PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features

Xiaolei Zhu[1] and Julie C. Mitchell[2,3]*

[1] BACTER Institute, University of Wisconsin-Madison, Madison, Wisconsin 53706
[2] Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706
[3] Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin 53706

ABSTRACT

Hot spots constitute a small fraction of protein–protein interface residues, yet they account for a large fraction of the binding affinity. Based on our previous method (KFC), we present two new methods (KFC2a and KFC2b) that outperform other methods at hot spot prediction. A number of improvements were made in developing these new methods. First, we created a training data set that contained a similar number of hot spot and non-hot spot residues. In addition, we generated 47 different features, and different numbers of features were used to train the models to avoid over-fitting. Finally, two feature combinations were selected: One (used in KFC2a) is composed of eight features that are mainly related to solvent accessible surface area and local plasticity; the other (KFC2b) is composed of seven features, only two of which are identical to those used in KFC2a. The two models were built using support vector machines (SVM). The two KFC2 models were then tested on a mixed independent test set, and compared with other methods such as Robetta, FOLDEF, HotPoint, MINERVA, and KFC. KFC2a showed the highest predictive accuracy for hot spot residues (True Positive Rate: TPR = 0.85); however, the false positive rate was somewhat higher than for other models. KFC2b showed the best predictive accuracy for hot spot residues (True Positive Rate: TPR = 0.62) among all methods other than KFC2a, and the False Positive Rate (FPR = 0.15) was comparable with other highly predictive methods.

Proteins 2011; 79:2671–2683.

Key words: protein-protein interface; hot spot; SVM; plasticity; atomic density; solvent accessible surface area.

INTRODUCTION

Protein–protein interactions are important for many different cellular processes such as signal transduction, transport, cellular motion, and regulatory mechanisms. In many respects, these protein–protein interactions are not independent; they are generally integrated into larger biological interaction networks. Because of this interdependence, any erroneous or disrupted protein–protein interaction can cause disease.[1] Although many works have explored the fundamental principles that govern protein–protein interactions,[2–9] none of them entirely characterize the determinants of protein association. This is because many features contribute to protein–protein interactions, such as hydrogen bonds, salt bridges, and hydrophobic and van der Waals interactions. Simultaneously, protein–protein interactions are influenced by differences in the interface size, shape, and the number and type of amino acids within different protein–protein interfaces.

One way to understand the principles of protein–protein interactions is to study the individual interactions between residues of the protein–protein interface. The binding specificity and affinity of protein–protein interactions are important for protein function; however, the contributions of individual residues to protein–protein binding specificity and affinity are not homogeneous. A small subset of interface residues, called "hot spots," account for most of the binding affinity, and often binding specificity, of protein–protein interactions.[10–12] Identifying these hot spot residues within protein–protein interfaces can help us better understand protein–protein interactions and may also help us to modulate protein–protein binding.[13]

Experimental methods have been used to identify hot spot residues at protein–protein interfaces. Alanine scanning mutagenesis has been used extensively as an experimental method to identify protein–protein interface hot spots.[14] In this method, interface residues are systematically replaced with alanine, and the binding free energy difference is measured.

Additional Supporting Information may be found in the online version of this article.
Grant sponsors: US Department of Energy SciDAC and GTL: Genomics Programs; Grant numbers: DE-FG02-04ER25627, DE-FC02-07ER64494.
*Correspondence to: Julie C. Mitchell, Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706. E-mail: jcmitchell@wisc.edu.
Received 6 December 2010; Revised 3 April 2011; Accepted 27 April 2011
Published online 20 May 2011 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.23094

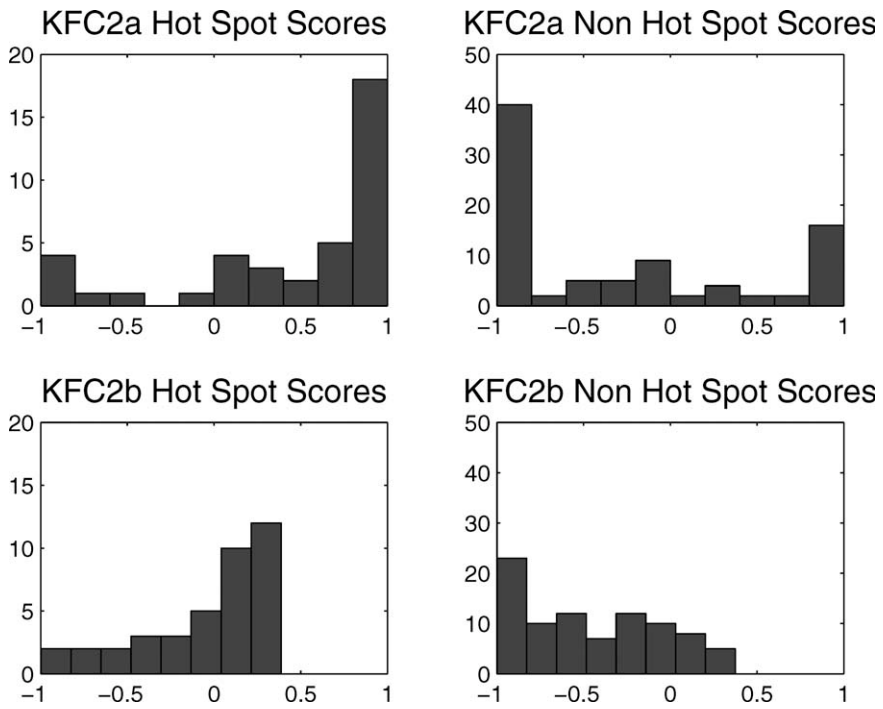

Figure 1

The histograms give score distributions of hot spot residues and non-hot spot residues for the KFC2a and KFC2b models on the mixed independent test set. The KFC2a model is seen to strongly distinguish between positive and negative examples. The KFC2b model is more conservative in scoring its hot spot predictions, but the positive predictions are overwhelmingly true hot spots. Below −0.5, the negative predictions are predominantly non-hot spots.

This is a direct and effective method for protein–protein interface hot spot identification; however, it involves considerable time and expense to express, purify, and test many protein sequences.

In recent years, several computational methods, which simulate experimental alanine scanning, have been developed to predict protein–protein interface hot spots.[15–21] Some of these methods are based on molecular dynamics simulations[17] while others are based on different energy functions.[15,16,18] Machine learning approaches,[22–26] such as KFC[26,27] and MINERVA,[25] have also been developed to identify hot spots at protein–protein interfaces based on geometric and biochemical features of residue–residue contacts across binding interfaces. Theoretical methods[28,29] are also used to predict interface hot spots. More recently, a simple empirical method, based only on residue-residue pairwise potentials and solvent accessibility,[30,31] has been developed and shown to achieve good predictive accuracy. Another interesting method, developed by Grosdidier and Fernandez-Recio,[32] predicts interface hot spots using protein docking tools. These computational methods provide a complement to experimental methods and can reduce the number of mutations that experimental researchers must pursue when attempting to modulate binding affinity within a protein interface.

Although these computational methods have made great progress in predicting protein–protein interface hot spot residues, there is still the potential for improvements in calculation speed and predictive accuracy. Compared with the original Knowledge-based FADE and Contacts (KFC) approach developed by Darnell *et al.*,[26] the methods of other groups have clearly improved predictive ability. However, the true positive rate on hot spot residues is still quite low for many of these methods. For an experimental scientist, it is often crucial to achieve good accuracy on true hot spot residues. A few false positives can be later eliminated during the course of experimentation, whereas hot spots that are not identified, and thus not tested, become lost information.

In this article, we propose two improved knowledge-based hot spot prediction methods based on the KFC approach.[26] One of them, KFC2a, has superior predictive accuracy and calculation speed. This is particularly true when its accuracy at predicting hot spot residues is compared to other methods. The other method we developed, KFC2b, has impressive predictive accuracy for both hot spot residues and non-hot spot residues. We see from the score distributions for both models on the mixed independent test data that they reliably distinguish between hot spot and non hot spot residues (see Fig. 1).

The training data set is an important factor in building a model. The training data sets used in creating prior machine learning methods, such as KFC and MINERVA2, contain more non-hot spot residues than hot spot residues, which

can bias their models to predict non-hot spot residues preferentially. To avoid biased predictions, we generated a training data set of interface residues that contained nearly equal numbers of hot spot residues and non-hot spot residues.

Features for machine learning methods are another key factor for building a knowledge-based model. Several new features have been added to KFC2 compared with the original KFC model, such as features related to solvent accessible surface areas and properties of amino acid neighbors of the target residue. Additional features like π–π and cation–π interactions are also considered in KFC2. In all, 47 features were generated, and combinations with good predictive accuracy were identified. Two feature combinations were ultimately selected for the final models. The KFC2a model trained on eight features gives excellent predictive accuracy for both our training and mixed independent data sets. The KFC2b model, trained on seven features, also has good predictive ability and a lower false positive rate than KFC2a. Providing two models allows researchers to tailor predictions to their needs, as some may prefer a model with the highest true positive rate, while others may wish to minimize false positive predictions.

Both KFC2a and KFC2b were compared with existing computational hot spot prediction methods. Parameters such as recall, precision, and the F1 score have been used previously to create comparisons between methods. However, the recall and F1 score are related to the composition of the data set, making it difficult to compare performance for methods trained and tested on different data. For this reason, we use the accuracy on hot spot residues (Sensitivity, or True Positive Rate) and accuracy on non-hot spot residues (Specificity, or True Negative Rate) to evaluate our models, as these are independent of the ratio between positive and negative examples. The Supporting Information also gives the recall, precision and F1 for all methods, along with a discussion of how the data composition affects the F1 score.

METHODS

Data sets

Balanced training data set for cross-validation

The training data is a subset of that used by Cho *et al.*,[25] which includes 265 interface residues with observed alanine scanning energy differences. These data were obtained from ASEdb[33] and the published data of Kortemme and Baker.[16] To achieve a balanced training data set, we selected all 65 amino acids in 17 protein complexes for which a mutation to alanine produces a $\Delta\Delta G \geq 2.0$ kcal mol^{-1}. Then 67 (or 66) of the remaining 200 non-hot spot residues were selected as part of the balanced training data set. We divided the 200 non-hotspot residues by assigning them to three groups in a sequential fashion (1, 4, 7,... into one group; 2, 5, 8,... into another; and 3, 6, 9,... into the third). It was felt that this maximized the diversity within the data set, because data were taken from many different proteins. The first group, containing 67 data points, produced slightly better results than the other two groups, and so our final model is based on this training data. Thus, our training data set consisted of 132 interface residues, 65 of which were hot-spot residues and 67 of which were non-hot spot residues. The remaining 133 non-hot spot residues were used as an additional independent test set. Table I lists the protein structures from which the data are derived, and the experimentally determined $\Delta\Delta G$ values for these residues are listed in the Supporting Information (Table SI).

Table I
The 17 Complexes Used in the Training Data Set

PDB code	Molecule 1 name	Molecule 2 name
1A4Y	RNase inhibitor	Angiogenin
1A22	Human growth hormone	Human growth hormone binding protein
1AHW	Immunoglobulin Fab5G9	Tissue factor
1BRS	Barnase	Barstar
1BXI	Colicin E9 immunity Im9	Colicin E9 DNase
1CBW	BPTI trypsin inhibitor	Chymotrypsin
1DAN	Blood coagulation factor VIIA	Tissue factor
1DVF	Idiotopic antibody FV D1.3	Anti-idiotopic antibody FV E5.2
1F47	Cell division protein ZIPA	Cell division protein FTSZ
1FC2	Fc fragment	Protein G
1FCC	Fc(IGG1)	Protein G
1GC1	Envelope protein GP120	CD4
1JRH	Antibody A6	Interferon-gamma receptor
1NMB	N9 neuramidase	Fab NC10
1VFB	Mouse monoclonal antibody D1.3	Hen egg lysozyme
2PTC	BPTI	Trypsin
3HFM	Hen egg lysozyme	Ig FAB fragment HyHEL-10

Negative independent test set

As mentioned above, 67 non-hot spot residues out of 200 non-hot spot residues were selected as part of our training data set, and the remaining 133 non-hot spot residues were used as an additional independent test set (Supporting Information Table SII). As with the training data, these data have reported $\Delta\Delta G$ values.

Mixed independent test set

We used the same independent test set as Cho *et al.*[25] This data set is constructed from the Binding Interface Database (BID).[34] There is numerical alanine mutation data in the BID database for some examples; however only "strong," "intermediate," "weak," and "insignificant" were used to denote the effect of each alanine mutation in these cases. In addition, only "strong" mutations are considered to be hot spots in this independent test set. This test set includes 126 alanine-mutated data coming from 18 complexes (Supporting Information Tables SIII and SIV), of which 39 residues are identified as hot spots.

Features

Definition of interface residues

We used the change in buried solvent accessible surface area upon complex formation to define the interface residues. Solvent accessible surface areas of residues in the protein–protein complex and residues in the monomers were calculated using NACCESS.[35] If the difference between these values was greater than zero, the corresponding residue was defined as an interface residue.

Hydrophobicity

A residue's size and hydrophobicity were considered as features of the residue. We used the residue's molecular weight and solvent accessible surface area to characterize its size. The hydrophobicity of a residue was a useful feature in the model learning, especially when considered in conjunction with the *CORE_RIM* feature defined below. We used the Fauchere and Pliska[36] scale to score hydrophobicity.

Features related to solvent accessibility

The solvent accessibility of an interface residue is important for determining if the residue may be an interface hot spot. We calculated the solvent accessible surface area using the program NACCESS[35] and generated nine different features related to the residues' solvent accessibility: *DELTA_TOT, DELTA_SC, SA_RATIO1, SA_RATIO2, SA_RATIO3, SA_RATIO4, SA_RATIO5, CORE_RIM, POS_PER.*

DELTA_TOT is the difference in a residue's solvent accessible surface areas in the bound and unbound structures, which gives the buried solvent accessible surface area of a residue upon complexation.

$$DELTA\_TOT = ASAunb - ASAbnd. \quad (1)$$

DELTA_SC has a similar definition to *DELTA_TOT*, except that we calculated the difference between a residue's side chain solvent accessible surface areas in the bound and unbound structures. This definition excludes any changes in solvent accessibility for backbone atoms but is otherwise identical to *DELTA_TOT.*

$$DELTA\_SC = ASAunb(side\ chain) - ASAbnd(side). \quad (2)$$

SA_RATIO1 is the ratio between a residue's solvent accessible surface area when bound in a complex and the residue's maximum solvent accessible surface area as a tripeptide,[37] which shows to what extent the residue is buried in the complex.

$$SA\_RATIO1 = \frac{ASAbnd}{maxASA}. \quad (3)$$

SA_RATIO2 is the ratio between *DELTA_TOT* and the residue's maximum solvent accessible area in a tripeptide,[37] which reflects whether the buried absolute solvent accessible surface area of a residue is large enough to be considered important.

$$SA\_RATIO2 = \frac{DELTA\_TOT}{\max ASA}. \quad (4)$$

SA_RATIO3 is the ratio between a residue's side chain solvent accessible area in complex and the residue's maximum side chain solvent accessible area in a tripeptide[37]:

$$SA\_RATIO3 = \frac{ASAbnd(side\ chain)}{\max ASA(side\ chain)}. \quad (5)$$

SA_RATIO4 is the ratio between DELTA_SC (buried side chain solvent accessible surface area of a residue when the complex forms) and the residue's maximum side chain solvent accessible area in a tripeptide[37]:

$$SA\_RATIO4 = \frac{DELTA\_SC}{\max ASA(side\ chain)}. \quad (6)$$

SA_RATIO5 reflects whether a residue is well surrounded in the unbound state and whether it is well buried in the complex.

$$SA\_RATIO5 = \frac{DELTA\_TOT^{*}\max ASA}{ASAunb}. \quad (7)$$

CORE_RIM indicates a residue's position at the protein–protein interface, in particular whether it lies at the rim or core of the interface.

$$CORE\_RIM = \frac{DELTA\_TOT}{ASAunb}. \quad (8)$$

POS_PER is a second measure of position within the protein interface. *CORE_RIM* reflects a residues' absolute position at the protein–protein interface, but for some protein–protein interfaces, only a small fraction of interface residues can be considered "core" residues while most interface residues are more accurately classified as being at the boundary between the core and the rim. If we only use *CORE_RIM*, these boundary interface residues are less likely to be predicted as hot spot residues, despite the fact that many boundary residues are hot spots. To address this issue, we defined the *POS_PER* feature to be a complement of *CORE_RIM*. We arranged the *CORE_RIM* values of residues in the same protein–protein interface from smallest to largest and assigned each *CORE_RIM* value a sequence number (from 1 to *N*, where *N* is the total number of residues of the interface), and then normalized the sequence number of these interface residues by dividing each sequence number by the total number of the interface residues (*N*).

Features related to neighbors of the target residue

The environment of a target interface residue can affect its importance to the protein interaction. We

defined environmental features of a target residue with two distance cutoffs 4.0 and 5.0 Å. Within this environment, we calculated the number of residues and the number of atoms around the side chain of the target residue. The number of rotatable single bonds within the side chain was also calculated to reflect the flexibility of the residue. Additionally, we calculated the weighted rotatable single bond number, which is the number of rotatable single bonds divided by the number of atoms in the side chain. We can think of these bond number features as measuring the degree of local plasticity within the region of the interface that surrounds the target residue, which suggests the degree to which the interface might adapt to a mutation. Finally, the total hydrophobicity of residues around the side chain of the target residue was also calculated to reflect the environmental hydrophobicity of the target residue.

ATMN4 and *ATMN5* give the total number of neighboring atoms of a target residue's side chain. *ATMN4* is for the distance cutoff 4.0 Å, and *ATMN5* is for the distance cutoff 5.0 Å. The neighbor atoms of a target residue are calculated by considering the whole complex.

RESN4 and *RESN5* are defined similarly to *ATMN4* and *ATMN5*, except that they tabulate the number of neighboring residues instead of the number of neighboring atoms. A neighbor residue is included if any of its atoms makes a close contact with a side chain atom in the target residue. *RESN4* uses the distance cutoff 4.0 Å, and RESN5 uses the distance cutoff 5.0 Å.

ROT4 and *ROT5* give the total number of side chain rotatable single bonds to target residues for residues within distance cutoffs 4.0 and 5.0 Å. These features measure the potential of the interface to remold itself locally in response to mutation.

WT_ROT4 and *WT_ROT5* give weighted rotatable single bond numbers of a residue's side chain. These are calculated by dividing the rotatable single bond numbers by the number of atoms in the residue's side chain, using distance cutoffs of 4.0 or 5.0 Å.

HP4 and *HP5* give the summation of hydrophobic scores of the neighboring side chains of a target residue using a distance cutoff of 4.0 or 5.0 Å. The Fauchere and Pliska hydrophobic scale[36] was used to calculate each residue's hydrophobicity.

Biochemical contact features

Biochemical contact features were calculated following Darnell *et al.*[26] Specifically, we used the molecular modeling package WHAT-IF[38] to identify non-covalent interactions within protein complexes. Three kinds of non-covalent interactions were calculated: atomic contacts, hydrogen bonds, and salt bridges.

An atomic contact was recorded if the distance between the van der Waals surfaces of two atoms was less than 0.25 Å. Atomic contacts were classified using three categories: polar, nonpolar, and generic. Only contacts across the interface (versus contacts internal to one of the interacting partners) were tabulated.

Hydrogen bonds were identified by the WHAT-IF program using an optimized hydrogen bond network model.[39] Hydrogen bond scores between 0 and 1 are assigned based on the strength of the hydrogen bond. The hydrogen bond score of a residue is the sum of the hydrogen bond scores of its atoms.

Salt bridges were identified by the WHAT-IF program when the distance between the centers of a basic nitrogen and an acidic oxygen was less than 4.5 Å. The salt bridge feature of a residue is the number of salt bridges it makes with its binding partner.

Features related to FADE points

Local atomic density was shown to be an important feature of interface residues in our prior studies of hot spots, and a rapid method for calculating atomic density was previously developed.[40] In this method, grid points are generated in the space around a protein–protein complex, and each grid point that is within 3.0 Å of both binding partners is identified as a FADE point. Then a spherical volume with a radius of 10.0 Å centered at each interface residue's center of mass is partitioned into 10 nested spherical shells each with a thickness of 1.0 Å. FADE points in different shells are summed, resulting in a FADE point distribution surrounding the target residue. This distribution represents the local packing density of the target residue within concentric shells. Integrals of these shells provide density information within a 3D region. In total, we created 17 features related to FADE points based on shells or volumetric regions with different upper and lower bounds for the radii. *FADE_POINT_9*, which was selected as one feature of KFC2b, has a lower bound of 8.0 Å and an upper bound of 9.0 Å. *FADE_POINT_10*, which was selected as one feature of KFC2a, has a lower bound of 9.0 Å and an upper bound of 10.0 Å.

π–π and π–cation interaction features

It is known that π–π and π–cation interactions are important in protein-protein binding, and we calculated these interactions according to the criteria in McGaughey *et al.*[41] and Gallivan and Dougherty.[42]

Derived features

Additional features were derived from the individual features defined above.

HBdRATIO3 indicates the presence of a well-buried hydrogen bond. If a hydrogen bond is formed across the protein–protein interface, and this hydrogen bond is not solvent exposed, the participating residues are more likely to be hot spots.[43] Thus, we defined a feature proportional to the hydrogen bond score defined using WHAT-IF and inversely proportional to the side chain solvent accessibility.

$$HBdRATIO3 = \frac{HBscore}{SA\_RATIO3}. \quad (9)$$

PLAST4 and *PLAST5* measure the potential for local deformations within the protein interface. In alanine scanning experiments, a residue's side chain is replaced by a single carbon atom, and the conformations of its neighboring residues may or may not adapt to this change. In the case where adaptation is not possible, a hot spot is more likely to occur.

$$PLAST4 = \frac{WT\_ROT4}{ATMN4^{*}\max ASA(side\ chain)}, \quad (10)$$

$$PLAST5 = \frac{WT\_ROT5}{ATMN5^{*}\max ASA(side\ chain)}. \quad (11)$$

The ratio *WT_ROT/ATMN* measures the average flexibility of the residues around the target residue and the *maxASA* (side chain) represents the size of the side chain of the target residue. This local measure of potential adaptability is proportional to the average flexibility and inversely proportional to the size of the side chain of the target residue. Alternately, this can be thought of as the balance between the local plasticity in the region surrounding a residue and the void created by replacing the side chain of the residue by a single atom.

Training with SVM

Many different training methods can be used to build a knowledge-based model for predicting protein–protein interface hot spots. On the basis of the efficiency and robust output of support vector machines (SVM),[44] we used the program SVM^{light}[45] to train our models. We also tried the decision tree method C5.0[46] used to create the original KFC model, but the models trained with SVM exhibited higher predictive accuracy than those trained using decision trees.

Feature combinations

Many features were generated for testing, but using all of them would likely lead to over-fitting of the models. We tried different total feature numbers (such as 4, 6, 8, 10, 12, and 14). We divided features into different classes, such as features related to solvent accessibility, features related to FADE points and so on. After this, we combined features from different classes to create diverse feature combination sets.

Kernel functions

Different kernel functions, such as polynomial kernel functions, linear kernel functions, and Gaussian Radial Basis Function (RBF) kernel functions, were used to train the model. We tried different C factors (ranging from 0 to 64.0) different G factors (ranging from 0 to 2) for Gaussian RBF kernel functions. The value of C controls the trade-off between allowing training errors and forcing strict margins, while the value of G determines the Gaussian RBF width. Depending on the composition of the data set and the number of features used, the optimal values for C and G can vary tremendously.

Cross-validation

Three different strategies for cross-validation were used to evaluate the performance of our model. Considering the correlation of residues in the same interface, Darnell *et al.*[26] used a "leave one-protein-complex out" cross-validation to estimate the performance of their model. Cho *et al.* used a 10-fold cross validation to validate their model,[25] dividing their training data into 10 roughly equal subsets, then successively training on 90% of the data and testing on the remaining 10% over the course of 10 training runs. We applied both "leave-one-protein out" and 10-fold cross validation as well as the standard "leave one out" cross-validation, to estimate the performance of our model. All three methods produced nearly identical results, which are given in the Supporting Information (Table SV). We will present the results obtained by standard leave-one-out cross validation, which are identical to the 10-fold cross-validation.

Based on different feature combinations, different kernel functions were used to train different models. The leave-one-out cross validation was then used to check whether the models were over-fitted, and the independent test set was used to check the predictive accuracy of different models. According to the results of the cross-validation and the predictive accuracy of the mixed independent test set, models with better predictive accuracies for hot and non-hot spot residues were selected, along with the corresponding feature combinations of those models. From the results of these models, we determined that the Gaussian RBF kernel functions produced the best models. For the feature combinations we selected, the G factors were optimized to refine the models. MATLAB code was then used to visualize the distribution of scores for SVM training in different data sets. The score distributions of both hot spot residues and non-hot spot residues showed the ability of different models to distinguish hot spot residues from non-hot spot residues (see Fig. 1). Finally, using the Gaussian RBF kernel function with $C = 8.0$ and $G = 0.00002$, KFC2a, which showed the highest true positive rate (TPR or sensitivity), was trained using eight features: hydrophobicity, *DELTA_TOT*, *CORE_RIM*, *POS_PER*, *ROT5*, *PLAST4*, *PLAST5*, and *FADE_POINT_10*. The first four of these features relate primarily to solvation effects and burial within the interface, while the final four features relate to local plasticity and atomic density. Using the Gaussian RBF kernel function with $C = 0.03$ and $G = 0.000065$, KFC2b, which showed a

high sensitivity and lower false positive rate (FPR, which is equal to 1-specificity) than KFC2a, were trained using seven features: residue size, *POS_PER*, *SA_RATIO5*, *ROT4*, *ROT5*, *HP5*, *FADE_POINT_9*. Compared with the features for KFC2a, residue size and hydrophobicity of residues around target residue are specific features of KFC2b.

Evaluation parameters: Sensitivity, specificity, and accuracy

Although many measures have been used for the assessment of protein–protein interface hot spot prediction methods, these parameters may not be convincing in some situations. Simple but basic parameters, such as sensitivity, specificity, and accuracy, can show the predictability of different models. The numbers of true positives, false positives, true negatives, and false negatives are designated, respectively as TP, FP, TN, and FN. Then:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{12}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{13}$$

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

In addition to these measures, the area under the curve (AUC) for the receiver operating characteristic (ROC) curve is another important parameter with which to assess our methods.

RESULTS

Feature selection

Because we did not know in advance how many features would result in over-fitting the model to our data set, different total numbers of features (such as 4, 6, 8, 10, 12, and 14) were specified. Given some total number of features, we then maximized the number of categories (such as features related to solvent accessibility, neighbors of the target residue, and so on) that could be sampled. This was an effective way to decrease the combinational complexity of feature sampling while creating models with diverse feature combinations.

According to the models trained using different feature combinations, we found that the following eight features achieved the best model (KFC2a) among those we examined: hydrophobicity, *DELTA_TOT*, *CORE_RIM*, *POS_PER*, *ROT5*, *PLAST4*, *PLAST5*, and *FADE_POINT_10*. Many more features and combinations were tried, but no better models were obtained. This feature combination implies that the solvent accessibility of residues, atomic density, relative position (core vs. rim) within the interface, and local interface plasticity can be used to predict protein–protein interface hot spots and non-hot spots with high accuracy. Our approach incorporates local plasticity features that estimate the potential for structural adaptation to a mutation, which are unique to our model.

Table II
The Confusion Matrices of Different Models for the Same Training Data Set

		Predicted[b]	
	Actual[a]	Hot spot	Other
KFC2a	Hot spot	51 (TP)	14 (FN)
	Other	15 (FP)	52 (TN)
KFC2b	Hot spot	36 (TP)	29 (FN)
	Other	9 (FP)	58 (TN)
Robetta	Hot spot	33 (TP)	32 (FN)
	Other	8 (FP)	59 (TN)
FOLDEF	Hot spot	20 (TP)	45 (FN)
	Other	5 (FP)	62 (TN)
KFC	Hot spot	36 (TP)	29 (FN)
	Other	8 (FP)	59 (TN)
MINERVA2	Hot spot	38 (TP)	27 (FN)
	Other	3 (FP)	64 (TN)
HotPoint	Hot spot	34 (TP)	31 (FN)
	Other	9 (FP)	58 (TN)

[a]Class determined by experiment.
[b]Class predicted by the model.

KFC2a predicted hot spot residues with the highest sensitivity (Sensitivity = 0.78 for the training data set and Sensitivity = 0.85 for the mixed independent test data set) and the highest accuracy for the training data set (Accuracy = 0.78); however, the specificity of KFC2a was a somewhat lower than for several other methods (Specificity = 0.78 for the training data set, and Specificity = 0.71 for the mixed independent test set. See Table III and Table V). We also developed a second model (KFC2b), which had comparable predictive accuracies with other methods for both hot spot residues and non-hot spot residues and a higher specificity than KFC2a (Specificity = 0.87 for the training data set, and Specificity = 0.85 for the mixed independent test data set). This model was trained on seven features: residue size, *POS_PER*, *SA_RATIO5*, *ROT4*, *ROT5*, *HP5*, *FADE_POINT_9*. In this way, we can provide users of KFC2 with two sets of predictions that allow them to choose whether they prefer a model to predict more hot spots with a somewhat lower specificity (KFC2a), or whether they prefer fewer hot spots with a higher specificity (KFC2b). We also looked at adjusting the cutoff for the KFC2a score to achieve this goal, but we ultimately achieved a better result by retraining a different model.

Cross-validation using the balanced training data set

To make sure we did not over-train both of our models, KFC2a and KFC2b were tested by the standard leave-one out cross validation (the leave-one protein out cross validation and the 10-fold cross-validation were also used

Table III
Evaluation of the Prediction Results of the Cross Validations of Different Models Using Sensitivity (TPR or Recall), Specificity (TNR), and Accuracy

	Sensitivity	Specificity	Accuracy
KFC2a	0.78	0.78	0.78
KFC2b	0.55	0.87	0.71
MINERVA2[a]	0.58	0.96	0.77
HotPoint	0.52	0.87	0.70
KFC[a]	0.55	0.88	0.72
Robetta	0.51	0.88	0.70
FOLDEF	0.31	0.93	0.62

[a]The predicted results are from the original cross-validation.

to test the model), and the results of these cross-validations are shown in Supporting Information Table SV. KFC2a and KFC2b were compared with other methods (HotPoint,[30,31] MINERVA2,[25] KFC,[26,27] FOLDEF,[15] and Robetta[16]), by assessing residues from the same training data set using the other methods. Table II shows the confusion matrices for our methods compared with other methods. Of these methods, KFC2a predicted the most actual hot spots as hot spots (51/65). We also found that the predictive accuracy of KFC2b was comparable to that of other methods, and only slightly outperformed by MINERVA2 on the cross-validated training data (however, note that KFC2b outperformed MINERVA2 on the mixed independent test set as will be discussed in the next section.) We calculated the sensitivity, specificity and accuracy for all these methods (Table III). According to these parameters, KFC2b had a similar sensitivity to other methods (Sensitivity = 0.55), and this value was only lower than that of MINERVA2 (Sensitivity = 0.58). The specificity and accuracy of KFC2b were also comparable with other methods except MINERVA2. KFC2a had the highest sensitivity (Sensitivity = 0.78), which represents the predictive ability for actual hot spot residues. KFC2a also had the highest accuracy (Accuracy = 0.78) compared with other methods. The predictive ROC curves of our models KFC2a and KFC2b for cross validation data set were shown in Figures 2 and 3. The AUC (area under ROC curve) of these two curves were 0.89 for KFC2a and 0.87 for KFC2b. Both of the areas gave two-tailed *P*-values less than 0.0001, which means that the differences between our two models and a random model are statistically significant.

It is important to note that our balanced training data set is a subset of the data used to train MINERVA2 and the original KFC model. Because those models were trained to do well on data in our training set, we cannot simply apply their final models (such as MINERVA2 and KFC) to the examples in our training data set and obtain a fair comparison. Instead, we used the results reported from the Supporting Information in their papers, in which the cross-validated predictions were explicitly reported, to estimate the performance of their models on

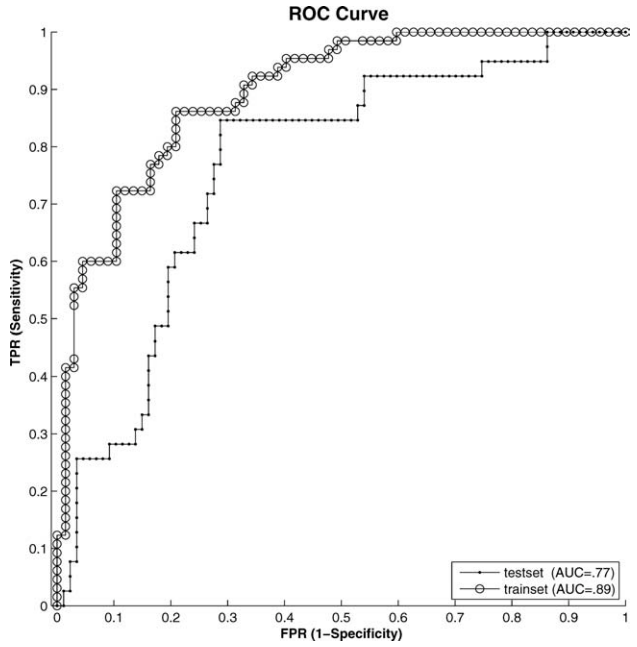

Figure 2
ROC curves for predictions made by KFC2a on the balanced training mixed independent test sets. The AUC for the training data set is 0.89, and the AUC for the test set is 0.77.

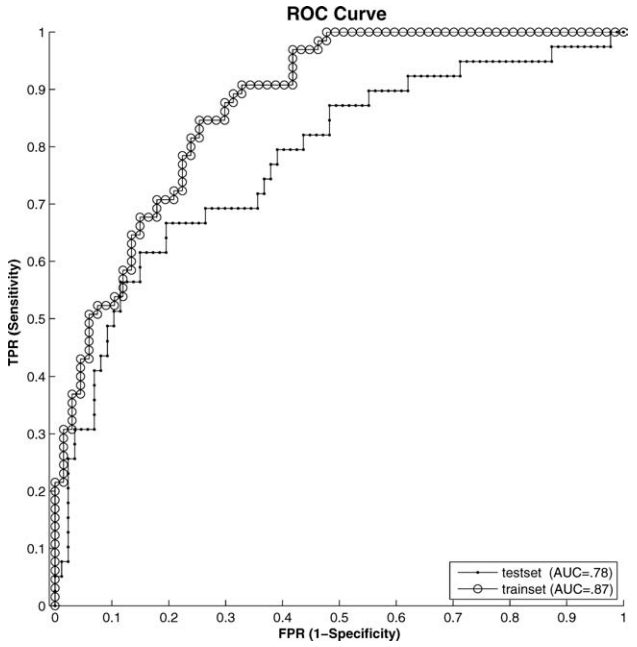

Figure 3
ROC curves for predictions made by KFC2b on the balanced training mixed independent test sets. The AUC for the training data set is 0.87, and the AUC for the test set is 0.78.

Table IV
Confusion Matrices of Different Models for the Same Independent Test Set

		Predicted[b]	
	Actual[a]	Hot spot	Other
KFC2a	Hot spot	33 (TP)	6 (FN)
	Other	25 (FP)	62 (TN)
KFC2b	Hot spot	24 (TP)	15 (FN)
	Other	13 (FP)	74 (TN)
Robetta	Hot spot	13 (TP)	26 (FN)
	Other	12 (FP)	75 (TN)
FOLDEF	Hot spot	10 (TP)	29 (FN)
	Other	11 (FP)	76 (TN)
KFC	Hot spot	12 (TP)	27 (FN)
	Other	13 (FP)	74 (TN)
MINERVA2	Hot spot	18 (TP)	21 (FN)
	Other	8 (FP)	79 (TN)
HotPoint	Hot spot	24 (TP)	15 (FN)
	Other	25 (FP)	62 (TN)

[a]Class determined by experiment.
[b]Class predicted by the model.

our balanced training data. This approach is the fairest possible comparison as it is based on the unbiased assessment of their models on our training data, whereas direct application of the final models would lead to extremely high rates of predictive accuracy due to having been trained to perform well on these particular data.

Prediction performance for the mixed independent test set

The predictive performance of our models on the mixed independent test set is shown in Table IV. As with the training data, the performance of our models was compared to other models. Table V shows the assessment results of different models for the mixed independent test set. From Table V, we can see that KFC2a showed the highest sensitivity (Sensitivity = 0.85), which outperformed all other methods. This means that KFC2a can identify most of interface hot spot residues, which is a meaningful result for experimental scientists. KFC2b displayed the best accuracy of the mixed independent test set among all the methods. Its sensitivity (Sensitivity = 0.62) was higher than other methods except KFC2a and HotPoint. The specificity of KFC2b was comparable with other methods. The predictive ROC curves of KFC2a and KFC2b for this independent test set were also shown in Figures 2 and 3; the AUC (area under ROC curve) of these two curves were 0.77 for KFC2a and 0.78 for KFC2b. As with the training cross-validation, *P*-values of the AUC areas were less than 0.0001, implying the results were statistically significant.

Prediction performance for the negative independent test set

We also used the non-hot spot residues that were omitted from the training data used by Cho *et al.* as an additional test of predictive accuracy. For these data, the specificity of KFC2a was 0.68 (91/133), which is similar to the specificity (0.71) of the non-hot spot residues in the mixed independent test shown in Table V. The specificity of KFC2b was 0.82 (109/133), which is also similar to the specificity (0.85) of the non-hot spot residues in the mixed independent test shown in Table V.

Table V
Evaluation of the Hot Spot Prediction by Different Methods for the Mixed Independent Test Set

	Sensitivity	Specificity	Accuracy
KFC2a	0.85	0.71	0.75
KFC2b	0.62	0.85	0.78
MINERVA2	0.46	0.91	0.77
HotPoint	0.62	0.71	0.68
KFC	0.31	0.85	0.68
Robetta	0.33	0.86	0.70
FOLDEF	0.26	0.87	0.68

DISCUSSION

Based on our balanced training data set, we built two new models: KFC2a and KFC2b. For the mixed independent test set, KFC2a had the best predictive accuracy for hot spot residues (sensitivity) but lower predictive accuracy for non-hot spot residues (specificity) in com-

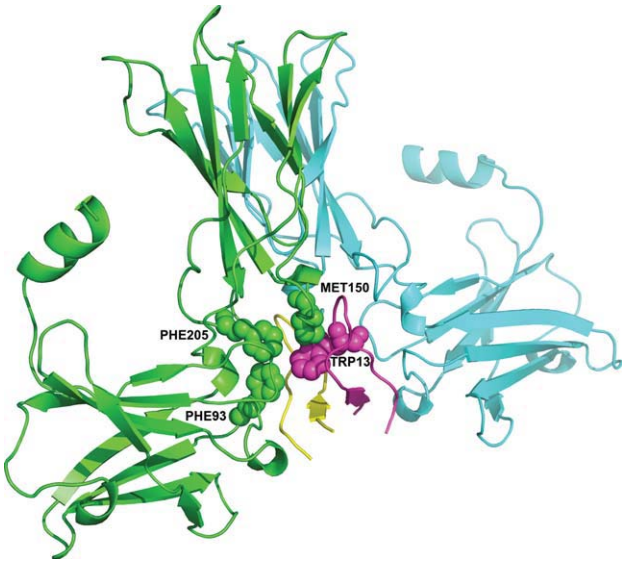

Figure 4
Interface hot spot residues of erythropoietin receptor and erythropoietin mimetic peptide (PDB 1ebp), for which all hot spot residues were correctly predicted by KFC2a but not by the original KFC model (see Supporting Information Table SIV). Green: chain A; Cyan: chain B; Magenta: chain C; Yellow: chain D.

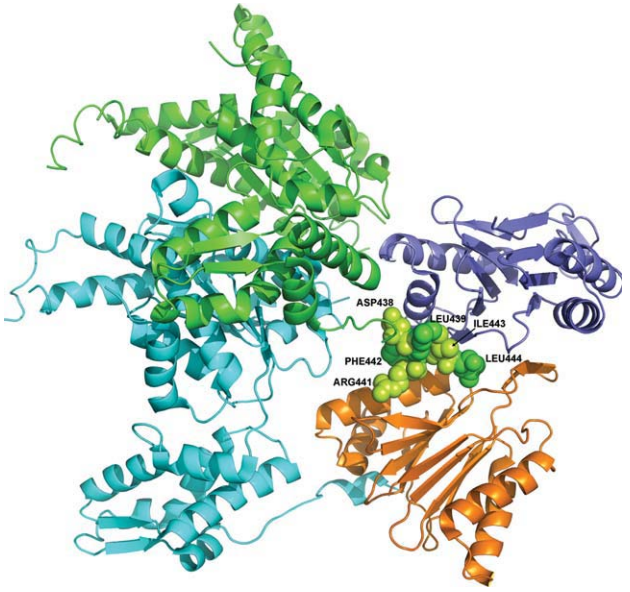

Figure 5

Interface hot spot residues of ATP-dependent HSLU protease ATP-binding subunit and ATP-dependent protease HSLV (PDB 1g3i), for which all hot spot residues were correctly predicted by KFC2a but not by the original KFC model (see Supporting Information Table SIV). Green: chain A; Cyan: chain B; Orange: chain G; Blue: chain H.

parison with other methods. As is seen in Figures 4 and 5, important hot spots were missed by our original KFC model but are now identified by KFC2a. KFC2b also had better predictive accuracy for hot spot residues than other methods (except KFC2a and HotPoint), and its predictive accuracy for non-hot spot residues was comparable with other methods. The SVM score distributions for the mixed independent test set are shown in Figure 1. The concentration of predictions of near 1.0 for hot spots and −1.0 for non-hot spots suggest that KFC2a can discriminate these hot spot residues and non-hot spot residues with great confidence. The KFC2b model performs better for predicting non-hot spots than hot spots, as is true for other published methods.

Balanced training data set

A balanced training data set, which contains nearly equal numbers of hot spot residues and non-hot spot residues, avoids training a model that is biased toward predicting the type of data (positive or negative examples) that are overrepresented in the data. For both our models, the predictive accuracy does not significantly incline to one type of the data. As experimental alanine scanning data (especially hot spot residue data) increases, it is more reasonable to use balanced training data sets to build hot spot prediction models. In a real biological system, only 5–10% of residues are hot spots,[10,12,24] and some researchers may prefer to build models based on this skewed distribution. It is likely that this will cause problems for machine learning techniques, and to us, it is more logical to strive for equal predictive accuracy on positive and negative examples.

Feature selection

We calculated 47 different features, and a subset of these was selected to train the final models. The features for KFC2a were primarily related to solvent accessibility, position within the interface, packing density, and local plasticity. Features for KFC2b were related to residue size, solvent accessibility, packing density, and the flexibility and hydrophobicity of residues around the target residue.

Features related to the solvent accessibility of residues have been shown to be an important factor for hot spot prediction in our models. These observations are consistent with those of Tuncbag *et al.*, whose HotPoint[31] method is among the best current hot spot prediction methods and for which the main features relate to solvent accessibility and position within the interface. The change of solvent accessible surface area when a complex forms is related to the solvation energy, which is one of the most important factors for protein–protein binding.[47] It has also been observed that residues at the core of an interface have a higher probability of being hot spot residues compared with residues at the rim of the interface.[12,48–50]

Given that the sizes of protein interfaces and the relative balance between core and rim residues can vary greatly between different protein–protein interfaces, we ranked our *CORE_RIM* values from smallest to largest and created a normalized score (*POS_PER*) by dividing the rank by the total number of the interface residues, in order to allow comparison of *CORE_RIM* values between different interfaces.

Among those features related to neighbors of target residues, the most predictive features measured the potential flexibility of residues surrounding the target residues. This local plasticity suggests the degree to which the interface can remold itself in response to a mutation. Plasticity features were considered in a different way by Lise *et al.*,[23] and it has been observed that hot spot residues are inclined to be more rigid during the course of molecular dynamics simulations, serving as "anchor residues" for the interaction.[43,49,51] If residues around the target residues are also more rigid, this suggests that little entropy is lost upon binding. Additionally, when neighboring residues are able to adapt to a side chain substitution, the presence of a hot spot is less likely because that portion of the interface can accommodate a range of side chain geometries, whereas a hot spot is more likely in the presence of rigidity. Several other features expected to

be helpful for discriminating hot spot residues and non-hot spot residues were not selected such as π–π, cation–π, and hydrogen bond interactions. This does not imply that these interactions are unimportant, but rather suggests something about predictive accuracy. For example, if the inclusion of hydrogen bonds causes a model to generate too many false positives, the model will not ultimately be viewed as a good model, even if it identifies many true positives.

Evaluation parameters

Parameters such as F1 score have been considered sufficient parameters to evaluate different models; however, the F1 score is related to the ratio of hot spot residues to non-hot spot residues in the training data set. For example, given fixed values of sensitivity and specificity, the F1 score will be different for different ratios between positive examples and negative examples. The precision, recall, and F1 scores for all methods are reported in Supporting Information Tables SVI and SVII, and Supporting Information Table SVIII demonstrates the effect of data composition on the F1 score.

Here we used the accuracy for positive examples (sensitivity) and the accuracy for negative examples (specificity) to evaluate different models. These two measures are often used to select good models, and, in principle, they should be independent of the ratio of positive to negative examples within the data set. These measures also allow the model developer to choose good models according to different objectives. A model with higher sensitivity can be selected if the goal is to predict as many actual positive examples as possible. If the desire is to minimize the false positive rate, models with higher specificity would be preferred.

Comparison with other methods

Our models were compared with other methods such as Robetta,[16] FOLDEF,[15] KFC,[26,27] MINERVA2,[25] and HotPoint.[30,31] For the cross-validation, the model KFC2a showed the highest sensitivity (Sensitivity = 0.78), while the highest reported sensitivity for other models was 0.58 (MINERVA2), suggesting that KFC2a is a superior model for predicting hot spot residues (see Table III). For the mixed independent test set, the highest sensitivity of other methods was 0.62 (HotPoint), while the sensitivity of KFC2a was 0.85 (see Table V). Note that for the HotPoint method, a distance cutoff of 6.0 Å was used, since the default cutoff of 5.0 Å does not produce predictions for all cases in the training and testing data sets. This would have unfairly reduced the reported sensitivity value.

KFC2b is comparable with other methods on both the sensitivity and the specificity for both the cross-validation and the mixed independent test set. KFC2b showed higher specificity for both the cross-validation and the mixed independent data set compared with KFC2a, demonstrating that it is a good complement to KFC2a. As mentioned before, method selection depends on the demand for either higher sensitivity or higher specificity. It should be noted that Robetta and FOLDEF solved a slightly different problem, which is to estimate exact $\Delta\Delta G$ values, which is arguably more challenging than estimating the probability that $\Delta\Delta G$ exceeds a given threshold.

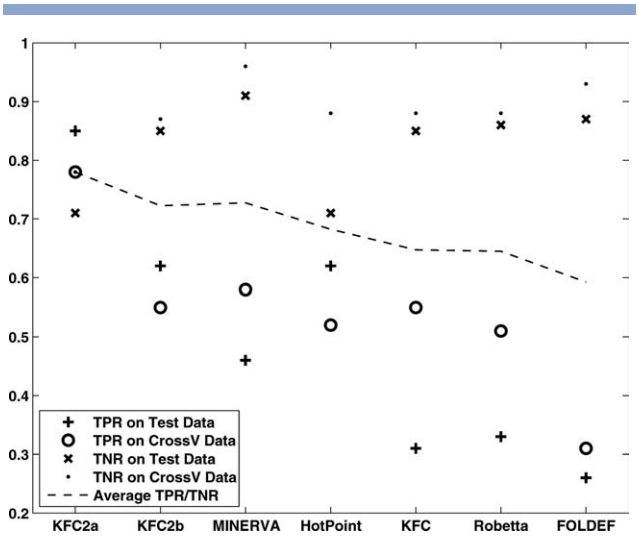

Figure 6

The figure shows the true positive (TPR) and true negative (TNR) rates for the mixed independent test data and cross-validated training data for each method. All methods except HotPoint displayed similar TNR between the training and test data sets. All methods except the original KFC method and Robetta had a similar TPR on both data sets. Their average TPR values are similar to one another, and these are somewhat lower than other methods except FOLDEF. Finally, the average between TPR and TNR was 0.70 or less for most methods when considered on both data sets. We see that KFC2a is the only method that scores higher than 0.70 accuracy in all categories of prediction.

For better comparisons of the predictive accuracies for both hot spot residues and non-hot spot residues, Figure 6 was plotted to provide a comprehensive assessment of each method relative to its true sensitivity and specificity. The rates are computed for both the cross-validation and the mixed independent test data set (see Tables III and V). In this way, the similarities and differences between methods are highlighted. The consistency of sensitivity and specificity across different data sets, the difference in their values, and the average accuracy become apparent in this figure.

The sensitivity and specificity values are similar for the cross-validated training data and the mixed independent test data, with three exceptions. For HotPoint, specificity on the training data is quite high, while it is much lower on the mixed independent test data. For the original KFC method and Robetta, sensitivity was significantly different between the two data sets. This is perhaps due to compositional differences in the two data sets relative

to the preferences of the models, and additional studies may suggest average accuracy rates that lie between the observed values. If we simply average the inconsistent values, KFC and Robetta appear statistically similar.

The average of sensitivity and specificity is an expected total accuracy on balanced data, and this was between 0.60 and 0.70 for most of the methods, with the exception of KFC2a, for which it was nearly 0.80. Expected total accuracy on data that is distributed as 5–10% hot spots and 90–95% non-hot spots strongly favors methods with the highest values for specificity. By this measure, FOLDEF has one of the best average accuracies of any method, yet it only predicts 25–30% of hot spots. Specificity dominates the calculation, and it is possible to have a statistically accurate hot spot prediction method that never predicts any hot spots.

We can adjust cutoffs in our models to change the sensitivity and specificity. It might seem that we could trade a 5% increase in specificity for a 5% decrease in sensitivity, but it is not so simple. For example, changing the cutoffs in the KFC2a model in order to achieve the high specificity of KFC2b (0.87) results in a model that predicts only 30% of hot spots. Conversely, if one adjusts the cutoff value for the KFC2b model to achieve the high sensitivity of KFC2a (0.78), the specificity drops to 0.56, resulting in large numbers of false positive predictions. Adjusting the score cutoff to achieve equality between sensitivity and specificity resulted in values of 0.71 in the case of KFC2a and 0.69 in the case of KFC2b. The use of ROC curves[52] is common to find optimal tradeoffs. For both the KFC2a and KFC2b models, an ROC analysis set the cutoff value very near to zero and calculated an area under the curve between 0.77 and 0.89, which is statistically significant.

Impact of accessibility and plasticity feature groups

It is also useful to assess the relative contribution of different categories of features. Our analysis suggests that features related to solvent accessibility enhance sensitivity of the SVM model, while those related to plasticity more significantly impact the specificity. In particular, using only three solvent accessibility features (DELTA_TOT, CORE_RIM, and POS_PER), it is possible to obtain a model having Sensitivity = 0.87 and Specificity = 0.56 for the mixed independent test set. Using only three plasticity features (ROT5, PLAST4, and PLAST5), it is possible to obtain a model having Sensitivity = 0.51 and Specificity = 0.71. In comparison, the KFC2a model has Sensitivity = 0.85, and Specificity = 0.71 on the mixed independent data set.

Since the majority of hot spots are found within the core of the interface, it makes sense that a model distinguishing core from rim can identify many hot spots. However, such a model has no basis upon which to distinguish between core residues that are true hot spots and ones that are non-hot spots. Results for the plasticity analysis suggest that a lack of plasticity is strongly indicative of a hot spot, but it is not a required characteristic of hot spots. For example, a residue in a flexible region of the interface might still be a hot spot if it participates in a strong non-covalent interaction. By combining groups of features that capture a range of contributing effects, models having both high sensitivity and specificity can be obtained.

CONCLUSIONS

Hot spot residues at the interface comprise a small fraction of the interface residues that account for most of the binding affinity and can be vital to binding specificity. Experimental identification of hot spot residues is costly, and computational methods can thus be helpful in suggesting residues for possible experimentation.

In this study, we present two improved knowledge-based hot spot prediction methods. KFC2a has the highest sensitivity compared to other methods. KFC2b has comparable predictive ability compared to other methods, but it has a higher specificity than KFC2a. A balanced training data set was used to train the models, which avoids creating a method biased to favor prediction of either hot spot residues or non-hot spot residues. Several features were introduced into this new model, such as solvent accessible surface area and flexibility of residues around the target residue. The SVM learning method was used to train our model, and our best model (KFC2a) predicted hot spot residues with an accuracy of 0.85 and non-hot spot residues with an accuracy of 0.71 when applied to the mixed independent test data. We compared our models with other computational hot spot prediction models, and demonstrated that KFC2a offered significant improvement in predicting hot spots. The KFC2 model will be available through http://kfc.mitchell-lab.org concurrently with the publication of this manuscript.

ACKNOWLEDGMENTS

The authors thank Gary Wesenberg for his considerable work on the KFC Server. They thank Gerald Stoecklein for assistance with manuscript preparation and revision.

REFERENCES

1. Kann MG. Protein interactions and disease: computational approaches to uncover the etiology of diseases. Brief Bioinform 2007;8:333–346.
2. Chothia C, Janin J. Principles of protein–protein recognition. Nature 1975;256:705–708.
3. Janin J. Protein–protein recognition. Prog Biophys Mol Biol 1995;64:145–166.
4. Janin J. Principles of protein–protein recognition from structure to thermodynamics. Biochimie 1995;77:497–505.

5. Janin J. Elusive affinities. Proteins 1995;21:30–39.
6. Janin J, Chothia C. The structure of protein-protein recognition sites. J Biol Chem 1990;265:16027–16030.
7. Jones S, Thornton JM. Protein–protein interactions: a review of protein dimer structures. Prog Biophys Mol Biol 1995;63:31–65.
8. Jones S, Thornton JM. Principles of protein–protein interactions. Proc Natl Acad Sci USA 1996;93:13–20.
9. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. J Mol Biol 1993;234:946–950.
10. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol 1998;280:1–9.
11. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science 1995;267:383–386.
12. Moreira IS, Fernandes PA, Ramos MJ. Hot spots—a review of the protein–protein interface determinant amino-acid residues. Proteins 2007;68:803–812.
13. Gonzalez-Ruiz D, Gohlke H. Targeting protein–protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. Curr Med Chem 2006;13:2607–2625.
14. Cunningham BC, Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 1989;244:1081–1085.
15. 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.
16. 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.
17. 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;121:8133–8139.
18. Diller DJ, Humblet C, Zhang X, Westerhoff LM. Computational alanine scanning with linear scaling semiempirical quantum mechanical methods. Proteins 2010;78:2329–2337.
19. Moreira IS, Fernandes PA, Ramos MJ. Computational alanine scanning mutagenesis—an improved methodological approach. J Comput Chem 2007;28:644–654.
20. Huo S, Massova I, Kollman PA. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J Comput Chem 2002;23:15–27.
21. Guharoy M, Chakrabarti P. Empirical estimation of the energetic contribution of individual interface residues in structures of protein–protein complexes. J Comput Aided Mol Des 2009;23:645–654.
22. Assi SA, Tanaka T, Rabbitts TH, Fernandez-Fuentes N. PCRPi: presaging critical residues in protein interfaces, a new computational tool to chart hot spots in protein interfaces. Nucleic Acids Res 2010;38:e86.
23. Lise S, Archambeau C, Pontil M, Jones DT. Prediction of hot spot residues at protein–protein interfaces by combining machine learning and energy-based methods. BMC Bioinformatics 2009; 10:365.
24. Ofran Y, Rost B. Protein–protein interaction hotspots carved into sequences. PLo S Comput Biol 2007;3:e119.
25. Cho KI, Kim D, Lee D. A feature-based approach to modeling protein–protein interaction hot spots. Nucleic Acids Res 2009;37:2672–2687.
26. Darnell SJ, Page D, Mitchell JC. An automated decision-tree approach to predicting protein interaction hot spots. Proteins 2007;68:813–823.
27. Darnell SJ, LeGault L, Mitchell JC. KFC Server: interactive forecasting of protein interaction hot spots. Nucleic Acids Res 2008;36(Web Server issue):W265–W269.
28. Li L, Zhao B, Cui Z, Gan J, Sakharkar MK, Kangueane P. Identification of hot spot residues at protein–protein interface. Bioinformation 2006;1:121–126.
29. Burgoyne NJ, Jackson RM. Predicting protein interaction sites: binding hot-spots in protein–protein and protein–ligand interfaces. Bioinformatics 2006;22:1335–1342.
30. 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.
31. Tuncbag N, Keskin O, Gursoy A. HotPoint: hot spot prediction server for protein interfaces. Nucleic Acids Res 2010;38 (Suppl):W402–W406.
32. Grosdidier S, Fernandez-Recio J. Identification of hot-spot residues in protein–protein interactions by computational docking. BMC Bioinformatics 2008;9:447.
33. Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 2001;17:284–285.
34. Fischer TB, Arunachalam KV, Bailey D, Mangual V, Bakhru S, Russo R, Huang D, Paczkowski M, Lalchandani V, Ramachandra C, Ellison B, Galer S, Shapley J, Fuentes E, Tsai J. The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. Bioinformatics 2003;19:1453–1454.
35. Hubbard SJ, Thornton JM. NACCESS. London: Department of Biochemistry and Molecular Biology, University College; 1993.
36. Fauchere JL, Pliska V. Hydrophobic parameters pi of amino-acid side chains from the partitioning of *N*-acetyl-amino-acid amides. Eur J Med Chem 1983;18:369–375.
37. Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. Nature 1987;328:834–836.
38. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56.
39. Hooft RW, Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. Proteins 1996;26:363–376.
40. Mitchell JC, Kerr R, Ten Eyck LF. Rapid atomic density methods for molecular shape characterization. J Mol Graph Model 2001;19: 325–330, 388–390.
41. McGaughey GB, Gagne M, Rappe AK. pi-Stacking interactions. Alive and well in proteins. J Biol Chem 1998;273:15458–15463.
42. Gallivan JP, Dougherty DA. Cation-pi interactions in structural biology. Proc Natl Acad Sci USA 1999;96:9459–9464.
43. Li X, Keskin O, Ma B, Nussinov R, Liang J. Protein–protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. J Mol Biol 2004;344:781–795.
44. Noble WS. What is a support vector machine? Nat Biotechnol 2006;24:1565–1567.
45. Joachims T. Learning to classify text using support vector machines, Dissertation, Kluwer; 2002.
46. Quinlan JR. C4.5: Programs for machine learning. 1993.
47. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
48. Guharoy M, Chakrabarti P. Conservation and relative importance of residues across protein–protein interfaces. Proc Natl Acad Sci USA 2005;102:15447–15452.
49. Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein–protein interactions. Proc Natl Acad Sci USA 2004;101: 11287–11292.
50. Keskin O, Ma B, Nussinov R. Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol 2005;345:1281–1294.
51. Yogurtcu ON, Erdemli SB, Nussinov R, Turkay M, Keskin O. Restricted mobility of conserved residues in protein–protein interfaces in molecular simulations. Biophys J 2008;94:3475–3485.
52. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143: 29–36.