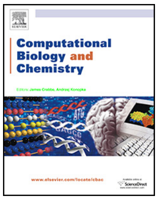
Computational Biology and Chemistry 56 (2015) 7–12

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem



Structural properties and interaction energies affecting drug design. An approach combining molecular simulations, statistics, interaction energies and neural networks

Dimitris Ioannidis [a], Georgios E. Papadopoulos [b], Georgios Anastassopoulos [c], Alexandros Kortsaris [a], Konstantinos Anagnostopoulos [a,*]

[a] Laboratory of Biochemistry, Department of Medicine, Democritus University of Thrace, 68100 Alexandroupolis, Greece
[b] Department of Biochemistry and Biotechnology, University of Thessaly, Ploutonos 26 & Aeolou Greece, 41221 Larisa, Greece
[c] Laboratory of Medical Informatics, Department of Medicine, Democritus University of Thrace, 68100 Alexandroupolis, Greece

ARTICLE INFO

Article history:
Received 18 October 2014
Received in revised form 17 February 2015
Accepted 22 February 2015
Available online 25 February 2015

Keywords:
Drug design
Molecular dynamics simulation
Interaction energy
Neural networks

ABSTRACT

In order to elucidate some basic principles for protein–ligand interactions, a subset of 87 structures of human proteins with their ligands was obtained from the PDB databank. After a short molecular dynamics simulation (to ensure structure stability), a variety of interaction energies and structural parameters were extracted. Linear regression was performed to determine which of these parameters have a potentially significant contribution to the protein–ligand interaction. The parameters exhibiting relatively high correlation coefficients were selected. Important factors seem to be the number of ligand atoms, the ratio of N, O and S atoms to total ligand atoms, the hydrophobic/polar aminoacid ratio and the ratio of cavity size to the sum of ligand plus water atoms in the cavity. An important factor also seems to be the immobile water molecules in the cavity. Nine of these parameters were used as known inputs to train a neural network in the prediction of seven other. Eight structures were left out of the training to test the quality of the predictions. After optimization of the neural network, the predictions were fairly accurate given the relatively small number of structures, especially in the prediction of the number of nitrogen and sulfur atoms of the ligand.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Since the 1980s, computer-aided drug design has been employed to help with the design of new drugs (determination of the molecular target of the drug and its structure, determination of the interaction mechanism, etc.). The field of protein structure–function has contributed significantly to this (Ghersi and Sanchez, 2011). An example is the use of the complementarity function (Sobolev et al., 1997; Sobolev et al., 1999) for protein–protein interactions, where the potential for a stable interaction of two biomolecules is determined by the frequency of favorable atomic interactions versus the frequency of unfavorable ones. A trend is to use the four basic non-covalent interactions: electrostatic, hydrogen bonds, van der Waals and hydrophobic interactions (Lins and Brasseur, 1995). Relevant to this is the effort to statistically study the various types of aminoacids of the interaction hotspots (Boganand and Thorn, 1998; Halperin et al., 2004; Jones and Thornton, 1996).

To understand the above interactions we must also use topological criteria, since these interactions depend on the distance and atom position. Initially, the lock-and-key model was used (Durrant and McCammon, 2011; Whitesides and Krishnamurthy, 2005) for the interaction of proteins with receptor agonists and antagonists and enzyme inhibitors and activators. This approach was complemented using molecular simulations, due to the increase in computing power and the experimental and theoretical determination of the behavior of atoms in semi-empirical force fields, like CHARMM (MacKerell et al., 1988, 1998). An important point is the handling of the hydrophobic interactions (Lins and Brasseur, 1995; Meyer et al., 2006).

Very useful are the in silico screening methods (Schmidt et al., 2009; Wishart et al., 2008), where a chemical substance can be checked for probable interactions against a database of structures. Also, open is the question of the role of neighboring relatively conserved water molecules that regulate and enhance the

---

Abbreviation: PDB, Protein Data Bank; MDS, molecular dynamics simulation; EM, energy minimization; MDS + EM, molecular dynamics simulation followed by energy minimization.

* Corresponding author. Tel.: +30 25510 30502; fax: +30 25510 30502.
E-mail addresses: kanagnos@med.duth.gr, kanagno@gmail.com (K. Anagnostopoulos).

http://dx.doi.org/10.1016/j.compbiolchem.2015.02.016
1476-9271/© 2015 Elsevier Ltd. All rights reserved.

interaction of proteins, like e.g., hsp90 (Yan et al., 2008). Water molecules are of interest for many aspects relevant to computational drug design such as structure, thermodynamics, binding free energy calculation, molecular docking and molecular simulation studies (deBeer et al., 2010). As replacing or conserving a crystal water molecule remains a critical problem for effective drug design, the immobile water molecules emerge as potential targets. Furthermore, other atoms have also been implicated that allosterically regulate receptors, e.g., sodium ions in dopamine receptor (Selent et al., 2010).

One method of structure-based or direct drug design is attempting to build a ligand that fits to a known receptor 3D structure (receptor-based drug design) (Schneider and Fechner, 2005). Such approaches include fragment-based drug discovery (Merour et al., 2014) or combinatorial chemistry by Monte Carlo methods (Grzybowski et al., 2002). The use of 3D-QSAR (quantitative structure-activity relationships) is also widespread (Verma et al., 2010).

Of particular importance is the inclusion of oxygen, nitrogen and sulfur atoms in the ligand. These are atoms with high electronegativity (especially oxygen and nitrogen) and therefore they can significantly influence the properties and interactions of the ligand. Hence the importance of the prediction of how many there should be in it.

In this paper, we attempted an empirical analysis of the protein–ligand interaction, using PDB databank structures and combining various approaches using statistical measurements. Various parameters of the interactions were determined to check which of these could be useful during in silico drug design. Also, the role of immobile water molecules was studied. Finally, a neural network was created from some of the interaction parameters determined, and this was used to predict other interaction parameters.

2. Materials and methods

2.1. Databases used

To obtain the starting material the Protein Data Bank PDB (Berman et al., 2000) was used. All the entries of the PDB database with *Homo sapiens* as source organism were downloaded (a total of 20,761 entries). A script in the PERL programming language was written to facilitate the examination and selection of the appropriate PDB entries. Using this script, only the PDB entries that met the following criteria were included: (I) to contain a HETSYN record (indication of the existence of a ligand), (II) to contain a TER record only once or twice (so that the selected files will contain only one or two protein chains, (III) not to contain heme or acetylglucosamine in the HETSYN record and (IV) not to contain nucleic acids. After this selection 1528 PDB entries were left, which were further examined manually and more entries were excluded if the ligands were sugars or lipids. After this round of selection, 400 PDB records were left. Next, these 400 entries were examined and those that were similar were excluded (similarity was defined as >95% identity in the protein sequence except when two or more PDB entries with similar proteins contained different ligands. In such cases the structures were retained, e.g PDB IDs 2J4A, 1XZX and 1Q4X, supplement IDs 6, 7 and 8). In the end, 95 entries were left. Of these 8 were further excluded due to many discontinuities in the structure, to yield a total of 87 final entries for the analysis. The general aim was to select simple entries consisting mostly of a protein and a ligand. These entries with the corresponding PDB IDs are shown in the Supplementary material.

These 87 entries were classified in four categories (color coded in the Supplementary material): receptor agonists/antagonists, enzyme inhibitors, serum blockers–transporter inhibitors and conformation blockers–chaperone inhibitors.

2.2. Molecular dynamics simulations

In order to to relax the protein–ligand system, energy minimization and molecular dynamics simulations (MDS) were performed using the PDB files as the initial state. The MDS were performed using the programs NAMD (Phillips et al., 2005) with the CHARMM force field (MacKerell et al., 1988; Brooks et al., 1983) (version 27). File preparation and analysis of the MDS trajectories were performed with the VMD (Humphrey et al., 1996) program.

The standard topology files of the CHARMM force field were insufficient to generate the structure file to be used in the MDS for all the PDB entries because no topologies for the ligands of the sample exist in the CHARMM force field. For this reason, the PRODRG program (Schüttelkopf and Aalten, 2004) was used to generate the missing topologies and force field parameters. The force field parameters for the covalent interactions (bonds, angles, dihedral angles and impropers) were also normalized to be of the same order of magnitude as those of CHARMM. E.g., in CHARMM the bond potential function is $K_b(b–b_0)^2$. For a C—H bond, b is the bond length, b_0 (the equilibrium bond length) is 1.090 Å and K_b (the spring constant) is 367.6 kcal mol^{-1} Å^{-2}. When PRODRG gave a value of 13,971.0 kcal mol^{-1} Å^{-2} for K_b, this was divided by a factor of 40 to be of the same order of magnitude as that of CHARMM. Without this normalization, the simulations would crash shortly after they started.

For the MDS, each system was placed in a box without periodic boundary conditions and water molecules were added using the “add Solvation box” plugin of VMD. Next, ions were added (only Na^+ and Cl^-) using the VMD plugin “Add Ions”. The dimensions of the water and ion-filled box were 5 Å from the coordinates of the most extreme atom of the protein–ligand structure in each dimension. The simulations were run at 300 Kelvin with a 2 femtosec timestep. The basic parameters for the MDS were: 1–4 scaling = 1.0, cutoff = 12.0, switchdist = 10.0, pairlistdist = 13.5 without harmonic or periodic boundary conditions, with constant temperature control and Langevin dynamics. A representative configuration file containing the parameters of the simulation can be found in the Supplementary materials. Before the actual simulation, an energy minimization (EM) was performed for 3000 steps. The number of steps of the actual simulation were chosen at about 60,000 steps depending on each system size and time consumption required for each trajectory. The aim of the short MDS was to relax the crystal structure in a way compatible with the force field used, thus making the energy calculations more reliable for the purpose they were used. In order to verify this, after the simulation one more energy minimization was performed for 1000 steps and the non-covalent energies of the two stages (end of MDS and final MDS + EM) were compared. In all but two cases (ID 56 and 57, PDB ID 1ZD2 and 1ZD4, Supplementary material) the differences (comparing only the protein + ligand system) were small (<13 kcal/mol in absolute values). Furthermore, the root mean square deviation (RMSD) for both the protein backbone and the ligand were determined and compared to the initial structure. In all cases the RMSD remained stable and with values of <0.8 Å (for the protein) and <2.0 Å (for the ligand). Even in the two exceptions mentioned above (ID 56 and 57, PDB ID 1ZD2 and 1ZD4) the RMSD for the ligand was stable at about 6.0 Å and 3.5 Å, respectively. For this reason, these two cases were retained for subsequent study.

After the simulation, many parameters related to the structure and stability of the protein–ligand systems were extracted; e.g., the total non-bonding interaction energies before and after the simulation, the number of water molecules in the protein–ligand

Download English Version:

https://daneshyari.com/en/article/15051

Download Persian Version:

https://daneshyari.com/article/15051

Daneshyari.com