

Protein–protein interface hot spots



**ASSIGN
BUSTER**

LITERATURE REVIEW

Protein-protein interactions

Proteins cooperate with other proteins to perform biological function. They physically interact with their partners through their interfaces with weak and non-covalent bonds. Understanding how proteins interact is crucial to predict unknown protein-protein interaction (PPI) and can help drug discovery. Thus, to discriminate the interface region from the rest of the surface, myriad of studies have analyzed some characteristics of interface region like hydrophobicity, solvent accessibility, shape and electrostatic complementarity, evolutionarily conservation, flexibility, residue propensities and hydrogen bonding.[1-3]

Hot Spots

Studies on protein interfaces have revealed that the free energy contributions of interfacial residues to binding are not uniformly distributed. A small subset of interfacial residues named hot spots is accounting for the majority of the binding free energy [4, 5]. Alanine Scanning Mutagenesis is an experimental method which is based on the fact that when a residue is mutated to alanine if it causes a significant drop in the binding free energy ($\Delta\Delta G$) then it is a hot spot. Hot spot information from experimental studies is deposited in the several databases such as the Alanine Scanning Energetics database (ASEdb) [6] and the binding interface database (BID) [7]. Unfortunately, experimental determination of hot spots is time-consuming, labor-intensive and has high economic costs. Therefore, there is a certain need for developing computational methods to identify hot spots [8].

Characteristics of Hot Spots

Bogan and Thorn reported that hot spots are abundant in Tryptophan (Trp) , Arginine (Arg) and Tyrosine(Tyr). On the other hand, some residues like Leucine (Leu), Methionine (Met), Serine (Ser), Threonine (Thr) and Valine (Val) are disfavored. The hot spots are surrounded by a ring of residues that are energetically less important and occlude bulk solvent from the hot spots (O-ring theory) [6]. “ Double water exclusion” theory refines the “ O-ring theory” and reveals that hot spots themselves are water-free.[9]

Keskin et al. showed that computational hot spots are not homogeneously distributed along the protein interfaces; rather they are clustered within locally tightly packed regions, called “ *hot regions* ”[10] and Cho et al. reached similar conclusion by showing significantly higher atomic packing density values of hot spots.[11] , del Sol and O’Meara and Keskin et al. illustrated that there is a correlation between hot spot residues and structurally conserved residues so the conserved and buried residues, are either experimental hotspot or in direct contact with an experimental hotspot.[10, 12]

Kozakov et al. have recently demonstrated that hot spots are clustered in specific regions that are distinguishable due to their concave topology and those regions are patterned with hydrophobic and polar residues [13].

Understanding microenvironment of hot spots is also significant. Ye et al. demonstrated that Alanine (Ala), Aspartic Acid (D), Glycine (G), Histidine (H), Isoleucine (I), Asparagine (N), Serine (S) and Tyrosine (Y) are more likely to occur close by hot spots than non-hot spots.

General definition of hot spot is that it is a residue that causes significant change in the binding free energy when it is mutated to Alanine. However, there is no consensus in the literature for what value of change is significant. In the most recent studies on hot spot prediction, they used 1.0 kcal/mol ([14, 15]) and mostly 2.0 kcal/mol ([16–23]) as the threshold to differentiate hot spots and nonhot spots, whereas Ofra and Rost chose 2.5 kcal/mol threshold to define hot spots and 0 kcal/mol to define nonhot spots [24]. Tuncbag et al. defined hot spots to be interface residues with higher than 2.0 kcal/mol and nonhot spots to be the ones with lower than 0.4 kcal/mol. Chen et al. and Xia et al. also used same definition of Tuncbag et al. [25, 26] while Cho et al. used two different cutoff values for the definition in their study, that is, 1.0 and 2.0 kcal/mol. [11]

Hot Spot Prediction

In recent years, several computational methods have been developed to predict protein-protein interface hot spots. Some of these methods are energy-based such as Robetta [27] and FOLDEF function [28] while others are based on molecular dynamics simulations [29][30].

Ofra and Rost developed a knowledge-based method (ISIS) that was based on neural networks with features extracted from sequence environment and evolutionary profile and used ISIS to predict hot spots [24, 31]. Darnell et al. combined decision tree approach based on atomic contacts, physicochemical properties, and shape specificity with computational alanine scanning [17]. Grosdidier and Fernandez-Recio developed a method that predicted interface hot spots using protein-docking tools without protein complex knowledge [32]. Recently, Wang et al. presented structure-based computational <https://assignbuster.com/proteinprotein-interface-hot-spots/>

approach that determines hot spots through the docking of a compound known as the inhibitor of the specific Protein-Protein Interaction (PPI).[16]

Feature-based method of Guney et al. identified hot spots using solvent accessible surface areas, residue conservation and residue propensity [33] and Tuncbag et al. presented empirical feature-based method by combining solvent accessibility and statistical pairwise residue potentials [34, 35]

Lise et al. combined the strengths of machine learning and energy-based methods by considering the basic energy terms as input features of machine learning models such as Support Vector Machines(SVMs) and Gaussian Processes[19, 36]. In their recent study, they integrated their approach with two additional SVM classifiers to overcome limitations on predictions involving arginine or glutamic acid residues.[19]

Machine learning approaches have also developed for hot spot prediction. Cho et al. proposed a hot spot prediction method based on protein structure, sequence and molecular interaction that used decision trees for feature selection and SVMs for classification.[11] Xia et al. also employed SVM classifiers with features such as protrusion index and solvent accessibility to predict hot spots[25]. Assi et al. applied Bayesian networks with energetic, structure-based and sequence-based features to predict hot spots [20]. Chen et al. generated sequence-based SVM model that utilized physicochemical features, Position Specific Scoring Matrix (PSSM), evolutionary conservation score, and sequence entropy[18]. Zhu and Mitchell built two knowledge-based hot spot prediction methods (KFC2a and KFC2b) based on different feature combinations using SVMs. [21] Wang et al. proposed a random forest

model that took hybrid features of residues that were defined with residues itself and its interacting residues of the opposite chain [37]. Chen et al. predicts hot spots using IBk algorithm as a classifier with only physicochemical characteristics of residues. [26]. Furthermore, Ye et al. constructed the SVM model using network features and microenvironment features [38]

Ozbek et al. applied Gaussian Network Model (GNM) both on unbound and complex structures to predict hot spots.[39]