Progress in Biophysics and Molecular Biology xxx (2014) 1–9

Contents lists available at ScienceDirect

Progress in Biophysics and Molecular Biology

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

Original research

Hot spots in protein–protein interfaces: Towards drug discovery

Engin Cukuroglu, H. Billur Engin, Attila Gursoy, Ozlem Keskin*

Center for Computational Biology and Bioinformatics and College of Engineering, Koc University, Istanbul, Turkey

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Hot spot
Protein–protein interface
Drug design
Protein–protein interaction

ABSTRACT

Identification of drug-like small molecules that alter protein–protein interactions might be a key step in drug discovery. However, it is very challenging to find such molecules that target interface regions in protein complexes. Recent findings indicate that such molecules usually target specifically energetically favored residues (hot spots) in protein–protein interfaces. These residues contribute to the stability of protein–protein complexes. Computational prediction of hot spots on bound and unbound structures might be useful to find druggable sites on target interfaces. We review the recent advances in computational hot spot prediction methods in the first part of the review and then provide examples on how hot spots might be crucial in drug design.

1. Introduction

Protein–protein interactions play crucial roles in regulating biological processes, cellular and signaling pathways. Protein–protein binding sites are called interfaces. Residue properties in interfaces are the key elements in protein–protein recognition, binding and affinity. Residue based analysis can help revealing protein–protein binding mechanisms. Alterations in native protein–protein interactions may lead to several diseases. Therefore, targeting the interfaces between proteins has an enormous potential in drug discovery (Kar et al., 2012; Thangudu et al., 2012; Wells and McClendon, 2007). Drugs targeting protein–protein interactions should ultimately bind to protein interfaces if not to allosteric sites. However, targeting interfaces is more challenging than targeting active sites of enzymes or G protein-coupled receptors in drug discovery since interfaces are relatively large, often flat without specific ligand binding pockets.

Residues in protein–protein interfaces do not equally contribute to the binding energies. There are critical residues called hot spots which contribute most to the binding energy (Bogan and Thorn, 1998; Clackson and Wells, 1995). Hot spot residues can be detected by alanine scanning mutagenesis experiments (Clackson and Wells, 1995). If the binding energy difference is more than 2 kcal/mol after mutating a residue to an alanine, it is labeled as a hot spot. Bogan and Thorn (Bogan and Thorn, 1998) analyzed amino acid compositions of hot spots and concluded that some residues are more favorable as hot spots in protein–protein interfaces. Tyr, Arg and Trp are the most frequent ones which are critical due to their size and conformation. They also reported that hot spots are surrounded by a set of residues, that are energetically less important resembling to an O-ring in a pipe fitting, to occlude hot spots from water molecules. There is a correlation between change in the accessible surface area and energy contribution of residues (Guharoy and Chakrabarti, 2005). Moreira et al. (2007a) also supported O-ring hypothesis using Molecular Dynamic (MD) simulations. Further, these hot spots are not randomly distributed in the interfaces but rather clustered. Hot spots are assembled within densely packed regions. These modular assembly regions are called hot regions (Keskin et al., 2005). This binding site organization justify how a given protein molecule may bind to different protein partners. Kleanthous and coworkers (Meenan et al., 2010) showed that a limited number of mutations at the interface of cognate complex of colicin E9 endonuclease and immunity protein 9 provide high-affinity binding of E9 to immunity protein 2, although at a weaker affinity compared to the cognate complex. These experimental findings were also studied by computational hot spot organizations (Cukuroglu et al., 2012). The organization of hot spot residues provides a mechanism to obtain binding affinity and specificity to different partners. Therefore, cooperativity of these residues can reveal the complex binding organizations in specificity (Cukuroglu et al., 2010; Shulman-Peleg et al., 2007).

Hot spots might also be important to find kinetic behavior of protein–protein complexes. Agius et al. (2013) made use of hot spot energetics and architectures of hot spots/hot regions to predict changes in dissociation rates upon a mutation. They used a set of biophysical and statistical descriptors to estimate hot spot energies.

* Corresponding author.
E-mail address: okeskin@ku.edu.tr (O. Keskin).

http://dx.doi.org/10.1016/j.pbiomolbio.2014.06.003
0079-6107/

These descriptors are then used as features to estimate successfully off-rate changes of single − or multi-point mutations.

Determining hot spot residues by experimental techniques is costly and time consuming, so computational methods have been developed to predict hot spot residues in bound and unbound protein structures (Cho et al., 2009; Darnell et al., 2007; del Sol and O'Meara, 2005; Grosdidier and Fernandez-Recio, 2008; Guerois et al., 2002; Huo et al., 2002; Kortemme and Baker, 2002; Li and Liu, 2009; Moreira et al., 2007a; Ofran and Rost, 2007; Tuncbag et al., 2009). Prediction results show that prediction accuracies are improving over the years. Prediction methods will be a complement to experimental studies to find hot spot residues which have an important role on binding affinity and specificity. In addition, computational methods can guide experimental mutational studies and explain functional and mechanistic aspects of protein binding (Fernández-Recio, 2011).

Last but not least, there is an increasing number of studies showing that hot spots may be important in drug design (Fry, 2012; Jubb et al., 2012; Thangudu et al., 2012; Wells and McClendon, 2007). Drug-like small molecules prefer to bind to hot spots at protein–protein interfaces (Arkin and Wells, 2004). Below, we will review first the recent advances in computational hot spot prediction and then protein-drug interactions and how hot spots are crucial in drug design.

2. Predicting hot spot residues

Hot spot prediction methods can be classified into two categories: based on unbound, monomeric structures and protein–-protein complex structures. Previous studies used different data sets and thresholds to predict hot spot residues. A complete list of the available tools and their properties are tabulated in SupplementaryFile_1.

3. Hot spot predictions on bound protein structures

Most of the hot spot prediction tools concentrated on bound protein–protein interactions in order to detect hot spot residues in protein interfaces. The analysis of hot spot residues detected by experimental techniques is limited to a small number of complexes. Information on experimentally determined hot spot residues has been deposited in databases. ASEdb is the first alanine mutation database developed by Thorn and Bogan (2001) and later BID was formed by Fischer et al. (2003). The cost and difficulty of determining hot spot residues experimentally led to prediction of these residues by computational approaches. Hence, Kortemme and Baker (Kortemme and Baker, 2002; Kortemme et al., 2004) proposed a computational alanine scanning method that uses energies of packing interactions, hydrogen bonds and solvation. Guerois et al. (2002) used FOLD-X energies to predict the hot spot residues. Another energy based method was developed by Gao et al. (2004) using hydrogen bond, hydrophobic and VdW interactions (three major non-covalent interactions) in order to estimate the individual contribution of each interfacial residue to the binding energy. The calculated energy changes of mutations were compatible with experimental results.

MD simulations are suitable for detailed analysis of protein–-protein interactions at the atomic level and they can be used for prediction of hot spot residues (Gonzalez-Ruiz and Gohlke, 2006; Grosdidier and Fernandez-Recio, 2008; Huo et al., 2002; Landon et al., 2007; Moreira et al., 2007a; Rajamani et al., 2004; Wang et al., 2013; Yogurtcu et al., 2008). Both energy and MD based hot spot prediction methods have high accuracy rates, but they are computationally expensive and difficult to apply on large scale studies.

Knowledge-based methods form another approach to predict hot spot residues. They usually use machine learning methods to learn from known hot spot data. The major advantage of knowledge-based methods is their computational efficiency. However, they are very sensitive to the selection of features such as residue type, size, hydrophobicity, accessible surface areas, etc. to characterize hot spot residues, and it is hard to find the best feature combination. Most of the studies use diverse features in order to increase their prediction accuracies even if they use similar machine learning algorithms such as support vector machine (SVM), linear regression, neural networks, Bayesian networks, and random forest models. Assi et al. (2010) used sequence conservation, energy scores and contact number information to predict hot spot residues. Lise et al. built up a prediction method based on machine learning (Lise et al., 2009) but their method was not working well on Arg and Glu, so they improved their approach adding two additional classifiers specific for these two amino acids (Lise et al., 2011). They mainly used van der Waals potentials, desolvation, hydrogen bonds and electrostatistics energies in order to predict hot spot residues. Koes and Camacho (Koes and Camacho, 2012) used machine learning approach with accessible surface area (ASA), relative ASA (RASA), evolutionary rate, conservation score, free energy of complexation and change in free energy of the alanine mutation values. They reported that ASA, RASA and per residues estimate of the free energy values were the most informative features and had good classification accuracies. Also, Xia et al. (2010) exhaustively searched different features of protein structures in order to increase the hot spot prediction accuracy and they concluded that ASA related features showed better discriminative power as suggested by Cho et al. (2009). According to the work of Xia et al. (2010) protrusion index was also a good discriminator of hot spots which was also shown in Li et al.'s work (Li et al., 2004). Unlike these features, Cho et al. (2009) found that weighted atomic packing density and weighted hydrophobicity had a discriminative power on hot spot residue predictions. Mitchell and her colleagues proposed two distinct methods: KFC (Darnell et al., 2007) which used shape specificity and biochemical contact features of the interface residues and they updated their approach with KFC2 (Zhu and Mitchell, 2011) which used interface solvation, atomic density and plasticity features. They concluded that lack of plasticity was strongly indicative of a hot spot residue but it was not a requirement. Wang et al. (2012a) used mass, polarizability and isoelectric point of residues, the relative side-chain accessible surface area and the average depth index to predict hot spot residues. As shown in previous studies, using accessible surface area, energy, atomic packing density and plasticity related features in a suitable combination increases the hot spot prediction accuracy.

Empirical formula based methods are also used instead of machine learning algorithms in order to predict hot spot residues. Pavelka et al. (2009) used only conservation scores to identify hot spot residues. Guney et al. (2008) used conservation score and ASA values with an empirical formula. Tuncbag et al. (2009) showed that RASA and pair wise potentials are much more discriminative than conservation scores for predictions. Indeed, it should be the family of proteins that determine whether conservation is discriminative. For antibody/antigen complexes, conservation is not a good feature (Assi et al., 2010). Kruger and Gohlke (2010) used pair wise potentials with degree of buriedness to predict hot spot residues. Geppert et al. (2011) used pair wise potentials, atom types and residue properties to generate an empirical formula with an additional voting system in order to find the functional hot spot residues. Shulman-Peleg et al. (2007) performed structural alignment of functionally similar protein–protein complexes in order to find spatial chemical conservation of the residues which correspond to hot spot residues. Hot regions in protein–protein

Please cite this article in press as: Cukuroglu, E., et al., Hot spots in protein–protein interfaces: Towards drug discovery, Progress in Biophysics and Molecular Biology (2014), http://dx.doi.org/10.1016/j.pbiomolbio.2014.06.003