Finite absorbing Markov chain as a model of small-ligand binding process

Marcin Pacholczyk[1], Damian Borys[1] and Marek Kimmel[1,2]

[1]Institute of Automatic Control, Silesian University of Technology, 44-100 Gliwice, Akademicka 16
{marcin.pacholczyk, damian.borys}@polsl.pl
[2]Rice University, Department of Statistics, Houston, TX, USA
kimmel@rice.edu

Abstract. Analysis of protein - small molecule interactions is crucial in the discovery of new drug candidates and lead structure optimization. Small biomolecules (ligands) are highly flexible and may adopt numerous conformations upon binding to the protein. Scoring functions are traditionally used in many docking protocols and have key impact on a quality of structure-based virtual screening. A correct scoring function should be able to guide search algorithm to find and recognize native-like docking poses. In ideal case scoring function should be able to predict binding affinity. Despite extensive research, scoring remains a major challenge in structure-based virtual screening. We apply Stochastic Roadmap Simulation (SRS) and finite absorbing Markov chain theory to build a model of protein-ligand binding process [1, 2]. We propose a computational quantity – *time to escape* (TTE) from a funnel of attraction around binding site as a measure of binding affinity. The results based on PDBBind CoreSet [3] show statistically significant correlation between actual binding affinity and calculated TTE.

1 Model of protein-ligand interaction

The model of electrostatics associated with Poisson-Boltzmann equation is far more accurate in this case than simple Coulombic models and incorporates features such as location dependent dielectric constant and mobile ions contribution to the electrostatic potential (natural environment for proteins is usually salty aquatic solution). Protein is considered a rigid body limited by solvent accessible surface [5]. In order to solve PBE on 3D grid (Fig. 1.) we use two computer programs DelPhi [6] and APBS [7].

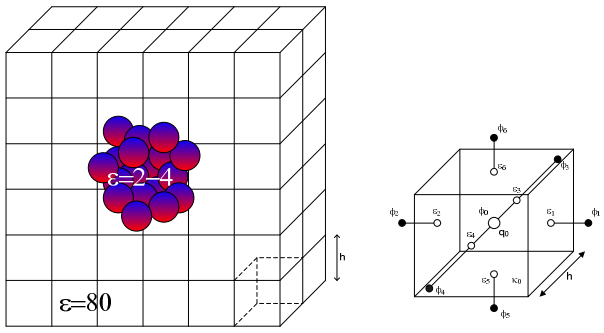

Fig. 1. Poisson-Boltzmann model of electrostatics

2 Stochastic Roadmap Simulation

Each node of a roadmap represents one conformation of a ligand. Formally, each conformation of n parameters is represented by a vector q. The set of all possible conformations forms the conformational space C. SRS assumes that the interactions are described by an energy function $E(q)$, which depends only on the conformation q of the ligand. A pathway in C represents motion of the ligand around protein. A roadmap may be considered a directed graph G encoding many pathways in C. Each node of a roadmap is a randomly selected conformation q from C with associated energy $E(q)$. Each directed edge between two nodes v_i and v_j has associated weight, which is equal to the probability of transition between the two nodes. In order to construct a roadmap the algorithm samples n conformations, randomly and independently from C. Then for each node v_i one finds k nearest neighbors of that node according to selected metric (i.e. RMSD or Euclidean). After that a transition probability P_{ij} is computed for every pair of neighboring nodes (Fig. 2.). P_{ij} calculation is based on difference of energy:

$$\Delta E_{ij} = E\left(v_i\right) - E\left(v_j\right)$$

between nodes v_i and v_j according to the formula:

$$P_{ij} = \frac{1}{N_i} e^{-(\Delta E_{ij}/k_B T)}, \quad \Delta E_{ij} > 0$$

or

$$P_{ij} = \frac{1}{N_i}, \quad \Delta E_{ij} \le 0$$

where k_B - Boltzmann constant, T - system temperature, N_i- number of neighbors of node. The self-transition probability is defined as:

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}$$

which ensures that the transition probabilities from any node sum up to 1.

3 The Time to Escape

Although it is possible to perform a simulation on a roadmap, which corresponds to a discrete version of the standard Monte Carlo method (discretization is defined by a roadmap) Apaydin et al. [1] suggest that usually it is not needed to generate individual trajectories on a roadmap but rather evaluate a parameter of interest. Time to escape (expressed as a number of simulation steps) from the funnel of attraction around the protein binding site is given as an example. Apaydin et al. propose the escape time as a measure of affinity of a ligand to a putative binding site. The funnel of attraction F_i is defined as the set of conformations within 10 Å RMSD of the bound conformation (Fig. 3.). Expected value of the time to escape can be easily calculated using the first step analysis technique [1], from Markov chain theory [4] by solving the following system of equations [1]:

$$\tau_i = 1 + \sum_{v_i \in F_i} P_{ij} \tau_j$$

where τ_i– time to escape starting from *i*–th node, *F*i– funnel of attraction around i–th binding site, v_i – *i*–th node.

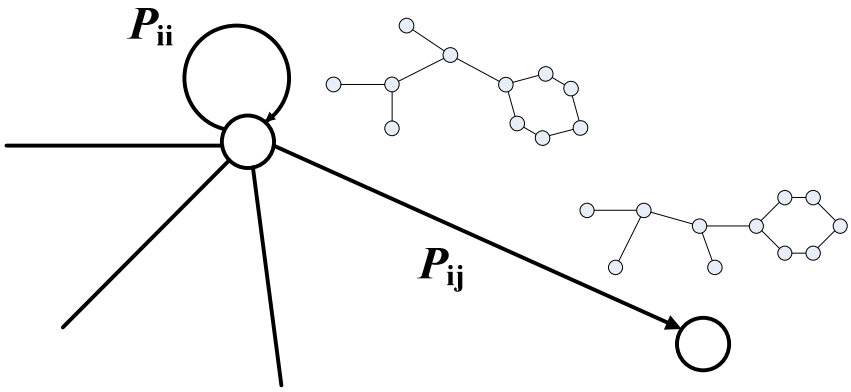

Fig. 2. Map building and assignment of transition probability

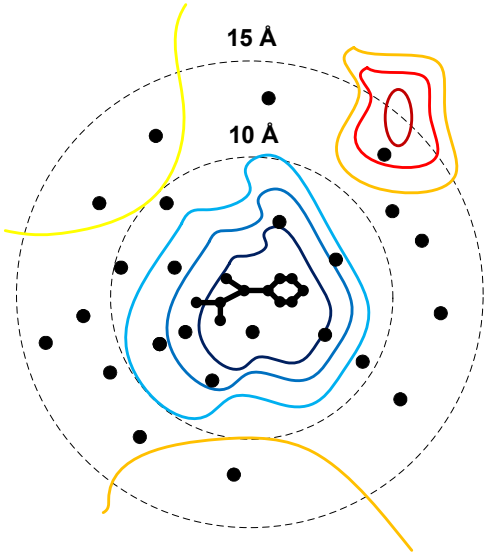

Fig. 3. Funnel of attraction around binding site

4 Results and Discussion

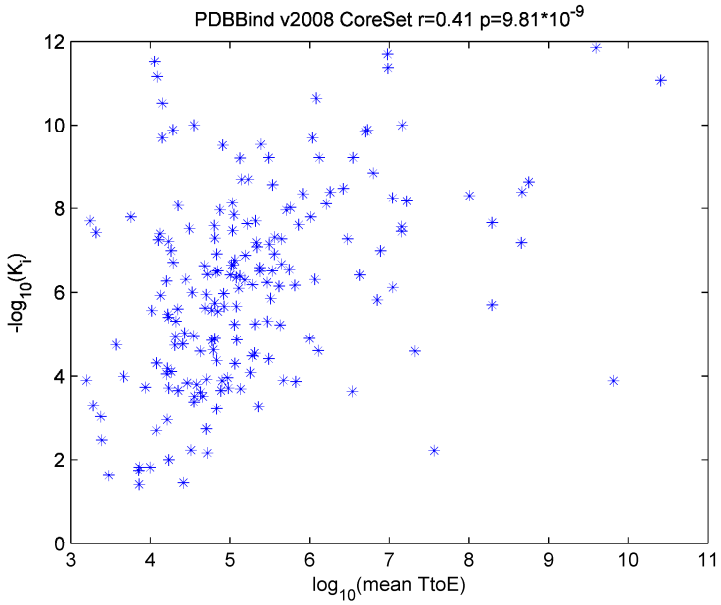

Fig. 4. Correlation between time to escape and experimental K_i obtained for PDBBind v2008 CoreSet

We applied the approach described in the paper to enzyme-inhibitor complexes with experimentally determined affinity data deposited in the PDBBind v2008 CoreSet [3]. The results show significant correlation between the computed mean time to escape and experimentally determined binding constant K_i. Pearson's correlation coefficient R=0.41.

Acknowledgements

This work has been partially supported by Polish National Science Centre funds under grant Maestro DEC-2012/04/A/ST7/00353 based at the Institute of Automatic Control, Silesian University of Technology, Gliwice, Poland.

References

1. Apaydin MS, Brutlag DL, Guestrin C, Hsu D, Latombe JC, Varma C: Stochastic roadmap simulation: An efficient representation and algorithm for analyzing molecular motion. J Comp Biol 2003, 10:257-281.
2. Pacholczyk M, Kimmel M: Exploring the landscape of protein-ligand interaction energy using probabilistic approach. J Comp Biol 2011, 18(6):843-850.
3. Wang R, Fang X, Lu Y, Wang S: The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. J Med Chem 2004, 47(12): 2977-2980.
4. Taylor HM, Karlin S. Markov Chains: Introduction. In An Introduction to Stochastic Modelling, Academic Press 1998, San Diego. pp 95-198.
5. Connolly ML: Solvent-Accessible Surfaces of Proteins and Nucleic-Acids. Science 1983, 221,709-713.
6. Rocchia W, Alexov E, Honig B: Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. Journal of Physical Chemistry B 2001, 105, 6507-6514.
7. Holst MJ, Saied F: Numerical-Solution of the Nonlinear Poisson-Boltzmann Equation - Developing More Robust and Efficient Methods. Journal of Computational Chemistry 1995, 16, 337-364.