBI GI number OR RefSeq ID, wild-type protein FASTA sequences, and wild and new residue after mutation (single-letter amino acid code) for making their predictions. In order to quantify the destabilization

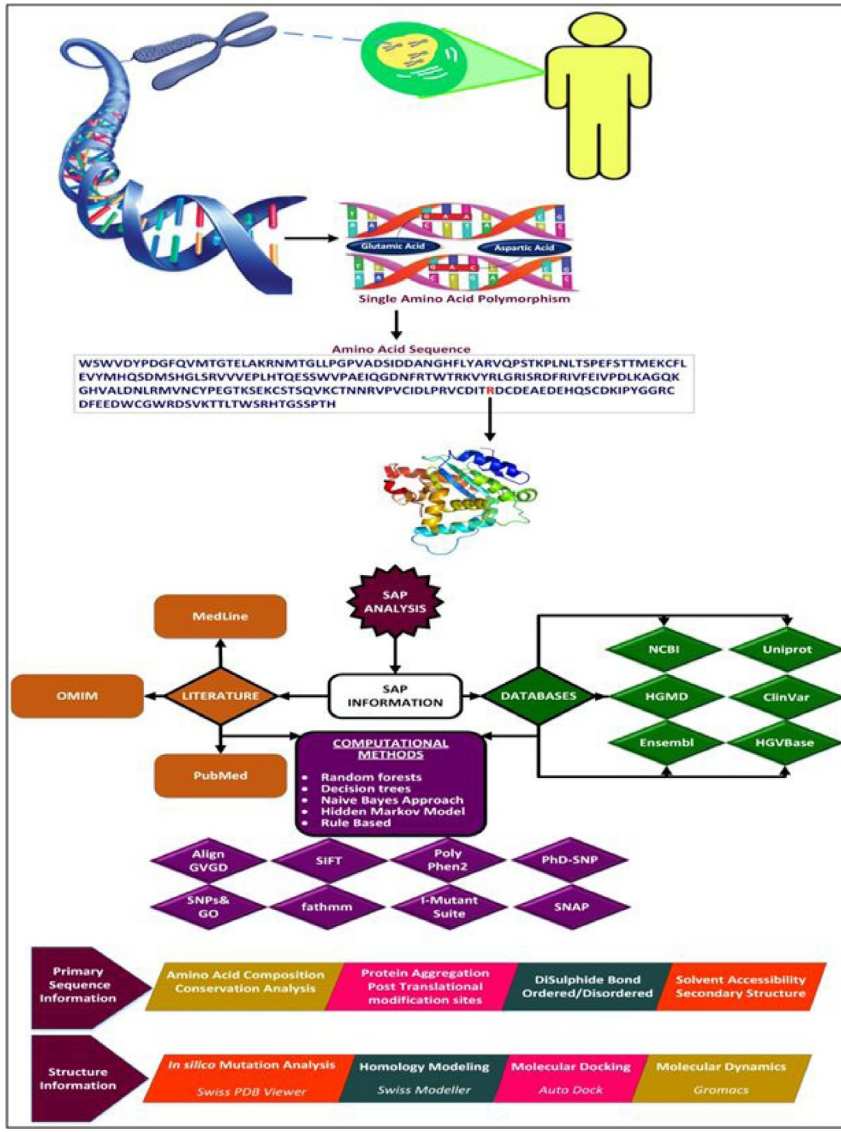

Figure 10.1 Computational pipeline in Single Amino Acid Substitutions analysis.

effects of SAPs, the protein stability change upon mutations can be evaluated by calculating the difference in folding free energy change between wild-type and mutant proteins (ΔΔG or ddG) without performing any experimental results (Capriotti et al., 2005; Chen, Lin, & Chu, 2013; Cheng,

Randall, & Baldi, 2006; Dehouck et al., 2009; Dosztanyi, Magyar, Tusnady, & Simon, 2003; Magyar, Gromiha, Pujadas, Tusnády, & Simon, 2005; Masso & Vaisman, 2008; Parthiban, Gromiha, & Schomburg, 2006). SAP prediction tools can be classified for better understanding into sequence-, structure-, and sequence- and structure-based methods.

3. DATABASE RESOURCES FOR SAPs

As of November 7, 2013, about 62,676,337 SNPs of Homo sapiens were identified and deposited in the major repository National Center for Biotechnology Information (NCBI) database (Sherry et al., 2001) (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?view+summary=view+summary&build_id=138). In addition, there are now a few other extensive databases that provide information about the DNA variations. This includes UniProt database (Apweiler et al., 2010), ClinVar database (Maglott et al., 2013), human genome variation database, HGVBase (Fredman et al., 2002), Human Gene Mutation Database (HGMD) (Stenson et al., 2012), Online Mendelian Inheritance in Man (OMIM) (Amberger, Bocchini, & Hamosh, 2011), and MutDB (Mooney & Altman, 2003). Repositories like dbSNP and UniProt contain information about the experimentally proved SAPs, but a few of these are annotated with respect to function. OMIM contains disease-related literature information; HGMD contains disease variants with one or more references in the literature, and ClinVar database contains clinically significant variant information. The availability of a comprehensive SNP catalog offers the possibility of identifying many disease loci and eventually pinpointing functionally important variants in which the nucleotide change alters the function or expression of a gene that directly influences a disease outcome. The study of the distribution of SNPs particularly in different populations is also valuable for investigating molecular events that underlie the evolution, namely, genetic drift, mutation, recombination, and selection. These illustrate important changes in human history, for example, tracing the origin of populations and their migrations. In addition, disease-specific databases were initiated to specifically collect variant information pertaining to one disease (Basu, Kollu, & Banerjee-Basu, 2009; caBIG Strategic Planning Workspace (caBIG), 2007; George Priya Doss, Nagasundaram, Srajan, & Chiranjib, 2012; Ingman & Gyllensten, 2006; Nuytemans, Theuns, Cruts, & Van Broeckhoven, 2010; Ruiz-Pesini et al., 2007) and in defining the molecular basis of the disease (Giardine et al., 2007; Kawabata, Ota, & Nishikawa,

1999) and structural stability (Gromiha et al., 1999). In addition, visualization tools were developed to analyze the effect of SAP on protein structure along with conservation and physicochemical properties (Chang & Fujita, 2001; Han et al., 2006; Luu, Rusu, Walter, Linard, et al., 2012; Luu, Rusu, Walter, Ripp, et al., 2012; Uzun, Leslin, Abyzov, & Ilyin, 2007; Venselaar, Te Beek, Kuipers, Hekkelman, & Vriend, 2010). Several researchers have reviewed in detail about the computational methods and also compared the performance of each computational method over the other existing methods, which are available online (Castellana & Mazza, 2013; Frousios, Iliopoulos, Schlitt, & Simpson, 2013; George Priya Doss, Rajasekaran, Arjun, & Sethumadhavan, 2010; George Priya Doss & Sethumadhavan, 2009a, 2009b; George Priya Doss et al., 2008; Gnad, Baucom, Mukhyala, Manning, & Zhang, 2013; Gray, Kukurba, & Kumar, 2012; Khan & Vihinen, 2010; Peterson, Doughty, & Kann, 2013; Reumers, Schymkowitz, & Rousseau, 2009; Thusberg, Olatubosun, & Vihinen, 2011; Tiffin, Okpechi, Perez-Iratxeta, Andrade-Navarro, & Ramesar, 2008; Wang, Eickholt, & Cheng, 2010; Wang, Sun, Akutsu, & Song, 2013).

4. MOLECULAR PHENOTYPIC EFFECT ANALYSIS

SAPs can alter the biophysical properties of amino acid residues such as size, charge, hydrogen bonding, hydrophobic contacts, disulfide bonds, van der Waals, and electrostatic interactions at critical folding positions, which will have an impact on residue contacts and thereby lead to loss of protein stability as well as folding, flexibility, and aggregation of the protein (Betz, 1993; Dill, Ozkan, Weikl, Chodera, & Voelz, 2007; Eriksson et al., 1992; Horovitz, Serrano, Avron, Bycroft, & Fersht, 1990; Pace et al., 2011). Most often the change in protein stability results in increased propensity for protein aggregation. As a consequence, SAPs can have a considerable effect in the solubility and aggregation propensity of a protein (Karplus & Kuriyan, 2005; Keage et al., 2009; Valerio et al., 2005; Wong et al., 2005). Several computational algorithms have been proposed to predict the aggregation-nucleating sequences in proteins using either sequence-based or structural bioinformatics tools (Conchillo-Sole et al., 2007; De Baets et al., 2012; Fernandez-Escamilla, Rousseau, Schymkowitz, & Serrano, 2004; Garbuzynskiy, Lobanov, & Galzitskaya, 2010; Maurer-Stroh et al., 2010; Trovato, Seno, & Tosatto, 2007; Tsolis, Papandreou, Iconomidou, & Hamodrakas, 2013; Van Durme et al., 2009). SAPs also introduce disorder

predisposition in a target protein, which may affect the protein conformation, increase the flexibility, and lead to alterations in function. SAP effects on structural disorders can be analyzed by providing a sequence as the input (Cheng, Randall, Sweredoski, & Baldi, 2005; Cheng, Sweredoski, & Baldi, 2005; Dosztányi, CsizmLok, Tompa, & Simon, 2005; Galzitskaya, Garbuzynskiy, & Lobanov, 2006; Ishida & Kinoshita, 2007, 2008; Linding, Russell, Neduva, & Gibson, 2003; Prilusky et al., 2005; Sickmeier et al., 2007; Vullo, Bortolami, Pollastri, & Tosatto, 2006). There are significant numbers of reports that explain the involvement of missense mutations in posttranslational target sites leading to diseases (Grasbon-Frodl et al., 2004; Radivojac et al., 2008; Vogt et al., 2007). Posttranslation modification sites (PTMs) are implicated in many cellular processes and have a vital role in regulating the functional and structural properties of protein (Walsh, 2006). Different PTMs of known protein like phosphorylation, glycosylation, methylation, acetylation, and sumoylation can be analyzed by various computational methods (Blom, Sicheritz-Ponten, Gupta, Gammeltoft, & Brunak, 2004; Chang et al., 2009; Gupta et al., 1999; Huang, Lee, Tseng, & Horng, 2005; Kiemer, Bendtsen, & Blom, 2005).

5. SEQUENCE INFORMATION ANALYSIS

Protein primary sequence provides the most direct and readily available information regarding the clues for functional mutation sites that can be extracted from the amino acid sequence in cases where no structural information is available. Population genetic studies describe that a significant fraction of functional SAPs was present in a conserved region. Residues that evolve under strong selective pressure are found to be significantly associated with human diseases (Arbiza et al., 2006). The importance of residue for maintaining the structure and function of a protein can usually be inferred from how conserved it appears in a multiple sequence alignment of that protein and its homologues. A comparative analysis of amino acid conservation from multiple species by protein sequence alignments gives an indication of which amino acid residues are truly conserved and which of them represent localized evolution. It is assumed that SAPs in the protein sequences that are observed among living organisms have survived natural selection. Disease causing, or deleterious mutations are most likely to correspond to evolutionarily conserved positions in protein sequence due to their functional importance (Tavtigian et al., 2006; Thusberg & Vihinen, 2009). Several methods were made available online predict the conservation

analysis of multiple sequence alignments (Ashkenazy, Erez, Martz, Pupko, & Ben-Tal, 2010; Berezin et al., 2004; Gu & Vander Velden, 2002; Pupko, Bell, Mayrose, Glaser, & Ben-Tal, 2002; Siepel et al., 2005). ConSurf (Ashkenazy et al., 2010) is a tool based on comparative analysis of amino acid conservation from multiple species by protein sequence alignments and provides an indication of which amino acid residues are truly conserved and which of them represent localized evolution. There is extensive research indicating the function of substituted amino acids in causing diseases (Dobson, Munroe, Caulfield, & Saqi, 2006; Khan & Vihinen, 2007; Vitkup, Sander, & Church, 2003) and Trp and Cys residues in determining protein stability (Arbiza et al., 2006). Few studies have illustrated the importance of Cys residues in a protein sequence since most of the protein foldings are dependent on disulfide bonds (Song, Geng, et al., 2009; Song, Lim, & Tong, 2009). SNP-associated residue changes to or from Cys will likely destabilize the protein structure. Composition of each amino acid can be calculated by statistical analysis of protein sequences (Brendel, Bucher, Nourbakhsh, Blais-dell, & Karlin, 1992; Cheng et al., 2006). Solvent accessibility, considered as a discriminating feature in disease-associated SAPs, tends to occur at buried sites; benign substitutions tend to occur at solvent accessible sites (Ferrer-Costa et al., 2002; Sunyaev et al., 2001). Solvent accessibility from an exposed to buried state could be considered functionally significant in the mutant protein at the structural level (Chen & Zhou, 2005), which can be accessed by ACCpro (Pollastri, Baldi, Fariselli, & Casadio, 2002), WHAT IF (Vriend, 1990) and WESA (Chen & Zhou, 2005). Secondary structure elements can be analyzed using NetSurfP-1.1 (Petersen, Petersen, Andersen, Nielsen, & Lundegaard, 2009), Jpred (Cole, Barber, & Barton, 2008), YASPIN (Lin, Simossis, Taylor, & Heringa, 2005), STRIDE (Heinig & Frishman, 2004), DSSP (Kabsch & Sander, 1983), and SSPro (Cheng, Randall, et al., 2005; Cheng, Sweredoski, et al., 2005).

6. COMPUTATIONAL METHODS FOR STRUCTURE DETERMINATION

As of November 2013, PDB contains 95280 entries (Bernstein et al., 1997) of experimentally solved structures, which includes multiple structures of the same protein, while UniProtKB/Swiss-Prot contains 541561 sequence entries (Apweiler et al., 2004) and NCBI RefSeq database contains 33,139,144 protein entries (Pruitt, Tatusova, Brown, & Maglott, 2012).

Due to the advent of cost-effective high-throughput gene sequencing technologies, the number of sequence entries in the aforementioned databases is increasing. Furthermore, the number of solved structure determination will tend to increase; the number of newly discovered sequences grows much faster than the number of structures solved (Levitt, 2007). The protein structures are solved by X-ray crystallography, nuclear magnetic resonance (NMR), and high-resolution molecular microscopy (EM). These methods of structural determination are limited by cost, time consumption, and requirement of specialized instruments, which leads to a large gap between the solved structures and available protein sequences in the databases. Due to this, the application of computational-based methods in 3D structure predictions has increased and also become a valuable resource in defining protein function (Hermann et al., 2007) and studying the impact of mutation at structural level (George Priya Doss, Chakraborty, Rajith, & Nagasundaram, 2013; Kosinski, Hinrichsen, Bujnicki, Friedhoff, & Plotz, 2010) and drug discovery (Liu, Tang, & Capriotti, 2011). Existing computation methods for structure determination fall into two categories (Zhang, 2008a, 2008b): templates-based comparative (or homology) and threading methods, which utilize structures of known homologous proteins as starting templates (Kolinski, Rotkiewicz, Ilkowski, & Skolnick, 1999; Rost, Fariselli, & Casadio, 1996), and free modeling methods (*de novo* and *ab initio*), which apply the principles of physical chemistry in protein folding, often in combination with efficient fragment searching techniques (Jothi, 2012; Lesk, 1997; Zemla, Venclovas, Reinhardt, Fidelis, & Hubbard, 1997). These computational structure prediction methods were discussed in detail (Kryshtafovych & Fidelis, 2009; Pierri, Parisi, & Porcelli, 2010; Werner, Morris, Dastmalchi, & Church, 2012; Zhang, 2009). In this section, we have discussed the steps followed in homology modeling. In template-based methods, the tertiary structure of an unknown protein can be modeled using a known 3D structure of protein with the homologous sequence (homology modeling), while, in fold recognition, the protein structure was modeled based on the proteins with known structures having the same fold but no homology to the proteins with known structure (Daga, Patel, & Doerksen, 2010; Martí-Renom et al., 2000; Qu, Swanson, Day, & Tsai, 2009). Homology modeling uses only sequence similarity, whereas fold recognition uses both structure and sequence relationship. Homology modeling consists of four major steps: (a) template identification, (b) alignment of target sequence with template structures, (c) model building, and (d) model evaluation.

Template identification is one of the most important steps in homology modeling, which is performed by searching the target sequence in databases such as PDB, which includes solved structures. The commonly used searched methods BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990), FASTA (Lipman & Pearson, 1985; Pearson & Lipman, 1988), PSI-BLAST (Altschul et al., 1997), HHSearch (Koehl & Delarue, 1994), HHpred (Söding, Biegert, & Lupas, 2005), and Phyre (Kelley & Sternberg, 2009) provide a ranking of templates along with "*E*-value" alignment scores. A good candidate template is selected based on the *E* value equal to zero, with the highest similarity and a template of solved structures by X-ray crystallography. A sequence identity cutoff of 30% is considered as the standard threshold in homology modeling. In the case of low identity, Doolittle (1986) formulated three rules for template selection: sequences longer than 100 amino acids with <25% identities (with gaps) are probably related; sequences with 15–25% identity might be related ("twilight zone") but need additional statistical analyses to help establish this with confidence; sequences with 15% identity are most likely not related. Choosing multiple templates can improve the quality of the model when compared to use a single template (Fernandez-Fuentes, Rai, Madrid-Aliste, Fajardo, & Fiser, 2007; Kosinski, Tkaczuk, Kasprzak, & Bujnicki, 2008; Wallner, Lindahl, & Elofsson, 2008). Once a final template is selected, target-template alignment should be performed by using unique pair-wise or multiple sequence alignment tools such as Clustal Omega (Sievers et al., 2011), T-Coffee (Notredame, Higgins, & Heringa, 2000), ClustalW2 (Larkin et al., 2007), 3DCoffee (O'Sullivan, Suhre, Abergel, Higgins, & Notredame, 2004), and Muscle (Edgar, 2004). Tress, Jones, and Valencia (2003) proposed that inclusion of regions sharing highest sequence similarity along with common motifs can be considered as correctly aligned because they tend to be evolutionarily conserved. Additional sequence information, localization of hydrophobic regions, secondary structure elements, and disulphide bonds are considered to improve alignment. The next step in homology modeling involves model building, where a 3D structure model is built based on the given target-template alignment and template structures. Nowadays, this procedure has become fully automated. Basically, model building methods are grouped as follows (Wallner & Elofsson, 2005; Xiang, 2006): rigid-body assembly methods, which build a model from the structurally conserved regions of the template that align to the target sequence, like 3D-JIGSAW (Bates, Kelley, MacCallum, & Sternberg, 2001), BUILDER (Koehl & Delarue, 1995), and SWISS-MODEL (Arnold, Bordoli, Kopp, & Schwede, 2006);

segment matching methods like SegMod/ENDCAD (Levitt, 1992); spatial restraint methods like MODELLER (Eswar et al., 2006); and artificial evolution methods like NEST (Petrey et al., 2003). A spatial restraint method utilizes satisfying restraints such as bond lengths and angles, van der Waals contact distances, and dihedral angles to map onto the target-template structure alignments. Studies comparing the model building methods have rated MODELLER as the best among the existing model building methods (Dalton & Jackson, 2007; Wallner & Elofsson, 2005). Model evaluation remains as the fundamental and most important step in homology modeling and in defining whether the model created is of good quality. The error in homology structure comes from the side chains and loops. For this, many approaches including hybrid methods have been proposed to rectify these errors (Arnold et al., 2006; Das & Baker, 2008; Deane & Blundell, 2001; Fernandez-Fuentes, Zhai, & Fiser, 2006; Holm & Sander, 1992; Hwang & Liao, 1995; Koehl & Delarue, 1994; Krivov, Shapovalov, & Dunbrack, 2009; Lee & Subbiah, 1991; Liang, Zheng, Zhang, & Standley, 2011; Rohl, Strauss, Misura, & Baker, 2004; Samudrala & Moult, 1998; Sippl, 1993; Xiang, Soto, & Honig, 2002; Xu & Berger, 2006). Additionally, many model quality assessment programs are made available, which define a scoring function that is capable of discriminating good and bad models (Bowie, Lüthy, & Eisenberg, 1991; Davis et al., 2007; Hooft, Vriend, Sander, & Abola, 1996; Laskowski, MacArthur, Moss, & Thornton, 1993; Lüthy, Bowie, & Eisenberg, 1992; Sippl, 1993; Wiederstein & Sippl, 2007). Threading methods can be used in place of homology modeling, when template structures share less than ~30% sequence identity with the target sequence with evolutionary relationship (Miller, Jones, & Thornton, 1996; Xu, Li, Kim, & Xu, 2003). The structures that are not modeled by homology modeling and threading can be performed using free modeling servers (Das et al., 2007; Jayaram et al., 2006; Kim, Chivian, & Baker, 2004; Kinch et al., 2011; Rohl et al., 2004; Wang, Yang, Li, Liu, & Zhou, 2010; Zhang, 2008a, 2008b; Zhang, Wang, et al., 2010). The abovementioned structure prediction methods have their own advantages and disadvantages, and their consistency varies for each structural problem. Choosing appropriate methods depends mainly on the availability of a suitable template and computational resources.

7. DOCKING

All the drugs presently available in the market have gone through years of clinical research and drug trials (Jorgensen, 2004). Even a single disease

can have many options for a drug. It is estimated that a typical drug-discovery cycle, from lead identification to clinical trials, can take 14 years with a cost of 800 million US dollars before it can be sold in the market to the general public. It takes the scientists even longer to finally select that one "potential" drug from among the many other options. Rapid advancements in the fields of genomics, proteomics, and biotechnology have fuelled the drug-discovery process. This constant expansion has led to the approval of 18 drugs for human use by the US Food and Drug Administration, with approximately four acting on novel target structures (Rask-Andersen, Almen, & Schioth, 2011). In spite of this success, an "efficacy–effectiveness gap" exists and is, ultimately, the result of variability in patient–drug responses. Also, marketed drugs exhibit limited efficacy of about 30–60% (Sadee & Dai, 2005; Wilkinson, 2005) and may also exhibit drug toxicity (Wilke & Dolan, 2011) toward the patient. Drug discovery and drug research have contributed more to the progress in the field of medicine than any other scientific factor. Many studies have highlighted that drug discovery and designing is a complex, highly expensive, risky, and quite cumbersome *process* (Congreve, Murray, & Blundell, 2005; DiMasi & Grabowski, 2007; DiMasi, Grabowski, & Vernon, 2004; DiMasi, Hansen, Grabowski, & Lasagna, 1991; Kolb & Sharpless, 2003). The drug-discovery process aids in designing a candidate drug or lead compound, which binds with the target-specific protein whose function is thought to be essential for the disease phenotype. The traditional drug-discovery process involves three major steps: target identification and validation, lead identification, and lead optimization. This process is expensive and takes a very long time in identifying the drug. Modern drug discovery involves application of computational methods in an efficient manner to predict drug–target interaction, stability, and activity. Bioinformatics, through various tools and techniques, can eliminate the numerous drug choices to just 10 drugs or even fewer. Virtual screening is the screening of many compounds that might dock with a particular target macromolecule and lead to the formulation of a potential drug (Taboureau, Baell, Fernández-Recio, & Villoutreix, 2012). It is a high-throughput screening of millions of compound databases in the hopes of finding a unique compound or a drug that can replace an existing drug or that can shed light on diseases with no drugs. All the compounds that are screened do not necessarily exist (Lavecchia & Di Giovanni, 2013). Millions of compounds can be screened easily without having to spend much time, effort, or money. Virtual screening can be divided into two broad categories: ligand-based and structure-based. There

are stringent rules that have to be followed in virtual screening. Many *in silico* tools can be used to design libraries of compounds with drug-like properties (Villoutreix et al., 2007). These are predominantly biophysical properties based on empirical rules. A well-known example is Lipinski's "rule of five" (Lipinski, 2000), which states that a compound is likely to be "nondrug-like" if it has more than 5 hydrogen bond donors and more than 10 hydrogen bond acceptors; molecular mass is greater than 500 and lipophilicity is above 5. New pharmacokinetic data found in rats have caused this rule to be revisited (Ridder, Wang, de Vlieg, & Wagener, 2011). Many related rules have been subsequently modified and proposed as the "rule of three" (Rees, Congreve, Murray, & Carr, 2004), which defines fragment properties with an average molecular weight $\leq$300 Da, a *C*log $P \leq 3$, the number of hydrogen bond donors ≤ 3, the number of hydrogen bond acceptors ≤ 3, and the number of rotatable bonds <3. Recently, Pfizer's "rule of 3/75" has been described, which states that compounds with a calculated partition coefficient (*C*log *P*) of <3 and topological polar surface area (TPSA) >75 have the best chances of being well tolerated from a safety perspective *in vivo* (Hughes et al., 2008).

There are various new drugs whose development was heavily influenced by computational methods and screening strategies (Durrant & McCammon, 2011). One such important example is the HIV protease inhibitor. Most drugs are ligand–protein complexes, where the ligand enhances the function of the protein, which in turn helps it to fight against the disease. The other alternative is when the ligand helps in downregulating the expression of a particular protein it binds to. Both the scenarios depend on how well and easily the ligand and protein bind to each other. Such results of ligand–protein binding can be obtained by performing molecular docking (Gulati, Cheng, & Bates, 2013). There are many sources for the structure of the protein and drugs or ligands separately (Pak & Wang, 2000). There are not much data for protein and drug complexes.

There are three major advances in the field of drug designing (Rao & Srinivas, 2011):

- The first one is the conformational modeling of all small molecules, ligands, macromolecules, and their complexes; these are molecules that are potential drug candidates.
- The next is to determine their physical, chemical, and biological properties called property modeling.
- The last is to optimize the chemical, physical, and biological properties or molecular design.

Docking is not the only process that is important. Much before docking, the search and retrieval of sequences or 3D structures of target proteins are more important. Some of the databases for retrieval of sequences include NCBI, GenBank, and UniProt. Drug banks have formed an integral part of today's drug discovery and designing. A drug bank is a unique database that consolidates information on drugs. Drug banks combine the sequence, structure, and pathway information of drugs or target drug molecules with chemical, physical, biological, pharmacological, and pharmaceutical information (Wishart, 2007). Several other databases also exist with known 3D structures of potential drug–targets, ligands, diseases, and their associated pathways.

Molecular docking is an invaluable tool in the field of molecular biology, computational structural biology, computer-aided drug designing, and pharmacogenomics. Docking also plays a vital role in virtual screening of huge libraries. The subsequent results are ranked accordingly, and structural hypothesis of how the ligands inhibit/activate the target macromolecule can be deduced (Azam & Abbasi, 2013). This proves to be an invaluable part in lead optimization of drugs. Most importantly, docking provides certain information that is difficult to deduce through conventional experimental methods. The docking procedure fits two molecules together, protein and ligand in 3D space. Their binding complementarity is then evaluated, and the results are displayed with the best fit "scores." This method is widely used in hit identification and lead optimization of drugs. Docking, simply put, tries to find the best "fit" between two molecules. It can be a protein–ligand docking or a protein–protein docking or a nucleic acid–macromolecule docking. Docking aims at finding out if two molecules can interact. If they do interact, docking attempts to find out the orientation in which the interactions are at its maximum and where the binding energy is at its minimum (Gold, 2007). In a ligand–protein docking, the main goal is to predict the predominant binding pockets; most effective docking softwares perform this by searching high-dimensional spaces efficiently and thoroughly. The protein used for docking purpose should have previously solved 3D structure or a constructed structure by homology modeling (Alonso, Bliznyuk, & Gready, 2006). The two main aspects considered for molecular docking include accurate structural modeling and rendition of active binding site(s). Basically, docking consists of four main steps: (a) preparation of ligand and the receptor, (b) identification of active sites, (c) generation of putative complexes, and (d) evaluation of the complexes by scoring.

The initial step in docking is to identify or define the binding sites by using experimental information derived from the mutagenesis or cross-linking studies or from homologous structures whose binding site is already known. Any molecule, whether a protein or a ligand, may undergo slight structural or conformational changes after binding with another molecule. This makes binding site analysis slightly difficult. Structure-based algorithms are used for binding site analysis (Lahti et al., 2012) (Ghersi & Sanchez, 2011). For the prediction of binding sites, numerous softwares are available, which are categorized into structural similarity approach, geometric approach, and energy-based approach and docking. In the next step of docking, the ligand pose has to be predicted in defined pocket. Several various molecular docking algorithms (Table 10.1) are now available that can fit or "dock" small molecules like ligands into pockets of macromolecules like proteins or sometimes DNA, with different scoring and search algorithms (Abagyan, Totrov, & Kuznetsov, 1994; B-Rao, Subramanian, & Sharma, 2009; Claussen, Buning, Rarey, & Lengauer, 2001; Ewing, Makino, Skillman, & Kuntz, 2001; Friesner et al., 2004; Goodsell, Morris, & Olson, 1996; Jones, Willett, Glen, Leach, & Taylor, 1997; McGann, Almond, Nicholls, Grant, & Brown, 2003; Morris et al., 1998; Rarey, Kramer, Lengauer, & Klebe, 1996), which can predict in an accurate and fast manner. These algorithms determine all possible optimal conformations for a given complex (protein–protein and protein–ligand) in an environment, where each conformation is linked with a final score. In addition, each algorithm calculates the energy of all the resulting conformations of each individual interaction (Sushma & Suresh, 2012). Most docking softwares use scoring techniques that correctly rank the docking conformations (McConkey et al., 2002). Scoring is an important component of docking. The work is futile if there are many conformations but no ranking system. When docking is performed, it is very important to have the right conformation of the docked molecule and also to have each individual conformation to be correctly ranked. This helps to identify the most probable biological conformations (Kollman, 1993). The scoring functions should be able to differentiate between different orientations of the same receptor and ligand. The scoring functions are mathematical functions that assign a value based on the strength of the interaction between the two docked molecules. Each docked conformation is scored for best fit (Zsoldos, Reid, Simon, Sadjad, & Johnson, 2007). This scoring process is repeated the number of times defined by the user or according to the maximum iterations supported by the program. Scoring functions predict factors like van der

Table 10.1 List of computational methods employed in docking analysis

Database name	Resource
Drug banks	
Therapeutic Target Database (Chen et al., 2002)	http://bidd.nus.edu.sg/group/cjttd/
DrugBank (Knox et al., 2011)	http://www.drugbank.ca/
PubChem (Bolton, Wang, Thiessen, & Bryant, 2008; Wang et al., 2009)	http://pubchem.ncbi.nlm.nih.gov/
Binding MOAD (Hu, Benson, Smith, Lerner, & Carlson, 2005)	http://bindingmoad.org/
PDBbind (Wang, Fang, Lu, & Wang, 2004; Wang, Wolf, Caldwell, Kollman, & Case, 2004)	http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp
PDTD (Gao et al., 2008)	http://www.dddc.ac.cn/pdtd/
DGIdb (Griffith et al., 2013)	http://dgidb.genome.wustl.edu/
TDR Targets (Magariños et al., 2012)	http://tdrtargets.org/
SuperDrug (Goede, Dunkel, Mester, Frommel, & Preissner, 2005)	http://bioinf.charite.de/superdrug/
ChemBank (Seiler et al., 2008)	http://chembank.broadinstitute.org/
BindingDB (Liu, Lin, Wen, Jorissen, & Gilson, 2007)	http://www.bindingdb.org/bind/index.jsp
CancerDR (Kumar et al., 2013)	http://crdd.osdd.net/raghava/cancerdr/
Binding site prediction tools	
CASTp (Dundas et al., 2006)	http://sts.bioengr.uic.edu/castp/
LIGSITE (Hendlich, Rippmann, & Barnickel, 1997; Huang & Schroeder, 2006)	http://projects.biotec.tu-dresden.de/pocket/
SURFNET (Laskowski, 1995)	http://www.ebi.ac.uk/thornton-srv/software/SURFNET/
SMAP-WS (Ren, Xie, Li, & Bourne, 2010)	http://nbcr-222.ucsd.edu/smap_ws/
PocketPicker (Weisel, Proschak, & Schneider, 2007)	http://gecco.org.chemie.uni-frankfurt.de/pocketpicker/
FINDSITE (Brylinski & Skolnick, 2008)	http://cssb.biology.gatech.edu/findsite

Continued

Table 10.1 List of computational methods employed in docking analysis—cont'd

Database name	Resource
PBBinder (Hooft et al., 1996)	http://160.80.35.80/PDBinder/
PDBSite (Ivanisenko, Grigorovich, & Kolchanov, 2000)	http://wwwmgs.bionet.nsc.ru/mgs/gnw/pdbsite/
LigAsite (Dessailly, Lensink, Orengo, & Wodak, 2008)	http://www.bigre.ulb.ac.be/Users/benoit/LigASite/index.php?home
3DLigandSite (Wass et al., 2010)	http://www.sbg.bio.ic.ac.uk/~3dligandsite/
PocketAnnotate (Anand, Yeturu, & Chandra, 2012)	http://proline.biochem.iisc.ernet.in/pocketannotate/reference.php
Active Site Prediction (Singh, Biswas, & Jayaram, 2011)	http://www.scfbio-iitd.res.in/dock/ActiveSite_new.jsp
Docking softwares	
AutoDock4.2 (Morris et al., 2009)	http://autodock.scripps.edu
PatchDock (Schneidman-Duhovny et al., 2005a)	http://bioinfo3d.cs.tau.ac.il/PatchDock
ClusPro (Comeau, Gatchell, Vajda, & Camacho, 2004)	http://cluspro.bu.eduhttp://nrc.bu.edu/cluster
DockingServer (Bikadi & Hazai, 2009)	http://www.dockingserver.com
DOCK 6.6 (Brozell et al., 2012)	http://dock.compbio.ucsf.edu
3DLigandSite (Wass, Kelley, & Sternberg, 2010)	http://www.sbg.bio.ic.ac.uk/~3dligandsite
@TOME (Pons & Labesse, 2009)	http://atome.cbs.cnrs.fr/AT2/meta.html
AutoDock Vina (Trott & Olson, 2010)	http://vina.scripps.edu
BSP-SLIM (Lee & Zhang, 2012)	http://zhanglab.ccmb.med.umich.edu/BSP-SLIM
FiberDock—Flexible induced-fit backbone refinement in molecular docking (Mashiach, Nussinov, & Wolfson, 2009)	http://bioinfo3d.cs.tau.ac.il/FiberDock
GEMDOCK—Generic evolutionary method for molecular docking (Yang & Chen, 2004)	http://gemdock.life.nctu.edu.tw/dock

Table 10.1 List of computational methods employed in docking analysis—cont'd

Database name	Resource
Hex (Ghoorah, Devignes, Smaïl-Tabbone, & Ritchie, 2013)	http://hex.loria.fr
idTarget (Wang et al., 2012)	http://idtarget.rcas.sinica.edu.tw
iGEMDOCK (Yang & Chen, 2004)	http://gemdock.life.nctu.edu.tw/dock/igemdock.php
iScreen (Tsai, Chang, & Chen, 2011)	http://iscreen.cmu.edu.tw
ParDOCK (Gupta, Gandhimathi, Sharma, & Jayaram, 2007)	http://www.scfbio-iitd.res.in/dock/pardock.jsp
Surflex-Dock (Jain, 2003)	http://www.tripos.com/index.php?family=modules,SimplePage,&page=Surflex_Dock
AuPosSOM (Mantsyzov, Bouvier, Evrard-Todeschi, & Bertho, 2012)	https://www.biomedicale.univ-paris5.fr/aupossom
BetaDock (Kim, Kim, et al., 2011)	http://voronoi.hanyang.ac.kr/software.htm
DOCK Blaster (Irwin et al., 2009)	http://blaster.docking.org
eHiTS—Electronic high-throughput screening (Zsoldos, Reid, Simon, Sadjad, & Johnson, 2006)	http://www.simbiosys.ca/ehits
FITTED—Flexibility induced through targeted evolutionary description (De Cesco et al., 2012)	http://fitted.ca/index.php?option=com_content&task=view&id=50&Itemid=40
Fleksy (Nabuurs, Wagener, & de Vlieg, 2007)	http://www.cmbi.ru.nl/software/fleksy
FlexX (Sousa, Fernandes, & Ramos, 2006)	http://www.biosolveit.de/flexx
FLIPDock (Zhao & Sanner, 2007)	http://flipdock.scripps.edu/what-is-flipdock
FRED—Fast exhaustive docking (McConkey, Sobolev, & Edelman, 2002)	http://www.eyesopen.com/docs/oedocking/current/html/fred.html
GlamDock (Tietze & Apostolakis, 2007)	http://www.chil2.de/Glamdock.html

Continued

Table 10.1 List of computational methods employed in docking analysis—cont'd

Database name	Resource
GOLD (Verdonk, Cole, Hartshorn, Murray, & Taylor, 2003)	http://www.ccdc.cam.ac.uk/products/life_sciences/gold
GPCRautomodel (Launay et al., 2012)	http://genome.jouy.inra.fr/GPCRautomdl/cgi-bin/welcome.pl
GRAMM-X (Tovchigrechko & Vakser, 2006)	http://vakser.bioinformatics.ku.edu/resources/gramm/grammx
HADDOCK—High ambiguity-driven biomolecular docking (Dominguez, Boelens, & Bonvin, 2003)	http://www.nmr.chem.uu.nl/haddock
HomDock (Marialke, Tietze, & Apostolakis, 2008)	http://www.chil2.de/HomDock.html
ICM-Docking (Fernandez-Recio, Totrov, & Abagyan, 2003)	http://www.molsoft.com/docking.html
kinDOCK—A ligand transposition server (Martin, Catherinot, & Labesse, 2006)	http://abcis.cbs.cnrs.fr/LIGBASE_SERV_WEB/PHP/kindock.php
Lead Finder (Novikov et al., 2012)	http://www.moltech.ru
MVD—Molegro Virtual Docker (Thomsen & Christensen, 2006)	http://www.molegro.com/mvd-product.php
ParaDocks—Parallel Docking Suite (Muegge, 2006)	http://www.paradocks.org
PLANTS—Protein–Ligand ANT System (Korb, Stützle, & Exner, 2006)	http://www.tcd.uni-konstanz.de/research/plants.php
Rosetta FlexPepDock—High-resolution modeling of peptide–protein interactions (London, Raveh, Cohen, Fathi, & Schueler-Furman, 2011)	http://flexpepdock.furmanlab.cs.huji.ac.il/index.php
RosettaLigand (Hirst, Alexander, McHaourab, & Meiler, 2011)	http://www.rosettacommons.org/software
SwissDock (Grosdidier, Zoete, & Michielin, 2011)	http://www.swissdock.ch/docking
SymmDock—Prediction of complexes with Cn symmetry by geometry-based docking (Schneidman-Duhovny et al., 2005b)	http://bioinfo3d.cs.tau.ac.il/SymmDock

Table 10.1 List of computational methods employed in docking analysis—cont'd

Database name	Resource
TarFisDock—Target fishing dock (Li et al., 2006)	http://www.dddc.ac.cn/tarfisdock
VEGA ZZ (Pedretti, Villa, & Vistoli, 2004)	http://www.vegazz.net
VLifeDock	http://www.vlifesciences.com/products/VLifeMDS/VLifeDock.php
Visualization tools	
BALL View (Moll, Hildebrandt, Lenhof, & Kohlbacher, 2006)	http://www.ball-project.org/
Visual Molecular Dynamics (Humphrey, Dalke, & Schulten, 1996)	http://www.ks.uiuc.edu/Research/vmd/
UCSF Chimera (Pettersen et al., 2004)	http://www.cgl.ucsf.edu/chimera/
PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC)	http://pymol.org/
RasMol (Sayle & White, 1995)	http://www.rasmol.org/

Waals force, binding energy, the number of hydrogen bonds formed between the docked receptor and ligand, and other types of intermolecular interactions in the final docked molecule. Scoring only decides the best conformations for further research or for the purpose of creating a new drug (Alonso et al., 2006). Currently, scoring functions are utilized by docking softwares in one of two ways. In the first approach, the docked molecule, for example, the ligand–protein conformations, is ranked completely by the scoring functions (Huang, Grinter, & Zou, 2010). The search algorithms then modify this arrangement a little. These rearranged conformations are again ranked by the scoring functions. The second approach uses a two-stage scoring function. First, the reduced function is used to direct the search, and then more a vigorous function is used to rank the resulting structures. There are drawbacks with this directed approach. These directed methods make assumptions about the energy hyper surface, often omitting computationally expensive terms such as electrostatics, and considering only a few types of interactions such as hydrogen bonds. Such algorithms are, therefore, directed to areas of importance as determined by the reduced scoring

function (Taylor, Jewsbury, & Essex, 2002). There are many algorithms used for docking, which include search algorithm, molecular dynamics, Monte Carlo, genetic algorithms, fragment-based methods, point complementary methods, distance geometry methods, and systemic searches. It is important to note that many docking technologies incorporate multiple or blended approaches into their techniques. In rigid-body docking, ligand and target protein are rigid, whereas in flexible docking, the flexibility of the ligand and/or the target protein (receptor) is considered. Rigid-body docking is fast but does not consider induced fit and is cheaper in comparison to flexible docking. Major types of scoring functions are listed in the succeeding text.

7.1. Force-field scoring functions

The force-field scoring function decides the binding energy by calculating the sum on nonbonding interactions such as van der Waals, electrostatics, bond stretching/bending/torsional forces, and entropy contributions (Taylor et al., 2002). The electrostatic force is calculated by Coulombic formulation; electrostatic potential energy is represented as a pair-wise summation of Coulombic interactions (Kitchen, Decornez, Furr, & Bajorath, 2004), and van der Waals terms are described by a Lennard-Jones potential function. Changing the potential increases or decreases the acceptance threshold of the score between the protein and the ligand. Changing the potential also determines the proximity of the ligand to the protein (Meng, Zhang, Mezei, & Cui, 2011).

7.2. Empirical free-energy scoring functions

In an empirical-based scoring function, the elements such as hydrogen bond, ionic interaction, the hydrophobic effect, and binding entropy are taken individually. These components are then individually multiplied by a coefficient to give an individual score. All the individual scores are then summed up to get a final score (Campbell, Gold, Jackson, & Westhead, 2003).

7.3. Knowledge-based scoring functions

Knowledge-based scoring functions use statistical analysis of the ligand–protein interactions. The function is based on the theory that if there is a high interaction between a particular protein and a ligand, then the probability of combining them together or occurring together is very high (Meng et al., 2011).

7.4. Consensus-based scoring functions

Scoring functions are prone to errors. The consensus scoring function is more recent trend. This scoring function combines the scores of all the scoring programs and gives one final score in order to minimize the errors and to find the perfect ligand for a given target macromolecule (Meng et al., 2011). There is a potential limitation to this method; sometimes instead of the error being minimized, there can be amplification in the calculation errors, which can void the balance of this scoring function aims at (Kitchen et al., 2004).

8. TYPES OF DOCKING

Molecular docking is also referred to as small molecular docking. Molecular docking is a study of how two or more molecular structures, for instance, drug and catalyst or macromolecule receptor, match along to be a perfect fit (Gane & Dean, 2000). Binding orientation of small-molecule drug candidates to their macromolecular targets predicts the affinity and activity of a given small molecule (Hakes, Lovell, Oliver, & Robertson, 2007).

Protein–protein docking is a simple procedure, which involves docking of two protein molecules without any need of experimental measurement. Flexible and rigid docking is followed in this type of docking (Ehrlich & Wadey, 2003). Shape complementarity is the most essential ingredient of the scoring functions for protein–protein docking (Chen & Weng, 2003). The steady rise in the number of protein structures elucidated has boosted the number of protein–protein docking studies, and intensive research is being carried out in the field. Many proteins that remain rigid after forming a complex can also be docked (Hakes et al., 2007).

Protein–ligand docking is the most commonly used docking technique. It predicts the position of a ligand when it is bound to its receptor molecule, in this case, a protein. The ligand might act as an inhibitor or a promoter. Large libraries of ligands are scanned to choose potential drug candidates (Smith, Engdahl, Dunbar, & Carlson, 2012).

AutoDock is a molecular docking suite consisting of automated docking tools. AutoDock consists of two main programs: AutoDock and AutoGrid. AutoDock docks the two molecules according to the grid, which is precalculated and set by AutoGrid. AutoDock is considered one of the best programs when it comes to docking and virtual screening (Park, Lee, & Lee, 2006). This section will give a brief overview of the steps followed in AutoDock 4.2 (Fig. 10.2). Various possible problems must be resolved

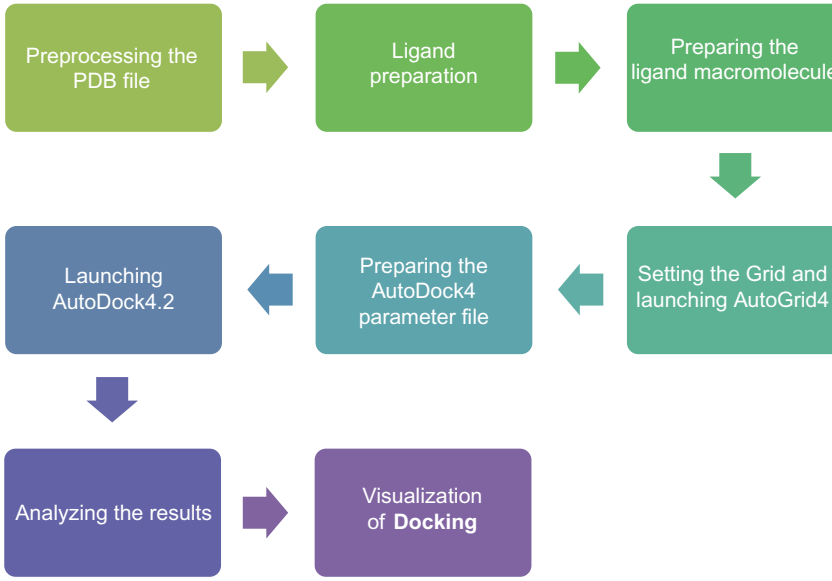

Figure 10.2 Steps in followed in docking analysis.

before a protein can be used for AutoDock. This includes missing atoms, chain breaks, and alternate locations. Potential energy grids are used by different docking programs. These grids represent the energy calculations, and in their most basic form, the grid stashes two types of potentials: the electrostatic and the van der Waals force. The grid was formulated so that the information about the receptor's energy contributions could be stored on grid points. This allowed the necessity of it being read only during ligand scoring. More options can be explored in AutoDock, and the options may vary depending on the complexes that are being docked and also the complexity of the problem in hand.

8.1. Advantages of docking

- The application of docking in a targeted drug-delivery system is a huge benefit. One can study the size, shape, charge distribution, polarity, hydrogen bonding, and hydrophobic interactions of both ligand (drug) and receptor (target site).
- Molecular docking helps in the identification of target sites of the ligand and the receptor molecule.

- Docking also helps in understanding of different enzymes and their mechanism of action.
- The "scoring" feature in docking helps in selecting the best fit or the best drug from an array of options.
- Not everything can be proved experimentally as traditional experimental methods for drug discovery take a long time. Molecular docking helps in moving the process of computer-aided drug designing faster and also provides every conformation possible based on the receptor and ligand molecule.
- Docking has a huge advantage when it comes to the study of protein interactions.
- There are millions of compounds, ligands, drugs, and receptors, the 3D structure of which has been crystallized. Virtual screening of these compounds can be made.

8.2. Limitations of docking

- In protein–small-molecule docking, there can be problems in the receptor structure. A reliable resolution value for small-molecule docking is below 1.2 Å (Gohlke & Klebe, 2002), while most crystallographic structures have a resolution between 1.5 and 2.5 Å. Increasing the use of homology models in docking should be looked at with care as they have even poorer resolution (Mihasan, 2010). Most applications accept and yield good results for structures below 2.2 Å. All the same, care should be taken while picking a structure.
- The scoring functions used in docking, almost all of them, do not take into account the role played by covalently bound inhibitors or ions (Mihasan, 2012).
- The methodology and research in protein–protein docking have to be greatly increased as the success in this field is greatly hampered by many false positives and false negatives (Moreira, Fernandes, & Ramos, 2010).

9. MOLECULAR DYNAMICS

In the late 1950s and the early 1960s, Alder, Wainwright, and Rahman first developed molecular dynamics (MD) to understand the atomic movement of liquids (Maginn & Elliott, 2010). With the advancement of computer science, MD has consequently become a precious and powerful tool in many domains. Since the 1970s, MD is widely applied to study the structure and dynamics of the complexity of chemical components of life, especially

proteins or nucleic acids. Computer-aided drug design has come into more focus due to the development of computer science. Improvements in computer science and algorithm have played a vital role in the development of next-generation tools in computer-aided drug design. Currently, MD is used for simulations of biomolecular systems comprising many thousands of atoms within a very short span of time such as nanoseconds (Plimpton, 1995). Therefore, this method is used to solve a number of biological problems including drug discovery and development (Yang, 2010). However, in this process, more computational resources are involved. Conventional methods for the experimental determination of protein structures have relied on X-ray crystallographic and NMR techniques that have been considered to decide the structure of a protein with high precision. However, dynamic properties of molecules, especially proteins and DNA, are connected with backbone and side-chain moieties. MD enables us to understand the crucial determinants of many aspects of protein actions especially stability, folding, and function (Benkovic & Hammes-Schiffer, 2003; Eisenmesser, Bosco, Akke, & Kern, 2002; Frauenfelder, Sligar, & Wolynes, 1991; Karplus & McCammon, 2002; Rasmussen, Stock, Ringe, & Petsko, 1992; Wong & McCammon, 2003). Such dynamic processes can be understood by a variety of different techniques, particularly detecting NMR through relaxation phenomena (Kay, 1998). During drug-discovery process, crystallographic studies and NMR techniques are very important in understanding the role of protein–ligand binding as well as protein dynamics. It also helps in understanding the flexibility and motion of the proteins in the ligand-binding assay, where emerging computational methods have made the process simpler. During the prediction of protein motions, the calculations of the quantum-mechanical motion are needed, but it is difficult to explain the chemical reactions and motions in large molecular systems. MD helps to describe the chemical reactions and motion of proteins (Carloni, Rothlisberger, & Parrinello, 2002). This method helps to understand the interactions of protein and ligand or protein–drug interactions, which have multifaceted use in drug discovery and development (Utesch, Daminelli, & Mroginski, 2011).

According to quantum mechanics, physical quantities are represented by averages over microscopic conformations of the system, which are distributed in harmony with a particular statistical assembly (Weaire & Aste, 2010). Newtonian dynamics involves the conservation of energy and molecular dynamic trajectories, which can present a number of arrangements distributed according to the microcanonical ensemble. This is why physical

quantity can be calculated through MD with the help of the arithmetic average over instantaneous values of that quantity obtained from the trajectories (Tuckerman, Yarne, Samuelson, Hughes, & Martyna, 2000). In the limit of infinite, simulation time of the true value of the measured thermodynamical properties. *In vitro*, the quality of sampling and accuracy of the interatomic potentials used in simulations are always limited. In fact, the quality of sampling may not be proper, especially for processes of the timescale larger than typical molecular dynamic simulations, and caution should be exerted when drawing conclusions from such computer experiments (Phillips et al., 2005). MD simulations illustrate the physical movement of atoms and molecules as they interact over time. This is accomplished by the potential energy function, also so-called force fields, used in most of the standard molecular simulation programs for biological systems. It takes the form of the summation of different additive terms that correspond to bond distance stretching (E_{bonds}), bond-angle bending (E_{angles}), bond dihedral or torsion angle ($E_{\text{dihedrals}}$), van der Waals potential (E_{vdw}), and electrostatic potential (E_{elect}). The first three terms are considered to be the intramolecular bonding interactions, and each term involves a multitude of atoms connected by chemical bonds. The other two terms represent the nonbonded interactions between atoms. The most common force fields are OPLS-AA (Jorgensen, Maxwell, & Tirado-Rives, 1996), CHARMM (MacKerell et al., 1998), GROMOS (Christen et al., 2005), and AMBER (Wang, Fang, et al., 2004; Wang, Wolf, et al., 2004), which can perform energy calculations, energy minimization, and dynamic calculations. These tools are regularly used in biomolecular simulation and principally vary as per their parameters. However, these tools usually give similar results. It is very interesting that the forces acting on every one of the system atoms are calculated; the site of these atoms is stimulated according to Newton's laws of motion, and therefore, Newtonian dynamics is applicable. The simulation time is then higher, often by only 1 or 2 quadrillionths of a second, and in this way, the procedure is replicated millions of times on average. Since several calculations are necessary, MD simulations are executed on computer clusters or supercomputer systems, which run hundreds of parallel processors at a time to calculate the process (Herschbach, 1987). Some well-accepted simulation software packages are AMBER (Case et al., 2005), CHARMM (MacKerell et al., 1998), and NAMD (Kalé et al., 1999; Phillips et al., 2005), which apply their default force fields. In the current era, pharmacogenomics is a prominent field, and drug discovery is moving in this direction. There are many studies where MD is applied in studying the conformational space accessible to proteins

and protein–ligand interactions and refined experimental or modeled protein structures (Schlick, Collepardo-Guevara, Halvorsen, Jung, & Xiao, 2011). Modeled protein structures (Fan et al., 2009) reveal transient binding sites (Ivetac & McCammon, 2010), examine the stability and strength of docked protein–ligand conformations (B-Rao et al., 2009), aid drug discovery (Salsbury, 2010), and explore altered drug binding profiles of protein variants (Shan et al., 2011). In addition to this, MD simulations were carried out on both the native and mutant proteins to show its flexibility and effects on the protein (George Priya Doss, Nagasundaram, Chakraborty, Chen, & Zhu, 2013; George Priya Doss, Rajith, & Chakraborty, 2013; George Priya Doss, Rajith, Rajasekaran, et al., 2013; John et al., 2013; Miteva, Brugge, Rosing, Nicolaes, & Villoutreix, 2004; Steen, Miteva, Villoutreix, Yamazaki, & Dahlback, 2003; Witham, Takano, Schwartz, & Alexov, 2011; Zhang, Teng, Wang, Schwartz, & Alexov, 2010), stability on protein–protein interactions (George Priya Doss & Nagasundaram, 2013), and protein–ligand (George Priya Doss, Rajith, Chakraborty, Balaji, et al., 2013; Nagasundaram & George Priya Doss, 2013) and protein–DNA interactions (George Priya Doss & Nagasundaram, 2012). There are studies showing agreement between computational and experimental measurements of macromolecular dynamics (Bruschweiler & Showalter, 2007; LaConte, Voelz, Nelson, & Thomas, 2002; Markwick et al., 2010; Peter et al., 2003). Combining this information, MD analysis has the potential to be a vital resource in elucidating the molecular effects of mutations and variable drug responses in the context of target protein structures with genetic variants. Here, molecular dynamics has an advantage over experimental methods. Along with the molecular dynamic power, advances in computer technology and algorithm design will definitely act as a driving force in computer-aided drug design in the development of novel pharmacological drugs.

10. CONCLUDING REMARKS

The current world is fast paced; the number of diseases crippling the human race is on the rise. This has caused a never before seen urgency for results for drugs to combat the various diseases that plague the human race. In such a demanding era, science has to rely on bioinformatics, an emerging field that has helped to cut down on the huge amount of time and resources that are otherwise wasted. Single-nucleotide polymorphisms are the most common single amino acid substitutions found in the human genome.

There exist quite a few traditional methods of SNP analysis. Few examples of such experimental methods include DNA sequencing, capillary electrophoresis, single-strand conformation polymorphism, and restriction fragment length polymorphism. These traditional methods have many limitations. A few of these methods require a large sample of DNA, require expensive equipments, and also face difficulty in handling large sequences and are also not practically easy in the process of complicated drug designing. The completion of the Human Genome Project has changed the each core of bioinformatics. With a plethora of data available at the click of a mouse, and very little time in obtaining results, it is impossible to rely only on traditional experimental methods. Drug designing in the late twentieth century has been on the rise. This called for the integration of computational methods as they help in relaying whether the compound or a ligand of which is being screened as a potential drug can be an asset or liability to a particular research. Quality of information can also be improved with the use of computational methods. Computers can integrate results from both *in vivo* and *in vitro* studies from many different labs. This information database can be used by scientists and researchers from all over the world to come up with better and faster results, and newer technologies. Computational methods now allow scientists to cut down on animal testing, as the data provided are accurate, and further testing can be done with computer modeling. This saves wastage of chemicals and also reduces animal suffering.

Personalized medicine is another upcoming field; the scientific community is looking at now. It is common knowledge that a drug that is effective for one person may not be as effective or may even be harmful to another individual. All this is caused by the tiniest changes in our genomes, which make us different from one another (SNPs). Identifying significant SAPs that produce clinically relevant phenotype is important to provide personalized diagnosis and treatment. Personalized medicine would require sequencing each person's genome, analyzing the SNPs, and then formulating which drug would be best suited for a particular individual. Even though computational methods give a faster access to the solution, they are still only predictions. These methods can only predict which SAPs are deleterious and particular disease phenotypes. These methods cannot provide insight how the disruption happens on a molecular scale. Hence, bridging this gap between computational and clinical methods proves to be a challenging task for personalized and precision medicine. There are also limitations to abovementioned computational methods. All the data produced might not be useful, if the results or observations made during experiments vary.

Most of the methods used in *in silico* still have gaping holes that are yet to be filled. More research is required to improve the quality of results derived from such computational research. Existing SAP prediction methods, homology modeling methods, docking, and molecular dynamics softwares have different algorithms and varying protocols. The scoring algorithms of *in silico* tools have to be looked into, as even now, researchers have to rely on multiple scoring algorithms and manually cross-check to see if the results hold true. Therefore, only computational methods cannot be used to validate a drug. Most of the computational prediction methods available online were benchmarked by the curators with their known datasets and were shown to perform well. There are mounting studies that compare the prediction scores from a set of methods and the results bring to light that no single method can be rated as the best predictor. The results obtained from these indicate that a combination of different methods with sequence and structure information may provide a wider coverage and accurate prediction in the study of SAPs to be either deleterious or neutral. Basically, computational methods used different algorithms to predict the impact of deleterious variants, and therefore, the outcome may differ for each tool. However, the positive predictions overlap in all the computational disease prediction methods, which show that they have a high possibility to behave in a similar fashion. The variation in their prediction scores might be due to the difference in features utilized by the methods or the trained datasets.

Computational research has to be complemented with traditional experimental methods. Computational methods help in narrowing down the possible choices of drugs or target molecules, and the experimental biologist can take the given output and check how a drug or any compound will work in a biological system. It has been more than a decade, since the completion of the Human Genome Project. This has helped us to gain insight into the human genome. For instance, the recent ENCODE project has helped to bring before us the biochemical functions of almost 80% of the human genome. In spite of this, a revolution in the medical field is yet to be seen. Cost and time reduction methods can sequence the human genome points to a near future where DNA sequencing would be routinely used in medical practices. A human genome sequence generates a large amount of data, and additional data will be generated in the case of sequencing carried out for personalized medicines. There is a need for next-generation computers and faster networks that can easily handle and process many terabytes of data. When clinical practices will be amalgamated with computational power in the future, the systems should be able to handle the influx of a large amount

of data from all the patients. The computers must also be able to take on each task based on their nature of urgency. Moreover, prioritizing the disease genes can also be incorporated. All these issues need to be dealt with so that the populace receives the best medical attention. The promise of personalized medicine and precision medicine will be made possible by the computational methods discussed in this chapter and will be a useful resource for researchers looking to widen their research scope.

ACKNOWLEDGMENTS

The authors take this opportunity to thank the management of VIT and Galgotias University for providing the facilities and encouragement to carry out this work.

REFERENCES

1000 Genomes Project Consortium, Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061–1073.

Abagyan, R., Totrov, M., & Kuznetsov, D. (1994). ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *Journal of Computational Chemistry*, *15*, 488–506.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*, 248–249.

Akhavan, S., Miteva, M. A., Villoutreix, B. O., Venisse, L., Peyvandi, F., Mannucci, P. M., et al. (2005). A critical role for Gly25 in the B chain of human thrombin. *Journal of Thrombosis and Haemostasis*, *3*, 139–145.

Alexov, E. (2004). Numerical calculations of the pH of maximal protein stability: The effect of the sequence composition and three-dimensional structure. *European Journal of Biochemistry*, *271*(1), 173–185.

Alonso, H., Bliznyuk, A. A., & Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*, *26*(5), 531–568.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*, 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Amberger, J., Bocchini, C., & Hamosh, A. (2011). A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). *Human Mutation*, *32*, 564–567.

Anand, P., Yeturu, K., & Chandra, N. (2012). PocketAnnotate: Towards site-based function annotation. *Nucleic Acids Research*, *40*, W400–W408.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: The Universal Protein knowledgebase. *Nucleic Acids Research*, *32*, D115–D119.

Apweiler, R., Martin, M. J., O'Donovan, C., Magrane, M., Alam-Faruque, Y., Antunes, R., et al. (2010). The universal protein resource (UniProt) in 2010. *Nucleic Acids Research*, *38*, D142–D148.

Arbiza, L., Duchi, S., Montaner, D., Burguet, J., Pantoja-Uceda, D., Pineda-Lucena, A., et al. (2006). Selective pressures at a codon level predict deleterious mutations in human disease genes. *Journal of Molecular Biology*, *358*, 1390–1404.

Arnold, K., Bordoli, L., Kopp, J., & Schwede, T. (2006). The SWISS-MODEL workspace: A web based environment for protein structure homology modeling. *Bioinformatics*, *22*, 195–201.

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., & Ben-Tal, N. (2010). ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, *38*, W529–W533.

Azam, S. S., & Abbasi, S. W. (2013). Molecular docking studies for the identification of novel melatoninergic inhibitors for acetylserotonin-O-methyltransferase using different docking routines. *Theoretical Biology and Medical Modelling*, *10*(1), 63.

Bao, L., & Cui, Y. (2006). Functional impacts of non-synonymous single nucleotide polymorphisms: Selective constraint and structural environments. *FEBS Letters*, *580*, 1231–1234.

Bao, L., Zhou, M., & Cui, Y. (2005). nsSNPAnalyzer: Identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Research*, *33*, W480–W482.

Barroso, I., Gurnell, M., Crowley, V. E., Agostini, M., Schwabe, J. W., Soos, M. A., et al. (1999). Dominant negative mutations in human PPAR gamma associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*, *402*, 880–883.

Basu, S. N., Kollu, R., & Banerjee-Basu, S. (2009). AutDB: A gene reference resource for autism research. *Nucleic Acids Research*, *37*, D832–D836.

Bates, P. A., Kelley, L. A., MacCallum, R. M., & Sternberg, M. J. E. (2001). Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins*, *45*, 39–46.

Benkovic, S. J., & Hammes-Schiffer, S. (2003). A perspective on enzyme catalysis. *Science*, *301*, 1196–1202.

Berezin, C., Glaserm, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., et al. (2004). ConSeq: The identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, *20*, 1322–1324.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., et al. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, *112*, 535–542.

Betz, S. F. (1993). Disulfide bonds and the stability of globular proteins. *Protein Science*, *2*(10), 1551–1558.

Bikadi, Z., & Hazai, E. (2009). Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *Journal of Cheminformatics*, *1*, 15.

Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., & Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, *4*(6), 1633–1649.

Board, P. G., Pierce, K., & Coggan, M. (1990). Expression of functional coagulation factor XIII in Escherichia coli. *Thrombosis and Haemostasis*, *63*(2), 235–240.

Bolton, E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). PubChem: Integrated platform of small molecules and biological activities. *Annual Reports in Computational Chemistry*, *4*, 217–241.

Bonnardeaux, A., Davies, E., Jeunemaitre, X., Féry, I., Charru, A., Clauser, E., et al. (1994). Angiotensin II type 1 receptor gene polymorphisms in human essential hypertension. *Hypertension*, *24*, 63–69.

Boulling, A., le Marechal, C., Trouve, P., Raguenes, O., Chen, J. M., & Ferec, C. (2007). Functional analysis of pancreatitis associated missense mutations in the pancreatic secretory trypsin inhibitor (SPINK1) gene. *European Journal of Human Genetics*, *15*(9), 936–942.

Bowie, J. U., Lüthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, *253*, 164–170.
B-Rao, C., Subramanian, J., & Sharma, S. D. (2009). Managing protein flexibility in docking and its applications. *Drug Discovery Today*, *14*, 394–400.
Brendel, V., Bucher, P., Nourbakhsh, I., Blais-dell, B. E., & Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *89*, 2002–2006.
Bromberg, Y., & Rost, B. (2007). SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research*, *35*, 3823–3835.
Brozell, S. R., Mukherjee, S., Balius, T. E., Roe, D. R., Case, D. A., & Rizzo, R. C. (2012). Evaluation of DOCK 6 as a pose generation and database enrichment tool. *Journal of Computer-Aided Molecular Design*, *26*(6), 749–773.
Bruschweiler, R., & Showalter, S. A. (2007). Validation of molecular dynamics simulations of biomolecules using NMR spin relaxation as benchmarks: Application to the AMBER99SB force field. *Journal of Chemical Theory and Computation*, *3*, 961–975.
Brylinski, M., & Skolnick, J. (2008). A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 129–134.
caBIG Strategic Planning Workspace, (2007). The Cancer Biomedical Informatics Grid (caBIG): Infrastructure and applications for a worldwide research community. *Studies in Health Technology and Informatics*, *129*, 330–334.
Cai, Z., Tsung, E. F., Marinescu, V. D., Ramoni, M. F., Riva, A., & Kohane, I. S. (2004). Bayesian approach to discovering pathogenic SNPs in conserved protein domains. *Human Mutation*, *24*, 178–184.
Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*, *30*, 1237–1244.
Campbell, S. J., Gold, N. D., Jackson, R. M., & Westhead, D. R. (2003). Ligand binding: Functional site location, similarity and docking. *Current Opinion in Structural Biology*, *13*(3), 389–395.
Capriotti, E., & Altman, R. B. (2011). Improving the prediction of disease-related variants using protein three-dimensional structure. *BMC Bioinformatics*, *12*(4), S3.
Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., & Casadio, R. (2013). WS-SNPs&GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*, *14*, S6.
Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*, *33*, W306–W310.
Capriotti, E., Fariselli, P., Rossi, I., & Casadio, R. (2008). A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics*, *9*(2), S6.
Carloni, P., Rothlisberger, U., & Parrinello, M. (2002). The role and perspective of ab initio molecular dynamics in the study of biological systems. *Accounts of Chemical Research*, *35*(6), 455–464.
Cartegni, L., & Krainer, A. R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. *Nature Genetics*, *30*, 377–384.
Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., et al. (2005). The AMBER biomolecular simulation programs. *Journal of Computational Chemistry*, *26*, 1668–1688.
Castella, M., Pujol, R., Callén, E., Trujillo, J. P., Casado, J. A., Gille, H., et al. (2011). Origin, functional role, and clinical impact of Fanconi anemia FANCA mutations. *Blood*, *117*, 3759–3769.

Castellana, S., & Mazza, T. (2013). Congruency in the prediction of pathogenic missense mutations: State-of-the-art web-based tools. *Briefings in Bioinformatics*, *14*, 448–459.
Chang, H., & Fujita, T. (2001). PicSNP: A browsable catalog of nonsynonymous single nucleotide polymorphisms in the human genome. *Biochemical and Biophysical Research Communications*, *287*, 288–291.
Chang, W. C., Lee, T. Y., Shien, D. M., Hsu, J. B., Horng, J. T., Hsu, P. C., et al. (2009). Incorporating support vector machine for identifying protein tyrosine sulfation sites. *Journal of Computational Chemistry*, *30*(15), 2526–2537.
Chasman, D., & Adams, R. M. (2001). Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: Structure-based assessment of amino acid variation. *Journal of Molecular Biology*, *307*, 683–706.
Chen, X., Ji, Z. L., & Chen, Y. Z. (2002). TTD: Therapeutic target database. *Nucleic Acids Research*, *30*, 412–415.
Chen, C. W., Lin, J., & Chu, Y. W. (2013). iStable: Off-the-shelf predictor integration for predicting protein stability changes. *BMC Bioinformatics*, *14*(2), S5.
Chen, R., & Weng, Z. (2003). A novel shape complementarity scoring function for protein-protein docking. *Proteins: Structure, Function, and Bioinformatics*, *51*(3), 397–408.
Chen, H.-L., & Zhou, H.-X. (2005). Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Research*, *33*, 3193–3199.
Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, *62*(4), 1125–1132.
Cheng, J., Randall, A. Z., Sweredoski, M. J., & Baldi, P. (2005). SCRATCH: A protein structure and structural feature prediction server. *Nucleic Acids Research*, *33*, W72–W76.
Cheng, J., Sweredoski, M., & Baldi, P. (2005). Accurate prediction of protein disordered regions by mining protein structure data. *Data Mining and Knowledge Discovery*, *11*(3), 213–222.
Christen, M., Hünenberger, P. H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D. P., et al. (2005). The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, *26*, 1719–1751.
Claussen, H., Buning, C., Rarey, M., & Lengauer, T. (2001). FlexE: Efficient molecular docking considering protein structure variations. *Journal of Molecular Biology*, *308*, 377–395.
Cole, C., Barber, J. D., & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Research*, *35*(2), W197–W201.
Comeau, S. R., Gatchell, D. W., Vajda, S., & Camacho, C. J. (2004). ClusPro: A fully automated algorithm for protein-protein docking. *Nucleic Acids Research*, *32*(2), W96–W99.
Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X., & Ventura, S. (2007). AGGRESCAN: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics*, *8*, 65.
Conde, L., Vaquerizas, J. M., Santoyo, J., Al-Shahrour, F., Ruiz-Llorente, S., Robledo, M., et al. (2004). PupaSNP Finder: A web tool for finding SNPs with putative effect at transcriptional level. *Nucleic Acids Research*, *32*, W242–W248.
Congreve, M., Murray, C. W., & Blundell, T. L. (2005). Structural biology and drug discovery. *Drug Discovery Today*, *10*, 895–907.
Curtis, D., North, B. V., & Sham, P. C. (2001). Use of an artificial neural network to detect association between a disease and multiple marker genotypes. *Annals of Human Genetics*, *65*(Pt. 1), 95–107.
Daga, P. R., Patel, R. Y., & Doerksen, R. J. (2010). Template-based protein modeling: Recent methodological advances. *Current Topics in Medicinal Chemistry*, *10*, 84–94.
Dalton, J. A., & Jackson, R. M. (2007). An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics*, *23*(15), 1901–1908.

Dambosky, J., Prokop, M., & Koca, J. (2001). TRITON: Graphic software for rational engineering of enzymes. *Trends in Biochemical Sciences*, *26*, 71–73.
Das, R., & Baker, D. (2008). Macromolecular modeling with Rosetta. *Annual Review of Biochemistry*, *77*, 363–382.
Das, R., Qian, B., Raman, S., Vernon, R., Thompson, J., Bradley, P., et al. (2007). Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins*, *69*(8), 118–128.
Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., et al. (2007). MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, *35*, W375–W383.
De Baets, G., Van Durme, J., Reumers, J., Maurer-Stroh, S., Vanhee, P., Dopazo, J., et al. (2012). SNPeffect 4.0: On-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Research*, *40*, D935–D939.
De Cesco, S., Deslandes, S., Therrien, E., Levan, D., Cueto, M., Schmidt, R., et al. (2012). Virtual screening and computational optimization for the discovery of covalent prolyl oligopeptidase inhibitors with activity in human cells. *Journal of Medicinal Chemistry*, *55*, 6306–6315.
Deane, C. M., & Blundell, T. L. (2001). CODA: A combined algorithm for predicting the structurally variable regions of protein models. *Protein Science*, *10*, 599–612.
Dehouck, Y., Grosfils, A., Folch, B., Gilis, D., Bogaerts, P., & Rooman, M. (2009). Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. *Bioinformatics*, *25*, 2537–2543.
Dessailly, B. H., Lensink, M. F., Orengo, C. A., & Wodak, S. J. (2008). LigASite—A database of biologically relevant binding sites in proteins with known apo-structures. *Nucleic Acids Research*, *36*(1), D667–D673.
Dill, K. A., Ozkan, S. B., Weikl, T. R., Chodera, J. D., & Voelz, V. A. (2007). The protein folding problem: When will it be solved? *Current Opinion in Structural Biology*, *17*(3), 342–346.
DiMasi, J. A., & Grabowski, H. G. (2007). The cost of biopharmaceutical R&D: Is biotech different? *Managerial and Decision Economics*, *28*, 285–291.
DiMasi, J. A., Grabowski, H. G., & Vernon, J. (2004). R&D costs and returns by therapeutic category. *Drug Information Journal*, *38*, 211–223.
DiMasi, J. A., Hansen, R. W., Grabowski, H. G., & Lasagna, L. (1991). Cost of innovation in the pharmaceutical industry. *Journal of Health Economics*, *10*, 107–142.
Dimmic, M. W., Sunyaev, S., & Bustamante, C. D. (2005). Inferring SNP function using evolutionary, structural, and computational methods. *Pacific Symposium on Biocomputing*, *10*, 382–384.
Dixit, A., Torkamani, A., Schork, N. J., & Verkhivker, G. (2009). Computational modeling of structurally conserved cancer mutations in the RET and MET kinases: The impact on protein structure, dynamics, and stability. *Biophysical Journal*, *96*(3), 858–874.
Dobson, C. M. (2003). Protein folding and misfolding. *Nature*, *426*(6968), 884–890.
Dobson, R. J., Munroe, P. B., Caulfield, M. J., & Saqi, M. A. S. (2006). Predicting deleterious nsSNPs: An analysis of sequence and structural attributes. *BMC Bioinformatics*, *7*(1), 217.
Dominguez, C., Boelens, R., & Bonvin, A. M. (2003). HADDOCK: A protein-protein docking approach based on biochemical and/or biophysical information. *Journal of the American Chemical Society*, *125*, 1731–1737.
Doolittle, R. F. (1986). *Of Urfs and Orfs: A Primer on how to Analyze Derived Amino Acid*. (University Science Books) ISBN 0-935702-54-7.
Dosztányi, Z., CsizmLok, V., Tompa, P., & Simon, I. (2005). IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, *21*(16), 3433–3434.

Dosztanyi, Z., Magyar, C., Tusnady, G. E., & Simon, I. (2003). SCide: Identification of stabilization centers in proteins. *Bioinformatics*, *19*(7), 899–900.
Dryja, T. P., McGee, T. L., Hahn, L. B., Cowley, G. S., Olsson, J. E., Reichel, E., et al. (1990). Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. *The New England Journal of Medicine*, *323*(19), 1302–1307.
Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., & Liang, J. (2006). CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*, *34*, W116–W118.
Durrant, J., & McCammon, J. A. (2011). Molecular dynamics simulations and drug discovery. *BMC Biology*, *9*(1), 71.
Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.
Ehrlich, L. P., & Wadey, R. C. (2003). Protein–protein docking. *Reviews in Computational Chemistry*, *17*, 61.
Eisenmesser, E. Z., Bosco, D. A., Akke, M., & Kern, D. (2002). Enzyme dynamics during catalysis. *Science*, *295*, 1520–1523.
Elles, L. M. S., & Uhlenbeck, O. C. (2008). Mutation of the arginine finger in the active site of Escherichia coli DbpA abolishes ATPase and helicase activity and confers a dominant slow growth phenotype. *Nucleic Acids Research*, *36*(1), 41–50.
ENCODE Project Consortium, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., & Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74.
Erdin, S., Ward, R. M., Venner, E., & Lichtarge, O. (2010). Evolutionary trace annotation of protein function in the structural proteome. *Journal of Molecular Biology*, *396*(5), 1451–1473.
Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P., et al. (1992). Response of a protein structure to cavity creating mutations and its relation to the hydrophobic effect. *Science*, *255*, 178–183.
Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., et al. (2006). Comparative protein structure modeling using MODELLER. *Current Protocols in Bioinformatics,* Chapter 5, Unit 5.6.
Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design*, *15*, 411–428.
Fan, H., Irwin, J. J., Webb, B. M., Klebe, G., Shoichet, B. K., & Sali, A. (2009). Molecular docking screens using comparative models of proteins. *Journal of Chemical Information and Modeling*, *49*, 2512–2527.
Fernald, G. H., Capriotti, E., Daneshjou, R., Karczewski, K. J., & Altman, R. B. (2011). Bioinformatics challenges for personalized medicine. *Bioinformatics*, *27*, 1741–1748.
Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nature Biotechnology*, *22*, 1302–1306.
Fernandez-Fuentes, N., Rai, B. K., Madrid-Aliste, C. J., Fajardo, J. E., & Fiser, A. (2007). Comparative protein structure modeling by combining multiple templates and optimizing sequence-to-structure alignments. *Bioinformatics*, *23*, 2558–2565.
Fernandez-Fuentes, N., Zhai, J., & Fiser, A. (2006). ArchPRED: A template based loop structure prediction server. *Nucleic Acids Research*, *34*, W173–W176.
Fernandez-Recio, J., Totrov, M., & Abagyan, R. (2003). ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins*, *52*, 113–117.
Ferrer-Costa, C., Gelpi, J. L., Zamakola, L., Parraga, I., de la Cruz, X., & Orozco, M. (2005). PMUT: A web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*, *21*, 3176–3178.

Ferrer-Costa, C., Orozco, M., & de la Cruz, X. (2002). Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. *Journal of Molecular Biology*, *315*, 771–786.

Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). The energy landscapes and motions on proteins. *Science*, *254*, 1598–1603.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*, 851–861.

Frazer, K. A., Murray, S. S., Schork, N. J., & Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews. Genetics*, *10*, 241–251.

Fredman, D., Siegfried, M., Yuan, Y. P., Bork, P., Lehväslaiho, H., & Brookes, A. J. (2002). HGVbase: A human sequence variation database emphasizing data quality and a broad spectrum of data sources. *Nucleic Acids Research*, *30*(1), 387–391.

Freimuth, R. R., Stormo, G. D., & McLeod, H. L. (2005). PolyMAPr: Programs for polymorphism database mining, annotation, and functional analysis. *Human Mutation*, *25*, 110–117.

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, *47*, 1739–1749.

Frousios, K., Iliopoulos, C. S., Schlitt, T., & Simpson, M. A. (2013). Predicting the functional consequences of non-synonymous DNA sequence variants—Evaluation of bioinformatics tools and development of a consensus strategy. *Genomics*, *102*(4), 223–228.

Fujiwara, H., Tatsumi, K. I., Tanaka, S., Kimura, M., Nose, O., & Amino, N. (2000). A novel V59E missense mutation in the sodium iodide symporter gene in a family with iodide transport defect. *Thyroid*, *10*(6), 471–474.

Galzitskaya, O. V., Garbuzynskiy, S. O., & Lobanov, M. Y. (2006). FoldUnfold: Web server for the prediction of disordered regions in protein chain. *Bioinformatics*, *22*(23), 2948–2949.

Gane, P. J., & Dean, P. M. (2000). Recent advances in structure-based rational drug design. *Current Opinion in Structural Biology*, *10*, 401–404.

Gao, Z., Li, H., Zhang, H., Liu, X., Kang, L., Luo, X., et al. (2008). PDTD: A web-accessible protein database for drug target identification. *BMC Bioinformatics*, *19*(9), 104.

Garbuzynskiy, S. O., Lobanov, M. Y., & Galzitskaya, O. V. (2010). FoldAmyloid: A method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics*, *26*, 326–332.

George Priya Doss, C., Chakraborty, C., Rajith, B., & Nagasundaram, N. (2013). In silico discrimination of nsSNPs in hTERT gene by means of local DNA sequence context and regularity. *Journal of Molecular Modeling*, *19*(9), 3517–3527.

George Priya Doss, C., & Nagasundaram, N. (2012). Investigating the structural impacts of I64T and P311S mutations in APE1-DNA complex: A molecular dynamics approach. *PLoS One*, 7(2), e31677.

George Priya Doss, C., & Nagasundaram, N. (2013). Molecular docking and molecular dynamics study on the effect of ERCC1 deleterious polymorphisms in ERCC1-XPF heterodimer. *Applied Biochemistry and Biotechnology,* http://dx.doi.org/10.1007/s12010-013-0592-5. [Epub ahead of print].

George Priya Doss, C., Nagasundaram, N., Chakraborty, C., Chen, L., & Zhu, H. (2013). Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: A molecular dynamics approach. *Human Genomics*, 7, 10.

George Priya Doss, C., Nagasundaram, N., Srajan, J., & Chiranjib, C. (2012). LSHGD: A database for human leprosy susceptible genes. *Genomics*, *100*(3), 162–166.

George Priya Doss, C., Rajasekaran, R., Arjun, P., & Sethumadhavan, R. (2010). Prioritization of candidate SNPs in colon cancer using bioinformatics tools: An alternative approach for a cancer biologist. *Interdisciplinary Sciences*, *2*(4), 320–346.

George Priya Doss, C., Rajith, B., & Chakraborty, C. (2013). Predicting the impact of deleterious mutations in the protein kinase domain of FGFR2 in the context of function, structure, and pathogenesis—A bioinformatics approach. *Applied Biochemistry and Biotechnology*, *170*(8), 1853–1870.

George Priya Doss, C., Rajith, B., Chakraborty, C., Balaji, V., Magesh, R., Menon, S., et al. (2013). In silico profiling and structural insights of missense mutations in RET protein kinase domain by molecular dynamics and docking approach. *Molecular BioSystems*, http://dx.doi.org/10.1039/C3MB70427K.

George Priya Doss, C., Rajith, B., Rajasekaran, R., Srajan, J., Nagasundaram, N., & Debajyoti, C. (2013). In silico analysis of prion protein mutants: A comparative study by molecular dynamics approach. *Cell Biochemistry and Biophysics*, *67*(3), 1307–1318.

George Priya Doss, C., & Sethumadhavan, R. (2009a). Investigation on the role of nsSNPs in HNPCC genes—A bioinformatics approach. *Journal of Biomedical Science*, *24*(16), 42.

George Priya Doss, C., & Sethumadhavan, R. (2009b). Impact of single nucleotide polymorphisms in HBB gene causing haemoglobinopathies: In silico analysis. *New Biotechnology*, *25*(4), 214–219.

George Priya Doss, C., Sudandiradoss, C., Rajasekaran, R., Choudhury, P., Sinha, P., Hota, P., et al. (2008). Applications of computational algorithm tools to identify functional SNPs. *Functional & Integrative Genomics*, *8*(4), 309–316.

Ghersi, D., & Sanchez, R. (2011). Beyond structural genomics: Computational approaches for the identification of ligand binding sites in protein structures. *Journal of Structural and Functional Genomics*, *12*, 109–117.

Ghoorah, A. W., Devignes, M. D., Smaïl-Tabbone, M., & Ritchie, D. W. (2013). Protein docking using case-based reasoning. *Proteins,* http://dx.doi.org/10.1002/prot.24433.

Giardine, B., Riemer, C., Hefferon, T., Thomas, D., Hsu, F., Zielenski, J., et al. (2007). PhenCode: Connecting ENCODE data with mutations and phenotype. *Human Mutation*, *28*, 554–562.

Gnad, F., Baucom, A., Mukhyala, K., Manning, G., & Zhang, Z. (2013). Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics*, *14*(3), S7.

Goede, A., Dunkel, M., Mester, N., Frommel, C., & Preissner, R. (2005). SuperDrug: A conformational drug database. *Bioinformatics*, *21*(9), 1751–1753.

Gohlke, H., & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie (International Ed. in English)*, *41*, 2644–2676.

Gold, A. K. (2007). Cyber infrastructure, data, and libraries, part 2: Libraries and the data challenge: Roles and actions for libraries. *Office of the Dean (Library)*, *17*.

Goodsell, D. S., Morris, G. M., & Olson, A. J. (1996). Automated docking of flexible ligands: Applications of AutoDock. *Journal of Molecular Recognition*, *9*, 1–5.

Grasbon-Frodl, E., Lorenz, H., Mann, U., Nitsch, R. M., Windl, O., & Kretzschmar, H. A. (2004). Loss of glycosylation associated with the T183A mutation in human prion disease. *Acta Neuropathologica*, *108*, 476–484.

Gray, V. E., Kukurba, K. R., & Kumar, S. (2012). Performance of computational tools in evaluating the functional impact of laboratory-induced amino acid mutations. *Bioinformatics*, *28*(16), 2093–2096.

Griffith, M., Griffith, O. L., Coffman, A. C., Weible, J. V., McMichael, J. F., Spies, N. C., et al. (2013). DGIdb: Mining the druggable genome. *Nature Methods*, *10*(12), 1209–1210.

Gromiha, M. M., Oobatake, M., Kono, H., Uedaira, H., & Sarai, A. (1999). Role of structural and sequence information in the prediction of protein stability changes: Comparison between buried and partially buried mutations. *Protein Engineering*, *12*, 549–555.

Gromiha, M. M., & Ponnuswamy, P. K. (1993). Prediction of transmembrane beta-strands from hydrophobic characteristics of proteins. *International Journal of Peptide and Protein Research*, *42*(5), 420–431.

Grosdidier, A., Zoete, V., & Michielin, O. (2011). SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Research*, *39*, W270–W277.

Gu, X., & Vander Velden, K. (2002). DIVERGE: Phylogeny-based analysis for functional-structural divergence of a protein family. *Bioinformatics*, *18*, 500–501.

Gulati, S., Cheng, T. M. K., & Bates, P. A. (2013). Cancer networks and beyond: Interpreting mutations using the human interactome and protein structure. *Seminars in Cancer cell Biology*, *23*(4), 219–226.

Gupta, A., Gandhimathi, A., Sharma, P., & Jayaram, B. (2007). ParDOCK: An all atom energy based Monte Carlo docking protocol for protein-ligand complexes. *Protein and Peptide Letters*, *14*(7), 632–646.

Gupta, R., Jung, E., Gooley, A. A., Williams, K. L., Brunak, S., & Hansen, J. (1999). Scanning the available Dictyostelium discoideum proteome for O-linked GlcNAc glycosylation sites using neural networks. *Glycobiology*, *9*(10), 1009–1022.

Hakes, L., Lovell, S. C., Oliver, S. G., & Robertson, D. L. (2007). Specificity in protein interactions and its relationship with sequence diversity and coevolution. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7999–8004.

Han, A., Kang, H. J., Cho, Y., Lee, S., Kim, Y. J., & Gong, S. (2006). SNP@Domain: A web resource of single nucleotide polymorphisms (SNPs) within protein domain structures and sequences. *Nucleic Acids Research*, *34*, W642–W644.

Hanemann, C. O., D'Urso, D., Gabreëls-Festen, A. A., & Müller, H. W. (2000). Mutation-dependent alteration in cellular distribution of peripheral myelin protein 22 in nerve biopsies from Charcot-Marie-Tooth type 1A. *Brain*, *123*(5), 1001–1006.

Hardt, M., & Laine, R. A. (2004). Mutation of active site residues in the chitin-binding domain ChBDChiA1 from chitinase A1 of Bacillus circulans alters substrate specificity: Use of a green fluorescent protein binding assay. *Archives of Biochemistry and Biophysics*, *426*(2), 286–297.

Hassett, C., Aicher, L., Sidhu, J. S., & Omiecinski, C. J. (1994). Human microsomal epoxide hydrolase: Genetic polymorphism and functional expression in vitro of amino acid variants. *Human Molecular Genetics*, *3*, 421–428.

Heinig, M., & Frishman, D. (2004). STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Research*, *32*, W500–W502.

Hendlich, M., Rippmann, F., & Barnickel, G. (1997). LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *Journal of Molecular Graphics & Modelling*, *15*, 359–389.

Hermann, J. C., Marti-Arbona, R., Fedorov, A. A., Fedorov, E., Almo, S. C., Shoichet, B. K., et al. (2007). Structure-based activity prediction for an enzyme of unknown function. *Nature*, *448*, 775–779.

Herschbach, D. R. (1987). Molecular dynamics of elementary chemical reactions (Nobel lecture). *Angewandte Chemie (International Ed. in English)*, *26*(12), 1221–1243.

Hirst, S. J., Alexander, N., McHaourab, H. S., & Meiler, J. (2011). RosettaEPR: An integrated tool for protein structure determination from sparse EPR data. *Journal of Structural Biology*, *173*, 506–514.

Holm, L., & Sander, C. (1992). Fast and simple Monte Carlo algorithm for side-chain optimization in proteins: Application to model. *Proteins: Structure, Function, and Genetics*, *14*, 213–223.

Hooft, R. W., Vriend, G., Sander, C., & Abola, E. E. (1996). Errors in protein structures. *Nature*, *381*, 272.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., & Fersht, A. R. (1990). Strength and cooperativity of contributions of surface salt bridges to protein stability. *Journal of Molecular Biology*, *216*, 103–144.

Hu, L., Benson, M. L., Smith, R. D., Lerner, M. G., & Carlson, H. A. (2005). Binding MOAD (Mother Of All Databases). *Proteins*, *60*, 333–340.

Huang, S. Y., Grinter, S. Z., & Zou, X. (2010). Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and future directions. *Physical Chemistry Chemical Physics*, *12*(40), 12899–12908.

Huang, H. D., Lee, T. Y., Tseng, S. W., & Horng, J. T. (2005). KinasePhos: A web tool for identifying protein kinase-specific phosphorylation sites. *Nucleic Acids Research*, *33*, W226–W229.

Huang, B., & Schroeder, M. (2006). LIGSITEcsc: Predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Structural Biology*, *6*, 19.

Hughes, J. D., Blagg, J., Price, D. A., Bailey, S., Decrescenzo, G. A., Devraj, R. V., et al. (2008). Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorganic & Medicinal Chemistry Letters*, *18*(17), 4872–4875.

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD—Visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38.

Hwang, J., & Liao, W. (1995). Side-chain by neural networks and simulated annealing optimization. *Protein Engineering*, *8*(4), 363–370.

Ingelman-Sundberg, M., Sim, S. C., Gomez, A., & Rodriguez-Antona, C. (2007). Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacology & Therapeutics*, *116*, 496–526.

Ingman, M., & Gyllensten, U. (2006). mtDB: Human Mitochondrial Genome Database, a resource for population genetics and medical sciences. *Nucleic Acids Research*, *34*, D749–D751.

International HapMap Consortium. (2003). The International HapMap Project. *Nature*, *426*, 789–796.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*, 931–945.

Irwin, J. J., Shoichet, B. K., Mysinger, M. M., Huang, N., Colizzi, F., Wassam, P., et al. (2009). Automated docking screens: A feasibility study. *Journal of Medicinal Chemistry*, *52*, 5712–5720.

Ishida, T., & Kinoshita, K. (2007). PrDOS: Prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research*, *35*, W460–W464.

Ishida, T., & Kinoshita, K. (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics*, *24*(11), 1344–1348.

Ivanisenko, V. A., Grigorovich, D. A., & Kolchanov, N. A. (2000). PDBSite: A database on biologically active sites and their spatial surroundings in proteins with known tertiary structure. In *The Second International Conference on Bioinformatics of Genome Regulation and Structure (BGRS'2000), Novosibirsk, Russia, 2* (pp. 171–174). .

Ivetac, A., & McCammon, J. A. (2010). Mapping the druggable allosteric space of G-protein coupled receptors: A fragment-based molecular dynamics approach. *Chemical Biology & Drug Design*, *76*, 201–217.

Jain, A. N. (2003). Surflex:? Fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of Medicinal Chemistry*, *46*(4), 499–511.

Jaruzelska, J., Abadie, V., d'Aubenton-Carafa, Y., Brody, E., Munnich, A., & Marie, J. (1995). In vitro splicing deficiency induced by a C to T mutation at position-3 in the intron 10 acceptor site of the phenylalanine hydroxylase gene in a patient with phenylketonuria. *Journal of Biological Chemistry*, *270*, 20370–20375.

Jayaram, B., Bhushan, K., Shenoy, S. R., Narang, P., Bose, S., Agrawal, P., et al. (2006). Bhageerath: An energy based web enabled computer software suite for limiting the

search space of tertiary structures of small globular proteins. *Nucleic Acids Research*, *34*(21), 6195–6204.
Jegga, A. G., Gowrisankar, S., Chen, J., & Aronow, B. J. (2007). PolyDoms: A whole genome database for the identification of nonsynonymous coding SNPs with the potential to impact disease. *Nucleic Acids Research*, *35*, D700–D706.
John, A. M., C, G. P., Ebenazer, A., Seshadri, M. S., Nair, A., Rajaratnam, S., et al. (2013). P. Arg82Leu von Hippel-Lindau (VHL) gene mutation among three members of a family with familial bilateral pheochromocytoma in India: Molecular analysis and in silico characterization. *PLoS One*, *8*(4), e61908.
Jones, R., Ruas, M., Gregory, F., Moulin, S., Delia, D., Manoukian, S., et al. (2007). CDKN2A mutation in familial melanoma that abrogates binding of p16INK4a to CDK4 but not CDK6. *Cancer Research*, *67*, 9134–9141.
Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology*, *267*, 727–748.
Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science*, *303*(5665), 1813–1818.
Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, *118*(11), 225–236.
Jothi, A. (2012). Principles, challenges and advances in ab initio protein structure prediction. *Protein and Peptide Letters*, *19*(11), 1194–1204.
Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637.
Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., et al. (1999). NAMD2: Greater scalability for parallel molecular dynamics. *Journal of Computational Physics*, *151*, 283–312.
Karchin, R., Diekhans, M., Kelly, L., Thomas, D. J., Pieper, U., Eswar, N., et al. (2005). LS-SNP: Large-scale annotation of coding non-synonymous SNPs based on multiple information sources. *Bioinformatics*, *21*, 2814–2820.
Karchin, R., Kelly, L., & Sali, A. (2005). Improving functional annotation of non-synonymous SNPs with information theory. *Pacific Symposium on Biocomputing*, *10*, 397–408.
Karplus, M., & Kuriyan, J. (2005). Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(19), 6679–6685.
Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural and Molecular Biology*, *9*, 646–652.
Kawabata, T., Ota, M., & Nishikawa, K. (1999). The protein mutant database. *Nucleic Acids Research*, *27*, 355–357.
Kay, L. E. (1998). Protein dynamics from NMR. *Nature Structural and Molecular Biology*, *5*, 513–517.
Keage, H. A., Carare, R. O., Friedland, R. P., Ince, P. G., Love, S., Nicoll, J. A., et al. (2009). Population studies of sporadic cerebral amyloid angiopathy and dementia: A systematic review. *BMC Neurology*, *9*, 3.
Kelley, L. A., & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: A case study using the Phyre server. *Nature Protocols*, *4*, 363–371.
Khan, S., & Vihinen, M. (2007). Spectrum of disease-causing mutations in protein secondary structures. *BMC Structural Biology*, 7(1), 1–18.
Khan, S., & Vihinen, M. (2010). Performance of protein stability predictors. *Human Mutation*, *31*(6), 675–684.
Kiemer, L., Bendtsen, J. D., & Blom, N. (2005). NetAcet: Prediction of N-terminal acetylation sites. *Bioinformatics*, *21*(7), 1269–1270.

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*, *32*, W526–W531.
Kim, E., Hyrc, K. L., Speck, J., Salles, F. T., Lundberg, Y. W., Goldberg, M. P., et al. (2011). Missense mutations in Otopetrin 1 affect subcellular localization and inhibition of purinergic signaling in vestibular supporting cells. *Molecular and Cellular Neurosciences*, *41*, 655–661.
Kim, D. S., Kim, C. M., Won, C. I., Kim, J. K., Ryu, J., Cho, Y., et al. (2011). BetaDock: Shape-priority docking method based on beta-complex. *Journal of Biomolecular Structure & Dynamics*, *29*(1), 219–242.
Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 111–120.
Kinch, L., Yong Shi, S., Cong, Q., Cheng, H., Liao, Y., & Grishin, N. V. (2011). CASP9 assessment of free modeling target predictions. *Proteins*, *79*(10), 59–73.
Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, *3*(11), 935–949.
Kleina, L. G., & Miller, J. H. (1990). Genetic studies of the lac repressor. XIII. Extensive amino acid replacements generated by the use of natural and synthetic nonsense suppressors. *Journal of Molecular Biology*, *212*, 295–318.
Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., et al. (2011). DrugBank 3.0: A comprehensive resource for 'Omics' research on drugs. *Nucleic Acids Research*, *39*, D1035–D1041.
Koehl, P., & Delarue, M. (1994). Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *Journal of Molecular Biology*, *239*(2), 249–275.
Koehl, P., & Delarue, M. (1995). A self consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modeling. *Nature Structural Biology*, *2*, 163–170.
Kolb, H. C., & Sharpless, K. B. (2003). The growing impact of click chemistry on drug discovery. *Drug Discovery Today*, *8*(24), 1128–1137.
Kolinski, A., Rotkiewicz, P., Ilkowski, B., & Skolnick, J. (1999). A method for the improvement of threading-based protein models. *Proteins*, *37*(4), 592–610.
Kollman, P. (1993). Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, *93*(7), 2395–2417.
Korb, O., Stützle, T., & Exner, T. E. (2006). PLANTS: Application of ant colony optimization to structure-based drug design. *Lecture Notes in Computer Science*, *4150*, 247–258.
Kosinski, J., Hinrichsen, I., Bujnicki, J. M., Friedhoff, P., & Plotz, G. (2010). Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. *Human Mutation*, *31*, 975–982.
Kosinski, J., Tkaczuk, K. L., Kasprzak, J. M., & Bujnicki, J. M. (2008). Template based prediction of three-dimensional protein structures: Fold recognition and comparative modeling. In J. M. Bujnicki (Ed.), *Prediction of protein structures, functions, and interactions*. Chichester: John Wiley & Sons, Ltd.
Koukouritaki, S. B., Poch, M. T., Henderson, M. C., Siddens, L. K., Krueger, S. K., VanDyke, J. E., et al. (2007). Identification and functional analysis of common human flavin-containing monooxygenase 3 genetic variants. *The Journal of Pharmacology and Experimental Therapeutics*, *320*, 266–273.
Krawczak, M., Ball, E. V., Fenton, I., Stenson, P. D., Abeysinghe, S., Thomas, N., et al. (2000). Human gene mutation database—A biomedical information and research resource. *Human Mutation*, *15*, 45–51.

Krishnan, V. G., & Westhead, D. R. (2003). A comparative study of machine-learning methods to predict the effects of single nucleotide polymorphisms on protein function. *Bioinformatics*, *19*(17), 2199–2209.

Krivov, G., Shapovalov, M., & Dunbrack, L. D., Jr. (2009). Improved prediction of protein side-chain conformations with SCWRL4. *Proteins: Structure, Function, and Bioinformatics*, *77*, 778–795.

Kryshtafovych, A., & Fidelis, K. (2009). Protein structure prediction and model quality assessment. *Drug Discovery Today*, *14*, 386–393.

Kumar, R., Chaudhary, K., Gupta, S., Singh, H., Kumar, S., Gautam, A., et al. (2013). CancerDR: Cancer drug resistance database. *Scientific Reports*, *3*, 1445.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, *4*(7), 1073–1081.

LaConte, L. E. W., Voelz, V. A., Nelson, W. D., & Thomas, D. D. (2002). Molecular dynamics simulation of site-directed spin labeling: Experimental validation in muscle fibers. *Biophysical Journal*, *83*(4), 1854–1866.

Lahti, J. L., Tang, G. W., Capriotti, E., Liu, T., & Altman, R. B. (2012). Bioinformatics and variability in drug response: A protein structural perspective. *Journal of the Royal Society, Interface*, *9*(72), 1409–1437.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948.

Laskowski, R. A. (1995). SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *Journal of Molecular Graphics*, *13*, 323–328.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, *26*, 283–291.

Launay, G., Téletchéa, S., Wade, F., Pajot-Augy, E., Gibrat, J. F., & Sanz, G. (2012). Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Engineering, Design & Selection*, *25*(8), 377–386.

Laurila, K., & Vihinen, M. (2009). Prediction of disease-related mutations affecting protein localization. *BMC Genomics*, *10*, 122.

Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: A critical review. *Current Medicinal Chemistry*, *20*(23), 2839–2860.

Lee, P. H., & Shatkay, H. (2008). F-SNP: Computationally predicted functional SNPs for disease association studies. *Nucleic Acids Research*, *36*, D820–D824.

Lee, C., & Subbiah, S. (1991). Prediction of protein side-chain conformation by packing optimization. *Journal of Molecular Biology*, *213*, 373–388.

Lee, H. S., & Zhang, Y. (2012). BSP-SLIM: A blind low-resolution ligand-protein docking approach using theoretically predicted protein structures. *Proteins*, *80*, 93–110.

Lesk, A. M. (1997). CASP2: Report on ab initio predictions. *Proteins: Structure, Function, and Genetics*, *1*, 151–166.

Levitt, M. (1992). Accurate modeling of protein conformation by automatic segment matching. *Journal of Molecular Biology*, *226*, 507–533.

Levitt, M. (2007). Growth of novel protein structural data. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 3183–3188.

Li, H., Gao, Z., Kang, L., Zhang, H., Yang, K., Yu, K., et al. (2006). TarFisDock: A web server for identifying drug targets with docking approach. *Nucleic Acids Research*, *34*, W219–W224.

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., et al. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, *25*, 2744–2750.

Liang, S., Zheng, D., Zhang, C., & Standley, D. M. (2011). Fast and accurate prediction of protein side-chain conformations. *Bioinformatics*, *27*(20), 2913–2914.
Lin, K., Simossis, V. A., Taylor, W. R., & Heringa, J. (2005). A simple and fast secondary structure prediction algorithm using hidden neural networks. *Bioinformatics*, *21*(2), 152–159.
Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., & Russell, R. B. (2003). Protein disorder prediction implications for structural proteomics. *Structure*, *11*, 1453–1459.
Linding, R., Russell, R. B., Neduva, V., & Gibson, T. J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Research*, *31*(13), 3701–3708.
Lipinski, C. A. (2000). Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods*, *44*(1), 235–249.
Lipman, D., & Pearson, W. (1985). Rapid and sensitive protein similarity searches. *Science*, *227*(4693), 1435–1441.
Liu, T., Lin, Y., Wen, X., Jorissen, R. N., & Gilson, M. K. (2007). BindingDB: A web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Research*, *35*, D198–D201.
Liu, T., Tang, G. W., & Capriotti, E. (2011). Comparative modeling: The state of the art and protein drug target structure prediction. *Combinatorial Chemistry & High Throughput Screening*, *14*(6), 532–547.
London, N., Raveh, B., Cohen, E., Fathi, G., & Schueler-Furman, O. (2011). Rosetta FlexPepDock web server—High resolution modeling of peptide-protein interactions. *Nucleic Acids Research*, *9*, W249–W253.
Lüthy, R., Bowie, J. U., & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature*, *356*, 83–85.
Luu, T. D., Rusu, A., Walter, V., Linard, B., Poidevin, L., Ripp, R., et al. (2012). KD4v: Comprehensible knowledge discovery system for missense variant. *Nucleic Acids Research*, *40*, W71–W75.
Luu, T. D., Rusu, A. M., Walter, V., Ripp, R., Moulinier, L., Muller, J., et al. (2012). MSV3d: Database of human missense variants mapped to 3D protein structure. *Database (Oxford)*, *2012*, bas018.
Ma, B., Elkayam, T., Wolfson, H., & Nussinov, R. (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 5772–5777.
MacKerell, A. D., Bashford, D., Bellott, M., Evanseck, R. L. J. D., Field, M. J., Fischer, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry. B*, *102*, 3586–3616.
Magariños, M. P., Carmona, S. J., Crowther, G. J., Ralph, S. A., Roos, D. S., Shanmugam, D., et al. (2012). TDR targets: A chemogenomics resource for neglected diseases. *Nucleic Acids Research*, *40*, D1118–D1127.
Maginn, E. J., & Elliott, J. R. (2010). Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Industrial & Engineering Chemistry Research*, *49*(7), 3059–3078.
Maglott, D., Chitipiralla, S., Church, D., Feolo, M., Garner, J., Jang, W., et al. (2013). ClinVar. NAR Molecular Biology Database Collection entry number 1570.
Magyar, C., Gromiha, M. M., Pujadas, G., Tusnády, G. E., & Simon, I. (2005). SRide: A server for identifying stabilizing residues in proteins. *Nucleic Acids Research*, *33*, W303–W305.
Mah, J. T. L., & Chia, K. S. (2007). A gentle introduction to SNP analysis: Resources and tools. *Journal of Bioinformatics and Computational Biology*, *5*, 1123–1138.

Mantsyzov, A. B., Bouvier, G., Evrard-Todeschi, N., & Bertho, G. (2012). Contact-based ligand-clustering approach for the identification of active compounds in virtual screening. *Advances and Applications in Bioinformatics and Chemistry*, *5*, 61–79.

Marialke, J., Tietze, S., & Apostolakis, J. (2008). Similarity based docking. *Journal of Chemical Information and Modeling*, *48*(1), 186–196.

Markwick, P. R. L., Cervantes, C. F., Abel, B. L., Komives, E. A., Blackledge, M., & McCammon, J. A. (2010). Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *Journal of the American Chemical Society*, *132*, 1220–1221.

Martin, L., Catherinot, V., & Labesse, G. (2006). kinDOCK: A tool for comparative docking of protein kinase ligands. *Nucleic Acids Research*, *34*(2), W325–W329.

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure*, *29*, 291–325.

Mashiach, E., Nussinov, R., & Wolfson, H. J. (2009). FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins*, *78*(6), 1503–1519.

Masso, M., & Vaisman, I. I. (2008). Accurate prediction of stability changes in protein mutants by combining machine learning with structure based computational mutagenesis. *Bioinformatics*, *24*(18), 2002–2009.

Masso, M., & Vaisman, I. I. (2010). Knowledge-based computational mutagenesis for predicting the disease potential of human non-synonymous single-nucleotide polymorphisms. *Journal of Theoretical Biology*, *266*(4), 560–568.

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P., & Tavtigian, S. V. (2006). Computational approaches for predicting the biological effect of p53 missense mutations: A comparison of three sequence analysis based methods. *Nucleic Acids Research*, *34*, 1317–1325.

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., et al. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature Methods*, *7*, 237–242.

McConkey, B. J., Sobolev, V., & Edelman, M. (2002). The performance of current methods in ligand-protein docking. *Current Science*, *83*(7), 845–856.

McGann, M. R., Almond, H. R., Nicholls, A., Grant, J. A., & Brown, F. K. (2003). Gaussian docking functions. *Biopolymers*, *68*, 76–90.

Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, *7*(2), 146.

Mi, H., Muruganujan, A., & Thomas, P. D. (2013). PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, *41*, D377–D386. http://dx.doi.org/10.1093/nar/gks1118.

Mihasan, M. (2010). Basic protein structure prediction for the biologist: A review. *Archives of Biological Sciences*, *62*, 857–871.

Mihasan, M. (2012). What in silico molecular docking can do for the 'bench-working biologists'. *Journal of Biosciences*, *37*(1), 1089–1095.

Miller, R. T., Jones, D. T., & Thornton, J. M. (1996). Protein fold recognition by sequence threading: Tools and assessment techniques. *The FASEB Journal*, *10*, 171–178.

Miteva, M. A., Brugge, J. M., Rosing, J., Nicolaes, G. A. F., & Villoutreix, B. O. (2004). Theoretical and experimental study of the D2194G mutation in the C2 domain of coagulation factor V. *Biophysical Journal*, *86*(1), 488–498.

Moll, A., Hildebrandt, A., Lenhof, H. P., & Kohlbacher, O. (2006). BALLView: a tool for research and education in molecular modeling. *Bioinformatics*, *22*(3), 365–366.

Mooney, S. (2005). Bioinformatics approaches and resources for single nucleotide polymorphism functional analysis. *Briefings in Bioinformatics*, *6*, 44–56.
Mooney, S. D., & Altman, R. B. (2003). MutDB: Annotating human variation with functionally relevant data. *Bioinformatics*, *19*(14), 1858–1860.
Moosawi, F., & Mohabatkar, H. (2009). Computer-assisted analysis of subcellular localization signals and post-translational modifications of human prion proteins. *Journal of Biomedical Science and Engineering*, *2*, 70–75.
Moreira, I. S., Fernandes, P. A., & Ramos, M. J. (2010). Protein–protein docking dealing with the unknown. *Journal of Computational Chemistry*, *31*, 317–342.
Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., et al. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, *19*, 1639–1662.
Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). Autodock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, *16*, 2785–2791.
Muegge, I. (2006). PMF scoring revisited. *Journal of Medicinal Chemistry*, *49*, 5895–5902.
Murphy, J. A., Barrantes-Reynolds, R., Kocherlakota, R., Bond, J. P., & Greenblatt, M. S. (2004). The CDKN2A database: Integrating allelic variants with evolution, structure, function, and disease association. *Human Mutation*, *24*, 296–304.
Nabuurs, S. B., Wagener, M., & de Vlieg, J. (2007). A flexible approach to induced fit docking. *Journal of Medicinal Chemistry*, *50*(26), 6507–6518.
Nagasundaram, N., & George Priya Doss, C. (2013). Predicting the impact of single-nucleotide polymorphisms in CDK2-flavopiridol complex by molecular dynamics analysis. *Cell Biochemistry and Biophysics*, *66*(3), 681–695.
Ng, P. C., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research*, *11*(5), 863–874.
Ng, P. C., & Henikoff, S. (2006). Predicting the effects of amino-acid substitutions on protein function. *Annual Review of Genomics and Human Genetics*, *7*, 61–80.
Notredame, C., Higgins, D. G., & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology*, *302*(1), 205–217.
Novikov, F. N., Stroylov, V. S., Zeifman, A. A., Stroganov, O. V., Kulkov, V., & Chilov, G. G. (2012). Lead Finder docking and virtual screening evaluation with Astex and DUD test sets. *Journal of Computer-Aided Molecular Design*, *26*(6), 725–735.
Nuytemans, K., Theuns, J., Cruts, M., & Van Broeckhoven, C. (2010). Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: A mutation update. *Human Mutation*, *31*, 763–780.
Ode, H., Matsuyama, S., Hata, M., Neya, S., Kakizawa, J., Sugiura, W., et al. (2007). Computational characterization of structural role of the non-active site mutation M36I of human immunodeficiency virus type 1 protease. *Journal of Molecular Biology*, *370*, 598–607.
Olatubosun, A., Valiaho, J., Harkonen, J., Thusberg, J., & Vihinen, M. (2012). PON-P: Integrated predictor for pathogenicity of missense variants. *Human Mutation*, *33*, 1166–1174.
Ortiz, M. A., Light, J., Maki, R. A., & Assa-Munt, N. (1999). Mutation analysis of the pip interaction domain reveals critical residues for protein-protein interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 2740–2745.
O'Sullivan, O., Suhre, K., Abergel, C., Higgins, D. G., & Notredame, C. (2004). 3DCoffee: Combining protein sequences and structures within multiple sequence alignments. *Journal of Molecular Biology*, *340*, 385–395.
Ozbabacan, S. E. A., Gursoy, A., Keskin, O., & Nussinov, R. (2010). Conformational ensembles, signal transduction and residue hot spots: Application to drug discovery. *Current Opinion in Drug Discovery and Development*, *13*(5), 527–537.

Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A., et al. (2011). Contributions of hydrophobic interactions to protein stability. *Journal of Molecular Biology*, *408*, 514–528.

Pak, Y., & Wang, S. (2000). Application of a molecular dynamics simulation method with a generalized effective potential to the flexible molecular docking problems. *The Journal of Physical Chemistry. B*, *104*(2), 354–359.

Park, H., Lee, J., & Lee, S. (2006). Critical assessment of the automated AutoDock as a new docking tool for virtual screening. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 549–554.

Parthiban, V., Gromiha, M. M., Hoppe, C., & Schomburg, D. (2007). Structural analysis and prediction of protein mutant stability using distance and torsion potentials: Role of secondary structure and solvent accessibility. *Proteins*, *66*, 41–52.

Parthiban, V., Gromiha, M. M., & Schomburg, D. (2006). CUPSAT: Prediction of protein stability upon point mutations. *Nucleic Acids Research*, *34*, W239–W242.

Pastinen, T., Ge, B., & Hudson, T. J. (2006). Influence of human genome polymorphism on gene expression. *Human Molecular Genetics*, *15*, 9–16.

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America*, *85*, 2444–2448.

Pedretti, A., Villa, L., & Vistoli, G. (2004). VEGA—An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script" programming. *Journal of Computer-Aided Molecular Design*, *18*, 167–173.

Peter, C., Rueping, M., Worner, H. J., Jaun, B., Seebach, D., & van Gunsteren, W. E. F. (2003). Molecular dynamics simulations of small peptides: Can one derive conformational preferences from ROESY spectra? *Chemistry*, *9*, 5838–5849.

Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., & Lundegaard, C. (2009). A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology*, *9*, 51.

Peterson, T. A., Doughty, E., & Kann, M. G. (2013). Towards precision medicine: Advances in computational approaches for the analysis of human variants. *Journal of Molecular Biology*, *425*(21), 4047–4063.

Petrey, D., Xiang, Z., Tang, C. L., Xie, L., Gimpelev, M., Mitros, T., et al. (2003). Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling. *Proteins*, *53*(6), 430–435.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, *26*(16), 1781–1802.

Pierri, C. L., Parisi, G., & Porcelli, V. (2010). Computational approaches for protein function prediction: A combined strategy from multiple sequence alignment to molecular docking-based virtual screening. *Biochimica et Biophysica Acta*, *1804*, 1695–1712.

Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, *117*(1), 1–19.

Pollastri, G., Baldi, P., Fariselli, P., & Casadio, R. (2002). Prediction of coordination number and relative solvent accessibility in proteins. *Proteins*, *47*, 142–153.

Pons, J. L., & Labesse, G. (2009). TOME-2: A new pipeline for comparative modeling of protein–ligand complexes. *Nucleic Acids Research*, *37*(2), W485–W491.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., et al. (2005). FoldIndex: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, *21*(16), 3435–3438.

Proia, R. L., & Neufeld, E. F. (1982). Synthesis of beta-hexosaminidase in cell-free translation and in intact fibroblasts: An insoluble precursor alpha chain in a rare form of Tay-Sachs disease. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 6360–6364.

Prokunina, L., & Alarcon-Riquelme, M. E. (2004). Regulatory SNPs in complex diseases: Their identification and functional validation. *Expert Reviews in Molecular Medicine*, *6*, 1–15.

Pruitt, K. D., Tatusova, T., Brown, G. R., & Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): Current status, new features and genome annotation policy. *Nucleic Acids Research*, *40*, D130–D135.

Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., & Ben-Tal, N. (2002). Rate4Site: An algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, *18*, S71–S77.

Qu, X., Swanson, R., Day, R., & Tsai, J. (2009). A guide to template based structure prediction. *Current Protein & Peptide Science*, *10*, 270–285.

Radivojac, P., Baenziger, P. H., Kann, M. G., Mort, M. E., Hahn, M. W., & Mooney, S. D. (2008). Gain and loss of phosphorylation sites in human cancer. *Bioinformatics*, *24*(16), i241–i247.

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: Server and survey. *Nucleic Acids Research*, *30*, 3894–3900.

Rao, V. S., & Srinivas, K. (2011). Modern drug discovery process: An in silico approach. *Journal of Bioinformatics and Sequence Analysis*, *2*(5), 89–94.

Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology*, *261*, 470–489.

Rask-Andersen, M., Almen, M. S., & Schioth, H. B. (2011). Trends in the exploitation of novel drug targets. *Nature Reviews. Drug Discovery*, *10*, 579–590.

Rasmussen, B. F., Stock, A. M., Ringe, D., & Petsko, G. A. (1992). Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature*, *357*, 423–424.

Raychaudhuri, S. (2011). Mapping rare and common causal alleles for complex human diseases. *Cell*, *147*, 57–69.

Rees, D. C., Congreve, M., Murray, C. W., & Carr, R. (2004). Fragment-based lead discovery. *Nature Reviews. Drug Discovery*, *3*(8), 660–672.

Ren, J., Xie, L., Li, W. W., & Bourne, P. E. (2010). SMAP-WS: A parallel web service for structural proteome-wide ligand-binding site comparison. *Nucleic Acids Research*, *38*, W441–W444.

Rennell, D., Bouvier, S. E., Hardy, L. W., & Poteete, A. R. (1991). Systematic mutation of bacteriophage T4 lysozyme. *Journal of Molecular Biology*, *222*, 67–88.

Reumers, J., Schymkowitz, J., & Rousseau, F. (2009). Using structural bioinformatics to investigate the impact of non-synonymous SNPs and disease mutations: Scope and limitations. *BMC Bioinformatics*, *27*, 10.

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*, *39*, e118.

Ridder, L., Wang, H., de Vlieg, J., & Wagener, M. (2011). Revisiting the rule of five on the basis of pharmacokinetic data from rat. *ChemMedChem*, *6*(11), 1967–1970.

Rignall, T. R., Baker, J. O., McCarter, S. L., Adney, W. S., Vinzant, T. B., Decker, S. R., et al. (2002). Effect of single active-site cleft mutation on product specificity in a thermostable bacterial cellulase. *Applied Biochemistry and Biotechnology*, *98–100*, 383–394.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature*, *405*, 847–856.

Rodriguez-Casado, A. (2012). In silico investigation of functional nsSNPs—An approach to rational drug design. *Research and Reports in Medicinal Chemistry*, *2*, 31–42.

Rohl, C. A., Strauss, C. E., Misura, K. M., & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods in Enzymology*, *383*, 66–93.
Rose, G. D., & Wolfenden, R. (1993). Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annual Review of Biophysics and Biomolecular Structure*, *22*, 381–415.
Rost, B., Fariselli, P., & Casadio, R. (1996). Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Science*, *5*(8), 1704–1718.
Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D., et al. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Research*, *35*, D823–D828.
Ryu, G. M., Song, P., Kim, K. W., Oh, K. S., Park, K. J., & Kim, J. H. (2009). Genome-wide analysis to predict protein sequence variations that change phosphorylation sites or their corresponding kinases. *Nucleic Acids Research*, *37*(4), 1297–1307.
Sadee, W., & Dai, Z. (2005). Pharmacogenetics/genomics and personalized medicine. *Human Molecular Genetics*, *14*, R207–R214.
Salsbury, F. R., Jr. (2010). Molecular dynamics simulations of protein dynamics and their relevance to drug discovery. *Current Opinion in Pharmacology*, *10*, 738–744.
Samudrala, R., & Moult, J. (1998). Determinants of side chain conformational preferences in protein structures. *Protein Engineering*, *11*(11), 991–997.
Saunders, C. T., & Baker, D. (2002). Evaluation of structural and evolutionary contributions to deleterious mutation prediction. *Journal of Molecular Biology*, *322*(4), 891–901.
Sayle, R., & White, E. J. M. (1995). RasMol: Biomolecular graphics for all. *Trends in Biochemical Sciences*, *20*(9), 374.
Schaefer, C., Meier, A., Rost, B., & Bromberg, Y. (2012). SNPdbe: Constructing an nsSNP functional impacts database. *Bioinformatics*, *28*(4), 601–602.
Schlick, T., Collepardo-Guevara, R., Halvorsen, L. A., Jung, S., & Xiao, X. (2011). Biomolecular modeling and simulation: A field coming of age. *Quarterly Reviews of Biophysics*, *44*, 191–228.
Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005a). Geometry based flexible and symmetric protein docking. *Proteins*, *60*, 224–231.
Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2005b). PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Research*, *33*, W363–W367.
Schwarz, J. M., Rodelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, *7*, 575–576.
Seiler, K. P., George, G. A., Happ, M. P., Bodycombe, N. E., Carrinski, H. A., Norton, S., et al. (2008). ChemBank: A small-molecule screening and cheminformatics resource database. *Nucleic Acids Research*, *36*, D351–D359.
Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., Seeliger, M. A., Shaw, D. E., et al. (2011). How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, *133*, 9181–9183.
Shastry, B. S. (2006a). Pharmacogenetics and the concept of individualized medicine. *The Pharmacogenomics Journal*, *6*, 16–21.
Shastry, B. S. (2006b). Role of SNPs and haplotypes in human disease and drug development. In M. Ozkan, M. J. Heller, & M. Ferrari (Eds.), *Micro/nano technology in genomics and proteomics: 2*, (pp. 447–458). New York: Springer.
Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*, 308–311.
Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L., Edwards, K. J., et al. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*, *34*(1), 57–65.

Shirley, B. A., Stanssens, P., Hahn, U., & Pace, C. N. (1992). Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry*, *31*, 725–732.
Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., et al. (2007). DisProt: The database of disordered proteins. *Nucleic Acids Research*, *35*, D786–D793.
Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*, 1034–1050.
Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, *7*, 539.
Singh, T., Biswas, D., & Jayaram, B. (2011). AADS—An automated active site identification, docking and scoring protocol for protein targets based on physico-chemical descriptors. *Journal of Chemical Information and Modeling*, *51*(10), 2515–2527.
Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins*, *17*, 355–362.
Smith, E. P., Boyd, J., Frank, G. R., Takahashi, H., Cohen, R. M., Specker, B., et al. (1994). Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *The New England Journal of Medicine*, *331*, 1056–1061.
Smith, R. D., Engdahl, A. L., Dunbar, J. B., Jr., & Carlson, H. A. (2012). Biophysical limits of protein–ligand binding. *Journal of Chemical Information and Modeling*, *52*(8), 2098–2106.
Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, *33*, W244–W248.
Song, E. S., Daily, A., Fried, M. G., Juliano, M. A., Juliano, L., & Hersh, L. B. (2005). Mutation of active site residues of insulin degrading enzyme alters allosteric interactions. *The Journal of Biological Chemistry*, *280*(18), 17701–17706.
Song, X., Geng, Z., Zhu, J., Li, C., Hu, X., Bian, N., et al. (2009). Structure-function roles of four cysteine residues in the human arsenic (+3 oxidation state) methyltransferase (hAS3MT) by site-directed mutagenesis. *Chemico-Biological Interactions*, *179*, 321–328.
Song, C. M., Lim, S. J., & Tong, J. C. (2009). Recent advances in computer-aided drug design. *Briefings in Bioinformatics*, *10*(5), 579–591.
Sousa, S. F., Fernandes, P. A., & Ramos, M. J. (2006). Protein-ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics*, *65*(1), 15–26.
Steen, M., Miteva, M., Villoutreix, B. O., Yamazaki, T., & Dahlback, B. (2003). Factor V new brunswick: Ala221Val associated with FV deficiency reproduced in vitro and functionally characterized. *Blood*, *102*(4), 1316–1322.
Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shaw, K., & Cooper, D. N. (2012). The human gene mutation database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. *Current Protocols in Bioinformatics*, chap: unit 6.
Stenson, P. D., Ball, E. V., Mort, M., Phillips, A. D., Shiel, J. A., Thomas, N. S., et al. (2003). Human gene mutation database (HGMD): 2003 update. *Human Mutation*, *21*, 577–581.
Stevanin, G., Hahn, V., Lohmann, E., Bouslam, N., Gouttard, M., Soumphonphakdy, C., et al. (2004). Mutation in the catalytic domain of protein kinase C γ and extension of the phenotype associated with spinocerebellar ataxia type 14. *Archives of Neurology*, *61*(8), 1242–1248.
Stitziel, N. O., Tseng, Y. Y., Pervouchine, D., Goddeau, D., Kasif, S., & Liang, J. (2003). Structural location of disease-associated single-nucleotide polymorphisms. *Journal of Molecular Biology*, *327*, 1021–1030.
Stryer, L. (1995). *Biochemistry* (4th ed.). New York: W.H. Freeman.
Sunyaev, S., Ramensky, V., Koch, I., Lathe, W., III, Kondrashov, A. S., & Bork, P. (2001). Prediction of deleterious human alleles. *Human Molecular Genetics*, *10*(6), 591–597.

Sushma, B., & Suresh, C. V. (2012). Docking—A review. *Journal of Applicable Chemistry*, *1*(2), 167–173.
Taboureau, O., Baell, J. B., Fernández-Recio, J., & Villoutreix, B. O. (2012). Established and emerging trends in computational drug discovery in the structural genomics era. *Chemistry & Biology*, *19*(1), 29–41.
Taillon-Miller, P., Gu, Z., Li, Q., Hillier, L., & Kwok, P. Y. (1998). Overlapping genomic sequences: A treasure trove of single-nucleotide polymorphisms. *Genome Research*, *8*(7), 748–754.
Takamiya, O., Seta, M., Tanaka, K., & Ishida, F. (2002). Human factor VII deficiency caused by S339C mutation located adjacent to the specificity pocket of the catalytic domain. *Clinical and Laboratory Haematology*, *24*(4), 233–238.
Tang, K. E. S., & Dill, K. A. (1998). Native protein fluctuations: The conformational-motion temperature and the inverse correlation of protein flexibility with protein stability,. *Journal of Biomolecular Structure and Dynamics*, *16*(2), 397–411.
Tavtigian, S. V., Deffenbaugh, A. M., Yin, L., Judkins, T., Scholl, T., Samollow, P. B., et al. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of Medical Genetics*, *43*(4), 295–305.
Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, *16*(3), 151–166.
Teng, S., Madej, T., Panchenko, A., & Alexov, E. (2009). Modeling effects of human single-nucleotide polymorphisms on protein-protein interactions. *Biophysical Journal*, *96*(6), 2178–2188.
Tennessen, J. A., Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., et al. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, *337*, 64–69.
The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
Thomas, M., Dadgar, N., Aphale, A., Harrell, J. M., Kunkel, R., Pratt, W. B., et al. (2004). Androgen receptor acetylation site mutations cause trafficking defects, misfolding, and aggregation similar to expanded glutamine tracts. *The Journal of Biological Chemistry*, *279*, 8389–8395.
Thomas, R., McConnell, R., Whittacker, J., Kirkpatrick, P., Bradley, J., & Sandford, R. (1999). Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene PKD1, by long-range PCR. *American Journal of Human Genetics*, *65*, 39–49.
Thomas, P. J., Qu, B. H., & Pedersen, P. L. (1995). Defective protein folding as a basis of human disease. *Trends in Biochemical Sciences*, *20*, 456–459.
Thomsen, R., & Christensen, M. H. (2006). MolDock:? A new technique for high-accuracy molecular docking. *Journal of Medicinal Chemistry*, *49*(11), 3315–3321.
Thusberg, J., Olatubosun, A., & Vihinen, M. (2011). Performance of mutation pathogenicity prediction methods on missense variants. *Human Mutation*, *32*(4), 358–368.
Thusberg, J., & Vihinen, M. (2009). Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods. *Human Mutation*, *30*, 703–714.
Tian, J., Wu, N., Guo, X., Guo, J., Zhang, J., & Fan, Y. (2007). Predicting the phenotypic effects of non-synonymous single nucleotide polymorphisms based on support vector machines. *BMC Bioinformatics*, *8*, 450.
Tietze, S., & Apostolakis, T. (2007). Glamdock: Development and validation of a new docking tool on several thousand protein-ligand complexes. *Journal of Chemical Information and Modeling*, *47*(4), 1657–1672.
Tiffin, N., Okpechi, I., Perez-Iratxeta, C., Andrade-Navarro, M. A., & Ramesar, R. (2008). Prioritization of candidate disease genes for metabolic syndrome by computational analysis of its defining phenotypes. *Physiological Genomics*, *35*(1), 55–64.

Tolkacheva, T., Boddapati, M., Sanfiz, A., Tsuchida, K., Kimmelman, A. C., & Chan, A. M. (2001). Regulation of PTEN binding to MAGI-2 by two putative phosphorylation sites at threonine 382 and 383. *Cancer Research*, *61*(13), 4985–4989.

Tomalik-Scharte, D., Lazar, A., Fuhr, U., & Kirchheiner, J. (2008). The clinical role of genetic polymorphisms in drug-metabolizing enzymes. *Pharmacogenomics*, *J8*, 4–15.

Tovchigrechko, A., & Vakser, I. A. (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research*, *34*, W310–W314.

Tress, M. L., Jones, D., & Valencia, A. (2003). Predicting reliable regions in protein alignments from sequence profiles. *Journal of Molecular Biology*, *330*(4), 705–718.

Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, *31*, 455–461.

Trovato, A., Seno, F., & Tosatto, S. C. (2007). The PASTA server for protein aggregation prediction. *Protein Engineering, Design & Selection*, *20*, 521–523.

Tsai, T. Y., Chang, K. W., & Chen, C. Y. (2011). iScreen: World's first cloud-computing web server for virtual screening and de novo drug design based on TCM database@Taiwan. *Journal of Computer-Aided Molecular Design*, *25*(6), 525–531.

Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A., & Hamodrakas, S. J. (2013). A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS ONE*, *8*(1), e54175.

Tuckerman, M. E., Yarne, D. A., Samuelson, S. O., Hughes, A. L., & Martyna, G. J. (2000). Exploiting multiple levels of parallelism in molecular dynamics based calculations via modern techniques and software paradigms on distributed memory computers. *Computer Physics Communications*, *128*(1), 333–376.

Ung, M. U., Lu, B., & McCammon, J. A. (2006). E230Q mutation of the catalytic subunit of cAMP-dependent protein kinase affects local structure and the binding of peptide inhibitor. *Biopolymers*, *81*, 428–439.

Utesch, T., Daminelli, G., & Mroginski, M. A. (2011). Molecular dynamics simulations of the adsorption of bone morphogenetic protein-2 on surfaces with medical relevance. *Langmuir*, *27*(21), 13144–13153.

Uzun, A., Leslin, C. M., Abyzov, A., & Ilyin, V. (2007). Structure SNP (StSNP): A web server for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways. *Nucleic Acids Research*, *35*, W384–W392.

Valerio, M., Colosimo, A., Conti, F., Giuliani, A., Grottesi, A., Manetti, C., et al. (2005). Early events in protein aggregation: Molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations. *Proteins*, *58*(1), 110–118.

Van Durme, J., Maurer-Stroh, S., Gallardo, R., Wilkinson, H., Rousseau, F., & Schymkowitz, J. (2009). Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Computational Biology*, *5*(8), e1000475.

van Wijk, R., Rijksen, G., Huizinga, E. G., Nieuwenhuis, H. K., & van Solinge, W. W. (2003). HK Utrecht: Missense mutation in the active site of human hexokinase associated with Hexo-kinase deficiency and severe nonspherocytic hemolytic anemia. *Blood*, *101*(1), 345–347.

Vatsis, K. P., Martell, K. J., & Weber, W. W. (1991). Diverse point mutations in the human gene for polymorphic N-acetyltransferase. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 6333–6337.

Vazquez, F. (2000). Phosphorylation of the PTEN tail regulates protein stability and function. *Molecular and Cellular Biology*, *20*, 5010–5018.

Venkatesan, R. N., Treuting, P. M., Fuller, E. D., Goldsby, R. E., Norwood, T. H., Gooley, T. A., et al. (2007). Mutation at the polymerase active site of mouse DNA

polymerase delta increases genomic instability and accelerates tumorigenesis. *Molecular and Cellular Biology*, *27*, 7669–7682.
Venselaar, H., Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L., & Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*, *11*, 548.
Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, *52*(4), 609–623.
Villoutreix, B. O., Renault, N., Lagorce, D., Sperandio, O., Montes, M., & Miteva, M. A. (2007). Free resources to assist structure-based virtual ligand screening experiments. *Current Protein and Peptide Science*, *8*(4), 381–411.
Vitkup, D., Sander, C., & Church, G. M. (2003). The amino acid mutational spectrum of human genetic disease. *Genome Biology*, *4*, R72–R80.
Vogt, G., Vogt, B., Chuzhanova, N., Julenius, K., Cooper, D. N., & Casanova, J. L. (2007). Gain-of-glycosylation mutations. *Current Opinion in Genetics & Development*, *17*, 245–251.
Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. *Journal of Molecular Graphics*, *8*, 52–56.
Vullo, A., Bortolami, O., Pollastri, G., & Tosatto, S. C. (2006). Spritz: A server for the prediction of intrinsically disordered regions in protein sequences using kernel machines. *Nucleic Acids Research*, *34*, W164–W168.
Wallner, B., & Elofsson, A. (2005). All are not equal: A benchmark of different homology modeling programs. *Protein Science*, *14*(5), 1315–1327.
Wallner, R. N., Lindahl, E., & Elofsson, A. (2008). Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Science*, *17*, 990–1002.
Walsh, C. T. (2006). *Posttranslational modification of proteins: Expanding nature's inventory*. Englewood, CO: Roberts and Company Publishers.
Wang, J. C., Chu, P. Y., Chen, C. M., & Lin, J. H. (2012). idTarget: A web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Research*, *40*, W393–W399.
Wang, Q., Curran, M. E., Splawski, I., Burn, T. C., Millholland, J. M., VanRaay, T. J., et al. (1996). Positional cloning of a novel potassium channel gene: KVLQT1 mutations cause cardiac arrhythmias. *Nature Genetics*, *12*(1), 17–23.
Wang, Z., Eickholt, J., & Cheng, J. (2010). MULTICOM: A multi-level combination approach to protein structure prediction and its assessments in CASP8. *Bioinformatics*, *26*, 882–888.
Wang, R., Fang, X., Lu, Y., & Wang, S. (2004). The PDBbind database: Collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *Journal of Medicinal Chemistry*, *47*, 2977–2980.
Wang, Z., & Moult, J. (2001). SNPs, protein structure, and disease. *Human Mutation*, *7*, 263–270.
Wang, Z., & Moult, J. (2003). Three-dimensional structural location and molecular functional effects of missense SNPs in the T cell receptor Vbeta domain. *Proteins*, *53*, 748–757.
Wang, J., Ronaghi, M., Chong, S. S., & Lee, C. G. (2011). pfSNP: An integrated potentially functional SNP resource that facilitates hypotheses generation through knowledge syntheses. *Human Mutation*, *32*, 19–24.
Wang, M., Sun, Z., Akutsu, T., & Song, J. (2013). Recent advances in predicting functional impact of single amino acid polymorphisms: A review of useful features, computational methods and available tools. *Current Bioinformatics*, *8*, 161–176.
Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of general amber force field. *Journal of Computational Chemistry*, *25*, 1157–1174.

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., & Bryant, S. H. (2009). PubChem: A public information system for analyzing bioactivities of small molecules. *Nucleic Acids Research*, *37*, W623–W633.
Wang, L. L., Yang, A. K., Li, Y., Liu, J. P., & Zhou, S. F. (2010). Phenotype prediction of deleterious nonsynonymous single nucleotide polymorphisms in human alcohol metabolism-related genes: A bioinformatics study. *Alcohol*, *44*(5), 425–438.
Wass, M. N., Kelley, L. A., & Sternberg, M. J. (2010). 3DLigandSite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Research*, *38*, W469–W473.
Weaire, D., & Aste, T. (2010). The pursuit of perfect packing. *Contemporary Physics*, *51*, 1.
Weigelt, J. (2010). Structural genomics—Impact on biomedicine and drug discovery. *Experimental Cell Research*, *316*, 1332–1338.
Weisel, M., Proschak, E., & Schneider, G. (2007). PocketPicker: Analysis of ligand binding sites with shape descriptors. *Chemistry Central Journal*, *13*(1), 7.
Werner, T., Morris, M. B., Dastmalchi, S., & Church, W. B. (2012). Structural modelling and dynamics of proteins for insights into drug interactions. *Advanced Drug Delivery Reviews*, *64*(4), 323–343.
Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, *35*, W407–W410.
Wilke, R. A., & Dolan, M. E. (2011). Genetics and variable drug response. *Journal of the American Medical Association*, *306*, 306–307.
Wilkinson, G. R. (2005). Drug metabolism and variability among patients in drug response. *The New England Journal of Medicine*, *352*, 2211–2221.
Wishart, D. S. (2007). Human Metabolome Database: Completing the 'human parts list'. *Pharmacogenomics*, *8*(7), 683–686.
Witham, S., Takano, K., Schwartz, C., & Alexov, E. (2011). missense mutation in CLIC2 associated with intellectual disability is predicted by in silico modeling to affect protein stability and dynamics. *Proteins*, *79*(8), 2444–2454.
Wjst, M. (2004). Target SNP selection in complex disease association studies. *BMC Bioinformatics*, *5*, 92.
Wong, P., Fritz, A., & Frishman, D. (2005). Designability, aggregation propensity and duplication of disease-associated proteins. *Protein Engineering, Design & Selection*, *18*, 503–508.
Wong, C. F., & McCammon, J. A. (2003). Protein flexibility and computer-aided drug design. *Annual Review of Pharmacology and Toxicology*, *43*, 31–45.
Wright, J. D., & Lim, C. (2007). Mechanism of DNA-binding loss upon single-point mutation in p53. *Journal of Biosciences*, *32*(5), 827–839.
Xiang, Z. (2006). Advances in homology protein structure modeling. *Current Protein & Peptide Science*, *7*(3), 217–227.
Xiang, Z., Soto, C. S., & Honig, B. (2002). Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(11), 7432–7437.
Xu, J., & Berger, B. (2006). Fast and accurate algorithms for protein side-chain packing. *Journal of the ACM*, *53*(4), 533–557.
Xu, H., Gregory, S. G., Hauser, E. R., Stenger, J. E., Pericak-Vance, M. A., Vance, J. M., et al. (2005). SNPselector: A web tool for selecting SNPs for genetic association studies. *Bioinformatics*, *21*, 4181–4186.
Xu, J., Li, M., Kim, D., & Xu, Y. (2003). RAPTOR: Optimal protein threading by linear programming. *Journal of Bioinformatics and Computational Biology*, *1*, 95–117.
Yamada, Y., Banno, Y., Yoshida, H., Kikuchi, R., Akao, Y., Murate, T., et al. (2006). Catalytic inactivation of human phospholipase D2 by a naturally occurring Gly901Asp mutation. *Archives of Medical Research*, *37*, 696–699.

Yang, S. Y. (2010). Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discovery Today*, *15*(11), 444–450.
Yang, J. M., & Chen, C. C. (2004). GEMDOCK: A generic evolutionary method for molecular docking. *Proteins: Structure, Function, and Bioinformatics*, *55*, 288–304.
Yang, Z. R., Thomson, R., McNeil, P., & Esnouf, R. M. (2005). RONN: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, *21*, 3369–3376.
Yoshida, A., Huang, I. Y., & Ikawa, M. (1984). Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *Proceedings of the National Academy of Sciences of the United States of America*, *81*, 258–261.
Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B., & Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C terminal tyrosine phosphorylation. *Cell*, *105*(1), 115–126.
Yuan, H. Y., Chiou, J. J., Tseng, W. H., Liu, C. H., Liu, C. K., Lin, Y. J., et al. (2006). FASTSNP: An always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Research*, *34*, 35–41.
Yue, P., & Moult, J. (2006). Identification and analysis of deleterious human SNPs. *Journal of Molecular Biology*, *356*, 1263–1274.
Zemla, A., Venclovas, C., Reinhardt, A., Fidelis, K., & Hubbard, T. J. (1997). Numerical criteria for the evaluation of ab initio predictions of protein structure. *Proteins*, *1*, 140–150.
Zhang, Y. (2008a). Progress and challenges in protein structure prediction. *Current Opinion in Structural Biology*, *18*(3), 342–348.
Zhang, Y. (2008b). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, *9*, 40.
Zhang, Y. (2009). Protein structure prediction: When is it useful? *Current Opinion in Structural Biology*, *19*, 145–155.
Zhang, Z., Norris, J., Schwartz, C., & Alexov, E. (2011). In-silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase. *PLoS One*, *6*(5), e20373.
Zhang, Z., Teng, S., Wang, L., Schwartz, C. E., & Alexov, E. (2010). Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Human Mutation*, *31*(9), 1043–1049.
Zhang, J., Wang, Q., Barz, B., He, Z., Kosztin, I., Shang, Y., et al. (2010). MUFOLD: A new solution for protein 3D structure prediction. *Proteins*, *78*, 1137–1152.
Zhao, Y., & Sanner, M. F. (2007). FLIPDock: Docking flexible ligands into flexible receptors. *Proteins: Structure, Function, and Bioinformatics*, *68*(3), 726–737.
Zhou, H., & Zhou, Y. (2002). Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science*, *11*(Suppl. 11), 2714–2726.
Zhou, S. F., Chan, E., Zhou, Z. W., Xue, C. C., Lai, X., & Duan, W. (2009). Insights into the structure, function, and regulation of human cytochrome P450 1A2. *Current Drug Metabolism*, *10*(7), 713–729.
Zsoldos, Z., Reid, D., Simon, A., Sadjad, B. S., & Johnson, A. P. (2006). eHiTS: An innovative approach to the docking and scoring function problems. *Current Protein and Peptide Science*, *7*(5), 421–435.
Zsoldos, Z., Reid, D., Simon, A., Sadjad, S. B., & Johnson, A. P. (2007). eHiTS: A new fast, exhaustive flexible ligand docking system. *Journal of Molecular Graphics and Modelling*, *26*(1), 198–212.