Geometrically centered region: a "wet" model of protein binding hot spots not excluding water molecules

Zhenhua Li and Jinyan Li*

Bioinformatics Research Center, School of Computer Engineering,

Nanyang Technological University, Nanyang Avenue, Singapore 639798

Short title: GCR: a model of protein binding hot spots

Keywords: GCR, hot spot, burial level, water-mediated contact, O-ring

Corresponding Author*:

Professor Jinyan Li

School of Computer Engineering

Nanyang Technological University

50 Nanyang Avenue, Singapore 639798

Tel: (65) 67906253 (office)

Fax: (65) 67926559

Email: jyli@ntu.edu.sg

Abstract

A protein interface can be as "wet" as a protein surface in terms of the number of immobilized water molecules. This important water information has not been explicitly taken by computational methods to model and identify protein binding hot spots, overlooking the water role in forming interface hydrogen bonds and in filing cavities. Hot spot residues are usually clustered at the core of the protein binding interfaces. However, traditional machine learning methods often identify the hot spot residues individually, breaking the cooperativity of the energetic contribution. Our idea in this work is to explore the role of immobilized water and meanwhile to capture two essential properties of hot spots: the compactness in contact and the far distance from bulk solvent.

Our model is named geometrically centered region (GCR). The detection of GCRs is based on novel tripartite graphs, and *atom burial levels* which are a concept more intuitive than SASA. Applying to a data set containing 355 mutations, we achieved an F measure of 0.6414 when $\Delta\Delta G \geqslant 1.0$ kcal/mol was used to define hot spots. This performance is better than Robetta, a benchmark method in the field. We found that all but only one of the GCRs contain water to a certain degree, and most of the outstanding hot spot residues have water-mediated contacts. If the water is excluded, the burial level values are poorly related to the $\Delta\Delta G$, and the model loses its performance remarkably. We also presented a definition for the O-ring of a GCR as the set of immediate neighbors of the residues in the GCR. Comparative analysis between the O-rings and GCRs reveals that the newly defined O-ring is indeed energetically less important than the GCR hot spot, confirming a long-standing hypothesis.

1 Introduction

Understanding the collective contribution of residues to protein association and identifying the energetically outstanding residues in protein binding interface are of great interest to biologists. Alanine scanning mutagenesis [1, 2], which systematically mutates a target residue into alanine and measures the binding free energy change ($\Delta\Delta$G), is widely used to examine the energetic importance of a residue in the binding of two proteins. Alanine does not have a side chain other than the C^{β} atom, thus mutating a residue (except glycine) into alanine is actually cutting off its side chain. Experimental results of alanine scanning show that the energetic contribution of residues to the binding is not evenly distributed over the interface residues; rather only a few mutations change the binding free energy significantly [3, 4]. These residues are called hot spot residues. The threshold of $\Delta\Delta$G is usually set to 1.0 kcal/mol or 2.0 kcal/mol.

Identification of hot spot residues is useful, from which some applications such as drug design, protein engineering and template-directed combinatorial chemistry can benefit very much [5]. However, the wet-lab experiments face the difficulty of high labor-intensity and time-consumption. This has pushed the growth of computational methods that are capable of predicting the protein binding hot spots [6, 7, 8, 9, 10, 11, 12], although these methods suffer either a poor performance or a lack of interpretability. For example, the methods by Cho *et al.* [11] and Lise *et al.* [12] used SVM to predict hot spot residues based on a set of features. Despite the fact that they achieved a relatively good performance, the contribution of the individual features and the way the features combine could not be easily comprehended.

Hot spots tend to have a low solvent accessible surface area (SASA), indicating that there is a ring of energetically less important residues that protect those hot spot residues from the attack of bulk solvent [13]. This proposition is the famous O-ring theory, and the ring of residues is collectively called an O-ring. The O-ring theory also suggests that the hot spots are located at the core of the interface while the O-ring resides at its rim. As the O-ring is inferred rather than defined, it lacks sufficient information for us to determine hot spots by using computational programs. Keskin *et al.* [14] agreed that the hot spot residues do not distribute uniformly across the whole interface but cluster into energetically independent regions. Unfortunately, this observation is based on structurally conserved computational hot spots, not on the original experimental data. Moreover, how and to what extent the structurally conserved computational hot spots are correlated with the true hot spots are not fully investigated by [14].

The water information has not been explicitly taken into consideration by the previous studies to model protein binding hot spots. It seems that the water is harmful to the hot spots due to a pretended reason that the hot spots are sheltered from the access by bulk solvent. Recently, Li and Liu [15] proposed a hypothesis named "double water exclusion" which even suggests that the hot spot itself is absolutely water-free. However, the role of water in forming interface hydrogen bonds and in filling cavities is biologically recognized [16, 17, 18]. A recent observation is that protein interfaces can be as "wet" as the protein surfaces [17]. A deeper study is by Ikura *et al.* [16] who classified the interface water molecules into three categories: cavity-filling water molecules, interaction-mediating water molecules and exposed water molecules. They found that the contribution of binding free energy by water mediated indirect hydrogen bonds can reach 4.4 kcal/mol on average.

In this work, we introduce a new model for protein binding hot spots that does not exclude immobilized water molecules. The model is named *geometrically centered region* (GCR). First, we differentiate the buried water that is immobilized in the protein complexes from the bulk solvent water by using the information of co-crystallized water. The buried water is treated as the integral part of the protein complexes, and so for the buried water in the interface in particular. Then an atom contact (or an atom sheltering relation) is newly defined, and an atom/residue burial level is introduced. From there, we model an interface by using a tripartite graph consisting of the partite of its immobilized water molecules in addition to the two chains, and we define the GCRs as those subgraphs of this interface graph which have dense contacts and which are located away from the bulk solvent with a deep atom burial level. To valid the long standing O-ring hypothesis, we further present a definition for the O-ring of a GCR as the set of immediate neighbors of the residues in the GCR.

We apply to a data set containing 355 mutations to evaluate our model. If $\Delta\Delta G \geqslant 1.0$ kcal/mol is used to define hot spot residues, there are 109 hot residues in this data set. If $\Delta\Delta G \geqslant 2.0$ kcal/mol is used, there are 50. Our model achieves an F measure of 0.6414 in the case of $\Delta\Delta G \geqslant 1.0$ kcal/mol or 0.5859 in the case of $\Delta\Delta G \geqslant 2.0$ kcal/mol. Comparing to the performance by the benchmark Robetta prediction method [7, 8], our model is better. More importantly, the results show that the GCRs are water-containing to a certain degree, implying that the hot spots are not, at least not absolutely, water-free. This result means that our GCR model is in agreement with the biological mechanism behind the 'wet' binding hot spots. We also conducted analysis on the residue compositions in the O-ring in comparison to that in the GCR. One of our findings is that the O-ring is truly less important

Table I: Summary of the data set used in this work.

Interaction partner 1	Interaction partner 2	PDB id	Resolution	#Water†	Hotspots ($\Delta\Delta G \geqslant 1.0$)	Hotspots ($\Delta\Delta G \geqslant 2.0$)	Mutations
GROWTH HORMONE	GROWTH HORMONE RECEPTOR	1A22	2.60	13/13/7	17	8	68
RIBONUCLEASE INHIBITOR	ANGIOGENIN	1A4Y	2.00	15/13/7	6	3	28
BARNASE	BARSTAR	1BRS	2.00	18/18/8	11	9	14
COLICIN E9 IMMUNITY PROTEIN*	COLICIN E9	1BXI	2.05	7/7/3	10	6	28
BPTI*	BOVINE CHYMOTRYPSIN	1CBW	2.60	6/5/4	1	1	9
SOLUBLE TISSUE FACTOR*	BLOOD COAGULATION FACTOR VIIA	1DAN	2.00	32/32/8	8	3	77
FV D1.3	FV E5.2	1DVF	1.90	16/16/8	22	9	25
THROMBIN*	THROMBOMODULIN	1DX5	2.30	13/13/4	11	5	17
CD4	ENVELOPE PROTEIN GP120	1GC1	2.50	14/14/6	3	0	48
BETA-LACTAMASE	BETA-LACTAMASE INHIBITORY PROTEIN	1JTG	1.73	30/29/18	7	2	10
NC10*	N9 NEURAMINIDASEFAB	1NMB	2.20	2/0/0	1	0	1
IGG1-KAPPA D1.3 FV	HEN EGG WHITE LYSOZYME	1VFB	1.80	22/22/14	11	3	29
TRYPSIN INHIBITOR*	BETA-TRYPSIN	2PTC	1.90	7/6/2	1	1	1
Total					109	50	355

*: mutations are done only on this partner.

†: The three numbers are the number of buried water molecules in the interface, the number of water molecules in GCR with burial level not less than 1.0 and the number of water molecules in GCR with burial level not less than 2.0.

in the binding free energy.

2 Materials and Methods

2.1 Data Set

Our data set of alanine scanning mutagenesis is compiled from ASEdb [19] and a few previously published data sets [4, 9, 20, 21, 22]. Specifically, a mutation satisfying the following criterions is included in our data set: (i) The corresponding structure of the wild type interaction complex has been resolved by X-ray crystallography, and it can be identified by a PDB [23] entry; (ii) The interaction is between two proteins having an extended interface; (iii) It is a mutation that mutates

the wild type residue specifically into alanine, that is, those mutations that mutate wild type residue into non alanine residue, glycine for example, or alanine shaving data are not included; and (iv) The corresponding PDB entry of the interaction has solvent information, i.e. the information of co-crystallized water molecules is detailed in the PDB file.

With these specifications, we obtained a data set containing 355 mutations, of which 109 are hot spots if $\Delta\Delta G \geqslant 1.0$ kcal/mol is set as the threshold, or 50 if $\Delta\Delta G \geqslant 2.0$ kcal/mol is used. These mutations are distributed in 13 protein-protein complexes (Table 1).

To make sure that our data set is diverse, we have checked the structural similarity of the interacting complexes by using the CE algorithm [24]. For any two interacting complexes in our data set, the two pairs of interacting chains do not have a significant structural similarity (RMSD$\geqslant$4.0, z-score$\leqslant$3.7) at the same time except the complexes bovine chymotrypsin-inhibitor (PDB id: 1CBW) and beta trypsin-inhibitor (PDB id: 2PTC). The interfaces of these two complexes are further checked. In fact, their interfaces are significantly different in size (37 and 60, number of amino acid residues at the interface), thus we keep both of them in our data set.

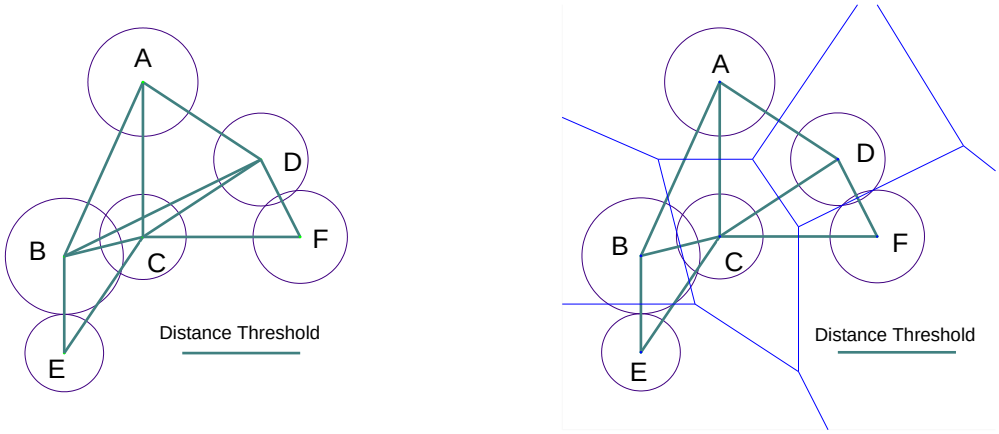
2.2 Definitions

2.2.1 Exposed atom, buried atom and atom contact

An atom is labeled as "exposed" or "buried" depending on its solvent accessible surface area (SASA). An atom is exposed if its SASA is higher than a threshold θ_S, otherwise it is buried. The θ_S here is set to 10.0Å^2 for all atoms. This is an empirical value. One can set this threshold to a lower value or even 0. We choose 10.0Å^2 in an aim to get more buried atoms. Actually when θ_S is set between 5.0 and 10.0, the performance varied only slightly, sometimes better sometimes not. All the exposed water molecules are removed from the structure by deleting the water molecules from the PDB structure iteratively when their oxygen atoms have an SASA greater than θ_S. We consider those removed water molecules somehow as part of the bulk solvent. The SASA of the remaining structure is calculated to determine the label of the atoms. In our study, all the hydrogen atoms are ignored and only heavy atoms are concerned. The SASA value is calculated by NACCESS which is based on the Lee and Richards algorithm [25].

Traditionally, atom contact is defined solely based on distance information. For example [11, 14], two atoms are said to be in contact if they have a distance less than a threshold. Some studies also use the van der Waals radii of atoms as an extra requirement in the definition. Under such

definitions, an atom can contact with all its nearby atoms that fall into the sphere with this atom as the center and the distance threshold as the radius. Thus, these definitions may cover some geometrically impossible atom contacts. Fig. 1(a) shows an example of the traditional definition for atom contact. Atom B and atom D are said to be a pair of contacting atoms, but they are actually separated by atom C from contacting each other.



(a) (b)

Figure 1: Traditional (a) and Voronoi diagram combined (b) definition of atom contact. The thick lines are the atom contacts defined and the thin lines in (b) show the corresponding Voronoi diagram of the point set. The van der Waals radii of atoms are taken into consideration, thus two contacting atoms must have a distance less than their van der Waals radii plus the distance threshold.

In this work, we define atom contact by combining distance information with Voronoi diagram. Two atoms are defined to be in contact only when they share a Voronoi facet (or a Delaunay edge, due to the duality between Voronoi and Delaunay diagrams). This guarantees that the contact is geometrically possible. Their distance is also required to be less than their van der Waals radii plus a threshold θ_L. Here, this threshold is set to the diameter of a water molecule, 2.75Å. Fig. 1(b) shows the atom contacts defined by our ideas. Atom B and atom D are not in contact any more because they do not share a Voronoi facet. Besides, atom E and atom F are not in contact because they are not close enough to each other, despite they do share a Voronoi facet.

The reason why we set θ_L to the diameter of a water molecule is that: by taking this value, an atom contact can then be regarded as a sheltering relation between two atoms. If two atoms contact with each other, they are somehow protected by each other from being accessed by bulk solvent, i.e. their SASA values are reduced due to this contact. Then we can further define atom burial

level rationally based on this definition of atom contact (sheltering relation). Another reason is that the use of this θ_L value yields a distance-tolerant definition of atom contact, and hence most of the specific atom-level contacts such as hydrogen bonds and salt bridges can be covered. This threshold was also adopted by Li and Liu [15] to define a water-free relation between two atoms. Nevertheless in our definition of atom contact, the Voronoi diagram plays a more important role. In all the atom pairs that share a Voronoi facet, only about 9.0% have a distance no less than their Van der Waals plus 2.75Å. And in all the Voronoi-facet-sharing atom pairs, more than 82% have a distance less than their Van der Waals plus 1.5Å. Thus in terms of the performance, when θ_L is larger than 1.5Å, the F measure of our model changes not much.

2.2.2 Atom burial level

When all of the atom contacts of a protein structure are known, an atom contact graph $G_A = (V_A, E_A)$ can be constructed straightforward by denoting V_A as the set of atoms and E_A as the set of atom contacts. The burial level of an atom is the extent to which the atom is buried. The burial level of an exposed atom is 0, and the burial level of a buried atom is equal to the length of the path from this atom to its nearest exposed atom, i.e. the minimum number of hops for this atom to reach an exposed atom.

To compute the burial levels of the heavy atoms in the atom contact graph, we add a pseudo atom node into the graph which represents the bulk solvent water. This pseudo atom node is connected to all of the exposed atoms, while the buried atoms are not connected to this node directly. The calculation of the atom burial levels is then changed into the single-source-shortest-path problem. By calculating the shortest paths from the pseudo node to all other nodes, the burial levels of all the atoms are obtained. The calculation was done by using Dijkstra's algorithm [26].

An alternative way to define the burial level is to use the concept of weighted alpha shape [27]. The atom contact is defined as the edges in the α-complex of the atoms; the α-exposed atoms can be directly defined as exposed atoms and hence the atom burial level can be defined. But the weighted alpha shape, as discussed by Kim *et al.* [28], does not fully follow the Euclidean metric, making the weight values of the atoms and the α value hard to adjust. We use the SASA value to determine the label (buried/exposed) of an atom for flexibility—By using a threshold (10Å^2) instead of using the geometrical position in α-shape directly, we can get more "buried" atoms and hence more differentiable burial level values of atoms.

Some existing concepts are related to our idea of atom burial level. Pintar *et al.* [29] proposed the concept "atom depth" as the Euclidean distance of an atom to its nearest exposed neighbor. However, this definition cannot capture the local sheltering relation between atoms. Bouvier *et al.* [30] used Voronoi diagram of the protein-protein interface to define Voronoi Shelling Order (VSO) of an atom. But their VSO, which is defined only on the interface, is merely based on geometry, thus it cannot capture the pocket or cavity structures.

The burial level of a residue is the average burial level of all the atoms in this residue.

2.2.3 Residue contact, residue contact graph and binding interface

Two residues are said to be in contact with each other if at least one pair of atoms from these two residues are in contact. In this work, a buried water molecule is also regarded as a "residue". The residue contact graph $G_R = (V_R, E_R)$ of a protein complex is a graph where the node set V_R is the set of residues and the edge set E_R is the set of residue contacts.

The interface buried water molecules are water molecules that are in contact with residues from both chains. We treat the interface buried water molecules as the integral part of the interface. With this in mind, we define an interface as a tripartite graph $G_I = (V_{I1} \bigcup V_{I2} \bigcup V_{IW},\ E_I)$. Here, G_I is a subgraph of G_R; V_{IW} is the set of interface buried water molecules; V_{I1} and V_{I2} are two sets of such residues that have at least one atom contact with a residue of the other chain or with some interface buried water molecule; The edge set E_I is the set of residue contacts among these residue sets, i.e. those contacts between the two chains and those between the interface residue and interface buried water.

2.2.4 Definition of geometrically centered region

The geometrically centered region (GCR) inside an interface is the region that is at the center of the interface. By "centered", we mean: (i) that the region must be in close contact with the two chains or with the interface buried water, and (ii) that the region must be located away from the rim of the interface. Computational procedure to identify the GCR in an interface consists of two steps. For a given interface tripartite graph G_I of a protein complex, we first identify its k-core, which is the maximum subgraph where every node has at least k neighbors. Second, we remove those slightly buried or exposed residues and those interface buried water molecules whose burial level is less than a threshold θ_B. The remaining subgraph is the "centered" part of G_I and is defined as the GCR of

this protein complex.

We use the GCR to model the hot spot in an interface. An amino acid residue is predicted to be a hot spot residue if it is covered by a GCR, or a non-hot spot residue otherwise.

The k-core step ensures that the GCR is in compact contact with the other chain or with interface buried water molecules. It is done by iteratively removing the nodes whose degree is less than k from G_I until all the nodes have no less than k neighbors. We set k to 3 in our model. A low k value will yield trivial k-cores which eliminate very few residues from the original graphs and a high k value will remove most residues from the original graph. The second step guarantees that the GCR is away from the attack of bulk solvent water. The θ_B is set according to the $\Delta\Delta$G threshold. We use GCRs with $\theta_B = 1.0$ to model $\Delta\Delta G \geqslant 1.0$ kcal/mol residues, and GCRs with $\theta_B = 2.0$ to model $\Delta\Delta G \geqslant 2.0$ kcal/mol residues. In the following sections, we refer to the value of $\Delta\Delta$G by only giving its value without the unit "kcal/mol" for convenience.

Note that the GCR of an interface may not be a connected graph. In rare cases a GCR may contain two or more connected components which are completely or partially isolated by the bulk solvent. These connected components are like the stably structured islands in unstable flowing bulk solvent.

2.3 Workflow of identifying GCR from protein complex structures

The computational steps to identify the GCR from a protein complex structure are summarized as follows:

1. Iteratively remove water molecules that have an SASA larger than θ_S from the protein complex structure;
2. Label atoms with "buried" or "exposed" according to the SASA and θ_S;
3. Identify all the Delaunay edges (equivalent to Voronoi facets) between atoms;
4. Determine the atom contact graph by using the Delaunay edges and the distance threshold θ_L;
5. Calculate the atom burial levels of all atoms;
6. Construct tripartite residue contact graph of the interface from the atom contact graph;
7. Iteratively remove residues/interface buried water molecules that have less than k neighbors;

8. Remove those residues/interface buried water molecules whose burial level is less than θ_B.

2.4 Metrics in performance evaluation

According to the prediction results, the residue mutations in our data set are divided into four types as traditionally: true positives (TPs), false positives (FPs), true negatives (TNs) and false negatives (FNs). TPs are the hot spot mutations that are covered by the GCRs. FPs are the non hot spot mutations that are wrongly covered by the GCRs. TNs are non hot spot mutations that are successfully filtered out by the GCRs. FNs are hot spot mutations that are not covered by the GCRs.

The performance of our model is evaluated by sensitivity (recall), specificity, precision, accuracy and F measure. Sensitivity is the extent to which the GCRs can cover the true hot spots: $sensitivity = \frac{|TP|}{|TP|+|FN|}$; specificity is the extent to which non hot spots residues are eliminated from the GCRs: $specificity = \frac{|TN|}{|TN|+|FP|}$; precision is the proportion of hot spot residues in the mutated residues in the GCRs: $precision = \frac{|TP|}{|TP|+|FP|}$; accuracy is the proportion of correctly classified mutations in all the mutations in our data set: $accuracy = \frac{|TP|+|TN|}{|TP|+|FP|+|TN|+|FN|}$. F measure (F1) combines the sensitivity and precision, which is a good indicator to reflect the overall performance of the model.

$$F1 = \frac{2 \times sensitivity \times precision}{sensitivity + precision} \tag{1}$$

The statistical significance of the GCR/GCRs is evaluated by randomly sampling sets of residues and buried waters in the interface tripartite graph having the same size as the GCR/GCRs. The corresponding performance, F measure, of these randomly sampled sets is first determined, and then the z-score of the GCR/GCRs is calculated by:

$$z - score_{GCR} = \frac{F1_{GCR} - \overline{F1}}{s_{F1}} \tag{2}$$

where $F1_{GCR}$ is the F measure value obtained by our GCR model, $\overline{F1}$ and s_{F1} are the average value and the sample standard derivation of the F measures obtained by the random sampling, respectively. The p-value of a z-score can be easily identified by normal cumulative distribution function.

To systematically compare the performance of our model and other methods, the performance on many subsets of the data set is evaluated. Each time half of the complexes are taken and the F1 value is calculated. The performances on all such subsets are calculated and the performance of different methods are compared by using Wilcoxon signed rank test [31]. For example, the comparison with

Table II: Performance of GCR in comparison to Robetta and FoldX when $\Delta\Delta G \geqslant 1.0$ is used to define hot spots. NA: the value is not applicable.

PDB id	TP	FP	TN	FN	Sensitivity	Precision	Specificity	Accuracy	z-score	F1	$F1_{Robetta}$	$F1_{FoldX}$
1A22	12	17	34	5	0.7059	0.4138	0.6667	0.6765	*2.074*	**0.5217**	0.4	0.4
1A4Y	6	5	17	0	1	0.5455	0.7727	0.8214	*2.913*	**0.7059**	0.5	0.5455
1BRS	8	0	3	3	0.7273	1	1	0.7857	*2.7696*	0.8421	**0.8571**	0.7059
1BXI	6	1	17	4	0.6	0.8571	0.9444	0.8214	1.5595	**0.7059**	0.533	0.6667
1CBW	1	1	7	0	1	0.5	0.875	0.8889	1.3377	0.6667	**1**	NA
1DAN	7	7	62	1	0.875	0.5	0.8986	0.8961	*3.454*	0.6364	**0.6667**	0.4348
1DVF	12	0	3	10	0.5455	1	1	0.6	*1.8177*	0.7059	0.5333	**0.7429**
1DX5	9	0	6	2	0.8182	1	1	0.8824	*2.89*	**0.9**	0.625	0.7273
1GC1	1	7	38	2	0.3333	0.125	0.8444	0.8125	-0.0017	0.1818	**0.25**	NA
1JTG	5	3	0	2	0.7143	0.625	0	0.5	0.3159	0.6667	0.7692	**0.8235**
1NMB	0	0	0	1	0	NA	NA	0	NA	NA	**1**	NA
1VFB	8	11	7	3	0.7273	0.4211	0.3889	0.5172	0.4742	0.5333	0.7407	**0.7619**
2PTC	1	0	0	0	1	1	NA	1	1.11	**1**	**1**	**1**
Total	76	52	194	33	0.6972	0.5938	0.7886	0.7606	*5.8139*	**0.6414** (±0.002)	0.5936 (±0.003)	0.6147 (±0.003)

Robetta is based on the whole data set (13 complexes). Each time half of the complexes (7 complexes) are taken and the F measure values of both GCR and Robetta are evaluated. There are 1716 such half-size subsets, thus the Wilcoxon signed rank test is done on the 1716 pairs of F measure values. The variance of the performances is also calculated to assess the robustness of the method.

3 Results and Discussion

3.1 Performance of GCR in hot spot prediction

As described, the interface of a protein-protein binding complex is defined as a tripartite graph where two parts represent the two protein chains and the third one stands for the interface water. By shrinking this graph into its k-core (k=3, a dense subgraph), only those residues that closely contact with the other side and with the interface buried water molecules are left. Some nodes in this k-core are further filtered by their burial levels. Subsequently, we get a region, namely geometrically centered region, that is away from the solvent bulk water. We use the GCRs whose burial level is not less than 1.0 or 2.0 to model the hot spots that have $\Delta\Delta G \geqslant 1.0$ or $\Delta\Delta G \geqslant 2.0$, respectively.

The performance of our GCR model on protein binding hot spot prediction is shown in Table II,

Table III: Performance of GCR in comparison to Robetta and FoldX when $\Delta\Delta G \geqslant 2.0$ is used to define hot spots. NA: the value is not applicable.

PDB id	TP	FP	TN	FN	Sensitivity	Precision	Specificity	Accuracy	z-score	F1	$F1_{Robetta}$	$F1_{FoldX}$
1A22	5	6	54	3	0.625	0.4545	0.9	0.8676	*3.3673*	**0.5263**	0.3333	0.4444
1A4Y	2	1	24	1	0.6667	0.6667	0.96	0.9286	*3.0762*	**0.6667**	0.3636	0.4
1BRS	5	0	5	4	0.5556	1	1	0.7143	*2.3271*	**0.7143**	**0.7143**	0.6154
1BXI	2	0	22	4	0.3333	1	1	0.8571	*1.6537*	**0.5**	0.4444	0.25
1CBW	1	0	8	0	1	1	1	1	*2.4132*	**1**	NA	NA
1DAN	2	1	73	1	0.6667	0.6667	0.9865	0.9740	*4.2331*	**0.6667**	NA	NA
1DVF	4	1	15	5	0.4444	0.8	0.9375	0.76	*2.2477*	0.5714	0.4615	**0.5882**
1DX5	4	0	12	1	0.8	1	1	0.9412	*3.8045*	**0.8889**	0.25	0.6667
1GC1	0	2	46	0	NA	0	0.9583	0.9583	NA	NA	NA	NA
1JTG	0	6	2	2	0	0	0.25	0.2	NA	NA	**0.25**	0.4444
1NMB	0	0	1	0	NA	NA	1	1	NA	NA	NA	NA
1VFB	3	3	23	0	1	0.5	0.8846	0.8965	*3.8247*	**0.6667**	0.4286	0.6
2PTC	1	0	0	0	1	1	NA	1	*2.3153*	**1**	**1**	**1**
Total	29	20	285	21	0.58	0.5918	0.9344	0.8845	*7.6556*	**0.5859** (±0.004)	0.3964 (±0.003)	0.4762 (±0.003)

where $\Delta\Delta G \geqslant 1.0$ is set as the threshold, and also in Table III when $\Delta\Delta G \geqslant 2.0$ is used as the threshold. Our method was compared to the Robetta server [7, 8] and FoldX method [32, 33]. Robetta uses an energetic idea to model computational alanine scanning and has been referred to as the gold standard in the benchmark comparison [12]. It is a web server and its interface alanine scanning service is available at `http://robetta.bakerlab.org/alascansubmit.jsp`. The mutations in our data set are submitted to the web server and the predicted $\Delta\Delta G$ values are retrieved. FoldX is also an energy based method and it can be used to study both protein stability and protein affinity. It is available at `http://foldx.crg.es/`. The latest version of FoldX, 3.0 beta 4, was download from this web site and the interface alanine scanning were performed by using the default parameters. We can see from the two tables that in both cases our model is better than those of Robetta and FoldX in terms of the F measure. Especially for $\Delta\Delta G \geqslant 2.0$ residues, our model is obviously better than the two energy based methods with an increase of F measure of 0.1895 (from Robetta) or 0.1097 (from FoldX). We have analyzed the significance of such difference in performance by using Wilcoxon signed rank test (as described in Methods). It turns out that the difference in performance is significant with p-values 2.6524×10^{-168} ($\Delta\Delta G \geqslant 1.0$) and 1.3268×10^{-281} ($\Delta\Delta G \geqslant 2.0$) when comparing with Robetta, or 6.6414×10^{-108} ($\Delta\Delta G \geqslant 1.0$) and 5.8429×10^{-272} ($\Delta\Delta G \geqslant 2.0$) when comparing with FoldX. These extremely small p-values indicate that the Robetta or FoldX method rarely outperforms our model

Table IV: Comparison of our model to CC/PBSA, EGAD and MINERVA.

Method	Common data*	ΔΔG	F1	$F1_{GCR}$	p-value
CC/PBSA	104	⩾1.0	0.5965	0.6207	0.2598
		⩾2.0	0.5185	0.5	0.7049
EGAD	196	⩾1.0	0.4961	0.6335	3.5595×10^{-13}
		⩾2.0	0.3929	0.5152	3.5591×10^{-13}
MINERVA	188	⩾1.0	0.6047	0.5988	0.1030
		⩾2.0	0.5625	0.6111	2.7963×10^{-34}

*: Due to the unavailability of the source code or limitations of the services of these methods, we perform the comparison by taking the common mutations shared by our data set and those published in their original papers. This column shows the number of common mutations.

in different subsets of the data. The low variances as shown in the tables also indicate that our model is robust.

As shown in rows 2 and 3 of Table IV, we also compared our model to another two energy-based methods, CC/PBSA[34] and EGAD[35] based on the common mutations shared by our data set and those reported in their work. Comparing with CC/PBSA, our methods achieved a better F measure for ΔΔG⩾1.0 residues and CC/PBSA achieves a better one for ΔΔG⩾2.0 residues. However, the large p-values indicate that the difference is not significant thus we cannot simply conclude which method is better in both cases. For EGAD, our method is significantly better in both cases as the F measure values by our method are larger and the p-values are low.

Although the training-testing protocol commonly used in the machine learning field is not easily adoptable to GCR, we compared our method to a recently published machine learning method MINERVA [11]. MINERVA uses SVM to predict the protein binding hot spots at the interfaces based on a set of features, which has achieved a relatively better performance than other two machine learning methods [9, 12] in terms of F1. As shown in row 4 of Table IV, for ΔΔG⩾1.0 residues, MINERVA has a slightly better F measure, however the p-value indicates that this is insignificant. For ΔΔG⩾2.0 residues, our method outperforms MINERVA significantly. Thus in terms of the performance, our straight forward model is competitive to machine learning methods.

Table V: Comparison between our model and HotSprint.

Model	GCR1.0*	GCR2.0†	HotSprint
Computational hot spots‡	94	37	48
p-value	2.12×10^{-15}	7.17×10^{-11}	3.01×10^{-4}

*: GCR1.0, GCR with burial level not less than 1.0.

†: GCR2.0, GCR with burial level not less than 2.0.

‡: Some mutations in computational hot spots or not are not clear in HotSprint due to the unavailability of the conservation information, thus the mutations used here are a subset of the original data set, containing 256 mutations.

In general, SVM based methods suffer the low-interpretability issue—the contribution of the individual features in the prediction is hardly known. Those energy-based methods either lack a good performance or are time-consuming. So we have compared our method to another method, HotSprint[10], which is as straightforward and as efficient as our method. HotSprint is also based on a simple empirical model and it combines conservation and SASA to identify computational hot spots. A residue is identified as a computational hot spot residue by HotSprint if its conservation score is larger than a threshold, its SASA change in complex formation is larger than a threshold and its SASA in the complex is smaller than a threshold. Because the $\Delta\Delta$G threshold they used is not clear, thus we compared our model to their model in terms of the significance of the difference of $\Delta\Delta$G of the mutations that are identified as hot spots versus that of the mutations identified as non hot spot residues. Mutations can be divided into two parts according to a model: computational hot spot residues and computational non hot spot residues. The statistical significance of the difference in $\Delta\Delta$G between the two parts is measured by using the Mann-Whitney test [36]. The p-values are shown in Table V. We can note that the p-value of HotSprint is much bigger than the both cases by our model. This indicates that there are more low $\Delta\Delta$G mutations in the computational hot spot residues or more high $\Delta\Delta$G mutations in the computational non hot spot residues by HotSprint than that by GCR. If the $\Delta\Delta$G threshold of HotSprint is mandatorily set as 1.0 or 2.0, its performance in F measure is 0.4098 ($\Delta\Delta G\geqslant 1.0$) or 0.3294 ($\Delta\Delta G\geqslant 2.0$), while the corresponding performance of GCR is 0.6429 ($\Delta\Delta G\geqslant 1.0$) or 0.5676 ($\Delta\Delta G\geqslant 2.0$).

The statistical significance of the results is shown in column 9 of Table II and Table III as

measured by z-score. Those z-scores higher than 1.65 (corresponding to p-values less than 0.05) are emphasized in italics. The overall z-scores indicate that the GCR model is significantly better at identifying hot spots in the interface than a random sampler for both $\Delta\Delta G \geqslant 1.0$ and $\Delta\Delta G \geqslant 2.0$ residues. In particular, the overall z-score of the GCRs in identifying $\Delta\Delta G \geqslant 1.0$ residues is 5.8139, corresponding to a p-value of 3.05×10^{-9}; and the overall z-score of the GCRs in identifying $\Delta\Delta G \geqslant 2.0$ residues is 7.6556, corresponding to a p-value of 9.66×10^{-15}. For individual structures, all of the GCRs corresponding to $\Delta\Delta G \geqslant 2.0$ residues are significant whenever the z-score is applicable. A few of the GCRs corresponding to $\Delta\Delta G \geqslant 1.0$ residues are not that significant but still show a performances better than random sampler, except GCR on 1GC1, whose z-score is close to 0 and thus the performance is just the same as a random sampler.

Fig. 2 shows two examples of the GCR hot spots in comparison to the true hot spot residues identified experimentally. As shown by the correspondence between the GCR (Fig. 2(a)) and the true hot spots (Fig. 2(b)) on the growth hormone receptor (binding with growth hormone, PDB id: 1A22, chain B), 8 of the 10 $\Delta\Delta G \geqslant 1.0$ residues are covered by the GCR, and 4 of the 5 $\Delta\Delta G \geqslant 2.0$ residues fall into the GCR. Again as shown in Fig. 2(c) and (d), where the GCR and true hot spot residues matches very well, 9 of the 11 $\Delta\Delta G \geqslant 1.0$ residues and 4 of the 5 $\Delta\Delta G \geqslant 2.0$ residues fall into the GCR on the thrombin heavy chain (binding with thrombomodulin, PDB id: 1DX5, chain M). The first GCR contains a total of 16 or 5 residues of the growth hormone receptor, where 6 or 1 is experimentally confirmed as non hot spots (false positives) for $\Delta\Delta G \geqslant 1.0$ or $\Delta\Delta G \geqslant 2.0$ residues , respectively. And the second contains a total of 9 or 4 residues of thrombin heavy chain for $\Delta\Delta G \geqslant 1.0$ or $\Delta\Delta G \geqslant 2.0$ residues, respectively. Thus all the residues in the GCR of thrombin heavy chain are true hot spots. There are two residues, PHE323 and GLY368, in the GCR (burial level not less than 1.0) of growth hormone receptor that have not been experimentally mutated, thus we cannot tell with 100% confidence whether these residues are true hot residues or not. However, both of them are identified as computational hot spots by HotSprint [10].

3.2 Burial level versus $\Delta\Delta G$

The burial level of a residue is aimed to measure the extent to which the residue is sheltered from the attack of bulk solvent water. It is more informative than the SASA value because those slightly buried residues and those deeply buried residues may both have a low SASA value. By using the burial level value only (without satisfying the interface and the k-core condition), we can identify

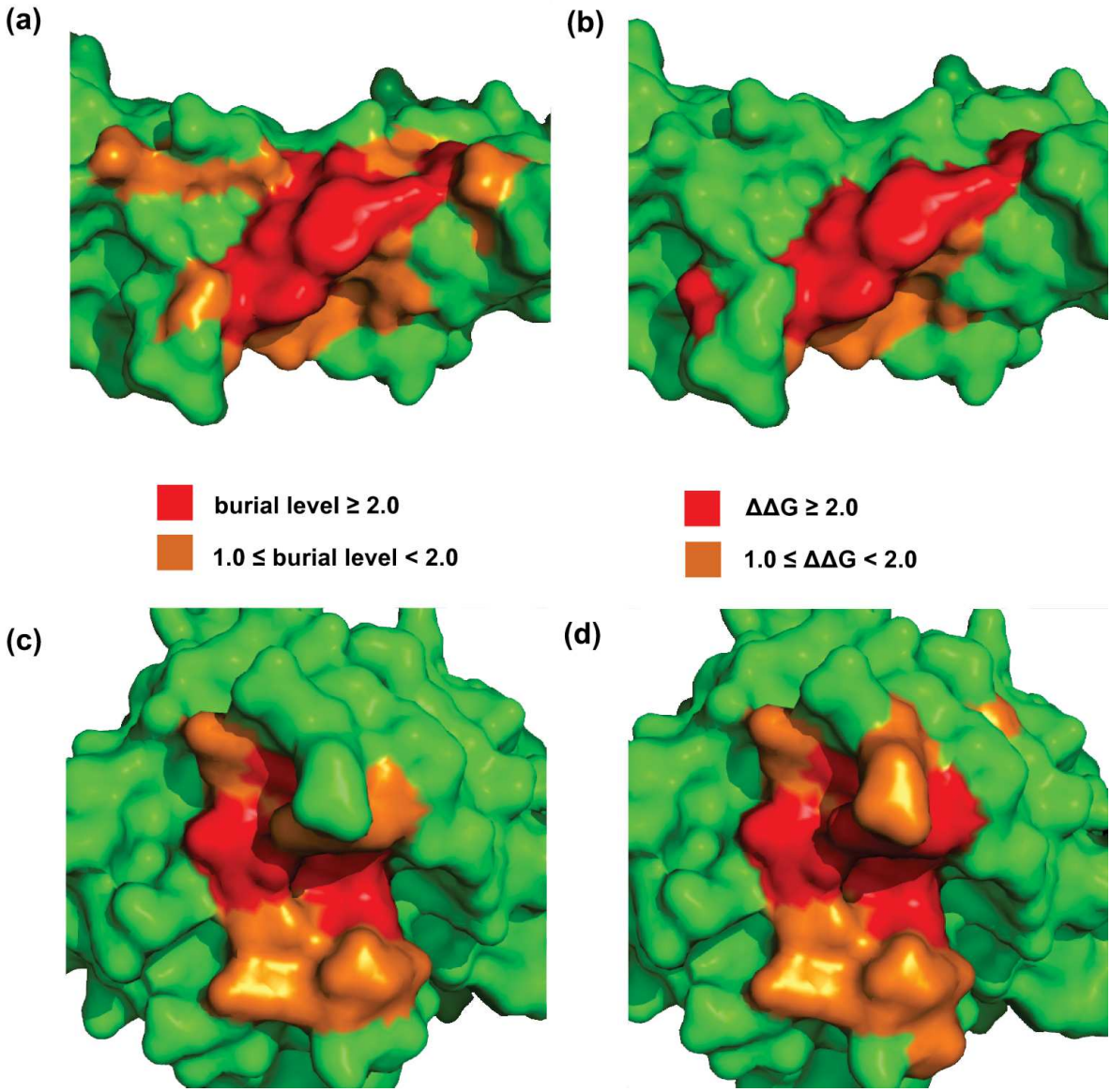
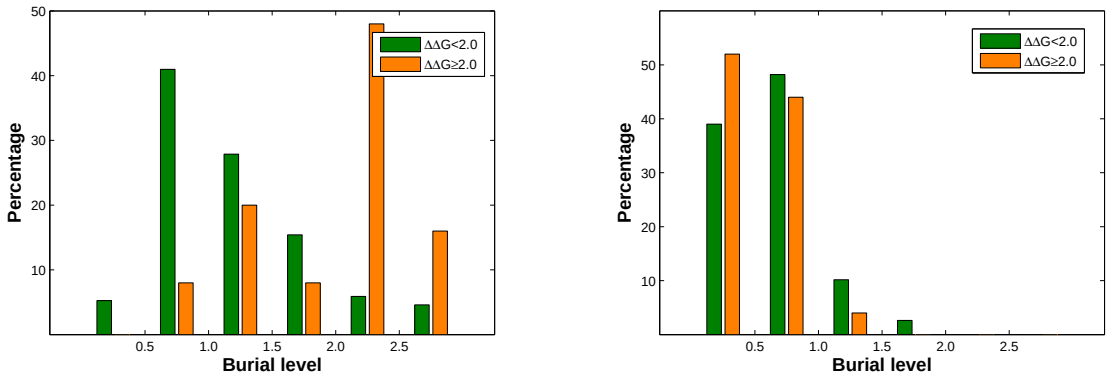


Figure 2: Two examples of GCRs and the true hot spot residues. The two sub-figures on the left panel show the GCRs, and the right-side ones show the true hot spot residues. (a) and (b): Growth hormone receptor binding with growth hormone, chain B of PDB structure 1A22. (c) and (d): Thrombin heavy chain binding with thrombomodulin, chain M of PDB structure 1DX5.

the $\Delta\Delta G \geqslant 1.0$ (or $\Delta\Delta G \geqslant 2.0$) residues with an F1 value of 0.5856 (or 0.5257), which is a not-too-bad performance although lower than that of the final model. Thus we conducted an analysis to understand the relation between the burial level and the $\Delta\Delta G$.

The distribution of burial level values in hot spot ($\Delta\Delta G \geqslant 2.0$) and non hot spot residues is depicted in Fig. 3(a). We can see that most of the hot spot residues ($\Delta\Delta G \geqslant 2.0$) have a large burial

level (⩾2.0). The non hot spot residues tend to have a low burial level, with nearly 90% of them having burial level below 2.0. This is why the GCR model can successfully eliminate most non hot spot residues.



(a) (b)

Figure 3: Distribution of the burial level values in hot spot and non hot spot residues. (a)Buried water as integral part of the complex. (b)Buried water as part of the bulk solvent.

3.3 Role of immobilized water in protein binding hot spots

Water molecules at the interface can fill the cavities and can form hydrogen bonds to strengthen and bridge the residue contacts [18]. It was found that protein interfaces can be as "wet" as the protein surfaces in terms of the number of immobilized water molecules [17]. Interface water molecules can be important both energetically [37] and functionally [38]; The explore of water for protein complex modeling has been proven to be helpful to protein docking [39]. However, very few protein-protein interaction studies have taken water information into consideration; the methods that explicitly use water information for protein binding hot spots prediction are even rarer. As introduced, we favor a "wet" model for protein binding hot spots that does not exclude water molecules.

We implemented this idea in two ways in the current work. First, the oxygen atoms of the buried water molecules are considered as part of the atom contact graph and hence the atom burial levels of all atoms can be affected by the water information. A buried water molecule is sheltered by other atoms, and it shelters other atoms as well. Second, the interface buried water molecules are treated as one of the three parts of the tripartite graph. This definition of interface considers those residues that contact with the other side indirectly through water molecules as interface residues, too. Thus for the model itself, the water information is required to get the right burial level values of

atoms/residues and to identify a proper region as the protein interface. Moreover, by allowing water molecules as nodes in the interface tripartite graph, we can apply the strict constraint, the 3-core, for the GCR to eliminate more non-hot spot residues. If water information is ignored in our model, only 5 $\Delta\Delta G \geqslant 1.0$ residues or 2 $\Delta\Delta G \geqslant 2.0$ residues can be correctly identified, indicating that water information is powerful in the hot spot prediction.

Most of the GCRs identified contain water molecules to a certain degree. This suggests that the hot spots are not absolutely water-free, instead they possess a proper amount of built-in and immobilized water molecules. Teyra and Pisabarro [40] had an interesting definition for interface residues. If a residue interacts with the other side only through water, it is defined as a wet spot; if a residue has both direct and indirect (interact via interface water) interactions with the other side, it is defined as a dual residue; and if a residue interacts only directly with the other side, it is defined as a dry residue. They found that the wet spots can contribute to the binding energetically and are different from surface residue in terms of their mobility [41]. This indicates that wet spots are part of the interface. We have examined the $\Delta\Delta G$ of interfacial mutations in our data set and their types according to this definition, as shown in Fig. 4. It can be noted that the wet spots have relatively low $\Delta\Delta G$, which is understandable because the direct contribution of water-mediated contacts are found not very strong by double mutant cycle analysis [18]. Surprisingly, those dry residues also do not have high $\Delta\Delta G$ values. For most of them the $\Delta\Delta G$ is lower than 2.0. In the 50 $\Delta\Delta G \geqslant 2.0$ residues in our data set, there are 46 dual residues, 1 wet spot, but only 3 dry residues. And in the 18 outstanding residues whose $\Delta\Delta G \geqslant 4.0$, there is only 1 dry residue with all the rest being dual residues. This clearly indicates that hot spot residues benefit a lot from nearby interface buried water molecules.

The buried water is critically important for the hot spot residues to make significant energetic contribution to the binding. Usually the binding site of two proteins do not share a 100% shape complementarity, thus the cavities and gaps between the two sides are filled with immobilized water molecules which can offer a stable local region for the buried hydrogen bonds and salt bridges that are vulnerable to bulk solvent [18, 38]. The ability of buried water molecules to stabilize the local environment can be viewed by the low mobility of wet spots [41] and the even lower mobility of dual residues [40]. These immobilized water molecules can also contact with both sides to form indirect residue contacts making insignificant yet positive contribution to the binding free energy [16]. Those deeply buried hydrogen bonds are the strongest hydrogen bonds [42], and the residues who possess

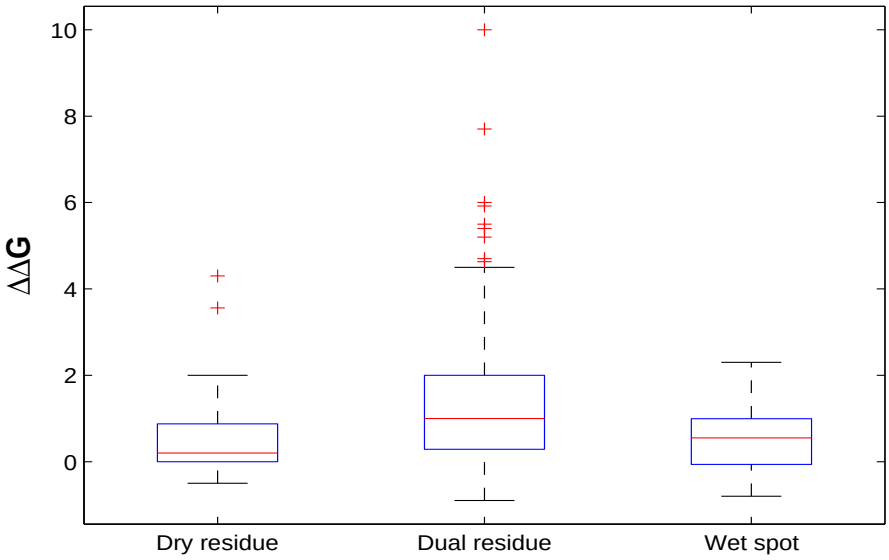
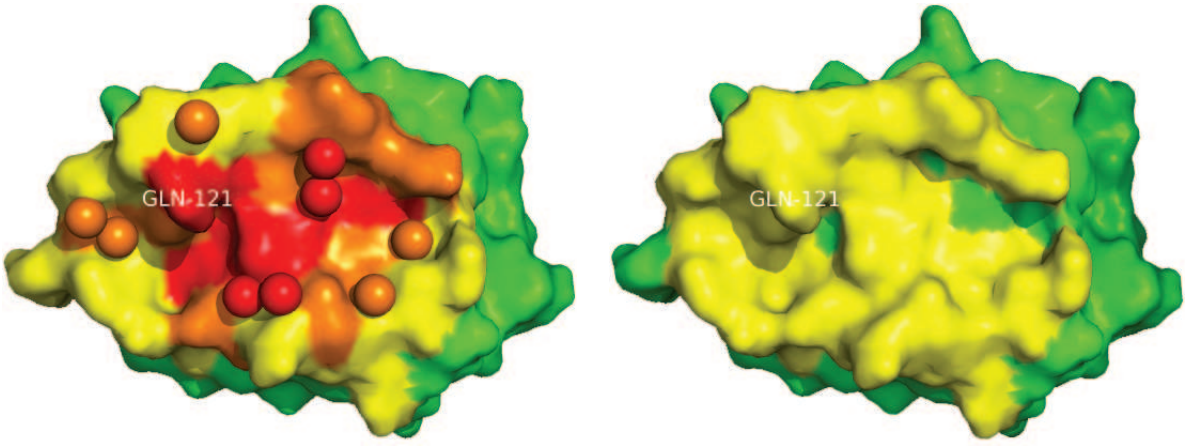


Figure 4: Distribution of $\Delta\Delta G$ of dry residues, dual residues and wet spots.

them are very likely to become hot spots. In our model, if we do not differentiate buried water from bulk solvent water, the burial level of the residues implies little information of the $\Delta\Delta G$. Fig. 3(b) shows the distribution of burial level values in hot spot ($\Delta\Delta G \geqslant 2.0$) and non hot spot residues if the buried water molecules are considered as the same as the bulk solvent. This is done by setting atoms contacting with buried water to contact with the pseudo atom. In this figure the hot spot residues no longer have a higher burial level, and the burial level can not tell the difference between hot spot and non hot spot any more.

The two pictures in Fig. 5 show the burial level patterns in the binding site of the hen egg white lysozyme (binding with IGG1-kappa D1.3 FV) when considering buried water as part of the protein complex or as part of the bulk solvent. When water is considered as part of bulk solvent, the interface residues tend to have lower values of burial level and subsequently the number of interface residues is smaller since the water-bridged contacts are eliminated. Residue GLN121, which is located at the upper left of the figure and stretches to the other side, has a $\Delta\Delta G$ of 2.9. This residue can be successfully identified as a hot spot residue according to our GCR model. But when the buried water is treated as part of bulk solvent, it is very poorly buried (burial level: 0.22) and hence it is not in the corresponding GCR any more.

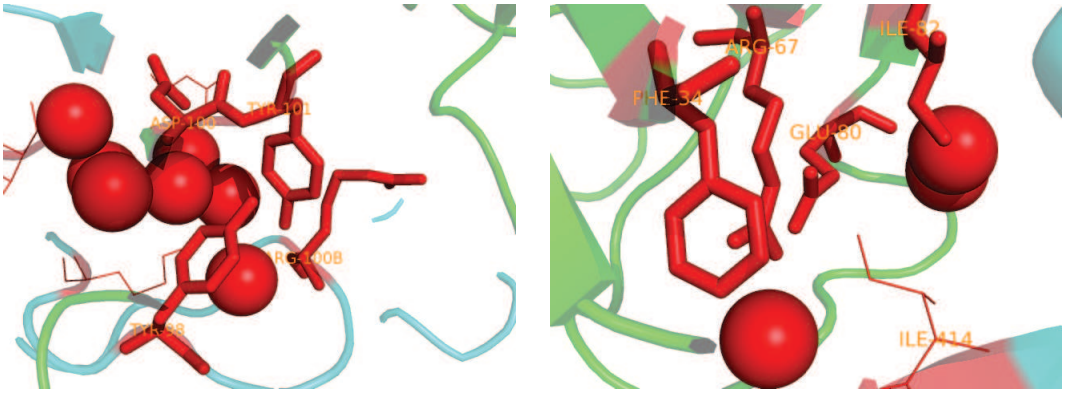
Fig. 6 shows two GCRs identified from the FV D1.3-FV D5.2 and thrombin-thrombomodulin



(a) (b)

Figure 5: Burial level of the residues on the binding site of hen egg white lysozyme (binding with IGG1-kappa D1.3 FV, PDB id: 1VFB, chain C) (a)when buried water is regarded as integral part of the complex and (b)when buried water is regarded as part of the bulk solvent. Non interface residues are in green, residues with burial level smaller than 1.0 are shown in yellow, residues with burial level no less than 1.0 and smaller than 2.0 are shown in orange and residue with burial level no less than 2.0 are in red. The spheres in (a) are water molecules.

interfaces. The first GCR (Fig. 6(a)) is relatively water-richer. The experimentally identified hot spot residues ($\Delta\Delta G \geqslant 2.0$) TYR101 and ASP100 from chain B and ARG100B and TYR98 from chain D are clustered together into a compact contact region while an amount of water molecules reside inside this hot spot cluster. The GCR in the second example (Fig. 6(b)) is relatively "dryer". There are only three water molecules in this GCR; the experimentally identified hot spot residues (PHE43, ARG67, ILE82 and GLU80 from chain M) are clustered into a local region all in a close contact with ILE414 from chain I directly. The three water molecules are not at the center of this region. This is an example that agrees with the results in [15], where hot spot clusters were assumed to be water-free biclique patterns. Again, interface buried water molecules and their immobilization during the association can strengthen the hydrogen bonds and salt bridges nearby, and hence the hot spots emerge. Structurally without water to model binding hot spots, the method would loss much accuracy in prediction.



(a) (b)

Figure 6: Examples of relatively water-richer (a, PDB id: 1DVF) and dryer (b, PDB id: 1DX5) GCRs with burial level not less than 2.0. The two interaction partners are in cyan and green, and GCRs are in red. Experimentally identified hot spot residues are shown in sticks, and the water molecules in the GCR are shown as spheres.

3.4 O-ring of protein binding hot spots

The existence of a ring of residues that protect the hot spot residues from the attack of solvent bulk water is inferred by the low level of SASA of hot spot residues as stated by the long-standing O-ring hypothesis [13]. Therefore, it is difficult to determine an O-ring without a clear definition of the hot spot. Now that we have defined GCRs as hot spots, it is easy to define the O-ring of a GCR as the immediate neighboring residues of the GCR. As described above, the atom/residue contact can be regarded as a sheltering relation between atoms/residues, thus an O-ring can also be regarded as the minimum set of residues that protects the GCR to preserve the current low-SASA state.

The O-ring hypothesis also says that the O-ring is energetically less important than the hot spot. Therefore, we examine whether the energetic contribution by the O-rings is truly less important than the hot spots. Fig. 7 shows a box plot of the $\Delta\Delta G$ of the mutations on the O-rings versus that on the GCRs with burial level no less than 2.0. It can be seen that the $\Delta\Delta G$ of the O-ring residues is significantly lower than that of the GCR residues. A few mutations on the O-rings have high values of $\Delta\Delta G$, but most of them are not. This result suggests that there indeed exists a ring of energetically less important residues that protect the GCR hot spots from the attack of bulk solvent water.

We have explained that the wet spots are not likely to be hot spot residues, thus only 7.8% of the GCR residues are wet spots. However, when we examined the occurrence of wet spots in the

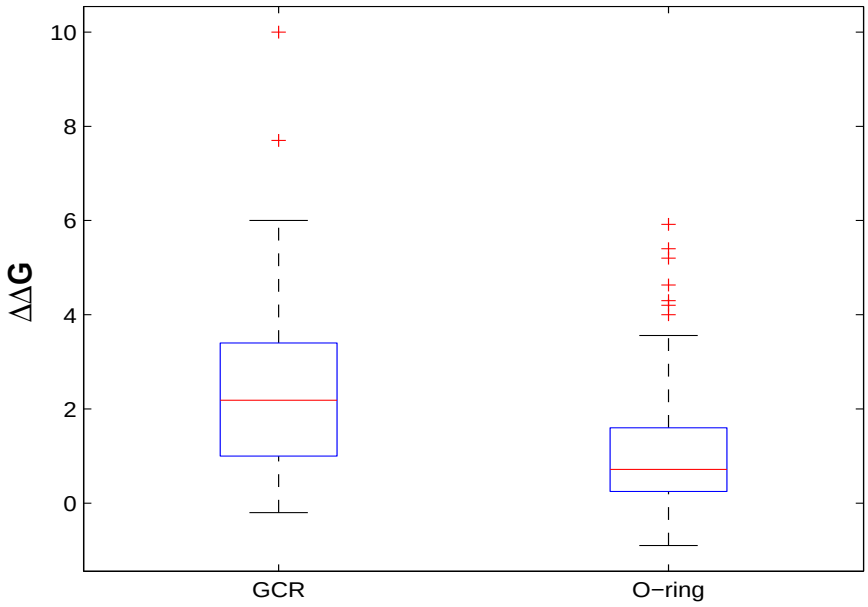


Figure 7: Distribution of $\Delta\Delta$G on the O-rings and the GCRs.

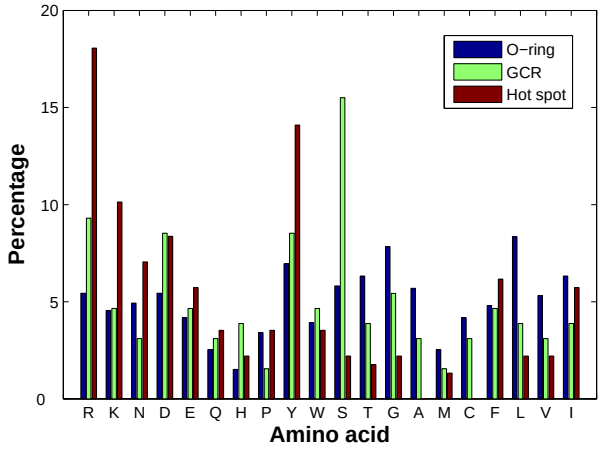


Figure 8: Residue compositions of the O-rings, GCRs and hot spots. The amino acids are ordered by the Kyte and Doolittle hydropathy index [43].

O-rings of the GCRs, we found that about 22.6% of the interfacial O-ring residues are wet spots. This shows that wet spots play an important role in the O-rings. They may not be able to become hot spot residues themselves but they are good protector of the hot spot residues. This emphasizes the indirect energetic contribution of wet spots as well as interface water molecules to the binding of the two proteins.

We also calculated the residue composition of the O-rings to compare with that of the GCRs (burial level not less than 2.0) and that of the $\Delta\Delta G \geqslant 2.0$ hot spot residues in ASEdb [19]. In ASEdb there are 3043 mutations, of which 227 are $\Delta\Delta G \geqslant 2.0$ hot spot residues. The result is shown in Fig. 8. In general, GCRs have a similar residue preference as true hot spots do—they both are rich in ARG, ASP, and TYR and both do not prefer hydrophobic residues, whereas the O-rings do not show such a tendency. It also can be noted from Fig. 8 that although the residue composition of GCRs is not perfectly correlated with that of hot spot residues, GCRs shows the same residue type preference as hot spot residues in comparison with O-ring residues most of the time. For example, the most hydrophillic residue ARG (as by Kyte and Doolittle hydropathy index [43]) is more preferred in hot spots than in O-rings, and it is also more preferred in GCRs than in O-rings. We have also compared these residue compositions to that of the Swiss-Prot database (`http://www.expasy.ch/sprot/relnotes/relstat.html`) for a more comprehensible understanding. One of our findings is that the residue composition of the O-rings is quite similar to that of the Swiss-Prot database with a correlation coefficient of 0.6404, while the compositions of the GCRs and the true hot spots are significantly different from that of the Swiss-Prot database, with a correlation coefficient of 0.1503 and -0.0490, respectively. This means that the O-rings possess a less characteristic amino acid preference, conversely implying that GCR hot spot is a different object from O-ring.

4 Conclusion

We have proposed a model named geometrically centered region to predict protein binding hot spots based on a novel definition of atom burial level that can be simply determined by Dijkstra's algorithm. This model emphasizes the nature of hot spots: the compactness in contact and the distance away from the bulk solvent. Our model achieved an F measure of 0.6414 or 0.5859 for the hot spots defined by $\Delta\Delta G \geqslant 1.0$ or $\Delta\Delta G \geqslant 2.0$, respectively. This performance is better than the benchmark method Robetta.

To our knowledge, the GCR model is the first hot spot prediction method that explicitly takes water information into consideration. The buried water is treated as the integral part of the protein structure, and so for the interface buried water in the interface. Our analysis result show that the presence of nearby buried water molecules is beneficial to the energetic contribution of hot spot residues. The immobilization of these buried water molecules during the association of proteins can stabilize the local region and strengthen the buried hydrogen bonds and salt bridges. Without considering these important roles of water, the performance of the prediction model is very poor.

With the definition of GCRs, we presented a clear definition for O-rings which enables an easy computation. The $\Delta\Delta G$ distribution and amino acid composition of the O-rings that surround the GCRs confirm that such O-rings are indeed energetically less important than the hot spots. Thus this result double confirms the good prediction performance by our GCR model for the binding hot spots.

Acknowledgement

This research work was funded by a Singapore MOE ARC Tier-2 grant (T208B2203).

References

[1] BC Cunningham and JA Wells. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science*, 244(4908):1081–1085, 1989.

[2] James A. Wells. Systematic mutational analyses of protein-protein interfaces. In *Molecular Design and Modeling: Concepts and Applications Part A: Proteins, Peptides, and Enzymes*, volume 202 of *Methods in Enzymology*, pages 390 – 411. Academic Press, 1991.

[3] T. Clackson and J. A. Wells. A hot spot of binding energy in a hormone-receptor interface. *Science*, 267(5196):383–386, January 1995.

[4] T. Clackson, M. H. Ultsch, J. A. Wells, and A. M. de Vos. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *Journal of Molecular Biology*, 277(5):1111–1128, April 1998.

[5] Irina S S. Moreira, Pedro A A. Fernandes, and Maria J J. Ramos. Hot spots-a review of the protein-protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*, 68(4):803–812, June 2007.

[6] Irina Massova and Peter A. Kollman. Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. *Journal of the American Chemical Society*, 121(36):8133–8143, September 1999.

[7] Tanja Kortemme and David Baker. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A*, 99(22):14116–14121, October 2002.

[8] Tanja Kortemme, David E. Kim, and David Baker. Computational alanine scanning of protein-protein interfaces. *Sci. STKE*, 2004(219):pl2+, February 2004.

[9] Steven J. Darnell, David Page, and Julie C. Mitchell. An automated decision-tree approach to predicting protein interaction hot spots. *Proteins: Structure, Function, and Bioinformatics*, 68(4):813–823, September 2007.

[10] E. Guney, N. Tuncbag, O. Keskin, and A. Gursoy. Hotsprint: database of computational hot spots in protein interfaces. *Nucleic Acids Research*, 36(Database):D662–D666, December 2007.

[11] Kyu-Il Cho, Dongsup Kim, and Doheon Lee. A feature-based approach to modeling protein-protein interaction hot spots. *Nucleic Acids Research*, 37(8):2672–2687, May 2009.

[12] Stefano Lise, Cedric Archambeau, Massimiliano Pontil, and David Jones. Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. *BMC Bioinformatics*, 10(1):365+, October 2009.

[13] Andrew A. Bogan and Kurt S. Thorn. Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, 280(1):1–9, July 1998.

[14] O. Keskin, B. Ma, and R. Nussinov. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. *Journal of Molecular Biology*, 345(5):1281–1294, February 2005.

[15] Jinyan Li and Qian Liu. "double water exclusion": a hypothesis refining the o-ring theory for the hot spots at protein interfaces. *Bioinformatics*, 25(6):743–750, March 2009.

[16] Teikichi Ikura, Yoshiaki Urakubo, and Nobutoshi Ito. Water-mediated interaction at a protein-protein interface. *Chemical Physics*, 307(2-3):111 – 119, 2004.

[17] F. Rodier, R. P. Bahadur, P. Chakrabarti, and J. Janin. Hydration of protein-protein interfaces. *Proteins: Structure, Function, and Bioinformatics*, 60(1):36–45, July 2005.

[18] Dana Reichmann, Yael Phillip, Asaf Carmi, and Gideon Schreiber. On the contribution of water-mediated interactions to protein-complex stability. *Biochemistry*, 47(3):1051–1060, January 2008.

[19] K. S. Thorn and A. A. Bogan. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, 17(3):284–285, March 2001.

[20] Brian C. Cunningham and James A. Wells. Comparison of a structural and a functional epitope. *Journal of Molecular Biology*, 234(3):554 – 563, 1993.

[21] DA Dougan, RL Malby, LC Gruen, AA Kortt, and PJ Hudson. Effects of substitutions in the binding surface of an antibody on antigen affinity. *Protein Engineering*, 11(1):65–74, 1998.

[22] A. O. Pineda, A. M. Cantwell, L. A. Bush, T. Rose, and E. Di Cera. The thrombin epitope recognizing thrombomodulin is a highly cooperative hot spot in exosite i. *The Journal of Biological Chemistry*, 277(35):32015–32019, August 2002.

[23] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, January 2000.

[24] I. N. Shindyalov and P. E. Bourne. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*, 11(9):739–747, September 1998.

[25] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. *Journal of Molecular Biology*, 55(3):379–400, February 1971.

[26] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, December 1959.

[27] Herbert Edelsbrunner. Weighted alpha shapes. Technical report, Champaign, IL, USA, 1992.

[28] D. Kim, J. Seo, D. Kim, J. Ryu, and C. Cho. Three-dimensional beta shapes. *Computer-Aided Design*, 38(11):1179–1191, November 2006.

[29] Alessandro Pintar, Oliviero Carugo, and Sandor Pongor. Atom depth as a descriptor of the protein interior. *Biophysical Journal*, 84(4):2553–2561, April 2003.

[30] Benjamin Bouvier, Raik Grünberg, Michael Nilges, and Frédéric Cazals. Shelling the voronoi interface of protein-protein complexes reveals patterns of residue conservation, dynamics, and composition. *Proteins: Structure, Function, and Bioinformatics*, 76(3):677–692, January 2009.

[31] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[32] Raphael Guerois, Jens E. Nielsen, and Luis Serrano. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *Journal of Molecular Biology*, 320(2):369–387, July 2002.

[33] Joost Schymkowitz, Jesper Borg, Francois Stricher, Robby Nys, Frederic Rousseau, and Luis Serrano. The foldx web server: an online force field. *Nucleic Acids Research*, 33, July 2005.

[34] Alexander Benedix, Caroline M. Becker, Bert L. de Groot, Amedeo Caflisch, and Rainer A. Bockmann. Predicting free energy changes using structural ensembles. *Nature Methods*, 6(1):3–4, January 2009.

[35] Navin Pokala and Tracy M. Handel. Energy functions for protein design: Adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *Journal of Molecular Biology*, 347(1):203–227, March 2005.

[36] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947.

[37] Y. Lu, C. Y. Yang, and S. Wang. Binding free energy contributions of interfacial waters in hiv-1 protease/inhibitor complexes. *J. Am. Chem. Soc.*, 128(36):11830–11839, September 2006.

[38] James D. R. Knight, Donald Hamelberg, Andrew J. Mccammon, and Rashmi Kothary. The role of conserved water molecules in the catalytic domain of protein kinases. *Proteins: Structure, Function, and Bioinformatics*, 76:527–535, 2009.

[39] Aalt D J D. van Dijk and Alexandre M J J M. Bonvin. Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics*, 22(19):2340–2347, 2006.

[40] Joan Teyra and M. T. Pisabarro. Characterization of interfacial solvent in protein complexes and contribution of wet spots to the interface description. *Proteins: Structure, Function, and Bioinformatics*, 67(4):1087–1095, 2007.

[41] Sergey Samsonov, Joan Teyra, and Teresa M. Pisabarro. A molecular dynamics approach to study the importance of solvent in protein interactions. *Proteins: Structure, Function, and Bioinformatics*, 73:515–525, 2008.

[42] A. Fernández and H. A. Scheraga. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc Natl Acad Sci U S A*, 100(1):113–118, January 2003.

[43] J. Kyte and R. Doolittle. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, 157(1):105–132, May 1982.