Are hot-spots occluded from water?

Irina Sousa Moreira*, Rui Miguel Ramos, Joao Miguel Martins, Pedro Alexandrino Fernandes and Maria Joao Ramos

Faculdade de Ciencias do Porto, Department of Chemistry and Biochemistry, University of Porto, Rua do Campo ALegre 687, Porto, Portugal

(Received 25 October 2012; final version received 11 December 2012)

Protein–protein interactions are the basis of many biological processes and are governed by focused regions with high binding affinities, the warm- and hot-spots. It was proposed that these regions are surrounded by areas with higher packing density leading to solvent exclusion around them – “the O-ring theory.” This important inference still lacks sufficient demonstration. We have used Molecular Dynamics (MD) simulations to investigate the validity of the O-ring theory in the context of the conformational flexibility of the proteins, which is critical for function, in general, and for interaction with water, in particular. The MD results were analyzed for a variety of solvent-accessible surface area (SASA) features, radial distribution functions (RDFs), protein–water distances, and water residence times. The measurement of the average solvent-accessible surface area features for the warm- and hot-spots and the null-spots, as well as data for corresponding RDFs, identify distinct properties for these two sets of residues. Warm- and hot-spots are found to be occluded from the solvent. However, it has to be borne in mind that water-mediated interactions have significant power to construct an extensive and strongly bonded interface. We observed that warm- and hot-spots tend to form hydrogen bond (H-bond) networks with water molecules that have an occupancy around 90%. This study provides strong evidence in support of the O-ring theory and the results show that hot-spots are indeed protected from the bulk solvent. Nevertheless, the warm- and hot-spots still make water-mediated contacts, which are also important for protein–protein binding.

Keywords: solvent-accessible surface area; hot-spots; interfacial water; O-ring theory; occlusion from water

1. Introduction

Protein–protein interactions (PPI) are the basis of many biological processes ranging from signal transduction and enzymatic regulation to the adhesion of cells. The huge importance of these interactions made protein–protein binding an area of significant interest in the pharmaceutical industry. PPI are governed by focused regions with high binding affinities, the hot-spots (Clackson, Ultsch, Wells, & de Vos, 1998; DeLano, Ultsch, de Vos, & Wells, 2000). The characterization of these structural key points is crucial to the development of means to inhibit protein–protein association, a therapeutic modality proposed for many diseases. Hot-spots were defined as residues whose mutation by alanine destabilizes the bound state ensemble relative to the unbound one (DeLano, 2002), generating a binding free energy difference ($\Delta G_{binding}$) higher than 4.0 kcal/mol. Warm-spots, also important for binding, upon alanine mutation generate a binding free energy difference between 2.0 and 4.0 kcal/mol; the definition for null-spots is that a mutation causes a binding free energy difference of $<2.0$ kcal/mol (Moreira, Fernandes, & Ramos 2007e). Hot- and warm-spots are compact, centralized regions, and mutational studies have shown that the tolerance for substitution is clearly lower for these residues than for others at the protein surface (Moreira et al., 2007). Experimental and computational methods have been highly used to identify these types of sites on protein surfaces (Huo, Massova, & Kollman, 2002; Kollman et al., 2000; Kortemme, Kim, & Baker, 2004; Massova & Kollman, 1999; Moreira, Fernandes, & Ramos, 2006a, 2006b; Moreira et al., 2007a; Moreira et al., 2007b; Moreira et al., 2007c; Moreira et al., 2007d; Moreira et al., 2007e; Moreira et al., 2007f; Moreira, Fernandes, & Ramos, 2008; Martins, Ramos, & Moreira, 2013). Results showed that the hot-spot of one monomer usually packs against the hot-spot of another, thereby establishing a region that is determinant for complex binding, and thus, may be a target site for drug discovery. The number of such hot-spots within densely packed regions is correlated essentially with the interface size, and the local organization of these hot-spots is a critical factor in stabilizing protein–protein interactions. Notably,
however, Chothia and Janin have proposed a very different theory: the Buried Surface model (BSA) (Chothia & Janin, 1975; Janin & Chothia, 1990; Janin, 2009) that states that the interaction is distributed more or less evenly over the surface of proteins (Chakrabarti & Janin, 2002). Yet, these two apparently opposed theories can be reconciled. Closer inspection shows that protein–protein interfaces can be split into a core and a rim region, according to the solvent accessibility of interface atoms. While the rim is very similar to the protein’s surface and displays very few hot-spots, the core contains the buried interface atoms and a larger number of hot-spots (Guharoy & Chakrabarti, 2005; Gharoy, Pal, Dasgupta, & Chakrabarti, 2011). Indeed, Gharoy and Chakrabarti clearly demonstrated a correlation between the $\Delta \Delta G_{\text{binding}}$ and the contribution to the BSA of the buried residues of the core, a relationship that is not discernible for the rim residues (Guharoy & Chakrabarti, 2005; Gharoy et al., 2011).

It had been proposed earlier that hot-spots are surrounded by regions with higher packing density leading to solvent exclusion around them, which results in a lower local dielectric constant and enhancement of specific electrostatic and hydrogen bond (H-bond) interactions. As these structures resembled an O-ring (see Figure 1), the idea became known as the O-ring theory, or the ‘Water Exclusion’ hypothesis (Bogan & Thorn, 1998). This theory led to the inference that warm- and hot-spots tend to have low solvent-accessible surface area (SASA) values. (Bogan & Thorn, 1998) also speculated that warm- and hot-spots tend to be surrounded by energetically less important residues that would function to occlude bulk solvent from them. This hypothesis was already proven not to be universal by other authors. Kosloff, Travis, Bosch, Siderovski, and Arshavsky (2011) have shown for the RGS-domain-G-protein interface that peripheral residues protect the hot-spots from the solvent and have essential roles in recognition, by optimizing the protein–protein specificity and activation (Kosloff et al., 2011). Li and colleagues also proposed a ‘double water exclusion’ theory in which they suggest that there is a ring of residues surrounding the warm- and hot-spots for avoiding the water invasion after binding and that these residues themselves are water free (Li & Liu, 2009). Rajamani et al. have also proposed that there are some residues at protein–protein interfaces that upon complex formation yield the largest decrease in SASA (Rajamani, Thiel, Vajda, & Camacho, 2004). These anchor residues were demonstrated to be functionally important (hot-spots) or kinetically important (Meireles, Domling, & Camacho, 2010; Rajamani et al., 2004). Bogan and Thorn’s (1998) original idea, the O-ring theory, was used by different authors to achieve new computational alanine scanning mutagenesis methods (Guharoy & Chakrabarti, 2005, 2010; Gharoy et al., 2011; Guney, Tuncbag, Keskin, & Gursoy, 2007; Li & Li, 2010; Liu & Li, 2010; Li, Wong, & Li, 2011; Tuncbag & Keskin, 2009; Xia, Zhao, Song, & Huang, 2010; Zhu & Mitchell, 2011), but the theory still lacks adequate experimental support.

Nowadays, it is accepted that the biological function of proteic macromolecular systems depends on a delicate interchange between the molecule itself and its function to occlude bulk solvent from them. This hypothesis was already proven not to be universal by other authors. Kosloff, Travis, Bosch, Siderovski, and Arshavsky (2011) have shown for the RGS-domain-G-protein interface that peripheral residues protect the hot-spots from the solvent and have essential roles in recognition, by optimizing the protein–protein specificity and activation (Kosloff et al., 2011). Li and colleagues also proposed a ‘double water exclusion’ theory in which they suggest that there is a ring of residues surrounding the warm- and hot-spots for avoiding the water invasion after binding and that these residues themselves are water free (Li & Liu, 2009). Rajamani et al. have also proposed that there are some residues at protein–protein interfaces that upon complex formation yield the largest decrease in SASA (Rajamani, Thiel, Vajda, & Camacho, 2004). These anchor residues were demonstrated to be functionally important (hot-spots) or kinetically important (Meireles, Domling, & Camacho, 2010; Rajamani et al., 2004). Bogan and Thorn’s (1998) original idea, the O-ring theory, was used by different authors to achieve new computational alanine scanning mutagenesis methods (Guharoy & Chakrabarti, 2005, 2010; Gharoy et al., 2011; Guney, Tuncbag, Keskin, & Gursoy, 2007; Li & Li, 2010; Liu & Li, 2010; Li, Wong, & Li, 2011; Tuncbag & Keskin, 2009; Xia, Zhao, Song, & Huang, 2010; Zhu & Mitchell, 2011), but the theory still lacks adequate experimental support.

Nowadays, it is accepted that the biological function of proteic macromolecular systems depends on a delicate interchange between the molecule itself and its

![O-ring structure](image-url)

Figure 1. Top view of the complex formed between IgG1 Kappa D1.3 FV and Hen Egg White Lysozyme. An O-ring structure can be seen with the warm- and hot-spots, colored in red, surrounded by null-spots, colored in light orange.
environment (Ball, 2008). Water plays a key role in understanding the interplay between structure and function, which is a central goal in protein science (Levy & Onuchic, 2006). It is imperative to understand the water’s role around the energetically important residues. Are these residues really occluded from solvent as stated by the O-ring theory? Are the warm- and hot-spots themselves water free as stated by the ‘double water exclusion’ theory (Li & Liu, 2009)? As proteins are dynamic polymers and their conformational flexibility is critical for function, particularly for interactions with other molecules (Chuang et al., 2010), the correct analysis of the PPI requires dynamical information. Thus, in order to achieve a so long needed answer to the questions raised above, we characterize different SASA features for a hot-spots database of various complexes subjected to explicit water MD simulation. We have also measured the RDF of water around them, water occupancy, and protein–water distances. This allowed us to provide strong structural evidence for the O-ring theory.

2. Material and methods

2.1. System setup

Nine different complexes for a total of 160 mutations were studied. Complexes were chosen based on their different interfacial characteristics (i.e. size, hydrophobicity, and amino-acid composition) in order to have a meaningful database. A complete list is given in Table 1. The protonation states of the different residues of the various proteins were determined using the PDB2PQR server at http://kryptonite.ncbr.net/pdb2pqr/ (Dolinsky, Nielsen, McCammon, & Baker, 2004) by the PROPKA methodology (Bas, Rogers, & Jensen, 2008; Li, Robertson, & Jensen, 2005; Olsson, Sondergaard, Rostkowski, & Jensen, 2011). The binding free energy differences values upon alanine mutation were taken out from the ASEdb (Alanine Scanning Energetics database) (Bogan & Thorn, 1998; Thorn & Bogan, 2001) and from references in Table 1.

2.2. Molecular dynamic simulations

The MD simulations were performed using the AMBER9 package (Case et al., 2004) with the Cornell force field (Cornell et al., 1995). Table 1 summarizes the composition of a variety of systems under study. The complexes were solvated by explicit TIP3P water molecules that extended 10 Å from any edge of the box to the protein atoms; the whole system can, therefore, be replicated in three dimensions and treated with periodic boundary conditions. Counter ions were added to the box to neutralize the systems. In each of the simulations, the system was initially energy minimized to remove bad contacts by steepest descent followed by conjugate gradient algorithms. The systems were then subjected to 2 ns of heating procedure (in NVT ensemble) in which the temperature was gradually raised to 300 K, followed by 6 ns runs in NPT ensemble. The Langevin thermostat (Jackson, Gabb, & Sternberg, 1998; Loncarich, Brooks, & Pastor, 1992) was used and the electrostatic interactions were calculated by using the particle mesh Ewald (PME) method (Darden, York, & Pedersen, 1993). Bond lengths involving hydrogen atoms were constrained to their equilibrium values using the SHAKE algorithm (Ryckaert, Ciccotti, & Berendsen, 1977). The equations of motion were integrated with a 2 fs time-step and the nonbonded interactions were truncated with a 10 Å cutoff.

2.3. Analysis

VMD (Humphrey, Dalke, & Schulten, 1996) and PTRAJ modules from the AMBER9 package (Case et al., 2004) were used in the different analyses carried out from the MD simulations. RMSDs (root-mean-square-deviations) were calculated for each simulation to ensure their equilibration and for the nine complexes ranged between 2.0 and 3.5 Å. Different SASA calculations were made to evaluate the importance of water molecules in the warm- and hot-spots microenvironment. The SASA of each interfacial residue within the complex (SASA_{cpx}) and within the separated monomers (SASA_{mon}) was averaged at the last 3 ns of the explicit MD simulation. SASA, as defined by Lee and Richards, is the area of the surface traced by the center of a probe sphere, whose radius is the nominal radius of the solvent, as it rolls over the van der Waals surface of the molecule (Lee & Richards, 1971). ΔSASA and SASArel defined by Equation (1) and (2), respectively, were also calculated. SASArel allows the differentiation of residues with equal ΔSASA but different solvent exposure. For example, a residue with a 50 Å² solvent accessibility in the monomer and 0 Å² in the complex and a residue that has a value of 150 Å² solvent accessibility in the monomer and 100 Å² in the complex are two very distinct situations. In both cases, ΔSASA is 50 Å² but solvent accessibility importance is strikingly different between the two. SASArel was already shown to be important in a previous work (Cho, Kim, & Lee, 2009).

\[
\Delta \text{SASA} = \text{SASA}_{\text{cpx}} - \text{SASA}_{\text{mon}} \quad (1)
\]

\[
\text{SASArel} = \frac{\text{SASA}_{\text{cpx}} - \text{SASA}_{\text{mon}}}{\text{SASA}_{\text{mon}}} = \frac{\Delta \text{SASA}}{\text{SASA}_{\text{mon}}} \quad (2)
\]

All the four SASA features were analyzed for the 160 residues for the nine explicit water MD simulations. We have also analyzed the RDF, $g(r)$, and the average
number of waters within a given distance, of all interfacial residues. \( G(r) \) gives the probability of finding an atom within a spherical shell of thickness \( \Delta r \) and central distance \( r \) from another atom, in relation to the probability expected for a bulk solvent distribution at the same density. It was calculated by compiling a histogram with a spacing of \( \Delta r = 0.02 \) and a range of 8 Å. Occupancy of waters in the first coordination shell was calculated by python in house scripts, which takes advantage of the ptraj module of the amber package. Tailor-made scripts were also used to evaluate the microenvironment surrounding each interfacial amino acid residue. Geometric criteria (donor–acceptor distance = 3.0 Å, angle cutoff 20°) were adopted to define the formation of an H-bond during the MD simulation. We have used the PISA web-server (Krissinel & Henrick, 2007; Krissinel, 2010) from the European Bioinformatics Institute to calculate interface area composition. This software defines the interface as the protein surface area which becomes inaccessible to solvent upon binding, and so is calculated as the difference in total accessible-surface areas of isolated and interacting structures divided by two:

\[
IA = \frac{ASA(MonomerA) + ASA(MonomerB) - ASA(Complex)}{2}
\] (3)

3. Results and discussion

It is now accepted that only a few residues at each PPI are responsible for the association driving force, named the warm- and hot-spots. The O-ring theory states that these residues tend to be clustered in the center of the interface surrounded by the null-spots to presumably occlude bulk solvent. To understand this phenomenon, we have carried out explicit water MD simulations of several known protein–protein complexes for which experimental \( \Delta \Delta G_{binding} \) values are available.

3.1. Molecular dynamic simulations

Proteins are not static structures and instead, they have an inherent flexibility that leads to a succession of structural changes that carve out a specific pathway. Full-atomistic MD simulations offer a unique opportunity to understand the structural and functional role of water around the warm- and hot-spots. MD stability was ensured by measuring the root-mean-square-deviation (RMSD) values of the backbone of all the residues during the MD simulation. We have used the PISA web-server (Krissinel & Henrick, 2007; Krissinel, 2010) from the European Bioinformatics Institute to calculate interface area composition. This software defines the interface as the protein surface area which becomes inaccessible to solvent upon binding, and so is calculated as the difference in total accessible-surface areas of isolated and interacting structures divided by two:

\[
IA = \frac{ASA(MonomerA) + ASA(MonomerB) - ASA(Complex)}{2}
\] (3)
3.2. Database analysis

Our database is described in the methodological section in Table 1. To better understand the chemical composition of the PPI, we plotted in Figure 3 the percentage of each type of amino acid residue in the protein, interface, and as a warm- and hot-spot. The warm- and hot-spots at the PPIs at our database are enriched in Tyr (25%), Asp (16%), Leu (9%), and Glu (9%). Nevertheless, Phe (7% vs. 3%), Trp (7% vs. 3%), and Met (5% vs. 2%) residues have higher percentages as energetic determinant residues than as PPI residues (% as hot spots vs. % as PPI residues). So, as reported by Bogan and Thorn (1998) tyrosine is over-represented as an energetic crucial residue and valine and threonines are under-represented. Arg is abundant as a warm- and hot-spot but does not exceed the expected level at the PPIs (5% vs. 7%). Cys, Pro, Ala, and Gly do not appear as warm- or hot-spots due to reasons related with the ASM itself or structural motifs. More concretely, Pro unlike other residues can exist in a cis-configuration that translates in a very different backbone conformation, and, therefore, Pro–Ala mutations are disruptive and can produce abnormal changes. Ala for the obvious reasons, and Gly and Cys are not usually mutated in alanine scans as they lead to structural disruption. Our database has 44 warm- and hot-spots that correspond to 25% of the protein–protein interfacial residues. Tyr and Asp residues are widely present in almost all the interfaces. Glu and Trp in smaller percentage and the rest of the warm- and hot-spots are characteristic of the specific protein–protein systems under study.

3.3. O-ring study

3.3.1. Solvent-accessible surface area

The O-ring theory states that occlusion of an amino acid from the bulk solvent plays a decisive role in determining the energetic contribution of that residue to protein–protein binding (Bogan & Thorn, 1998). Later, Guharoy and Chakrabarti (2005) illustrated that there is a correlation between energy change and decrease in the SASA of individual residues as a consequence of protein–protein binding. Solvent accessibility was also combined with conservation in an empirical formula to identify warm- and hot-spots computationally (Guney et al., 2007). We have supported the idea that hot spots are protected from solvent by a rim region (Moreira et al., 2007). Li, Keskin, Ma, Nussinov, and Liang (2004) suggested that warm- and hot-spots are either found on the complemented pockets or on the protruding surfaces to protect each other from the solvent. Various authors have previously focused on solvent accessibility to discriminate warm- and hot-spots residues but a clear understanding is still far from being achieved. In order to evaluate the relationship between solvent accessibility and $\Delta \Delta G_{\text{binding}}$, we measured four different SASA features: $\text{SASA}_{\text{cpx}}$, $\text{SASA}_{\text{mon}}$, $\Delta \text{SASA}$, and $\text{SASAreel}$ for the interfacial residues for which experimental data is available. The most used descriptors in the literature are $\text{SASA}_{\text{cpx}}$ and $\Delta \text{SASA}$. Table SI–1 from supplementary information summarizes the average and standard deviation values for each of the SASA features for the residues tested. Our results demonstrate that high $\Delta \text{SASA}$ and $\text{SASAreel}$ values are a necessary condition for a
residue to be a warm- and hot-spot residue. However, these features are not by themselves effective to distinguish between slightly buried residues, a large part of which are null-spots, and deeply buried residues that are very likely to be hot-spot residues. In fact, as hydrophilicity/hidrophobicity is a key physical property of a proteic surface, and one of the most relevant for PPI, we have also investigated if the nature of the residue has any influence on the qualitative nature of the relationship between SASA features and the $\Delta \Delta G_{\text{binding}}$ values. In Figure 4 are presented the average values for the four SASA features for all the amino acid residues at the database as well as for the negatively charged (Asp and Glu), positively charged (His, Lys and Arg), charged (Asp, Glu, Lys, Arg, His), polar (Ser, Thr, Asn, Gln, Tyr), nonpolar (Val, Ile, Leu, Met, Phe, Trp), and aromatic (Phe, Trp, Tyr, His) residues. Ala, Gly, Cys, and Pro were not considered for the motives mentioned above. There is a striking difference between the average SASA features values between the warm- and hot-spots and the null-spots except in the case of SASA$_{\text{mon}}$. It is clear that the warm- and hot-spots are more buried upon complexation than the null-spots as the SASA$_{\text{mon}}$, $\Delta$SASA, and SASArel are markedly different for these two groups. The values for the null-spots are almost half than for the warm- and hot-spots. This fact is even more stressed out for the positively charged residues. Although the standard deviation of these values go up to around 38% of the measured values, which highlights the difficulty in using SASA features as a sole feature for hot-spot detection, it clearly demonstrates that the energetically important residues are more occluded from solvent.

3.3.2. Radial distribution function

Although warm- and hot-spots are more deeply buried, it cannot be forgotten that water plays a pivotal role in determining the structure and dynamics of most biological systems by establishing specific interactions, screen efficiently coulombic interactions, mediate proton transfer, and can even be used as a structural component in protein secondary structure (Papoian, Ulander, & Wolynes, 2003; Rodier, Bahadur, Chakrabarti, & Janin, 2005). In this paper, we present an overview of the dynamical behavior of water within protein–protein interfaces. The water structure can be depicted in terms of a RDF, which is related to the liquid density $\rho(r)$ around a particle: $\rho(r) = \rho_0 g(r)$, where $\rho_0$ is the bulk density (Ball, 2008). The RDFs represent the probability density of finding any solvent molecule whose oxygen is at distance $r$ from the specific solute atom. Usually, the RDF of water presents an oscillatory profile and a peak due to the presence of H-bonds. The RDF profiles were measured for the residues at our database. As an example, in Figure 5 are plotted two RDF profiles, one for a representative hot-spot and another for a null-spot. They present very different behaviors. The peak around 2.5 Å seen in Figure 5(c) arises from the strong interaction between the water hydrogen atoms with the oxygen atoms of the carbonyl group of the
null-spot. The following peaks, whose location varies according to the system considered, are due to the interaction between water molecules and the nonhydrogen atoms of the amino acid residue. These peaks are less defined for the warm- and hot-spots. Panels b and d show that the typical number of waters around warm- and hot-spots and null-spots are also distinct. These values measured at 3, 4, 5, and 6 Å for the two groups are depicted in Figure 6. The average number of water molecules around warm- and hot-spots is lower than half of the number of water in the null-spots micro-environment. These values are even lower around the aromatic and nonpolar residues than around the charged residues. Figure SI-1 shows the average values by type of residue for the warm- and hot-spots as well as for the null-spots. Gln, His, Tyr, Trp, Phe, and Leu when acting as key energetic determinants have less than two waters around them at a distance of 5 Å. As null-spots, the water molecules that surround them can range on average from 1.48 up to 8.86. The number of waters triples when charged residues act as null-spots. The O-ring theory was proposed upon a statistical study of the \( \text{SASA}_{\text{cpp}} \) and \( \Delta \text{SASA} \) values for static structures. The results presented here, ranging from the SASA calculations to the RDF calculations, take into account the dynamics of protein–protein systems and clearly show that the average number of water molecules in the micro-environment around warm- and hot-spots is much lower than for the null-spots. This study provides strong evidence that indeed the warm- and hot-spots are protected from solvent, as predicted by the O-ring theory.

### 3.3.3. Water’s occupancy

Water-mediated contacts may provide an additional layer of recognition in protein binding. These can be highly specific and complement direct contacts in the task of recognizing a particular binding partner out of many competing proteins (Rodier et al., 2005). Interfacial
Figure 5. Water–protein radial distribution as a function of the distance between water oxygen atoms and the oxygen atoms of the carbonyl groups of Asp35 of Barstar (panel a) and Asp54 of Barnase (panel b), a typical hot-spot and null-spot, respectively, at the 1BRS interface. Panels c and d represent the number of waters around these two residues: c- Asp148 (hot-spot) and d- Asp52 (null-spot).

Figure 6. Average values for the RDF studied for the hot- and warm-spots as well as for the null-spots: 3 Å; 4 Å; 5 Å; and 6 Å (from top left to bottom right). The results are presented for the seven sets of residues analyzed (all, negatively charged, positively charged, charged, polar, aromatic, and nonpolar).
waters are often too disordered to be resolved crystallographically. Therefore, explicit MD simulations can be used to describe the structural and dynamical organization of the H-bonds formed between the solvent and their implications in the dynamical properties of the macromolecular systems. Table 2 shows the average and standard deviation values for these interactions as well as the water maximum occupancy around these residues.

We have focused on interactions involving the interfacial residues of our database. However, in certain cases, they interact with other residues for which no experimental ΔΔG binding information exist, and that for are not under study. These interactions were found for five of the nine complexes studied. The occupancy of water molecules in specific hydration sites does not seem to depend in any simple way on the nature of the residue to which they bind – its polarity or H-bonding ability, for example. In the cases shown, the water occupancy around the warm- and hot-spots is higher than 90%. These solvent molecules contribute by forming H-bonding networks with the warm- and hot-spots cluster. In the 1DFJ complex, Asp435, a hot-spot, establishes a high number of water-mediated H-bonds with the main and side-chain atoms that explain its energetically crucial role. For the 1DVF and 1BRS, the cluster of warm- and hot-spots is tightly packed, establishing H-bonds with waters with high occupancy. For the 1DVF complex, Tyr32_A interacts with Asp207 and Tyr208 by two water molecules. For the 1BRS complex, a two water-mediated interaction is made between Glu73 and Lys27. These interactions are extremely stable as seen in Table 2. In these three complexes, at least two water molecules mediated the H-bonds between warm- and hot-spots from different monomers. In the 1FQ9 and 1FLT complexes, the hot-spots also form water-mediated interactions. For example, in the 1FQ9 complex, the same water molecule connects the energetically important residue Glu96 with two different residues.

### Conclusions

Despite the ongoing effort to decipher the complex nature of protein interactions, these are not still entirely understood. Protein–protein binding is achieved by the connection of small regions with specific chemical and geometric properties that contribute significantly to the binding free energy (over 2.0 kcal/mol), the warm- and hot-spots. Bogan and Thorn proposed that these regions are surrounded by O-rings, which exclude bulk water molecules from them. A useful indicator of how a surrounding medium affects protein structures is represented by the surface area of protein atoms in contact with the solvent molecules (SASA). Although quantities derived from SASA are useful in many applications in protein design and structural biology, the computational cost of accurate SASA calculation in an ensemble of structures makes this calculation expensive. In order to test the O-ring theory, four SASA features were measured for the residues in our database. By measuring the average SASA features for the warm- and hot-spots as well as for the null-spots, it was possible to obtain a clear view of the different behavior of these two sets of residues. RDFs have allowed once more distinguishing between the two groups and undoubtedly warm- and hot-spots are occluded form the solvent. The number of water molecules in their micro-environment is much lower than for the null-spots. This work has finally provided strong evidence that warm- and hot-spots are protected from the bulk solvent validating the O-ring theory. However, we cannot forget that water-mediated interactions have significant power to construct an extensive and strongly bonded interface and indeed, warm- and hot-spots tend to protect the energetically important residues.

### Table 2. Average and standard deviation values for the network interactions mediated by water.

| Complex PDBID | Residue1-water interaction | Distance/Å | Water-residue2 interaction | Distance/Å | Occupancy/% |
|---------------|----------------------------|------------|----------------------------|------------|-------------|
| 1FQ9          | Glu96OE1-Water             | 2.8 ± 0.3  | Water-Arg250NH2            | 3.3 ± 0.4  | 95.8        |
|               |                           |            | Water-Asp282O              | 3.2 ± 0.7  | 69.0        |
|               | Tyr103O-Water              | 1.9 ± 0.2  | Water-Ala167O              | 2.8 ± 0.2  | 99.7        |
| 1FLT          | Tyr21OH-Water              | 2.9 ± 0.3  | Water-Leu204N              | 4.2 ± 0.4  | 99.8        |
| 1DVF          | Tyr32O-Water1              | 3.6 ± 0.5  | Water1-Tyr208O             | 2.8 ± 0.2  | 100.0       |
|               | Tyr32OH-Water2             | 4.0 ± 1.1  | Water2-Asp207OD2           | 3.3 ± 0.7  | 60.8        |
|               | Water1-Water2              |            |                            | 6.6 ± 0.7  | 100.0       |
| 1DFJ          | Asp435OD1-Water1           | 2.9 ± 0.3  | Water1-Val43O              | 4.2 ± 0.4  | 46.6        |
|               | Asp435OD2-Water1           | 3.1 ± 0.3  | Water2-Lys41NZ             | 5.1 ± 0.4  | 83.7        |
|               | Asp435OD1-Water2           | 3.4 ± 0.4  | Water2Phc120N              | 3.5 ± 0.3  | 99.6        |
|               |                           |            | Water2-His119NE2            | 3.7 ± 0.3  | 76.5        |
|               |                           |            | Water2-His12NE2            | 2.9 ± 0.2  | 99.8        |
| 1BRS          | Asp435OD2-Water 2          | 3.1 ± 0.3  | Water1-Lys27 N             | 3.4 ± 0.3  | 99.5        |
|               | Glu73OE1-Water1            | 2.8 ± 0.2  | Water1-Lys27 N             | 3.4 ± 0.3  | 99.5        |
|               | Glu73OE2-Water1            | 3.1 ± 0.3  |                             |            |             |
|               | Glu73OE1-Water2            | 4.3 ± 0.8  | Water1-Water2              | 3.1 ± 0.4  | 100.0       |
|               | Glu73OE2-Water2            | 3.2 ± 0.7  |                             |            |             |
to form H-bond networks with single water molecules that have an occupancy around 90%.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2012.758598.

References
Ball, P. (2008). Water as an active constituent in cell biology. Chemical Reviews, 108, 74–108.
Bas, D. C., Rogers, D. M., & Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein-ligand complexes. Proteins: Structure Function and Bioinformatics, 73, 765–783.
Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall’Acqua, W., … Poljak, R. J. (1994). Bound water molecules and conformational stabilization help mediate an antigen-antibody interaction. Proceedings of the National Academy of Sciences, 91, 1089–1093.
Bogan, A. A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. Journal of Molecular Biology, 280, 1–9.
Braden, B. C., Fields, B. A., Ysenn, X., Dall’Acqua, W., Goldbaum, F. A., Poljak, R. J., & Mariuzza, R. A. (1996). Crystal structure of an Fv–Fv idotipe-anti-idotope complex at 1.9 resolution. Journal of Molecular Biology, 264, 137–151.
Buckel, A. M., Schreiber, G., & Fersht, A. R. (1994). Protein–protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0 – ANG. resolution. Biochemistry, 33, 8878–8889.
Case, D. A., Darden, T. A., Cheatham, T. E. III, Simmerling, C. L., Wang, J., Duke, R. E., … Kollman, P. A. (2004). AMBER 9. San Francisco, CA: University of California.
Castro, M. J. M., & Anderson, S. (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: Effects on the kinetics and thermodynamics of binding to trypsin and chymotrypsin. Biochemistry, 35, 11435–11440.
Chakrabarti, P., & Janin, J. (2002). Dissecting protein–protein recognition sites. Proteins: Structure Function and Genetics, 47, 334–343.
Chen, C. Z., & Shapiro, R. (1997). Site-specific mutagenesis reveals differences in the structural bases for tight binding of RNase inhibitor to angiogenin and RNase A. Proceedings of the National Academy of Sciences of the United States of America, 94, 1761–1766.
Cho, K. J., Kim, D., & Lee, D. (2009). A feature-based approach to modeling protein–protein interaction hot spots. Nucleic Acids Research, 37, 2672–2687.
Chothia, C., & Janin, J. (1975). Principles of protein–protein recognition. Nature, 256, 705–708.
Chuang, G. Y., Mehra-Chaudhary, R., Ngaen, C. H., Zerbe, B. S., Kozakov, D., Vajda, S., & Beamer, L. J. (2010). Domain motion and interdomain hot spots in a multidomain enzyme. Protein Science, 19, 1662–1672.
Clackson, T., Ultsch, M. H., Wells, J. A., & de Vos, A. M. (1998). Structural and functional analysis of the 1: 1 growth hormone: Receptor complex reveals the molecular basis for receptor affinity. Journal of Molecular Biology, 277, 1111–1128.
Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., … Kollman, P. A. (1995). A 2nd generation force-field for the simulation of proteins, nucleic-acids and organic molecules. Journal of the American Chemical Society, 117, 5179–5197.
Dall’Acqua, W., Goldman, E. R., Eisenstein, E., & Mariuzza, R. A. (1996). A mutational analysis of the binding of two different proteins to the same antibody. Biochemistry, 35, 9667–9676.
Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald– an n.log(n) method for Ewald sums in large systems. Journal of Chemical Physics, 98, 10089–10092.
DeLano, W. L. (2002). Unraveling hot spots in binding interfaces: Progress and challenges. Current Opinion in Structural Biology, 12, 14–20.
DeLano, W. L., Ultsch, M. H., de Vos, A. M., & Wells, J. A. (2000). Convergent solutions to binding at a protein–protein interface. Science, 287, 1279–1283.
Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: An automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Research, 32, W665–W667.
Goldman, E. R., Dall’Acqua, W., Braden, B. C., & Mariuzza, R. A. (1997). Analysis of binding interactions in an idioype. Antidiototype protein–protein complex by double mutant cycles. Biochemistry, 36, 49–56.
Guharoy, M., & Chakrabarti, P. (2005). Conservation and relative importance of residues across protein–protein interfaces. Proceedings of the National Academy of Sciences of the United States of America, 102, 15447–15452.
Guharoy, M. & Chakrabarti, P. (2010). Conserved residue clusters at protein–protein interfaces and their use in binding site identification. BMC Bioinformatics 11, 286-303.
Guharoy, M., Pal, A., Dasgupta, M., & Chakrabarti, P. (2011). PRI CE (protein interface conservation and energetics): A server for the analysis of protein–protein interfaces. Journal of Structural and Functional Genomics, 12, 33–41.
Guney, E., Tuncbag, N., Keskin, O., & Gursoy, A. (2007). Hot sprint: Database of computational hot spots in protein interfaces. Nucleic Acids Research, 36, 662–666.
Hawkins, R. E., Russell, S. J., Baier, M., & Winter, G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen: The interaction of mutant D1.3 antibodies with lysozyme. Journal of Molecular Biology, 234, 958–964.
Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. Journal of Molecular Graphics, 14, 33–38.
Huo, S., Massova, I., & Kollman, P. A. (2002). Computational alanine scanning of the 1: 1 human growth hormone-receptor complex. Journal of Computational Chemistry, 23, 15–27.
Jackson, R. M., Gabb, H. A., & Sternberg, M. J. E. (1998). Rapid refinement of protein interfaces incorporating solvation: Application to the docking problem. Journal of Molecular Biology, 276, 265–285.
Janin J. (2009). Basic Principles of Protein–Protein Interaction, Chapter 1. In R. Nussinov & G. Schreiber (Eds.), Computational protein–protein interactions (pp. 1–19). Boca Raton, FL: CRC Press.
Janin, J., & Chothia, C. (1990). The structure of protein–protein recognition sites. Biological Chemistry, 265, 16027–16030.
Are hot-spots occluded from water? 197

Rodier, F., Bahadur, R. P., Chakrabarti, P., & Janin, J. (2005). Hydration of protein–protein interfaces. Proterins-Structure Function and Bioinformatics, 60, 36–45.

Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977). Numerical integration of cartesian equations of motion of a system with constraints – molecular dynamics of n-alkanes. Journal of Computational Physics, 23, 327–341.

Sauer-Eriksson, A. E., Kleywegt, G. J., Uhlén, M., & Jones, T. A. (1995). Crystal structure of the C2 fragment of staphylococcal protein G in complex with the Fc domain of human IgG. Structure London, England, 3, 265–278.

Schlessinger, J., Plotnikov, A. N., Ibrahimi, O. A., Eliseenkova, A. V., Yeh, B. K., Yayon, A., … Mohammadi, M. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Molecular Cell, 6, 743–750.

Schreiber, G., & Fersht, R. (1993). Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry, 32, 5145–5150.

Schreiber, G., & Fersht, A. R. (1995). Energetics of protein–protein interactions: Analysis of the Barnase-Barstar interface by single mutations and double mutant cycles. Journal of Molecular Biology, 248, 478–486.

Shapiro, R., Ruiz-Gutierrez, M., & Chen, C. Z. (2000). Analysis of the interactions of human ribonuclease inhibitor with angiogenin and ribonuclease A by mutagenesis: Importance of inhibitor residues inside versus outside the C-terminal “hot spot”. Journal of Molecular Biology, 302, 497–519.

Sloan, D. J., & Hellinga, H. W. (1999). Dissection of the protein G B1 domain binding site for human IgG Fc fragment. Protein Science, 8, 1543–1648.

Springer, B. A., Pantoliano, M. W., Barbera, F. A., Gunyuzlu, P. L., Thompson, L. D., Herblin, W. F., … Book, G. W. (1994). Identification and concerted function of two receptor binding surfaces on basic fibroblast growth factor required for mitogenesis. Journal of Biological Chemistry, 269, 26879–26884.

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

Tuncbag, N. G. A., & Keskin, O. (2009). Identification of computational hot spots in protein interfaces: Combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics, 25, 1513–1520.

Wiesmann, C., Fuh, G., Christinger, H. W., Eigenbrot, C., Wells, J. A., & de Vos, A. M. (1997). Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. Cell, 91, 695–704.

Xia, J. F., Zhao, X. M., Song, J. N., & Huang, D. S. (2010). APIS: Accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility. BMC Bioinformatics, 11, 174–188.

Zhu, X., & Mitchell, J. C. (2011). KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. Proteins: Structure, Function, and Bioinformatics, 79, 2671–2683.

Zhu, H., Ramnarayan, K., Anchin, J., Miao, W. Y., Sereno, A., Millman, L., … Wolff, M. E. (1995). Glu-96 of basic fibroblast growth factor is essential for high affinity receptor binding. Journal of Biological Chemistry, 270, 21869–21874.