Investigation of atomic level patterns in protein-ligand interactions

Ke Chen and Lukasz Kurgan

UNIVERSITY OF ALBERTA

biomine

Motivation

Protein-protein and protein-ligand interactions are among the central topics in structural biology. The former provide useful input for building protein-protein interaction networks and for understanding the protein's function, while the latter constitute a basis for rational drug discovery. To date, interactions between proteins and macromolecules, i.e., protein-protein [1], protein-DNA [2], and protein-RNA [3], have been systematically investigated while the driving forces that influence protein-ligand interaction are still ambiguous. To this end, we investigated the interactions between proteins and small ligands that exclude proteins, peptides, and nucleotides which were already investigated by other research groups. The purpose of this work is to investigate dominant forces in protein-ligand interaction.

Materials

Data set
The protein chains, which were taken from culledPDB list generated by PISCES server, are characterized by the following: 1) the chains share sequence identity of below 25%; 2) the resolution of the protein-ligand complex structure is below 2.0Å; and 3) the R factor is below 0.25. These criteria, which resulted in selection of 2320 chains, assure that the selected proteins shares low sequence identity the corresponding structures have sufficient quality. The protein and a ligand are assumed to interact with each other when at least one pair of heavy atoms, one from the protein and one from the ligand, can be found within 3.5Å distance. Excluding the water molecule, all molecules annotated as "HET" in Protein Data Bank (PDB), which includes organic compounds and ions, were taken as ligands. This excludes protein chains, peptides and nucleotides. As a result, 7759 pockets which have at least one contact with the considered ligand were extracted from the 2320 chains.

Classification of ligands
Among the 7759 complexes, some of the ligands appear multiple times, some are similar and could be grouped together. To facilitate analysis of the protein-ligand interactions we selected only these ligands that occur frequently and we group them into several categories. The ligands that bind to at least 100 pockets cover 59.4% of the considered complexes. Among these ligands, GOL, EDO, NAG, and ACT are organic compounds, Ca2+, Zn2+, Na+, Mg2+ and CD2+ are metal ions, SO42-, PO43-, Cl-, BR- and I- are inorganic anions. Additionally, some inorganic clusters, i.e., Fe-S cluster, also bind to a relatively large number of pockets. As a result, the considered ligands (including those that occur in less than 100 pockets) are grouped into four categories: organic compounds, metal ions, inorganic anions, and inorganic clusters. We analyze total of 3685 organic compounds (that include 560 distinct types), 1682 metal ions (25 types), 1837 inorganic anions (19 types), and 54 inorganic clusters (9 types), which cover (3685+1682+1837+54) / 7759 = 93.5% of all extracted pockets.
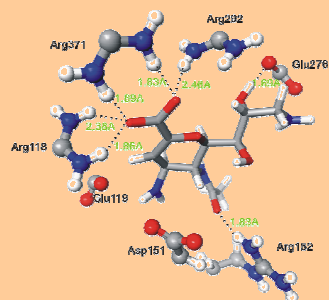
Figure 2. An example of hydrogen bonds formed between NH- group of a residue and oxygen atom of a ligand. The oxygen atom is colored red, nitrogen atom is blue, carbon atom is gray, and hydrogen atom is white. The residues in the pocket are in ball and stick form while the ligand is in stick form. The distance between the hydrogen atom (attached to donor) and the acceptor is shown next to the bond. The structure is taken from chain A of protein 1F8E, which interacts with 49A. The binding pocket contains four Arg residues and each residue contains 2 NH- groups. Three Arg residues (Arg118, Arg292, Arg371) are spatially adjacent, and they form five hydrogen bonds with the oxygen atoms of the oxygen atoms of the ligand.
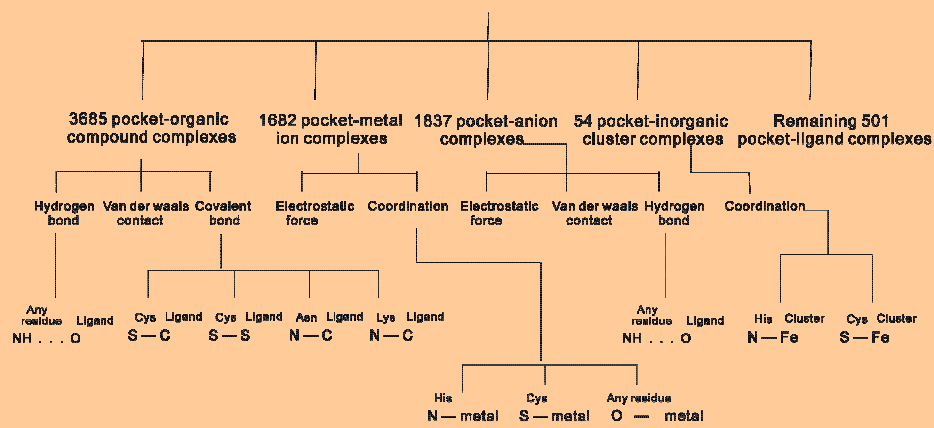

Figure 1. An overview of the protein pocket-ligand interactions. The top layer divides protein-ligand complexes into 5 major groups based on the type of the ligand. The second layer shows the major forces that are involved in formation of protein-ligand complexes for each type of the ligand. The bottom layer summarizes significant (frequently occurring) patterns for each force/bond type and each type of the ligand. The patterns are shown in XR…YL or XR – YL form where X denotes an atom type of residue R in the protein, Y denotes an atom type of the ligand L, strong interactions (covalent and coordination bonds) are depicted by "–", and weak interactions (hydrogen bond) are represented by "…".

Results

Summary of the interaction patterns
The protein pocket-ligand interactions are summarized in Figure 1. The top layer divides the 7759 protein-ligand complexes into 5 categories based on the ligand type. The second layer lists the major forces that are involved in formation of protein-ligand complexes for a given ligand type. For instance, protein-organic compound complexes are formed mainly by the means of covalent bonds, hydrogen bonds, and van der waals contacts. They accommodate for 99.9% of interactions. The remaining 0.1% of the contacts between a protein and the organic compound, which are omitted in the Figure 1, are due to electrostatic force. The bottom layer provides significant patterns that are associated with interactions for a given type of the ligand and a given type of bond/force, which are discussed in detail in the following sections.

Interaction patterns in protein-organic compound complexes
Organic compounds bind to proteins mainly by the van der waals contacts and hydrogen bonds. Total of 85771 contacts were observed between an organic compound and a protein and they include 77554 van der waals contacts, 7914 hydrogen bonds, and 246 covalent bonds.

Majority of the 246 covalent bonds formed between organic compounds and proteins could be summarized with four patterns: 1) 27 covalent contacts are formed between the thiol of Cys residue and the carbon atom of the organic compound (thioether bond); 2) 139 are formed between the nitrogen atom of Asn residue and the carbon atom of N-Acetyl-D-Glucosamine (NAG); 3) 28 concern the thiol of Cys residue and the sulfur atom of the organic compound (disulfide bond); and 4) 23 involve the nitrogen atom of Lys residue and the carbon atom of organic compound. The 4 types of covalent bonds cover (27+139+28+23)/246 = 88% of all investigated covalent bonds between proteins and organic compounds.

Hydrogen bonds are formed in 2466/3685 = 66.9% of the organic compound based complexes. The most frequently formed hydrogen bond involves NH- group (as the donor) of the AAs and the oxygen atom of the organic compounds. This type of the hydrogen bond covers 5206/7914 = 65.8% of all hydrogen bonds. The surface patch that is characteristic for NH- group has high potential to form hydrogen bonds with organic compounds. An example is given in Figure2.

Interaction patterns in protein-metal ion complexes
Among 1682 protein-metal ion complexes, 639 involve both coordination bonds and electrostatic force, 459 are based on electrostatic force but with no coordination bonds, and 499 incorporate coordination bonds with no electrostatic force. Overall, electrosta- tic force and coordination bonds are involved in (639+459+499) = 1597 complexes, which corresponds to 1597/1682 = 94.9% of all protein-metal ion complexes.

Total of 2345 coordination bonds are formed among the 1138 protein-metal ion complexes. The nitrogen atom in the side chain of His forms 787 coordination bonds with metal ions, sulfur atom of Cys forms 434 coordination bonds with metal ions, and oxygen atom (of any AA except Asp/Glu) forms 1039 coordination bonds. The bonds based on these three atoms correspond to (787+434+1039)/2345 = 96.4% of all coordination bonds.

Interaction patterns in protein-inorganic anion complexes
Inorganic anions bind to proteins mainly through electrostatic force, hydrogen bonds and van der waals contacts. Similarly to the protein-organic compound complexes, the most frequent hydrogen bond incorporates the NH- group of a residue that serves as the donor and the oxygen atom of a ligand. This pattern concerns 2777 hydrogen bonds which converts into 2777/3190 = 87.1% of all hydrogen bonds between a protein pocket and an inorganic anion.

Interaction patterns in protein-inorganic cluster complexes
Nine types of inorganic cluster, FS4, FES, SF4, F3S, CLF, FS3 CFN, FSO and NFS are found in PDB. We note that coordination bonds are involved in all 54 protein-inorganic cluster complexes. These bonds are usually formed between the iron atom of the cluster and the sulfur atom of Cys residue, and the iron atom of the cluster and the nitrogen atom in the side chain of His residue. These two coordination bond patterns cover 201/204 = 98.5% of all coordination bonds between inorganic cluster and a protein pocket.

Conclusion

We presented several strong interaction patterns, in terms of prevalent interactions between specific atoms of specific residue in the protein's pocket and specific atoms of the ligand. We discovered total of 11 strong interaction patterns that occur in 56% of all considered complexes. These results could be used to facilitate high-throughput identification of binding sites and in design of drug candidates.

References

[1] Jones S, Thornton JM. (1996) Principles of protein-protein interactions. Proc Natl Acad Sci U S A. 93:13-20.
[2] Luscombe NM, Laskowski RA, Thornton JM. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. 29: 2860-74.
[3] Ellis JJ, Broom M, Jones S. (2007) Protein-RNA interactions: structural analysis and functional classes. Proteins. 66:903-11.