PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Solvent-accessible surface area: How well can be applied to hot-spot detection?

João M. Martins, Rui M. Ramos, António C. Pimenta, and Irina S. Moreira*

REQUIMTE/Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal

ABSTRACT

A detailed comprehension of protein-based interfaces is essential for the rational drug development. One of the key features of these interfaces is their solvent accessible surface area profile. With that in mind, we tested a group of 12 SASA-based features for their ability to correlate and differentiate hot- and null-spots. These were tested in three different data sets, explicit water MD, implicit water MD, and static PDB structure. We found no discernible improvement with the use of more comprehensive data sets obtained from molecular dynamics. The features tested were shown to be capable of discerning between hot- and null-spots, while presenting low correlations. Residue standardization such as $_{rel}SASA_i$ or $_{rel/res}SASA_i$, improved the features as a tool to predict $\Delta\Delta G_{binding}$ values. A new method using support machine learning algorithms was developed: SBHD (Sasa-Based Hot-spot Detection). This method presents a precision, recall, and F1 score of 0.72, 0.81, and 0.76 for the training set and 0.91, 0.73, and 0.81 for an independent test set.

Proteins 2014; 82:479–490.

Key words: hot-spot; computational alanine scanning mutagenesis; solvent accessible surface area; feature based algorithms; support vector machine.

INTRODUCTION

Proteins and proteins-based interactions, like protein-protein interactions (PPI) or protein-DNA interactions (PDI), have been the main focus of interest of several research groups and pharmaceutical companies worldwide due to their importance in many biological processes, such as signal transduction pathways, enzymatic regulation or acting as catalysts, carriers, and many others.[1,2] Proteins tend to interact and bind with other macromolecules, through their interfaces, forming stable complexes. Studies have shown that these interfaces depend almost exclusively on central regions that account for the majority of the binding energy: the hot-spots (HS).[3–5] They can be experimentally found by measuring the binding free energy difference ($\Delta\Delta G_{binding}$) upon alanine mutation. Although this can be presented as the general description of a HS, there are different perspectives regarding the binding free energy difference associated with this kind of residues. Some authors define HS as residues with $\Delta\Delta G_{binding}$ higher than 2.0 kcal/mol; while residues that cause a binding free energy difference lower than 2.0 kcal/mol are defined as null-spots (NS).[5] Others use a $\Delta\Delta G_{binding}$ higher than 1.5 kcal/mol for HS and lower than 0.5 for NS.[6] In this work, we will differentiate residues in hot-spots (>2.0 kcal/mol) and null-spots (<2.0 kcal/mol).

An accurate detection of HS is of primal importance. Alanine Scanning Mutagenesis (ASM) is one of the common methods for the characterization of these residues, but since the experimental ASM methodology is time consuming and hard to execute, various computational methods have emerged. These can be classified in three groups: empirical functions or methods that use knowledge-based simplified models to evaluate complex association; fully atomistic methods that perform mutations of the interfacial residues to estimate binding free energies; and lastly, feature-based approaches, which tend to be more qualitative than quantitative. Thermodynamic Integration (TI) or Free Energy Perturbation (FEP)[7,8] are the most accurate methods available to calculate the binding strength of protein complexes, since it accurately

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: I. S. Moreira, REQUIMTE/Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre s/n, 4169-007 Porto, Portugal. E-mail: irina.moreira@fc.up.pt

Received 29 April 2013; Revised 25 August 2013; Accepted 2 September 2013
Published online 19 September 2013 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24413

predicts free energy differences from an initial to a final state. However, they are also the most demanding computational methods, which limit the screening of a large number of structural perturbations, and are commonly replaced by faster methods like the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) method.[9–12] As for the knowledge-based methods they generally use a set of simplified physical models for the characterization of the complex.[13] Such was the case in the work of Kortemme and Baker[14] that developed a quantitative model for the determination of binding energies, using an all-atom rotamer description of the side-chains and a free energy function consisting in a Lennard Jones potential, solvation interactions and hydrogen bonding. The feature-based methods use a variety of different chemical and physical characteristics of PPI. However, solvent-accessible surface area (SASA) features seem to be commonly used without a comprehensive knowledge why and even how well do they work.[15–25] The "O-Ring theory" or the "Water Exclusion" hypothesis proposed by Bogan and Thorn states was the first application of SASA features and states that the hot-spots are surrounded by a region of null-spots that lead to solvent and results in a lower local dielectric constant environment and enhancement of specific electrostatic and hydrogen bond interactions.[5] This theory was followed by many authors such as Guharoy and Chakrabarti that believed the interface could be divided in two different regions, a core and a rim. The rim would be formed by residues with partial accessibility to solvent, with few hot-spots, and the core would be formed by residues deeply buried in the interface and a high number of hot-spots. They also proposed a connection between the deeply buried surface area of core residues and its contribution to the binding free energy.[16,26,27] Many other authors have focused their attention in this subject, either by proposing modifications to the original one (Liang *et al.*[28]) or complementing it.[29–32] In this work we subjected several protein-protein (PP) and protein-DNA (PDNA) complexes to Molecular Dynamics (MD) simulations in explicit and implicit solvent. Several SASA features were measured and their use as hot-spot differentiators was statistically evaluated. The combination of these SASA features were also analyzed by a support vector machine learning (SVM) algorithm that produced an accurate new model for predicting hot-spots: SBHD (SASA-Based Hot-spot Detection) method.

MATERIALS AND METHODS

System setup

Fifteen different complexes for a total of 248 interfacial residues were studied: (i) Barnase and Barnstar (PDBID: 1BRS[33]); (ii) Igg1 Kappa D1.3 FV and Igg1 Kappa E5.2 FV (PDBID:1DVF[34]); (iii) Ribonuclease A and Ribonuclease Inhibitor (PDBID: 1DFJ[35]); (iv) Bacterial cell division ZipA and Ftsz (PDBID: 1F47[36]); (v) Vascular Endothelial Growth Factor and FLT-1 Receptor (PDBID: 1FLT[37]); (vi) Fibroblast Growth Factor 2 and Fibroblast Growth Factor Receptor 1 (PDBID: 1FQ9[38]); (vii) Igg1 Kappa D1.3 FV and Hen Egg white lysozyme (PDBID: 1VFB[39]); (viii) human placental RNase inhibitor (hRI) and human angiogenin (Ang) (PDBID: 1A4Y[40]); (ix) C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG (PDBID: 1FCC[41]), (x) bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI) (PDBID: 1CBW[42]), (xi) Nuclear Protein EBNA1 and DNA (PDBID: 1B3T[43]); (xii) Gene-regulating protein arc and DNA (PDBID: 1BDT[44]); (xiii) C-Myb DNA-Binding Domain and DNA (PDBID: 1MSE[45]); (xiv) High-Mobility Group Protein D and DNA (PDBID: 1QRV[46]); and (xv) the complex between the MCM1 Transcriptional Regulator and MAT Alpha-2 Transcriptional Repressor and DNA (PDBID: 1MNM[47]). These systems were selected based on the existence of experimental binding free energy ($\Delta\Delta G_{binding}$) values for the interfacial residues upon alanine mutations. The protonation state at physiological range of the different residues of the various proteins was determined using the PDB2PQR server at http://nbcr-222.ucsd.edu/pdb2pqr_1.8/[48] by the PROPKA methodology[49–51] that computes the pK_a values of the ionizable residues in a protein by determining a perturbation to the model pK_a value due to the protein environment. The number of hot- and null-spots of each system is shown in Supporting Information Table S1.

MD simulations

In this work the MD simulations were performed using the AMBER9[52] package with the modified Cornell force field, by Duan *et al.*—ff03[53,54] (PP complexes) and with the AMBER force field *ff99SB* (PDNA complexes). We performed two different types of simulations, with implicit and explicit solvent. The implicit solvent simulations were performed using the Generalized Born solvent method (GB^{OBC})[55] and the ionic strength was set to 0. As for the MD simulations in explicit solvent, each system was solvated using TIP3P explicit water molecules that extended 10 Å from any edge of the box to the protein atoms.[56] An appropriate amount of counter ions was added to neutralize the system. In each of the simulations we started with a minimization stage, to remove bad contacts, by steepest descent followed by conjugated gradient. Periodic boundary conditions were applied using the particle mesh Ewald (PME) method[57] to treat long-range electrostatic interactions and the non-bonded interactions were truncated with a 16 Å and a 10 Å cut-off, in the GB and in explicit solvent simulations, respectively. The systems were 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[58,59] algorithm was used to regulate the temperature of the system. Bond lengths involving hydrogen were constrained using the SHAKE algorithm[60] and the equations of motion were integrated with a 2 fs time step.

SASA features

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. Twelve different SASA features were determined for the interfacial residues with a known $\Delta\Delta G_{binding}$ for explicit water MD, implicit water MD, and solely from the PDB structure. ${}_{comp}SASA_i$ is the solvent accessible surface area of residue i in complex form, while ${}_{mon}SASA_i$ is the residue SASA in the monomer form. $\Delta SASA_i$, the SASA variation upon complexation, is determined using these features [Eq. (1)]. ${}_{rel}SASA_i$ is determined using the results from ΔSASA for each residue and dividing it by the ${}_{mon}SASA_i$ value for the same residue, providing a differentiation between residues with equal ΔSASA but different absolute monomer SASA values [Eq. (2)]. A further four features (${}_{comp/res}SASA_i$, ${}_{mon/res}SASA_i$, ${}_{\Delta/res}SASA_i$, and ${}_{rel/res}SASA_i$), defined by Eqs. (3) to (6), were determined employing amino-acid standardization by dividing the previous features by approximate average protein ${}_{res}SASA_r$ values as determined by Miller *et al.*[61,62], Gly = 85, Ala = 113, Cys = 140, Asp = 151, Glu = 183, Phe = 218, His = 194, Ile = 182, Lys = 211, Leu = 180, Met = 204, Asn = 158, Pro = 143, Gln = 189, Arg = 241, Ser = 122, Thr = 146, Val = 160, Trp = 259 and Tyr = 229, with *r* being the respective residue type. These values determined by Miller *et al.* were then replaced by our own protein average ${}_{ave}SASA_r$ values for each amino-acid type in its respective protein and used to assert amino-acid standardization, resulting in ${}_{comp/ave}SASA_i$, ${}_{mon/ave}SASA_i$, ${}_{\Delta/ave}SASA_i$, and ${}_{rel/ave}SASA_i$ defined in Eqs. (7) to (10).

$$\Delta SASA_i = |{}_{comp}SASA_i - {}_{mon}SASA_i| \quad (1)$$

$${}_{rel}SASA_i = \frac{\Delta SASA_i}{{}_{mon}SASA_i} \quad (2)$$

$${}_{comp/res}SASA_i = \frac{{}_{comp}SASA_i}{{}_{res}SASA_r} \quad (3)$$

$${}_{mon/res}SASA_i = \frac{{}_{mon}SASA_i}{{}_{res}SASA_r} \quad (4)$$

$${}_{\Delta/res}SASA_i = \frac{\Delta SASA_i}{{}_{res}SASA_r} \quad (5)$$

$${}_{rel/res}SASA_i = \frac{relSASA_i}{{}_{res}SASA_r} \quad (6)$$

$${}_{comp/ave}SASA_i = \frac{{}_{comp}SASA_i}{{}_{ave}SASA_r} \quad (7)$$

$${}_{mon/ave}SASA_i = \frac{{}_{mon}SASA_i}{{}_{ave}SASA_r} \quad (8)$$

$${}_{\Delta/ave}SASA_i = \frac{\Delta SASA_i}{{}_{ave}SASA_r} \quad (9)$$

$${}_{rel/ave}SASA_i = \frac{relSASA_i}{{}_{ave}SASA_r} \quad (10)$$

As the SASA features described from Eqs. (3) to (10) have very low order of magnitude, the results presented here were multiplied by a factor of 10^3.

Statistical treatment

Each feature generated earlier was divided into two categories, HS and NS according to the experimental $\Delta\Delta G_{binding}$ data available. Using a Mann-Whitney *U* test we determined whether the feature groups differ from one another, thus showing the feature as capable of discerning between residues belonging to either group. This is a nonparametric significance test which determines if a non-normal distributed group differs from another non-normal distributed group. This test was selected after determining, by box-plot analysis, that the hot- and null-groups for each feature didn't have a normal distribution and therefore a parametric test such as ANOVA could not be used. We also fitted, using a linear regression, each group data against each corresponding $\Delta\Delta G_{binding}$ value in order to ascertain the ability of each feature to predict variations in Gibbs free energy upon ASM. Average and standard deviation values were also calculated for the two groups for each feature.

Support vector machine (SVM)

Support vector machine is a supervised machine learning model in which two classes are classified according to a model developed by training data with known class-equivalence, in this case, HS and NS classes. Several kernel options are available to perform the two-class separation. In this work, and considering the most-widely used kernels, we selected the often used radial basis function kernel after both linear and polynomial kernels having worse performance in initial studies. The 248 interfacial residues with known experimental ASM results were split in a training set and a test set to better test our SVM model. Randomly, 32 hot-spots and 32 null-spots were selected to form the training set, resulting in a balanced 64-point training set, kept constant throughout the learning processes. Thus, the test set was comprised of the remaining data points, 33 hot-spots and 151 null-spots. The datasets were then divided, 12 features per data source available (PDB, explicit MD and implicit

MD) to combine between them. The length of the combination group varied between 1 and 12 features per algorithm training set, totaling 12,285 combinations between the three sets, 4095 per dataset. Finally, to better assess the MD cost-benefit relationship in hot-spot classification, we trained the algorithm with each feature standard deviation by itself or the calculated value and standard deviation of each feature in a 2-dimension SVM. The performance of the hot-spot detection method was assessed by the F1 score [Eq. (11)], which is defined as a function of Precision [*P*, Eq. (12)] and Recall [*R*, also called sensitivity, Eq. (13)]. F1 score, *P* and *R* can be defined as:

$$\mathrm{F1} = \frac{2PR}{P+R} \tag{11}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{12}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{13}$$

In which TP stands for true positive (predicted hot-spots that are actual hot-spots), FP stands for false positive (predicted hot-spots that are not actual hot-spots), and FN stands for false negative (non-predicted hot-spots that are actual hot-spots). Accuracy is defined as the ratio of number of correctly predicted residues to number of all predicted residues as in Eq. (14):

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}} \tag{14}$$

In which TN are the true negatives (correctly predicted null-spots). Specificity and NPV (negative predictive value) are other measures of performance and are formulated as:

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{15}$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{FN}+\mathrm{TN}} \tag{16}$$

To guarantee that our models were not over trained the statistical analysis values (F1 score, Recall, Precision) were estimated by a ninefold cross-validation performed on the training set for each of the SASA feature combinations.

The validity and performance of the method was determined by analyzing similar hot-spot prediction methodologies in terms of overall F1, specificity and accuracy over the same data used by us. The most widely available hot-spot prediction methodologies focus on protein-protein interfaces and therefore our tests comprise solely the protein-protein interface within our total data-set, with the F1, specificity and accuracy determined using our own developed methodology.

RESULTS

Dataset composition

Our database consists of 15 different protein based structures, 10 in which the interface is a PP interface and 5 in which the interface is a PD interface. From those, 248 interfacial residues were used for our study, for which experimental $\Delta\Delta G_{\mathrm{binding}}$ values upon alanine mutation were available. The tested data set consists of 65 hot-spots, 26% of the total residues, and 183 null-spots, 74% of the data, with a calculated *P* value between the two groups' experimental $\Delta\Delta G_{\mathrm{binding}}$ of 4.9×10^{-33} as determined by a Mann-Whitney *U* test, providing a large enough data set for statistical analysis. The amino-acid type and group type distribution is as follows: Glu (11%), Asp (8%), Phe (3%), Ile (6%), His (5%), Lys (10%), Met (4%), Leu (3%), Asn (4%), Gln (5%), Ser (5%), Arg (13%), Thr (4%), Trp (6%), Val (2%), Tyr (10%); charged residues (46%), in which 28% are positively charged residues and 18% are negatively charged residues; and 54% uncharged residues, with 29% of polar behavior and 25% nonpolar. Within hot-spot and null-spot datasets, group types were also determined, showing a prevalence of non-polar residues (31%) within the hot-spots group, being the negatively charged (18%) group the least represented. However, in the null group the prevalence is shared between positively charged and polar subgroups, both with 30%. It is easy to see the deviations from the general group behavior, with the marked decrease of positively charged residues within the hot-spot residues group, from 28% to 23%. The opposite is true for nonpolar residues, having a greater relative number within the hot-spot subgroup, from 25% to 31%. Conversely, within the null-spot subgroup there is a greater prevalence of positively charged residues in comparison with the general behavior, from 28% to 30%. The data set for experimental values spans a wide range of $\Delta\Delta G_{\mathrm{binding}}$, as shown in Figure 1, showing that the sample encompasses most expected experimental Gibbs free energy variations upon alanine scanning mutagenesis.

Hot-spot and null-spot differentiation

It is necessary to achieve an answer for three key questions: "What is the correlation between the $\Delta\Delta G_{\mathrm{binding}}$ and the SASA features values? What are their average values for hot-spots and null-spots? Can these features be used for hot-spot detection?" Therefore, in order to evaluate the relationship between solvent accessibility and binding free energy we have calculated the different SASA features described in the methodology. The most frequently used ones in the literature are $_{\mathrm{comp}}$SASA and ΔSASA. $_{\mathrm{rel}}$SASA was already introduced by Cho *et al.*[63] and Tuncbag *et al.*[25].as a SASA feature. To the best of our knowledge, the others ones were introduced and

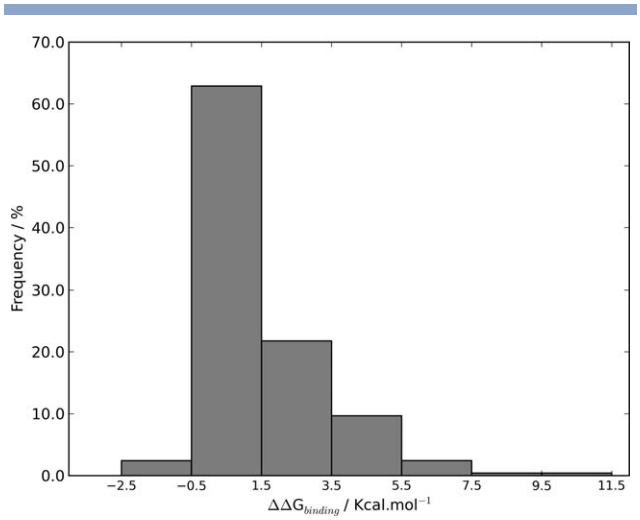

Figure 1

$\Delta\Delta G_{binding}$ distribution histogram.

applied to protein-protein interactions within this work. We have analyzed the SASA/$\Delta\Delta G_{binding}$ correlation for the PDB, explicit and implicit MD and found out very poor correlations between them for all SASA features (Supporting Information Table S2). This behavior is in agreement with the one described by Bogan and Thorn.[5] However, the PDB obtained data shows better correlation results throughout all the evaluated features. There is a significant increase of the correlation when focusing our attention in $_{rel}SASA_i$ and $_{rel/res}SASA_i$ features and for the charged residues. The energetic benefit of occluding bulk solvent from charged interactions is easily understandable as a lower effective dielectric increases the interaction strength of electrostatic and hydrogen bonding interactions.

Mann-Whitney U tests for the SASA features

The medians and standard deviations of each feature were determined for each type of data set; explicit water MD, implicit water MD, and PDB, as shown in Table I, Figure 2, and Supporting Information Table S3. The statistical Mann-Whitney *U* tests were done for the 12 features in order to determine which were capable of differentiating between the two groups, hot- and null-spot, spanning the three different datasets explained earlier, also present in the aforementioned tables. The Mann-Whitney *U* $_{comp}SASA_i$ tests show that the difference between the groups' medians is statistically valid, with 4.06×10^{-9}, 7.0×10^{-12}, and 6.32×10^{-12} for implicit water MD, explicit water MD, and PDB HS and NS group comparison. We can say that $_{comp}SASA_i$ medians have a significant difference when characterizing according to the $\Delta\Delta G_{binding}$ values. $_{mon}SASA_i$ in the implicit water data set shows a statistically irrelevant

Table I
Medians and Standard Deviations of Each Feature Determined for the PDB Dataset, Divided by HS, NS, and Total Groups

	#	$_{comp}SASA_i$	$_{mon}SASA_i$	$\Delta SASA_i$	$_{rel}SASA_i$	$_{comp/res}SASA_i$	$_{mon/res}SASA_i$	$_{res}\Delta SASA_i$	$_{rel/res}SASA_i$	$_{comp/ave}SASA_i$	$_{mon/ave}SASA_i$	$_{ave}\Delta SASA_i$	$_{rel/ave}SASA_i$
Total	248	40.95 ± 45.17	83.06 ± 48.71	29.42 ± 38.18	0.41 ± 11.39	0.20 ± 0.23	0.44 ± 0.24	0.15 ± 0.18	2.09 ± 44.97	0.75 ± 0.68	1.44 ± 1.31	0.47 ± 1.10	6.29 ± 229.84
NS	183	52.44 ± 46.59	84.07 ± 49.15	20.80 ± 35.64	0.30 ± 13.24	0.28 ± 0.23	0.45 ± 0.24	0.11 ± 0.17	1.59 ± 52.27	0.92 ± 0.66	1.40 ± 1.07	0.34 ± 0.88	4.55 ± 267.27
HS	65	17.39 ± 22.85	77.86 ± 46.27	55.84 ± 39.46	0.79 ± 0.88	0.08 ± 0.12	0.35 ± 0.23	0.28 ± 0.20	3.70 ± 3.74	0.35 ± 0.63	1.58 ± 1.78	1.25 ± 1.41	14.07 ± 16.31
p-value		6.32×10^{-12}	1.86×10^{-2}	6.93×10^{-5}	4.40×10^{-9}	9.78×10^{-13}	3.49×10^{-3}	4.16×10^{-4}	1.18×10^{-6}	2.87×10^{-9}	0.97	1.78×10^{-4}	1.08×10^{-5}

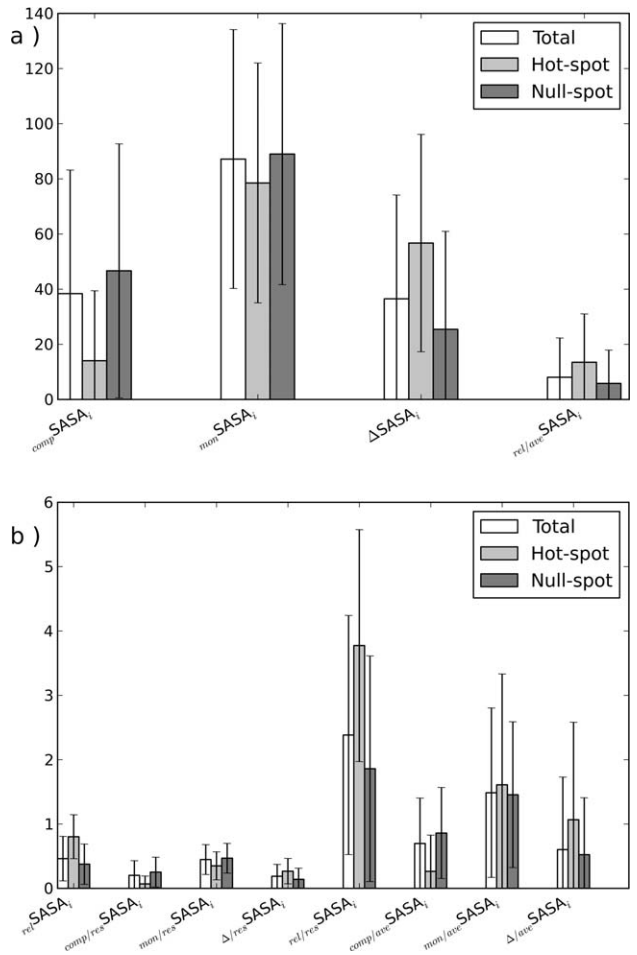

Figure 2

(**a**) Median value and standard deviation for $_{comp}SASA_i$, $_{mon}SASA_i$, $\Delta SASA_i$, and $_{rel/ave}SASA_i$ studied for hot- and null-spots (PDB dataset). (**b**) Median value and standard deviation for the $_{rel}SASA_i$, $_{com/res}SASA_i$, $_{mon/res}SASA_i$, $_{\Delta/res}SASA_i$, $_{rel/res}SASA_i$, $_{comp/ave}SASA_i$, $_{mon/ave}SASA_i$, and $_{\Delta/ave}SASA_i$ for hot- and null-spots (PDB dataset). The results are presented for the three groups of residues analyzed (Total, NS, and HS).

median difference between the two groups, evaluated by the statistical test as 6.01×10^{-2}, with the same being true for explicit water data. Their calculated median difference is once again statistically irrelevant, with 0.12 for the explicit water groups' difference. However, in the data gathered with the PDB dataset, the same is not true. We obtained a *P* value of 1.86×10^{-2}, below the statistical cut-off for group differentiation, with values of 88.95 and 78.49 for null- and hot-group. As expected, while solvent accessibility in the complex form is related with $\Delta\Delta G_{binding}$, solvent accessibility in the protein monomer form is not an indication of the residues' importance to the protein interaction. The ΔSASA feature can be more useful as it evaluates the variation between solvent accessible area in the monomer and in the complex form. For the implicit water data set, a *P* value of 2.38×10^{-4} between the hot- and null-spot groups shows statistical relevance between the two groups' medians. The same is true for the remaining data sets, with a *P*-value of 4.84×10^{-6} and 6.93×10^{-5} for explicit water data and PDB calculated data, respectively. This shows that ΔSASA is capable of differentiating between HS and NS to a high degree. However, SASA variation upon complex formation is not able to fully discriminate between HS and NS, as evidenced by the standard deviation for the null- and hot-groups calculated for implicit water MDs, 34.38 $Å^2$ and 37.60 $Å^2$ respectively, amounting to 1.77 and 0.70 times of the groups' value, with similar magnitudes for the other types of MD data origin. This is clearly unacceptable as a mean of predicting whether a residue belongs in either of the groups. For a better understanding of the difference between hot- and null-spot residues, one must look into amino acid standardization.

The amino acid standardization is implemented through the use of Eqs. (2) through (10). The results gathered show that the groups' difference is statistically relevant for all types of feature except for $_{mon/res}SASA_i$, represented in Eq. (4). The corresponding values for statistical group variance difference obtained from the implicit water data source are between 0.87, for $_{mon/ave}SASA_i$ and 6.11×10^{-10} for $_{comp/res}SASA_i$. The same is true for the other data sources, explicit water MD and PDB data, with the main difference being the extent of group differentiation that standardization by amino-acid SASA values achieves over the $_{mon}SASA_i$ dependent features such as $_{mon/res}SASA_i$ and $_{mon/ave}SASA_i$. In the implicit water MD features' *P* value, we have a decrease from 6.01×10^{-2} for $_{mon}SASA_i$, 8.81×10^{-3} for $_{mon/res}SASA_i$, and an increase to 0.87 for $_{mon/ave}SASA_i$. For the explicit water MD, the results are 0.12, 2.99×10^{-2}, and 0.40 for the groups above mentioned, respectively. In the case of the PDB gathered data, we found 1.86×10^{-2} for $_{mon}SASA_i$, 3.49×10^{-3} for $_{mon/res}SASA_i$, and 0.97 for $_{mon/ave}SASA_i$. This shows that standardization seems to have an impact over group differentiation, allowing the improvement of features which, by themselves, are unable to differentiate between hot- and null-spot residues. However, standardization seemed to improve group differentiation when the standardization values used where the previously published ones [Eqs. (3–6)]. When using the ones obtained from the individual protein analysis, the results seem to worsen [Eqs. (7–10)].

From this data analysis, a comparison is possible between data sources and the usefulness of more computational costly methods such as molecular dynamics with explicit water molecules. From observation of the statistical relevant group differentiation through the different data origins we can say that the data originated from the PDB is enough for the features tested to differentiate between residues belonging to either group. However, this test only refers to the feature capability of grouping hot- and null-spot residues correctly and whether this

grouping is in accordance to the experimental $\Delta\Delta G_{binding}$ data available. Considering only the Mann-Whitney *U* tests, it seems that the best features are the ones based on $_{rel}$SASA, regardless of data source, showing consistently the capability to differentiate between hot- and null-spot groups. This is probably due to the residue SASA standardization employed by these features, making them a good candidate for SASA-based hot-spot prediction.

Alternative hot- and null-spot definition analysis

The above analysis was also performed, with the single difference being the $\Delta\Delta G_{binding}$ for which a residue was counted as being hot- and null-spot. The alternate definition tested was $\Delta\Delta G_{binding}$ over or equal 1.5 kcal mol^{-1} for the hot-spot residues and below 0.5 kcal mol^{-1} for the null-spot residues. This reduces the total number of data set points to 173, with 87 being null-spots and 86 being hot-spots. The variation in the relative composition of the global data set shows a marked preference toward keeping more charged residues, while, conversely, excluding more uncharged residues. This is visible in the relative composition variation for each group and the group encompassing all the residues within each subgroup, with the general group losing 30% of its elements, while positively charged and negatively charged residues groups losing 23% and 28%. For this overall variation to occur there had to be larger losses in the remaining groups, as happened to polar and non-polar groups, by 32% and 38%. Regarding the Mann-Whitney *U* tests between null-spot and hot-spot groups, discernible differences were not found between the two groups as defined by either definition. The feature which seems to struggle to achieve hot and null-spot differentiation is, as expected, the $_{mon}$SASA$_i$ and, subsequently, those which depend on that value, $_{mon/res}$SASA$_i$ and $_{mon/ave}$SASA$_i$. Therefore, this alternative hot- and null-spot definition does not offer great advantages towards discerning between features' sub-groups, since the original definition tested was shown to be capable of already discerning between hot- and null-spots within the various feature calculated parameters. This lack of improvement over the first definition tested, coupled with the large number of dataset points ignored when employing this alternative definition makes this alternative hot- and null-spot definition a high-cost, low-advantage situation. This could possibly change in the future, depending mostly on the dataset composition and size, making this definition more attractive toward hot- and null-spot computational prediction and $\Delta\Delta G_{binding}$ prediction. We believe further tests are needed on different applications of hot-spot prediction towards computational goals, such as protein-protein docking scoring methods with well-defined reference proteins, allowing us to accurately determine the best correlated hot-spot definition with real-world protein behavior.

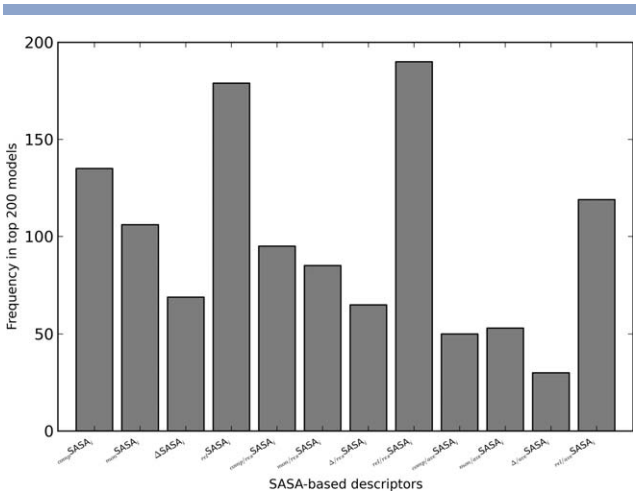

Figure 3

Frequencies (number of models) in the top 200 combinations of the various SASA features, ranked by F1 performance on the PDB training set.

Support vector machine (SVM)

Training set

We explore all possible combinations of the 12 features and a ninefold cross-validation test was performed on the training set for each of them. We have performed this test only for the $\geq$2.0 kcal/mol (HS) and $<$2.0 kcal/mol (NS) definition as the other HS definition would diminish our available training and testing sets to an extent that would impair the applicability of this SVM method. We then sorted the combination results by ninefold cross-validation F1 and from this sorted list we counted the feature frequency in the top 200 results, as shown for the PDB set in Figure 3 with a 10 occurrences cut-off, with the result of this test for the other two datasets shown in Supporting Information Figure S1. This way, we can easily identify the most common features present in the best combinations obtained. These are for the PDB dataset: $_{rel/res}$SASA$_i$ (190), $_{rel}$SASA$_i$ (179), $_{comp}$SASA$_i$, (135) and $_{rel/ave}$SASA$_i$ (119). The results are in agreement with our Mann-Whitney *U* tests that demonstrated that SASA standardization better discriminates null- and hot-spots. For the explicit MD, the four more populated are: ΔSASA$_i$ (140), $_{\Delta/res}$SASA$_i$ (113), $_{mon/res}$SASA$_i$ (103), and $_{mon}$SASA$_i$ (91); for implicit MD are: ΔSASA$_i$ (159), $_{rel/ave}$SASA$_i$ (136), $_{mon}$SASA$_i$ (133), and $_{\Delta/res}$SASA$_i$ (132). The top 20 feature combinations for the three data sets are shown in Tables (II–IV) For the PDB dataset (Table II), the top 20 cross-validation F1 sorted results from our SASA combinations have F1 scores in

Table II
Top 20 Highest Feature Combinations Ranked by F1 Score in the PDB Training Set

F1	Precision	Recall	Accuracy	Sensitivity	Descriptors used
0.76	0.72	0.81	0.75	0.81	5
0.76	0.72	0.81	0.75	0.81	6
0.76	0.72	0.81	0.75	0.81	4
0.76	0.72	0.81	0.75	0.81	4
0.75	0.70	0.81	0.73	0.81	7
0.75	0.71	0.78	0.73	0.78	5
0.75	0.71	0.78	0.73	0.78	5
0.75	0.71	0.78	0.73	0.78	5
0.75	0.71	0.78	0.74	0.78	5
0.75	0.71	0.78	0.73	0.78	6
0.75	0.71	0.78	0.73	0.78	5
0.75	0.71	0.78	0.73	0.78	6
0.74	0.68	0.81	0.72	0.81	6
0.74	0.68	0.81	0.72	0.81	7
0.74	0.68	0.81	0.72	0.81	6
0.74	0.68	0.81	0.72	0.81	7
0.74	0.73	0.75	0.73	0.75	4
0.74	0.69	0.78	0.72	0.78	7
0.74	0.69	0.78	0.72	0.78	6
0.74	0.69	0.78	0.72	0.78	8

Table IV
Top 20 Highest Feature Combinations Ranked by F1 Score in the Implicit MD Training Set

F1	Precision	Recall	Accuracy	Sensitivity	Descriptors used
0.71	0.67	0.75	0.69	0.75	4
0.70	0.75	0.66	0.72	0.66	6
0.70	0.75	0.66	0.72	0.66	7
0.69	0.72	0.66	0.70	0.66	4
0.69	0.72	0.66	0.70	0.66	4
0.69	0.72	0.66	0.70	0.66	6
0.69	0.69	0.69	0.69	0.69	4
0.68	0.74	0.63	0.70	0.63	7
0.68	0.74	0.63	0.70	0.63	4
0.68	0.70	0.66	0.69	0.66	4
0.68	0.67	0.69	0.67	0.69	7
0.68	0.64	0.72	0.65	0.72	4
0.67	0.68	0.66	0.67	0.66	6
0.67	0.65	0.69	0.65	0.69	6
0.67	0.71	0.63	0.69	0.63	6
0.67	0.71	0.63	0.69	0.63	5
0.66	0.61	0.72	0.63	0.72	3
0.66	0.66	0.66	0.65	0.66	4
0.66	0.66	0.66	0.66	0.66	7
0.66	0.69	0.63	0.67	0.63	8

the range 0.74 to 0.76, Precision 0.68 to 0.72, and Recall 0.75 to 0.81. For the top explicit MD results (Table III), F1 varied between 0.68 and 0.71, Precision between 0.65 and 0.77, and Recall between 0.62 and 0.75. For the implicit MD (Table IV), F1 score varied from 0.66 to 0.71, Precision from 0.64 to 0.74 and Recall from 0.63 to 0.75. The PDB results are slightly better and therefore, for the sole purpose of hot-spots detection based on SASA features, the computational cost of running an MD does not seem to be worth it. Several studies suggest that by using a small number of key features should be enough for hot-spot detection whereas a high number would lead to over-fitting. Having this in mind, we stress out that our first four models in the PDB dataset comprise four to six features, with $_{rel}SASA_i$ and $_{res/res}SASA_i$ being common to all. The 20 best results (Supporting Information Table S4) for the SVM algorithm are also presented for single SASA features without combination (F1: 0.46–0.64). Although, we clearly see a decrease of the statistical tests, the PDB dataset still presents the best results. MD influence over the SVM model capability was tested by training the algorithm with: a two-dimensional system containing each feature and its determined standard deviation; a single-dimensional system of each feature; and a single-dimensional system for each feature's standard deviation. The results (Supporting Information Tables S5 and S6) show that, either sorted by ninefold cross-validation F1 or sorted by test-set F1, the single-dimensional features' standard deviation sets consistently show worse performance than either single-dimensional feature values set or the two-dimensional set. Considering only the single feature value set and the feature value with its standard deviation, it appears that neither shows prevalence in the upper-scored tested models, leading us to believe that this features' standard deviation doesn't hold much information regarding the hot- and null-spot prediction using this methodology. This behavior could be predicted from inspection of Figure 2, which shows that the standard deviations are too large, and therefore their potential as hot-spot discriminant is very low.

Table III
Top 20 Highest Feature Combinations Ranked by F1 Score in the Explicit MD Training Set

F1	Precision	Recall	Accuracy	Sensitivity	Descriptors used
0.71	0.73	0.69	0.71	0.69	4
0.70	0.75	0.66	0.72	0.66	3
0.70	0.71	0.69	0.71	0.69	3
0.70	0.71	0.69	0.71	0.69	4
0.70	0.68	0.72	0.69	0.72	4
0.70	0.65	0.75	0.67	0.75	4
0.70	0.65	0.75	0.67	0.75	6
0.69	0.77	0.63	0.72	0.63	4
0.69	0.72	0.66	0.71	0.66	5
0.69	0.72	0.66	0.70	0.66	4
0.69	0.72	0.66	0.70	0.66	5
0.69	0.69	0.69	0.69	0.69	4
0.69	0.69	0.69	0.69	0.69	4
0.69	0.66	0.72	0.67	0.72	3
0.69	0.66	0.72	0.67	0.72	6
0.69	0.63	0.75	0.66	0.75	4
0.69	0.63	0.75	0.66	0.75	5
0.68	0.74	0.63	0.70	0.63	5
0.68	0.70	0.66	0.69	0.66	4
0.68	0.70	0.66	0.69	0.66	6

Table V
Performance of the 20 Best Combinations of the PDB Dataset on an Independent Test Set

F1	Precision	Recall	Accuracy	Sensitivity	Descriptors used
0.81	0.91	0.73	0.72	0.73	5
0.82	0.90	0.75	0.72	0.75	6
0.80	0.92	0.71	0.71	0.71	4
0.80	0.92	0.71	0.71	0.71	4
0.83	0.90	0.76	0.73	0.76	7
0.81	0.93	0.73	0.72	0.73	5
0.82	0.92	0.74	0.73	0.74	5
0.82	0.91	0.74	0.72	0.74	5
0.81	0.92	0.73	0.72	0.73	5
0.82	0.92	0.75	0.73	0.75	6
0.82	0.92	0.75	0.73	0.75	5
0.83	0.91	0.76	0.73	0.76	6
0.81	0.90	0.73	0.71	0.73	6
0.78	0.90	0.69	0.68	0.69	7
0.81	0.92	0.73	0.72	0.73	6
0.79	0.90	0.70	0.68	0.70	7
0.81	0.92	0.73	0.72	0.73	4
0.81	0.91	0.73	0.72	0.73	7
0.81	0.90	0.73	0.71	0.73	6
0.79	0.90	0.71	0.69	0.71	8

Table VII
Performance of the 20 Best Combinations of the Implicit MD Training Set on an Independent Test Set

F1	Precision	Recall	Accuracy	Sensitivity	Descriptors used
0.71	0.89	0.59	0.60	0.59	4
0.75	0.92	0.63	0.65	0.63	6
0.58	0.87	0.44	0.48	0.44	7
0.77	0.91	0.67	0.67	0.67	5
0.77	0.91	0.67	0.67	0.67	5
0.75	0.88	0.65	0.64	0.65	7
0.66	0.88	0.52	0.54	0.52	4
0.75	0.91	0.64	0.65	0.64	7
0.68	0.93	0.53	0.58	0.53	4
0.74	0.90	0.63	0.63	0.63	4
0.81	0.90	0.74	0.71	0.74	7
0.83	0.88	0.79	0.73	0.79	4
0.76	0.87	0.68	0.65	0.68	6
0.65	0.84	0.52	0.52	0.52	6
0.84	0.91	0.77	0.75	0.77	6
0.77	0.90	0.67	0.67	0.67	5
0.58	0.90	0.42	0.48	0.42	3
0.78	0.88	0.70	0.67	0.70	4
0.75	0.89	0.64	0.64	0.64	7
0.75	0.89	0.65	0.65	0.65	8

Test set

We have tested all our 20 top models described in the previous section on the remaining residues that constitute our test set to assess their predictive performance. Tables (V–VII) show the F1, Precision and Recall values for the various feature combinations from the different datasets. On the PDB dataset there is a slight increase of the F1 score to 0.78–0.82, due to the increase of the Precision to 0.90–0.93. These values are extremely high especially upon comparison with the available HS detection methods. The same trend of F1 scores increase is attained for the explicit (F1: 0.66–0.84) and implicit (F1: 0.58–0.84) MD datasets. Supporting Information Table S7 shows the results for the single feature in our independent test set. In this case we cannot find a solid common pattern as some features really misbehave in the test set whereas others present very high F1 score values. The same can be said about the standard deviations (Supporting Information Table S8). Supporting Information Table S9 shows the results for the average plus standard deviation of the SASA features within the MD datasets. The upper models show a clear F1 score increase in the test set (9% more than in the PDB dataset). We want to highlight that we do believe that MD simulations, by generating an ensemble of possible conformations of the systems in study, are very useful for a more complete comprehension of the HS and NS dynamic behavior. However, we cannot forget the computational time involved, and so the choice between MD/PDB should be

Table VI
Performance of the 20 Best Combinations of the Explicit MD Training Set on an Independent Test Set

F1	Precision	Recall	Accuracy	Sensitivity	Descriptors used
0.66	0.90	0.52	0.55	0.52	4
0.70	0.96	0.56	0.61	0.56	3
0.84	0.92	0.76	0.75	0.76	3
0.79	0.93	0.69	0.70	0.69	4
0.82	0.93	0.73	0.73	0.73	4
0.82	0.94	0.73	0.74	0.73	4
0.82	0.89	0.76	0.72	0.76	6
0.80	0.93	0.71	0.71	0.71	4
0.74	0.94	0.61	0.65	0.61	5
0.76	0.94	0.63	0.66	0.63	4
0.68	0.93	0.54	0.58	0.54	5
0.80	0.94	0.69	0.71	0.69	4
0.76	0.94	0.63	0.66	0.63	4
0.83	0.91	0.77	0.74	0.77	3
0.83	0.90	0.77	0.74	0.77	6
0.76	0.88	0.67	0.65	0.67	4
0.84	0.87	0.82	0.74	0.82	5
0.75	0.93	0.63	0.66	0.63	5
0.78	0.94	0.66	0.68	0.66	4
0.69	0.95	0.54	0.59	0.54	6

Table VIII
Comparison of Performance of the Proposed SBHD Method in the Test Set Against Other Available Hot-Spot Prediction Methodologies

Methodology	F1	Accuracy	Sensitivity	Precision	Specificity	NPV
SBHD	0.86	0.77	0.77	0.97	0.75	0.26
Promate	0.25	0.73	0.94	0.50	0.94	0.75
ROBETTA	0.51	0.76	0.87	0.57	0.87	0.81
KFC2-A	0.50	0.80	0.96	0.76	0.96	0.80
KFC2-B	0.51	0.77	0.89	0.61	0.89	0.81
HotPoint	0.58	0.77	0.84	0.58	0.84	0.84
ISIS	0.08	0.71	0.96	0.33	0.96	0.73

made taking into account the ratio quality/time that each researcher intends to achieve.

So, a combination of a low number of standardized features in a static structure could be the best way to get an accurate and fast method. Thus, we present here a new method: Sasa-Based Hot-spot Detection (SBHD). This method consists in the evaluation of 6 SASA parameters in a PDB structure: $_{comp}SASA_i$, $_{mon}SASA_i$, $_{rel}SASA_i$, $_{comp/res}SASA_i$, $_{rel/res}SASA_i$ and $_{rel/ave}SASA_i$. Its F1, Precision, and Recall values in the training and test set are: 0.76 versus 0.82, 0.72 versus 0.90, 0.81 versus 0.75, respectively. These values are higher than the ones reported until now in literature.

The external assessment of the developed method in comparison to other hot-spot prediction methods is summarized in Table VIII. As most available methods are only tested for protein-protein interface's hot-spot prediction, we limited our tested data to the protein-protein relevant subset and ran the prediction methodology for each available methodology as well as our own proposed methodology. It is clearly visible that, even though our machine-learning-assisted methodology only takes into account some of the SASA descriptors put forth in this work, it still achieves a comparable or better performance than the previously proposed methodologies. The F1 achieved using this method showed a big improvement over the available methods, while the accuracy kept close to the best available methods while losing some group sensitivity towards the other available methods. Table VIII clearly shows that the SBHD approach presents a precision well above the other methods, which shows its good performance. However, it presents a limitation. Although, it predicts true hot-spots very well, it also incorrectly predicts many null-spots as false hot-spots (low NPV value), showing an aspect in which further improvement can be achieved. It is, however, important to stress that our methodology showed even better overall results when composition of the test dataset is mixed between protein-protein interface sets and protein-DNA sets. This makes us confident that incorporating further interfacial residue descriptors such as CASM and genetic conservation could improve on the results already achieved by us.

DISCUSSION AND CONCLUSIONS

The study of this dataset improved the current knowledge about the use of Solvent Accessible Surface Area features as a way to predict hot- and null-spots within a protein-based interface. SASA features were shown to be capable of differentiating between hot- and null-spot groups, regardless of the hot-spot definition used. The improvement achieved through the use of more time-widespread data origins, explicit and implicit water molecular dynamics, was tested against the SASA features obtained from static PDB structures. We found that the computational costs associated with the bigger sample size and the more complex MD calculations did not provide enough improvement over the results obtained through the use of the static PDB structures. The various SASA features showed to be equally able to differentiate between the hot- and null-spot groups, making the increased computational cost unnecessary in this particular application. The correlation tests showed poor overall correlations between the experimental $\Delta\Delta G_{binding}$ and the calculated SASA features, regardless of the dataset source. Nonetheless, and in particular when specific residue types are selected, $_{rel}SASA_i$ and $_{rel/res}SASA_i$ were shown to achieve the best correlation values with the experimental values. This showed us that residue standardization improves the features as a tool to predict $\Delta\Delta G_{binding}$ values. To verify their use in hot-spot detection methods we have performed a SVM study of the features by themselves and in the various possible combinations. The performance of our approach was validated by a ninefold cross-validation and by using an independent test set. We showed that the same two features ($_{rel}SASA_i$ and $_{rel/res}SASA_i$) were highly represented in the top 200 models and in the best predictor models. We concluded that the PDB dataset shows the best results with F1 ranging up to 0.82. Therefore, we present our new method SBHD that consists in the evaluation of 6 features ($_{comp}SASA_i$, $_{mon}SASA_i$, $_{rel}SASA_i$, $_{comp/res}SASA_i$, $_{rel/res}SASA_i$, and $_{rel/ave}SASA_i$). We have also to stress out that the combination of the average and standard deviation value of the $_{rel/res}SASA_i$ in explicit MD give a F1 score of 0.91, which is surprisingly high but the computational time involved does not make the improvement over the PDB dataset worthy.

REFERENCES

1. Janin J. Elusive affinities. Proteins 1995;21:30–39.
2. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
3. Clackson T, Ultsch MH, Wells JA, de Vos AM. Structural and functional analysis of the 1:1 growth hormone: receptor complex reveals the molecular basis for receptor affinity. J Mol Biol 1998;277:1111–1128.
4. DeLano WL, Ultsch MH, de Vos AM, Wells JA. Convergent solutions to binding at a protein-protein interface. Science 2000;287: 1279–1283.
5. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
6. Gao Y, Wang R, Lai L. Structure-based method for analyzing protein-protein interfaces. J Mol Model 2004;10:44–54.
7. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct 2001;30:211–243.
8. Martins SA, Perez MAS, Moreira IS, Sousa SF, Ramos MJ, Fernandes PA. Computational alanine scanning mutagenesis: MM-PBSA vs. TI. J Chem Theor Comput 2013;9:1311–1319.
9. Kollman PA, Massova I, Reyes C, Kuhn B, Huo SH, Chong L, Lee M, Lee T, Duan Y, Wang W, et al. Calculating structures and free

energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
10. Massova I, Kollman PA. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. J Am Chem Soc 1999;121:8133–8143.
11. Moreira IS, Fernandes PA, Ramos MJ. Unravelling hot spots: a comprehensive computational mutagenesis study. Theor Chem Account 2007;117:99–113.
12. Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis—an improved methodological approach. J Comput Chem 2007;28:644–654.
13. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 2002;320:369–387.
14. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci USA 2002;99:14116–14121.
15. Guney E, Tuncbag N, Keskin O, Gursoy A. HotSprint: database of computational hot spots in protein interfaces. Nucleic Acids Res 2007;36:662–666.
16. Guharoy M, Chakrabarti P. Conservation and relative importance of residues across protein-protein interfaces. Proc Natl Acad Sci USA 2005;102:15447–15452.
17. Guharoy M, Chakrabarti P. Conserved residue clusters at protein-protein interfaces and their use in binding site identification. BMC Bioinformatics 2010;11:286.
18. Guharoy M, Pal A, Dasgupta M, Chakrabarti P. PRICE (PRotein Interface Conservation and Energetics): a server for the analysis of protein-protein interfaces. J Struct Funct Genomics 2011;12:33–41.
19. Li Z, Li J. Geometrically centered region: a "wet" model of protein binding hot spots not excluding water molecules. Proteins 2010;78:3304–3316.
20. Liu Q, Li J. Protein binding hot spots and the residue-residue pairing preference: a water exclusion perspective. BMC Bioinformatics 2010;11:244.
21. Xia JF, Zhao XM, Song JN, Huang DS. APIS: accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility. BMC Bioinformatics 2010;11:174.
22. Liu Q, Li J. Propensity vectors of low-ASA residue pairs in the distinction of protein interactions. Proteins 2010;78:589–602.
23. Li Z, Wong L, Li J. DBAC: a simple prediction method for protein binding hot spots based on burial levels and deeply buried atomic contacts. BMC Syst Biol 2011;5:5.
24. Zhu X, Mitchell JC. KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. Proteins 2011;79:2671–2683.
25. Tuncbag N, Gursoy A, Keskin O. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. Bioinformatics 2009;25:1513–1520.
26. Bahadur RP, Chakrabarti P, Rodier F, Janin J. Dissecting subunit interfaces in homodimeric proteins. Proteins 2003;53:708–719.
27. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. Proteins 2002;47:334–343.
28. Liang SD, Meroueh SO, Wang GC, Qiu C, Zhou YQ. Consensus scoring for enriching near-native structures from protein-protein docking decoys. Proteins 2009;75:397–403.
29. Martins JM, Ramos RM, Moreira IS. Structural determinats of a typical leucine-rich repeat protein. Commun Comput Phys 2013;13:238–255.
30. Moreira IS, Fernandes PA, Ramos MJ. Hot spot occlusion from bulk water: a comprehensive study of the complex between the lysozyme HEL and the antibody FVD1.3. J Phys Chem B 2007;111:2697–2706.
31. Moreira IS, Martins JM, Ramos MJ, Fernandes PA, Ramos MJ. Understanding the importance of the aromatic amino-acid residues as hot-spots. Biochem Biophys Acta 2013;1834:401–414.
32. Moreira IS, Ramos RM, Martins JM, Fernandes PA, Ramos MJ. Are hot-spots occluded from water? J Biomol Struct Dyn 2013; doi: 10.1080/07391102.2012.758598.
33. Buckle AM, Schreiber G, Fersht AR. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0A resolution. Biochemistry 1994;33:8878–8889.
34. Braden BC, Fields BA, Ysern X, Dall'Acqua W, Goldbaum FA, Poljak RJ, Mariuzza RA. Crystal structure of an Fv-Fv idiotope-anti-idiotope complex at 1.9 resolution. J Mol Biol 1996;264:137–151.
35. Kobe B, Deisenhofer J. A structural basis of the interactions between leucine-rich repeats and protein ligands. Nature 1995;374:183–186.
36. Mosyak L, Zhang Y, Glasfeld E, Haney S, Stahl M, Seehra J, Somers WS. The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. EMBO J 2000;19:3179–3191.
37. Wiesmann C, Fuh G, Christinger HW, Eigenbrot C, Wells JA, de Vos AM. Crystal structure at 1.7 Å resolution of VEGF in complex with domain 2 of the Flt-1 receptor. Cell 1997;91:695–704.
38. Schlessinger J, Plotnikov AN, Ibrahimi OA, Eliseenkova AV, Yeh BK, Yayon A, Linhardt RJ, Mohammadi M. Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol cell 2000;6:743–750.
39. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, Souchon H, Schwarz FP, Mariuzza RA, Poljak RJ. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. Proc Natl Acad Sci USA 1994;91:1089–1093.
40. Papageorgiou AC, Shapiro R, Acharya KR. Molecular recognition of human angiogenin by placental ribonuclease inhibitor[mdash]an X-ray crystallographic study at 2.0 A resolution. EMBO J 1997;16:5162–5177.
41. Sauer-Eriksson AE, Kleywegt GJ, Uhlen M, Jones TA. Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. Structure 1995;3:265–278.
42. Scheidig AJ, Hynes TR, Pelletier LA, Wells JA, Kossiakoff AA. Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of alzheimer's amyloid β-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): Engineering of inhibitors with altered specificities. Protein Sci 1997;6:1806–1824.
43. Bochkarev A, Bochkareva E, Frappier L, Edwards AM. The 2.2 angstrom structure of a permanganate-sensitive DNA site bound by the Epstein-Barr virus origin binding protein, EBNA1. J Mol Biol 1998;284:1273–1278.
44. Schildbach JF, Karzai AW, Raumann BE, Sauer RT. Origins of DNA-binding specificity: role of protein contacts with the DNA backbone. Proc Natl Acad Sci USA 1999;96:811–817.
45. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, Kanai H, Sarai A, Ishii S, Nishimura Y. Solution structure of a specific DNA complex of the MYB DNA-binding domain with cooperative recognition helices. Cell 1994;79:639–648.
46. Murphy FV, Sweet RM, Churchill MEA. The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. EMBO J 1999;18:6610–6618.
47. Tan S, Richmond TJ. Crystal structure of the yeast MAT alpha 2/MCM1/DNA ternary complex. Nature 1998;391:660–666.
48. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Research 2004;32:W665–W667.
49. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pKa values for protein–ligand complexes. Proteins 2008;73:765–783.
50. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. Proteins 2005;61:704–721.

51. Olsson MHM, Sondergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theor Comput 2011;7:525–537.
52. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, R.Luo, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA. AMBER 9. University of California, San Francisco; 2006.
53. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 1995;117:5179–5197.
54. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 2003;24:1999–2012.
55. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins 2004;55:383–394.
56. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935.
57. Darden T, York D, Pedersen L. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
58. Loncharich RJ, Brooks BR, Pastor RW. Langevin dynamics of peptides: the frictional dependence of isomerization rates of N-acetylalanyl-N'-methylamide. Biopolymers 1992;32:523–535.
59. Izaguirre JA, Catarello DP, Wozniak JM, Skeel RD. Langevin stabilization of molecular dynamics. J Chem Phys 2001;114:2090–2098.
60. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.
61. Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol 1987;196:641–656.
62. Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. Nature 1987;328:834–836.
63. Cho KI, Kim D, Lee D. A feature-based approach to modeling protein-protein interaction hot spots. Nucleic Acids Res 2009;37: 2672–2687.