REVIEW

In silico investigation of functional nsSNPs – an approach to rational drug design

Arantxa Rodriguez-Casado

IMDEA Food Institute, Madrid, Spain

Abstract: Nonsynonymous single-nucleotide polymorphisms (nsSNPs) of genes are coding variants that introduce amino acid changes to proteins. Because nsSNPs have diverse effects on protein structure and function, these variants can have a large impact on human health. Likewise, polymorphisms in genes encoding drug targets directly affect target protein function, drug–target interaction, or both to produce profound effects on drug response. Pharmacogenomics provides insight into how polymorphisms can affect drug responses that vary from potentially fatal adverse drug reactions to equally severe lack of therapeutic efficacy. This review focuses on the effects of deleterious nsSNPs at both functional and structural protein level, as well as on protein–protein interactions. It also explains current methods by which to predict the functional impact of nsSNPs based on physicochemical amino acid properties, sequence information, and structural attributes. Additionally, the review details the recent advances in the field of in silico pharmacogenomics, which provide insight into disease phenotypes and individual susceptibility to disease. In the near future, this information will be used to predict the most appropriate treatment and individualized drug therapy. Finally, this article provides an overview of in silico approaches for the assessment and development of safe and efficient tailored drugs.

Keywords: pharmacogenomics, nonsynonymous single-nucleotide polymorphisms, drug responses, drug-target protein, personalized medicine

Introduction

Triggered by the completion of the human genome sequence – along with the accumulated information on genetic polymorphisms – the relatively new science of pharmacogenomics provides, in principle, the possibility of personalized medicines with reduced side effects.[1,2] Millions of severe adverse drug reactions (ADRs) in humans are reported every year, which results in thousands of avoidable deaths.[3] Finding safe and effective drugs, however, is a complicated and slow process, mostly because each person responds differently to drugs and side effects are unpredictable.[2,4] Treatments for certain psychiatric disorders, pulmonary and cardiac diseases, and cancer (especially in geriatric patients) sometimes lead to polymedication, which, in turn, leads to harmful drug interactions or ADRs.[1,5] Currently, there is no easy way to anticipate an individual's response to drugs. Nevertheless, it is worth noting that genetic polymorphisms of drug-metabolizing enzymes and transporters can affect the absorption, distribution, metabolism, elimination, and toxicity (ADMET) of drugs and, thereby, contribute to the susceptibility of patients to treatment and to whether drugs do or do not cause adverse reactions.[2–6] Pharmacogenomics provides insight into how genetic polymorphisms affect response to drugs, and how drugs can be customized to match the genetically

Correspondence: Arantxa Rodriguez-Casado
C/Faraday 7, 28049 Madrid, Spain
Tel +34 91 279 6987
Fax +34 91 188 0756
Email arantxa.rodriguez@imdea.org

31

determined response of each person to drug therapies.[2–6] Indeed, pharmacogenomics can be used to improve drug efficacy, notably by overcoming drug resistance, finding new drug targets, and optimizing ADMET.[2–6]

Drug discovery involves screening strategies for hit-identification and optimization to increase affinity, selectivity, efficacy/potency, metabolic stability, and oral bioavailability. Once a compound that fulfills these requirements has been identified, the process of drug development begins to bring a new drug to the market. This process includes preclinical testing (on microorganisms/animals), clinical trials (on humans), and regulatory approval to market the drug. Numerous new chemical entities fail drug development due to a poor ADMET profile. The progress in computational techniques enables the use of in silico methods to predict ADMET properties in parallel with or prior to experimental investigations, so that huge numbers of compounds can be evaluated prior to their being synthesized and assayed.

The most common genetic variants are single-nucleotide polymorphisms (SNPs), which occur at a frequency of about 1 in 1650 bp throughout the human genome.[5] Since many diseases have a strong genetic component, the identification of SNPs and the understanding of their involvement in pathologic conditions would shed light on disease susceptibility and aid in the development of more effective targeting treatments. As of June 26, 2012, about 53,558,214 human SNPs were identified and deposited in the National Center for Biotechnology Information (NCBI) Database of Single-Nucleotide Polymorphism ([dbSNP] http://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary. cgi?view+summary=view+summary&build_id=137, dbSNP Build 137).

The most identifiable group of SNPs is constituted by the nonsynonymous SNPs (nsSNPs), a small ($<$1%) proportion of which are located within the coding regions of genes that lead to alterations in the amino-acid sequence of their corresponding proteins.[7,8] Nonsynonymous SNPs have the potential to affect the structure and function of expressed proteins and are, therefore, likely to represent modifiers of inherited susceptibility to disease.[7,8] Most nsSNPs have been identified by sequencing and genotyping DNA samples from general populations, particularly by the HapMap project (http://hapmap.ncbi.nlm.nih.gov/), a key resource for finding genetic variants that affect disease and response to drugs. Moreover, important databases that contain these variations include the Human Genome Variation Database ([HGVBase] https://www.gwascentral.org/) and the NCBI's dbSNP (http://www.ncbi.nlm.nih.gov/omim). Additionally, the Online

Mendelian Inheritance in Man ([OMIM] http://www.ncbi. nlm.nih.gov/omim) database organizes genetic disorders of inherited diseases mapped to human genes.

According to these databases, half of all genetic changes related to human diseases are attributable to nsSNPs. Hence, these polymorphisms are considered to be deleterious nsSNPs because they lead to dramatic phenotypic consequence.[9–11] However, not all nsSNPs are associated with disease. Some nsSNPs, called tolerant nsSNPs, maintain protein function even though the corresponding first-order structure has changed.[1,5,12,13] One of the current main problems of medical genetics is to identify nsSNPs disease-related phenotypes. Differentiation of deleterious nsSNPs from tolerant nsSNPs is important to characterize the genetic basis of human disease, and pathogenesis of disease and thus be able to assess individual susceptibility to disease.[14]

nsSNPs often alter function by disrupting protein structure and/or stability and by impacting functional binding sites.[10–12] Structural changes of receptors or active target-enzyme sites may affect drug–receptor or drug–enzyme interactions and, consequently, drug response.[15–18] Genetic polymorphisms of drug-metabolizing enzymes and transporters can affect the drug ADMET profile, thereby leading to an accumulation of unmetabolized drugs around the metabolic pathway and, as a result, increased ADR.[15–18] nsSNPs can be located in a pocket or in a void (type P), on a convex region (type S), or be completely buried inside the protein (type I). It has been found that 88% of pathogenic nsSNPs are of type P and rarely of type I.[19,20] By considering physicochemical differences of exchanged amino acids, evolutionary conservation, and structural features, numerous studies have found that about 70% of disease-associated nsSNPs correspond to highly conserved residues, usually buried inside the protein with a strong impact on protein structure, folding, stability, and function.[7,9,14,19–24]

There is a need to identify functional nsSNPs that may be deleterious. However, the identification of a single nsSNP may not be sufficient to relate the variation of a target protein to a disease or a drug response. Groups of genetic variants are inherited together in linkage disequilibrium and, thus, are particularly useful in genome–phenotype analyses. For this reason, in silico techniques are developed to integrate sets of nsSNPs across the entire genome to identify genetic loci that exist in linkage disequilibrium. These approaches increase the probability of success in identifying polymorphisms of drug targets, drug-metabolizing enzymes, drug transporters, and other genes that influence drug response,

as well as new disease susceptibility genes and pathways that are important in the etiology and pathogenesis of chronic diseases.[7,23,25–29]

Effects induced by nsSNPs on protein function

nsSNPs alter cellular functions in many ways. Indeed, nsSNPs often influence normal protein function by a combination of effects on protein stability, protein–protein interactions, and many others features as described below.

Characteristics of active-sites

If an amino acid is altered in the active-site region of a protein, it may decrease or destroy the ability of the protein to perform its function and lead to a disease phenotype. A critical site may be a catalytic residue or a residue involved in ligand-binding in an enzyme, or a residue involved in binding to partner molecules. In these cases, a disease phenotype may arise from loss/gain of function, altered binding specificity, or affinity in the protein, while the stability of the protein product is not affected. On the other hand, any conformational change altering the active sites will also affect the biochemical reaction. Thus, even if the nsSNP occurs close to the active site, the characteristics of the catalytic groups may be altered.[30–32] In such a case, the mutation may not completely abolish the biochemical reaction, but can change its kinetics and the native suitable cellular environment (pH, temperature), leading to a malfunctioning protein.[30–32] One example is the Snyder–Robinson syndrome (SRS) caused by three missense mutations (G56S, V132G, I150T) in the spermine synthase (SMS) gene that encodes a dimeric protein. Position G56 is at the periphery of the dimer binding interface, the V132 site is exactly at the dimer interface, and the I150 site is far away from the interface. Even if it is difficult to imagine any direct effect of the binding, almost any mutation at site I150 is expected to cause SRS. Site V132 is capable of having either disease-causing or harmless mutations. In terms of monomer stability and hydrogen-bond effect, site G56 is quite tolerable. Since the dimer formation is essential for the function of the SMS, such mutations are expected to be disease-causing.[30,31]

Solvent accessibility

A number of studies have shown that most of the deleterious nsSNPs affect protein function by disruption of the protein hydrophobic core.[9,10,12,20] Protein cores usually involve conserved residues, which are critical for protein stability

through hydrophobic contacts maintained with other residues.[33–36] The probability of the nsSNP being deleterious is increased when the volume, mass, and hydrophobicity difference between the original and mutated residues increases.[37] Therefore, introducing a hydrophobic residue on the surface could result in protein aggregation and, in turn, in a deleterious phenotype. Instead, a hydrophilic residue in the core protein would destabilize the folding processes due to vulnerable hydrogen side-chains that take part in hydrogen bonds.[38] The presence of a given nsSNP at solvent-accessible sites or at the interacting interface (protein–protein, protein–DNA) might alter interactions with other molecules or influence the protein solubility. Indeed, if the nsSNP is located in a highly conserved or a charged surface patch, possible alterations are of high biological significance.[9,10,27] In such cases, nsSNP effects are susceptible to alteration via binding to other molecules such as drugs.[35] Only a small number of nsSNPs are completely buried in the interior, and these are more likely to occur at conserved sites, whereas about 88% of disease-related nsSNPs are located in the surface pocket or an interior void of the protein, and do not exhibit the same tendency.[20]

Protein-folding, flexibility, and aggregation

Mechanisms of protein-folding are governed by the burial of side-chains inside the molecule, out of contact with water, and the formation of intramolecular interactions between amino-acid side-chains. Protein-folding involves the establishment of regular secondary structure, in which the protein's main structural domains (helices, strands, turns, and coils) are distributed through the most energetically favorable protein organization.[39] These structural domains, formed by hydrogen-bonding among main-chain polar groups and hydrogen bonds among side-chains, contribute to the stability of protein tertiary structure.[38] nsSNPs that modify the biophysical properties of residues at critical folding positions will have a profound impact on residue contacts and, thereby, may cause alterations in folding. Even if the mutant protein is correctly folded, it will be less stable or, if the folded mutant protein is conformationally fairly stable, in some cases it will be different enough to result in a dysfunctional protein.[40] As a consequence of the nsSNP, the most important conformational stabilizing forces such as hydrogen-bonding[36,38,41] and the hydrophobic[36,41,42] contacts may be disrupted, together with disulfide bonds[43] and van der Waals[42] and electrostatic[44] interactions. Furthermore, nsSNPs can also alter protein flexibility.[30,45]

When a protein performs its function, slight conformational changes often occur that are specifically suited for this particular function. If a mutation alters the protein's flexibility compared with its native structure, it will affect the protein's capacity to regulate these slight conformational changes, which are necessary for protein function.[30,45] Additionally, conformational flexibility is the principal mechanism affecting the tendency of the protein to aggregate.[46] The formation of aggregates is activated by the destabilization and opening of the native protein structure, which exposes aggregation-susceptible regions previously buried inside the structure. Aggregation involves the irreversible interaction of denatured protein molecules. The sequences of proteins determine – at least in part – aggregation propensities of proteins, which is why even slight alterations such as nsSNPs may have a considerable effect in the solubility and aggregation propensity of a protein.[46] Disease-related proteins have less-defined folds, isoelectric points closer to neutrality, contain more alternating hydrophilic/hydrophobic stretches compared to the average human protein, and have a higher tendency to aggregate.[47] Mutations that trigger protein aggregation are associated with molecular mechanisms involving amyloid formation being recognized as the cause of neurodegenerative disorders such as Alzheimer's disease.[48] Cystic fibrosis is an example of a genetic disease in which a variant protein is unable to fold correctly to a stable state in the endoplasmic reticulum, thereby failing to reach the cell membrane and be secreted in the quantity required for proper function. Other diseases, including some types of emphysema, result from mutations leading to the improper trafficking of proteins to the sites where they are needed.

Substitutions involving specific amino acids

About half of the mutated sites are involved in strong amino acid interactions; distributions of contact energy of residues are similar except for β-structures.[39] Due to their special characteristics, some residues can produce a distinct group of mutations. Glycine and proline residues are commonly found in regions in turns present in protein secondary structure, and their substitutions are expected to destabilize the protein or cause it to adopt a different fold.[39] Arginine and glycine are the most frequently mutated amino acids in secondary structures.[39] Tryptophan, tyrosine, and cysteine residues increase the chance of the nsSNPs being disease-related.[37,49]

Protein–protein, protein–DNA, and protein–membrane interactions

Cellular processes such as transcription, signal transduction, and transport, as well as most regulatory mechanisms, are mediated by protein–protein and protein–DNA interactions. Therefore, almost any change of an amino acid within the interface should affect the binding, often severely impacting the phenotype and raising the disease susceptibility.[50]

Binding recognition and specificity are altered by residue substitution involved in networking processes.[15,16] For instance, when substituting a small side-chain for a bulky side-chain in a narrow binding pocket, the entrance of the partner group will be blocked and the binding process will be completely or partially prevented.[16,31] Moreover, proteins with the largest number of interactions evolve slowly, suggesting that proteins with more interacting partners have a greater fraction of residues directly involved in their function. Hence, disease-associated nsSNPs are more likely to be conserved.[51,52] Protein–protein interfaces often contain binding hot spots – structurally conserved charged and polar residues surrounded by water – blocking hydrophobic residues that contribute mostly to binding.[53] Mutations that affect protein surface electrostatics may, thus, have diverse effects, ranging from changes in folding or stability, to alterations in partner/ligand binding affinity and specificity, and, in turn, protein function.[53] Similarly, nsSNPs at the protein–DNA interface can affect DNA regulation.[54] Those occurring at the protein–membrane interface can affect the signal processes across the membrane, the function of channels and pumps, and cellular adhesion.[55]

A typical example of disease caused by the disruption of protein-binding is Charcot-Marie-Tooth neuropathy, which can be triggered by the loss of interaction between myelin protein zero monomers that link adjacent membranes of the myelin sheath. In other cases, protein-binding is a means of allosteric regulation. To give an example, mutations in the binding interface of pantothenate kinase lead to inherited pantothenate kinase-associated neurodegeneration. Finally, there is also the possibility for mutations to change the binding specificity of a protein and thus lead to new and potentially disruptive interactions. Mutations in the family of human crystallin genes have been shown to alter affinity for binding partners. These erroneous interactions lead to congenital cataract.

Subcellular localization

Subcellular localization provides a specific environment for protein function and interactions with its biological partners for signaling pathways. A protein in a wrong subcellular location will have harmful effects on the other proteins functioning there and can affect DNA-transcription factors, resulting in altered expression of the corresponding protein, which in turn will disrupt the normal cell cycle, causing

diseases.[56] In about 1% of cases, the disease is caused by protein subcellular delocalization.[28] Indeed, a specific nsSNP within the signaling peptide – critical for protein transport across membranes – could result in a subcellular localization different from that of the native protein.[57,58] Even rare, severe inherited diseases are caused by nsSNPs in signal peptides, including familial hypoparathyroidism and coagulation factor X deficiency.[57,58]

Posttranslational modifications (PTMs)

PTMs are covalent modifications of side-chains in proteins after their translation. By adding functional groups (phosphorylation) to individual amino acids or by altering the chemical properties of amino acids (citrullination), PTMs create or disrupt covalent bonds to change the structure, localization, and function of proteins, playing an essential role in almost all cellular signaling and determining cellular dynamics and plasticity. A given nsSNP may abolish a PTM by replacing residues that cannot be modified, which leads to protein destabilization and changes in protein interactions, catalytic properties, or other protein functions.[59] Deleterious nsSNPs in PTM sites have been found in cancer[60] or immune deficiency[61] phenotypes.

In summary, alterations caused by nsSNPs on any of these features are likely to affect the structure and stability of the protein and, in turn, its function, contributing to disease susceptibility. Therefore, an accurate prediction of functional nsSNP effects and their consequences on related diseases is an important issue in improving drug discovery and drug development.

Approaches to predicting functional nsSNP effects

Many disease-causing protein nsSNPs have been characterized. Databases combined with available biochemical data on nsSNPs have encouraged the development of in silico methods to predict the effects of deleterious mutations on the corresponding proteins or assemblages. Most of them are based on a combination of physicochemical properties of amino acids, protein structure information, and evolutionary sequence conservation analysis.

Physicochemical properties-based methods

Amino-acid properties, such as size, flexibility and polarity side-chains, salt bridge and hydrogen-bond susceptibilities, and other geometrical considerations, play a crucial role in folding, stability, and function of proteins and in protein–protein interactions.[52] These amino-acid properties are dis-

tinguishable: therefore, the compatibility of a substitution can be used to predict its impact and to identify deleterious mutations.[11] A given nsSNP may change the physicochemical properties of the original residue to affect the protein stability and dynamics, or disrupt the interacting interface that prevents complex formation.[62]

Structure-based methods

Structure-based methods take an input sequence and find the best match against a protein structure database.[9,20,34] Because most predictive structure-based methods use general structural features surrounding the site of substitution and do not require detailed information at the atomic level, they can model the substitution onto the structure of a homologous protein rather than require the exact structure of the input sequence. These methods examine the position of nsSNPs and consider several structural properties such as solvent accessibility, secondary structure, or active-sites, and the difference in free energy between the exchanged amino acids.[9,10,21] If the nsSNP site is involved in ligand-binding, then the substitution may cause structural arrangements such as stability, affinity of receptor ligand complex, residue–residue contacts, and flexibility. Structural information is needed to fully understand the effects of mutations and disease phenotype.[63] Mapping of a specific nsSNP into the known three-dimensional (3D) structure can reveal whether the replacement is likely to have an impact on the normal folding or structure of the protein, or whether the amino-acid replacement destroys essential structure-maintaining contacts in the hydrophobic protein core, has a destabilizing impact on electrostatic interactions, or interacts with ligands.[9] Due to the difficulty in eliciting 3D structures experimentally, in silico approaches are useful in predicting the structural effects of functional nsSNPs.

Sequence homology-based methods

Sequence homology-based methods use evolutionary properties such as sequence conservation and phylogenetic trees. These methods calculate the probability of the substitutions based on multiple sequence alignments to identify evolutionarily related positions and distinguish between tolerant mutations and those involved in disease. The natural selection retains the more stable amino acids among species that are critical for protein function, stability, and interactions.[7,63] Among homologous proteins, the highly conserved residues are therefore evolving under strong selective pressure.[7] Such evolutionarily conserved residues are mapped on the representative structure. Based on the extraction of functionally important residues, these methods can identify active sites

and functional interfaces of proteins. Highly conserved and functionally important residues, such as those in active centers of enzymes as well as key residues for preservation of the protein stability, in particular those located in buried positions, have been shown to be frequent targets of disease-associated mutations.[49] By using disease-associated mutation data and multiple species of phylogenetic lineage, it has been shown that deleterious nsSNPs occur more frequently in evolutionarily conserved positions.[47,63] Genetic variants fixed during evolution may have selective effects for the organism, they may be neutral, or they may be deleterious and cause pathology. About 26%–32% of naturally occurring mutations have been predicted to affect protein function.[34]

Most functional nsSNP prediction methods are based on sequential or structural features.[14] However, the methods based on homology sequence can lead to incorrect annotations because there is no simple relationship between measures of sequence similarity and protein function. Structural analysis has been demonstrated to be more accurate, but studies have also reported ambiguous results from similarities based on structural comparisons alone.[9–12] Similar proteins can catalyze distinct reactions and different proteins can catalyze identical chemical reactions. Sequence analysis has proven to be valuable and can provide accurate functional inference for proteins sharing more than 70% sequence identity. More reliable is functional assignment using structure when a structural homology can be identified and active sites directly compared. Protein function is commonly associated with a particular 3D disposition of residues involved in binding and/or catalysis. Knowledge of protein 3D structure provides crucial insight into the function and activity of the protein. However, the structures of less than 1% of sequences have been experimentally solved. To face these limitations, numerous studies demonstrate that deleterious nsSNPs and their impact on protein function can be predicted by combined analysis of multiple sequence alignments and protein structural information.[9–12,21,34]

An increasing number of in silico approaches to discriminate between deleterious nsSNPs leading to a protein disorder and neutral polymorphisms that do not modify the phenotype have been implemented into webservers. The most representative are described below:[9,13,17,23,27–29,64–68]

1. Sorting Intolerant From Tolerant (SIFT) uses sequence homology to predict whether an amino acid substitution affects protein function and, hence, potentially alter phenotypes.
2. Polymorphism Phenotyping (PolyPhen) is a multiple sequence alignment webserver. The prediction is made through sequence-based characterization of the substitution site, calculation of position-specific independent count (PSIC) profile scores for two amino-acid variants, and calculation of structural parameters and contacts. It combines a variety of features such as sequences, evolutionary properties, and 3D structural information to predict if an nsSNP will affect the protein function and performs optimally if the structural information is available. More than 11,000 nsSNPs are annotated by this webserver. PolyPhen maps the amino-acid substitution to the known 3D structure of the protein to examine whether the substitution might destroy the protein's hydrophobic core, electrostatic interactions or interactions with ligands, or other important features of a protein based on the analysis of several structural parameters, and also on the analysis of several contact parameters.
3. SNPs3D is a website that assigns molecular functional effects of nsSNPs based on structure and sequence analysis. It provides various disease/gene relationships at the molecular level and models of gene–pathway–disease interaction. This server identifies the gene candidates involved in a specific disease, correlates the sets of candidate genes, and analyses the possible effects of nsSNPs on normal protein function.
4. SNPeffect 4.0 is another online tool that focuses on the molecular characterization and annotation of disease and polymorphism variants in the human proteome. It uses a combination of sequence- and structure-based bioinformatics tools that improve the accuracy of local nsSNPs functionality prediction.
5. Large-scale annotation of coding nsSNPs (LS-SNP) maps nsSNPs onto protein sequences, functional pathways, and comparative protein structure models and predicts the positions where nsSNPs cause the effects. The results can be used to find out the functional SNP candidates within a gene, haplotype, or pathway.
6. Structure SNP (StSNP) is a webserver for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways, which allows for examination of possible disease-related pathways associated with a particular nsSNP and linking of diseases with the current available molecular structure data.
7. AUTO-MUTE is a knowledge-based computational mutagenesis used to predict the disease potential of human nsSNPs.

These methods differ in the nature and properties of the variant they take into account. Some of them are based on evolutionary information. Others combine protein structural

and/or functional parameters. Information about protein 3D structures would increase prediction performance. The very best methods use also structural and functional information, whereas others that are solely based on sequence-level information perform rather well. These studies[9,13,17,23,27-29,64–68] compare the methods pair-wise, and whether the type of original or substituting amino acid residue, the structural class of the protein, or the structural environment of the amino acid substitution, had an effect on the prediction performance. Considering all the evaluation measures, no single method could be rated as the best.

Modern drug-discovery process

The endpoint of identification of disease-associated nsSNPs should be functional analysis, an understanding of the molecular mechanism of causation of the disease phenotype, and the development of drugs to target these mutants. Certain nsSNPs are associated with significant changes in drug efficacy and drug disposition. Given that these nsSNPs could occur in candidate drug-target proteins, they can influence treatment response and whether drugs produce adverse reactions.

The success of a drug is determined by the balance of target potency, selectivity, and ADMET profile.[69] Traditionally, drugs were discovered by synthesizing compounds in a time-consuming multi-step process against a battery of in vivo biological screens, then further investigating the candidates for their ADMET properties. Novel compounds entering clinical trial undergo years of rigorous preclinical testing, but only about 8% reach the market.[70] It has been estimated that 30% of compounds fail to show efficacy, while 50% of active compounds are removed from the pipeline due to toxicity.[71] Thus, early evaluation of ADMET properties has become mandatory to increase success in clinical studies. There are many factors that impede early determination of these factors, including large amounts of compound required for the in vivo studies, lack of reliable high-throughput in vitro assays, and inability of animal models to predict some human toxicity. Today, the process of drug discovery has been revolutionized by the advent of genomics, proteomics, bioinformatics, and efficient technologies such as combinatorial chemistry, high-throughput screening (HTS), virtual screening, de novo design, in silico ADMET screening, and structure-based drug design.

The drug-discovery process operates on a target-based approach in which a developed drug is designed to affect only specific genes and molecular mechanisms in order to selectively treat the disease without producing side effects. In silico techniques are used to simulate the chemical compound and design chemical structures that might work against it. These tools are helpful at any of the following stages of modern drug design process:

1. Target identification. Drug discovery begins with the identification of the function of a potential drug target and its role in the disease.[72] A drug target is a key molecule, such as a gene or protein, which is involved in a signaling pathway specific to a disease condition. Drugs may be designed that bind to the active region and inhibit this molecule, or may enhance the normal pathway by promoting specific molecules that may have been affected during disease development.[72,73] The specific molecular targets and groups of patients are identified by bioinformatics approaches such as homology-based, ligand-base, structures-based, and HTS technologies.

2. Target validation. After a drug target has been identified, it is necessary to demonstrate that it will yield the desired clinical outcome, specifically the improvement or elimination of a phenotype. In silico characterization can be performed by using approaches such as genetic-network mapping, protein-pathway mapping, disease-locus mapping, protein–protein interactions, and subcellular localization predictions.[74]

3. Hit-to-lead identification phase assures that the identified drug possesses favorable ADMET properties, improved potency, undesirable activities at other biological targets, and physiochemical/metabolic properties suggestive of reasonable in vivo pharmacokinetics.[74] The hit-to-lead phase can be identified by approaches like high-throughput biochemical, cellular, and natural product assays, structure-based design, and virtual HTS (vHTS).[74]

4. Lead optimization is the process of refining the chemical structure of a confirmed hit to improve its drug characteristics. Lead structures are optimized for target affinity and selectivity. Docking techniques are currently applied to aid in structure-based ADMET.[75]

5. Preclinical studies with and without the use of animals have the purpose of limiting risks whenever a new active substance is to be used as a human medicine. Preclinical testing involves: pharmacology, toxicology, preformulation, formulation analysis, and pharmacokinetics.[75]

6. Clinical trials are the fastest and safest way to find treatments that work in humans. Patients with predetermined characteristics are recruited, and data is collected on their health for a defined time period. Clinical trials are organized into prevention, screening, diagnostics, treatment, and quality of life.

7. The New Drug Application (NDA) and Food And Drug Administration (FDA) approval. The role of the NDA is to provide information from previous years of work, and submit manufacturing and labeling proposals of new pharmaceuticals, the goal of which is to gain FDA approval by demonstrating that the medicine is safe and effective for humans.[70]

In silico approaches to the drug discovery process

nsSNPs play important roles in the diverse responses in efficacy and toxicity of the human population to therapeutic agents by affecting drug-target proteins such as G-protein coupled receptors, enzymes, ion channels, and proteins involved in the detoxification pathways. In silico methods such as SIFT and PolyPhen allow screening for polymorphisms of various genes that may potentially cause diseases and altered drug-response or toxicity. Therapeutic targets of many disease-causing nsSNPs can be investigated by a computational approach with important implications for potential target discovery. Furthermore, the study of current therapeutic targets of drugs can lead to a new direction for future target identification. The identification of new therapeutic targets and the exploration of novel drugs show a promising future for the use of bioinformatics tools. This is valuable for the drug discovery and development process and, ultimately, for improving drug efficacy and safety profiles. The drug discovery process from identification of new selective, effective, and safe compounds to the development of drug candidates in clinical trials can be developed through a variety of in silico approaches such as ligand- and structure-based drug design, de novo design, and homology modeling, depending on how much information is available about drug targets and the potential drug.[76,77]

Virtual screening and de novo design approaches

These approaches use computer methods to select a set of new ligands on the basis of structures that are predicted to exhibit a biological activity on a given target. Virtual screening is divided into structural- and ligand-based methods.[78,79] The assumption of the ligand-based approach is that similar structures have similar biological activity.[80] These methods operate only on known active ligands and therefore they are adequate if the structure of a protein is not known. Docking is a computational tool of structure-based drug use to predict protein–ligand interaction geometries and binding affinities. It consists of a docking, which is the search pattern

for identifying suitable conformations, and a score, which is a measure of the affinity of various conformations.[81,82] Notable docking softwares are GOLD (The Cambridge Crystallographic Data Centre, Cambridge, UK), AutoDock (The Scripps Research Institute, La Jolla, CA), and DOCK (http://dock.compbio.ucsf.edu/). A complete webserver is DockingServer (Virtua Drug Ltd, Budapest, Hungary).

De novo design methods

These are automated computational procedures that create new molecules with drug-like properties by using genetic algorithms to improve the specified properties of the molecules.[83] De novo design is complementary to HTS[83] and, instead of searching for bioactive molecules in large collections of physically available compounds, de novo design creates chemical structures from scratch by assembling molecular fragments. Most approaches to de novo design consist of virtual synthesis of molecules assembled from fragments, evaluation of their biological activity by computational function, and optimization of the generated compounds. De novo design methods can be performed either by computing some similarity index of candidate compounds and known reference ligands (ligand-based approach) or be based on the 3D structure of a ligand-binding cavity (receptor-based approach).[83,84] The synthesis of combinatorial libraries is another important strategy in drug discovery that may be achieved with docking programs (LEA3D [Institut de Pharmacologie Moléculaire et Cellulaire, Valbonne, France], PLANTS [http://www.tcd.uni-konstanz.de/research/plants.php], FlexX [BioSolveIT GmbH, Sankt Augustin, Germany]) focused on building combinatorial libraries of molecules and performing virtual screenings.[78,79] As an alternative to docking methods and as a complementary modeling method to 3D structures at atomic level, pharmacophore is the spatial arrangement of features essential for a molecule to interact with a specific drug target receptor in a specific binding mode.[85,86] The PharmMapper webserver identifies potential drug targets from the pharmacophore database (namely PharmTargetDB) annotated from all the target information in TargetBank, BindingDB, DrugBank, and potential drug target databases, which provide information about disease-associated genes, genetic variations, and drugs or 3D structural models of drugs.[85,86]

Homology modeling

This refers to constructing an atomic-resolution model of the target protein from its amino-acid sequence and the experimental 3D structure of a related homologous protein.

It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structures.[87,88]

DNA microarrays

These are used to measure the expression patterns of thousands of genes in parallel, generating clues to gene function that can help to identify appropriate targets for therapeutic intervention. They can also be used to monitor changes in gene expression in response to drug treatments.[89]

vHTS

vHTS examines protein targets against databases of small-molecule compounds and chooses molecules that bind strongly to the target. If there is a "hit" with a particular compound, it can be extracted from the database for further testing.[88,90]

Drug lead optimization

This usually involves a series of modifications to the sequence and secondary structure of the potential drug. This process can be enhanced using software tools that explore related compounds to the lead candidate. The estimation of the activity is addressed by various ligand- and structure-based methods depending on the data on known ligands and the availability of the 3D structure of the target.[78,79]

Drug bioavailability and bioactivity (ADMET profile)

These are key characteristics of drugs. Traditionally, quantitative structure–activity relationships (QSAR) have been used to investigate the molecular features that influence the ADMET drug profile.[91] These methods use statistical approaches and experimental data to model complex biological processes.[92] As the 3D structures of major ADMET proteins become available, structure-based methods can be used to complement QSAR studies.[92] Several recent studies have shown how in silico ADMET predictions of drugs can minimize the number of compounds needed to be synthesized to obtain the required biochemical and/or physicochemical profile.[91,92]

Advances in genomics have triggered a shift in drug discovery, from strong single-target interaction to more global and comparative analysis of multi-target networks.[76] Numerous research works[72,88,91,92] have demonstrated that in silico methods offer the advantage of delivering new drug candidates more quickly and at a lower cost, with great potential across the pharmaceutical industry and also in other industries, such as consumer goods, where non-animal alternatives are being sought for assuring the safety of drugs.

Conclusion and future developments

Variations in protein sequence and function are mainly due to nsSNPs. The fraction of nsSNPs in the genome is relatively low (~10% of coding SNPs) but has a profound impact on the function of associated genes. About 30% of the nsSNPs constitute a set of genetic factors associated with disease predisposition. Most known disease-related nsSNPs in proteins of known 3D structure appear to affect important sites relevant for function, like ligand-binding surfaces and catalytic and regulatory sites. Structural mutations affect buried residues in the protein core, causing changes in amino-acid size and charge, hydrogen bonds, salt bridges, and S–S bridges. These changes cause loss of thermodynamic stability as well as aberrant folding and aggregation of proteins.

Certain nsSNPs are known to be associated with significant changes in drug efficacy and drug disposition. Given that these nsSNPs could be located in candidate drug-target proteins, they can influence treatment response and whether drugs produce adverse reactions. Variation in drug response is attributed to multiple genes rather than a single-gene mutation; therefore, it appears appropriate to perform pharmacogenomic studies comparing nsSNP maps and gene expression between normal and affected individuals. Pharmacogenomic screening of selected polymorphisms in clinical trial subjects should also allow clinical trials to be smaller, faster, and less expensive because they could be subjects from clinical trials whose genetic profile would result in ADRs or ineffective responses to a drug under evaluation. In silico pharmacogenomics is already identifying new mechanisms of drug action and novel therapeutic targets. Indeed, this field represents a powerful instrument with which to identify the relationship between genetics and chronic pain states. One example is the relationship between β-adrenergic receptors and opioid-induced hyperalgesia and the identification of novel targets of analgesic action, such as an inward rectifying potassium channel, which modulates analgesic response to multiple drugs. The utility of pharmacogenomics in drug therapy also manifests in its potential to translate into individualized medicine, drug development, and drug regulation which, like pharmacogenomics itself, need to cope with individual variability in drug therapy and which are only at the beginning of meeting this difficult and complex challenge.

Therefore, knowledge of effect of functional nsSNPs could improve drug therapy by optimizing efficacy and diagnostics and

reducing toxicity of drugs. The greatest challenge for the future is to understand the genotypic–environmental factor interaction, ethnicity, inheritance patterns in drug response, and how genetic variance responds to medicine. If the goals of pharmacogenomics are satisfied, physicians may genetically subdivide patients and treat each patient group according to their genetic profile, thereby prescribing appropriate medications aimed at the right target at the right dose, to achieve maximal therapeutic benefit with minimal, tolerable adverse effects.

Acknowledgments

Disclosure

The author reports no conflicts of interest in this work.

References

1. Shastry BS. Pharmacogenetics and the concept of individualized medicine. *Pharmacogenomics J*. 2006;6:16–21.
2. Ma Q, Lu AY. Pharmacogenetics, pharmacogenomics, and individualized medicine. *Pharmacol Rev*. 2011;63(2):437–459.
3. Pharmacovigilance: ensuring the safe use of medicines. In: *WHO Policy Perspectives on Medicines*. Geneva: World Health Organization; 2004. Available from: http://apps.who.int/medicinedocs/pdf/s6164e/s6164e. pdf. Accessed.
4. Suvarna BS. Role of pharmacogenomics in the drug development. *Kathmandu Univ Med J (KUMJ)*. 2009;7(2–26):172–176.
5. Shastry BS. Role of SNPs and haplotypes in human disease and drug development. In: Ozkan M, Heller MJ, Ferrari M, editors. *Micro/ Nano Technology in Genomics and Proteomics*. New York: Springer; 2006;2:447–458.
6. Pang GS, Wang J, Wang Z, Lee CG. Predicting potentially functional SNPs in drug-response genes. *Pharmacogenomics*. 2009;10(4): 639–653.
7. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002;30:3894–3900.
8. González-Castejón M, Marín F, Soler-Rivas C, Reglero G, Visioli F, Rodríguez-Casado A. Functional non-synonymous polymorphisms prediction methods: current approaches and future developments. *Curr Med Chem*. 2011;18:5095–5103.
9. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P. Prediction of deleterious human alleles. *Hum Mol Genet*. 2001;10(6): 591–597.
10. Wang Z, Moult J. SNPs, protein structure, and disease. *Hum Mutat*. 2001;7:263–270.
11. Yue P, Moult J. Identification and analysis of deleterious human SNPs. *J Mol Biol*. 2006;356(5):1263–1274.
12. Ng PC, Henikoff S. Predicting the effects of amino-acid substitutions on protein function. *Annu Rev Genomics Hum Genet*. 2006;7:61–80.
13. Masso M, Vaisman II. Knowledge-based computational mutagenesis for predicting the disease potential of human non-synonymous single-nucleotide polymorphisms. *J Theor Biol*. 2010;266(4):560–568.
14. Dimmic MW, Sunyaev S, Bustamante CD. Inferring SNP function using evolutionary, structural, and computational methods. *Pac Symp Biocomput*. 2005:382–384.
15. Ung MU, Lu B, McCammon JA. E230Q Mutation of the catalytic subunit of cAMP-dependent protein kinase affects local structure and the binding of peptide inhibitor. *Biopolymers*. 2006;81:428–439.
16. Rignall TR, Baker JO, McCarter SL, et al. Effect of single-active-site cleft mutation on product specificity in a thermostable bacterial cellulase. *Appl Biochem Biotechnol*. 2002;98–100:383–394.
17. Erdin S, Ward RM, Venner E, Lichtarge O. Evolutionary trace annotation of protein function in the structural proteome. *J Mol Biol*. 2010; 396(5):1451–1473.
18. Dean PM, Lewis RA, editors. *Molecular Diversity in Drug Design*. Dordrecht: Kluwer Academic Publishers; 1999.
19. Bao L, Zhou M, Cui Y. nsSNP analyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res*. 2005:W480–W482.
20. Stitziel NO, Tseng YY, Pervouchine D, Goddeau D, Kasif S, Liang J. Structural location of disease-associated single-nucleotide polymorphisms. *J Mol Biol*. 2003;327(5):1021–1030.
21. Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single-nucleotide polymorphisms. *Trends Genet*. 2000;16(5):198–200.
22. Sunyaev SR, Lathe III WC, Ramensky VE, Bork P. SNP frequencies in human genes: an excess of rare alleles and differing modes of selection. *Trends Genet*. 2000;16(8):335–337.
23. Yue P, Melamud E, Moult J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics*. 2006;7:166.
24. Cheng TM, Lu YE, Vendruscolo M, Lio P, Blundell TL. Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single-nucleotide polymorphisms. *PLoS Comput Biol*. 2008;4:e1000135.
25. Frederic MY, Lalande M, Boileau C, et al. UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity-application to four genes: FBN1, FBN2, TGFBR1, and TGFBR2. *Hum Mutat*. 2009;30:952–959.
26. Bromberg Y, Rost B. SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res*. 2007;35(11):3823–3835.
27. Ryan M, Diekhans M, Lien S, Liu Y, Karchin R. LS-SNP/PDB: annotated non-synonymous SNPs mapped to Protein Data Bank structures. *Bioinformatics*. 2009;25(11):1431–1432.
28. Reumers J, Maurer-Stroh S, Schymkowitz J, Rousseau F. SNPeffect v2.0: a new step in investigating the molecular phenotypic effects of human non-synonymous SNPs. *Bioinformatics*. 2006;22(17):2183–2185.
29. Wang P, Dai M, Xuan W, et al. SNP function portal: a web database for exploring the function implication of SNP alleles. *Bioinformatics*. 2006;22(14):E523–E529.
30. Zhang Z, Teng S, Wang L, Schwartz CE, Alexov E. Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Hum Mutat*. 2010;31(9):1043–1049.
31. Zhang Z, Norris J, Schwartz C, Alexov E. In-silico and in vitro investigations of the mutability of disease-causing missense mutation sites in spermine synthase. *PLoS One*. 2011;6(5):Article ID e20373.
32. Alexov E. Numerical calculations of the pH of maximal protein stability: the effect of the sequence composition and three-dimensional structure. *Eur J Biochem*. 2004;271(1):173–185.
33. Loladze VV, Ermolenko DN, Makhatadze GI. Thermodynamic consequences of burial of polar and non-polar amino-acid residues in the protein interior. *J Mol Biol*. 2002;320:343–357.
34. Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single-nucleotide polymorphisms: structure-based assessment of amino-acid variation. *J Mol Biol*. 2001;307:683–706.
35. Teng S, Michonova-Alexova E, Alexov E. Approaches and resources for prediction of the effects of non-synonymous single-nucleotide polymorphism on protein function and interactions. *Curr Pharm Biotechnol*. 2008;9(2):123–133.
36. Dill KA. Dominant forces in protein folding. *Biochemistry*. 1990;29(31): 7133–7155.
37. Dobson RJ, Munroe PB, Caulfield MJ, Saqi MAS. Predicting deleterious nsSNPs: an analysis of sequence and structural attributes. *BMC Bioinformatics*. 2006;7(1):217.
38. Pace CN, Fu H, Fryar KL, et al. Contribution of hydrophobic interactions to protein stability. *J Mol Biol*. 2011;408(3):514–528.

39. Khan S; Vihinen M. Spectrum of disease-causing mutations in protein secondary structures. *BMC Struct Biol*. 2007;7(1):1–18.

40. Thomas PJ, Qu BH, Pedersen PL. Defective protein folding as a basis of human disease. *Trends Biochem Sci*. 1995;20:456–459.

41. Dill KA, Ozkan SB, Weikl TR, Chodera JD, Voelz VA. The protein folding problem: when will it be solved? *Curr Opin Struct Biol*. 2007;17(3):342–346.

42. Eriksson AE, Baase WA, Zhang XJ, et al. Response of a protein structure to cavity creating mutations and its relation to the hydrophobic effect. *Science*. 1992;255:178–183.

43. Betz SF. Disulfide bonds and the stability of globular proteins. *Protein Sci*. 1993;2(10):1551–1558.

44. Horovitz A, Serrano L, Avron B, Bycroft M, Fersht AR. Strength and cooperativity of contributions of surface salt bridges to protein stability. *J Mol Biol*. 1990;216:103–144.

45. Karplus M, Kuriyan J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A*. 2005;102(19):6679–6685.

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

47. Wong P, Fritz A, Frishman D. Designability, aggregation propensity and duplication of disease-associated proteins. *Protein Eng Des Sel*. 2005;18:503–508.

48. Keage HA, Carare RO, Friedland RP, et al. Population studies of sporadic cerebral amyloid angiopathy and dementia: a systematic review. *BMC Neurol*. 2009;9:3.

49. Vitkup D, Sander C, Church G. The amino-acid mutational spectrum of human genetic disease. *Genome Biol*. 2003;4(11):R72.

50. Gong S, Blundell TL. Structural and functional restraints on the occurrence of single-amino-acid variations in human proteins. *PLoS ONE*. 2010;5(2):e9186.

51. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL. The human disease network. *Proc Natl Acad Sci U S A*. 2007;104(21): 8685–8690.

52. Teng S, Madej T, Panchenko A, Alexov E. Modeling effects of human single-nucleotide polymorphisms on protein-protein interactions. *Biophys J*. 2009;96(6):2178–2188.

53. Ma B, Elkayam T, Wolfson H, Nussinov R. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A*. 2003;100:5772–5777.

54. Wright JD, Lim C. Mechanism of DNA-binding loss upon single-point mutation in p53. *J Biosci*. 2007;32(5):827–839.

55. Kwa LG, Wegmann D, Brugger B, Wieland FT, Wanner G, Braun P. Mutation of a single-residue, β-glutamate-20, alters protein-lipid interactions of light harvesting complex II. *Mol Microbiol*. 2008;67(1): 63–77.

56. Hanemann CO, D'Urso D, Gabreëls-Festen AA, Müller HW. Mutation-dependent alteration in cellular distribution of peripheral myelin protein 22 in nerve biopsies from Charcot-Marie-Tooth type 1A. *Brain*. 2000;123(5):1001–1006.

57. Laurila K, Vihinen M. Prediction of disease-related mutations affecting protein localization. *BMC Genomics*. 2009;10:122.

58. Moosawi F, Mohabatkar H. Computer-assisted analysis of subcellular localization signals and post-translational modifications of human prion proteins. *J Biomedical Sci Eng*. 2009;2:70–75.

59. Ryu GM, Song P, Kim KW, Oh KS, Park KJ, Kim JH. Genome-wide analysis to predict protein sequence variations that change phosphorylation sites or their corresponding kinases. *Nucleic Acids Res*. 2009;37(4): 1297–1307.

60. Radivojac P, Baenziger PH, Kann MG, Mort ME, Hahn MW, Mooney SD. Gain and loss of phosphorylation sites in human cancer. *Bioinformatics*. 2008;24(16):i241–i247.

61. Vogt G, Vogt B, Chuzhanova N, Julenius K, Cooper DN, Casanova JL. Gain-of-glycosylation mutations. *Curr Opin Genet Dev*. 2007;17: 245–251.

62. Grantham R. Amino-acid difference formula to help explain protein evolution. *Science*. 1974;185:862–864.

63. Bao L, Cui Y. Prediction of the phenotypic effects of non-synonymous single-nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics*. 2005;21(10):2185–2190.

64. Lichtarge O, Sowa ME. Evolutionary predictions of binding surfaces and interactions. *Curr Opin Struct Biol*. 2002;12(1):21–27.

65. Ben-Shimon A, Eisenstein M. Computational mapping of anchoring spots on protein surfaces. *J Mol Biol*. 2010;402(1);259–277.

66. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4): 248–249.

67. Flanagan SE, Patch AM, Ellard S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers*. 2010;14(4):533–537. Erratum in: *Genet Test Mol Biomarkers*. 2010;14(5):730.

68. De Baets G, van Durme J, Reumers J, et al. SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants. *Nucleic Acids Res*. 2012;40:D935–D939.

69. Gleeson MP, Hersey A, Montanari D, Overington J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov*. 2011;10:197–208.

70. Marchetti S, Schellens JH. The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials. *Br J Cancer*. 2007;97:577–581.

71. Wang J, Urban L. The impact of early ADME profiling on drug discovery and development strategy. *Drug Discovery World*. 2004;5:73–86.

72. Kermani BG. Pharmacogenomics. In: Markel S, León D, editors. *Silico Technologies in Drug Target Identification and Validation*. Vol 6 of Drug Discovery Series. Boca Raton, FL: CRC Press, Taylor & Francis Group, 2006:323–343.

73. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol*. 2011;162(6):1239–1249.

74. Goodnow RA. Hit and lead identification: integrated technology-based approaches. *Drug Discov Today Technol*. 2006;3:367–375.

75. Olejniczak K, Guenzel P, Bass R. Preclinical testing strategies. *Drug Inf J*. 2001;35:321–336.

76. Bajorath J. Computational analysis of ligand relationships within target families. *Curr Opin Chem Biol*. 2008;12:352–358.

77. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today*. 2006;11:580–594.

78. Wilson GL, Lill MA. Integrating structure-based and ligand-based approaches for computational drug design. *Future Med Chem*. 2011; 3(6):735–750.

79. Sukumar N, Das S. Current trends in virtual high throughput screening using ligand-based and structure-based methods. *Comb Chem High Throughput Screen*. 2011;14(10):872–888.

80. Tresadern G, Bemporad D. Modeling approaches for ligand-based 3D similarity. *Future Med Chem*. 2010;2(10):1547–1561.

81. Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH. Structure-based virtual screening for drug discovery: a problem-centric review. *AAPS J*. 2012;14(1):133–141.

82. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*. 2011;7(2):146–157.

83. Hartenfeller M, Schneider G. De novo drug design. *Methods Mol Biol*. 2011;672:299–323.

84. Boda K, Johnson AP. Molecular complexity analysis of de novo designed ligands. *J Med Chem*. 2006;49:5869–5879.

85. Wolber G, Seidel T, Bendix F, Langer T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov Today*. 2008;13(1–2):23–29.

86. Steindl M, Schuster D, Wolber G, Laggner C, Langer C. High throughput structure-based pharmacophore modeling as a basis for successful parallel virtual screening. *J Comput Aided Mol Des*. 2006;20:703–715.

87. Wieman H, Tøndel K, Anderssen E, Drabløs F. Homology-based modeling of targets for rational drug design. *Mini Rev Med Chem*. 2004;4:793–804.

88. Satyajit D, Sovan S, Kapil S. Computer aided drug design – a new approach in drug design and discovery. *Int J Pharm Sci Rev Res*. 2010;4(3):146.

89. Liu Y, Verducci JS. Medicinal chemistry and drug design. In: Ekinci D, editor. *Microarray Analysis in Drug Discovery and Biomarker Identification*. InTech; 2012.

90. Subramaniam S, Mehrotra M, Gupta D. Virtual high throughput screening (vHTS) – A perspective. *Bioinformation*. 2008;3(1):14–17.

91. Gleeson MP, Hersey A, Hannongbua S. In-silico ADME models: a general assessment of their utility in drug discovery applications. *Curr Top Med Chem*. 2011;11:358–381.

92. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. Towards in silico structure-based ADMET prediction in drug discovery. *Drug Discov Today*. 2012;17(1–2):44–55.