Silesian University of Technology

Faculty of Automatic Control, Electronics and Computer Science

Doctoral dissertation

Algorithms for the analysis of molecular protein structures and drug-like ligands for modeling and simulation of residence time drug-molecular target

Author: mgr inż. Magdalena Ługowska
Supervisor: prof. dr hab. inż. Marek Kimmel
Co-supervisor: dr inż. Marcin Pacholczyk

Gliwice 2023

Dissertation abstract

"Algorithms for the analysis of molecular protein structures and drug-like ligands for modeling and simulation of residence time drug-molecular target"

mgr. inż. Magdalena Ługowska

Drug development is a complex process that remains subject to risks and uncertainties. In its early days, much emphasis was placed on the equilibrium binding affinity of a drug to a given target, which is described by the equilibrium dissociation constant (K_d). However, there are a large number of drugs that exhibit non-equilibrium binding properties. For this reason, optimization of other kinetic parameters such as dissociation rate constants (k_{off}) and association rate constants (k_{on}), has become increasingly important to improve accuracy in measuring \emph{in vivo} drug effects. To achieve this, the concept of drug-target residence time (τ) was developed to consider the continuous elimination of the drug, lack of equilibrium conditions and conformational dynamics of target molecules. Therefore, residence time has been shown to be a better estimate of lifetime potency than equilibrium binding affinity and is recognized as a key parameter in drug design. Nevertheless, being a single measure of drug potency, residence time provides a limited picture of binding kinetics and affinities. Due to the complex, labor-intensive, and costly nature of experimental methods for determining binding kinetic parameters, and the rapid advancement of technology, the demand for high-throughput *in silico* methods to estimate binding kinetic parameters and determine their key factors is increasing.

This thesis describes basic assumptions and applications of computational methods available for understanding and analysis of ligand binding and unbinding kinetics as well as determination of drug residence time in a target molecule. This includes molecular dynamics (MD), machine learning (ML) methods and their combination, as well as the use of Markov state models.

Because the existing bioinformatics databases do not contain complete information on the kinetic data of complexes with known crystallographic structure, the data were collected from published literature and transformed into a publicly available online database called PDBrt. Studies were performed for selected protein-ligand systems, the inhibitors of InhA (the enoyl acyl carrier protein reductase from Mycobacterium tuberculosis) - one of the key enzymes involved in the type II fatty acid biosynthetic pathway in M. tuberculosis. The heat shock protein inhibitor HSP90 and ligands of ENR, EGFR and HIV-1 proteins were also used to check the reproducibility of the Random τ Accelerated Molecular Dynamics (τRAMD) method.

The MD approach is used to analyze τRAMD to determine if the method is reproducible and applicable in calculating different relative residence times of drug-like compounds. When applied to

a series of similar compounds, τRAMD was found to provide the most accurate prediction of residence times. The reproducibility of τRAMD was demonstrated - the results obtained are similar to those published. However, the study showed that the τRAMD method did not allow estimating relative residence times that correlate well with experimental values for structurally diverse compounds. This suggests that the method has limited application and is not applicable to a wide range of compounds.

A machine-learning algorithm was proposed to identify molecular features affecting protein-ligand binding kinetics for a set of similar compounds. Molecular dynamics simulations of τRAMD results were used as model input. The study confirmed that τRAMD provides information about the characteristics of the dissociation pathway since the obtained dissociation trajectories can be used to identify the interactions that occur and the conformational changes of the system at subsequent time points. This information has been applied to further analyses, which led to the definition of key molecular properties for a series of InhA inhibitors. The proposed algorithm made it possible to obtain information on protein-ligand contacts that are specific to their residence times.