DRUG DISCOVERY TODAY BIOSILICO

Computer prediction of drug resistance mutations in proteins

Zhi Wei Cao, Lian Yi Han, Chan Juan Zheng, Zhi Lang Ji, Xin Chen, Hong Huang Lin and Yu Zong Chen

Drug resistance is of increasing concern in the treatment of infectious diseases and cancer. Mutation in drug-interacting disease proteins is one of the primary causes for resistance particularly against anti-infectious drugs. Prediction of resistance mutations in these proteins is valuable both for the molecular dissection of drug resistance mechanisms and for predicting features that guide the design of new agents to counter resistant strains. Several protein structure- and sequence-based computer methods have been explored for mechanistic study and prediction of resistance mutations. These methods and their usefulness are reviewed here.

Zhi Wei Cao
Lian Yi Han
Chan Juan Zheng
Xin Chen
Hong Huang Lin
Yu Zong Chen
Bioinformatics and Drug Design Group, Department of Computational Science, National University of Singapore, BLK SOC 1, level 7, Singapore 117543
*e-mail: csccyz@nus.edu.sg
Zhi Lang Ji
Department of Biology, School of Life Sciences, Xiamen University, Xiamen 361000, Fujian Province, PR China

Drug resistance is a global public health problem particularly for the treatment of infectious diseases [1,2] and cancer [3]. It has become increasingly possible for cross country transmission of drug-resistant organisms. There is an urgent need to develop resistance-evading drugs. Several mechanisms are responsible for drug resistance [4–8]. For infectious diseases, resistance is primarily mediated by mutations in the genes of infectious organisms that alter a drug's interaction with its corresponding target protein [4,5].

Considerable effort has been directed at the use of computational methods for studying the molecular mechanism of mutation-induced drug resistance and for developing predictive tools. The availability of the three-dimensional structure of drug targets involved in disease enables the use of molecular modeling and other structure-based approaches for evaluation of structural features, molecular interactions, solvation and dynamical properties of drug-protein binding and their correlation to resistance mutations [9–13]. Structure-derived binding energies [9,14–16] and binding-site volume-based fitness models [17] have also been used for facilitating the prediction of resistance mutations.

Structural information is available for a relatively small percentage of proteins. Thus methods for predicting resistance mutations directly from sequences are highly useful and are being developed. Interpretation programs have emerged for identifying and estimating the level of resistance [18,19]. Statistical learning methods such as neural networks [20,21], support vector machines (SVM) [22] and decision tree [23] have also shown promising potential for predicting resistance mutations.

Molecular modeling of drug resistance mutations

Structural analysis of proteins that contain resistance mutations indicated that mutations at drug-binding sites usually alter the tight packing between the binding drug and its receptor without substantial change in overall conformation [4,24]. Comparison of mutant and wild-type drug-receptor structures showed that, in the majority of cases, the only apparent change is in the pattern of local contact and hydrogen bonding at a mutation site [24,25]. Hydrophobic effects have been found to be important in several cases [10], but variation in local packing

1359-6446/04/$ – see front matter PII: S1359-6446(04)03377-5

interactions appears to be a major factor for the reduced drug binding affinity leading to resistance [4]. Thus molecular mechanics, molecular dynamics simulation, and Monte Carlo simulation methods are expected to be useful for structural optimization and structure-based energetic analysis of these mutations.

Molecular mechanics methods use atom–atom interaction energies for structural optimization and energetic analysis of a ligand–protein complex [9,14,15]. Bonded interactions are modeled by bond stretch, angle bending, and torsion terms. Non-bonded interactions are modeled by van der Waals and electrostatic terms. Hydrogen bonds can either be modeled by a separate term [14] or they can be included in van der Waals and electrostatic terms [9,15]. Moreover, solvation and entropic effects may be considered by using a simple solvation free energy model [14,26] and side-chain entropy model [15] respectively.

Structural optimization involves the selection of low energy molecular structures. One approach is to search molecular conformations for identifying the structure of the lowest energy. The other is energy minimization such that molecular structure is varied towards the direction of lower energy until it reaches the local minimum energy configuration. In most cases, ligand-protein binding is determined by non-bonded, hydrogen bond, solvation and side-chain torsion interactions. Thus ligand-binding affinities are frequently estimated by using these terms [9,14,15].

Molecular dynamics simulation methods derive trajectories of atomic positions of molecular motions and dynamical fluctuations by solving Newton's equations governed by the same sets of bonded and non-bonded interactions used in molecular mechanics methods. Solvation effect is described by using either explicit water molecules or continuous medium models. Entropic effect is derived from statistical analysis of trajectories of atomic positions. These methods are useful for structural optimization, motional and energetic analysis, and free energy computation [10–13].

The free energy difference between wild-type and mutant systems, which is useful for indicating drug resistance mutations, can be derived by using Monte Carlo simulation [27] as well as molecular dynamics simulation [10–13]. Monte Carlo simulation methods generate a series of molecular structures that are randomly distributed in the molecular conformational space and conform to a certain distribution pattern governed by the laws of statistical mechanics. These randomly generated structures can then be used to derive free energy difference between the wild-type and mutant structure by using the free energy perturbation method [27].

The structure of a ligand-protein complex and its mutants often needs to be modeled from a template. Such a template is usually the structure of a different mutational variant of the same protein or its complex with a different ligand [9–16]. The modeled structures of both wild-type and mutant HIV-1 protease–inhibitor complexes have been found to be consistent with the crystallographic structures, with root mean square differences ranging from 0.5A to 1.2A [9,14,15], which suggests that the quality of these modeled structures reaches the level useful for facilitating structure-based study and prediction of resistance mutations.

The ligand-protein binding energies computed from molecular mechanics and the free energies computed from molecular dynamics for several wild-type and mutant ligand–protein complexes showed significant correlation with the observed binding affinities [9–16,26], indicating that these energy functions are useful for facilitating energetic analysis and the prediction of resistance mutations.

Structure-based prediction of drug resistance mutations

Structure-base virtual screening has been widely used for designing new drugs [28,29]. To achieve high-speed screening of a large number of compounds, efficient computational procedures have been routinely applied to structural optimization and scoring of docked ligand-protein structures [30–34]. Some of these procedures are very similar to those used for the molecular study of drug resistance mutations [9,13–16], which raises the possibility of using virtual screening approaches to indicate possible drug resistance mutations [14]. While more accurate in modeling resistance mutations than those used in virtual screening studies [10–13], a full molecular dynamics procedure has yet to be employed in a general virtual screening process partly owing to its computationally intensive nature. Thus, procedures that either use molecular mechanics alone or molecular mechanics plus a small run of molecular dynamics for structural optimization has been the primary choice for structure-based prediction of resistance mutations in a virtual screening process [9,14].

Structural models and energy functions

Most models of mutant drug–protein complexes are based on a starting crystal structure of the wild-type protein complexed with the same drug [9,14,15]. Each mutation is introduced by stripping the amino acid down to the atom and replacing it by the side chain of the new amino acid. In some studies, the atom is also kept intact for mutations between amino acids R, K and Q, as these are relatively large amino acids and normally located at the protein surface [9]. The mutant structure is then optimized by conformation search for the local residues to release structural clash among them and the binding drug, which is conducted by variation of rotatable bonds of the drug and those of the side-chain of the amino acids in contact with the drug. This is followed by energy minimization to allow the mutant structure to find the local minimum energy conformation [9,15]. In some studies, molecular dynamics simulation is conducted to allow the mutant structure to reach a more appropriate local minimum energy conformation [14]. The procedure of this