OPEN ACCESS Freely available online

PLoS one

Predictions of Hot Spot Residues at Protein-Protein Interfaces Using Support Vector Machines

Stefano Lise, Daniel Buchan, Massimiliano Pontil, David T. Jones*

Department of Computer Science, University College London, London, United Kingdom

Abstract

Protein-protein interactions are critically dependent on just a few 'hot spot' residues at the interface. Hot spots make a dominant contribution to the free energy of binding and they can disrupt the interaction if mutated to alanine. Here, we present HSPred, a support vector machine(SVM)-based method to predict hot spot residues, given the structure of a complex. HSPred represents an improvement over a previously described approach (Lise *et al*, BMC Bioinformatics 2009, **10**:365). It achieves higher accuracy by treating separately predictions involving either an arginine or a glutamic acid residue. These are the amino acid types on which the original model did not perform well. We have therefore developed two additional SVM classifiers, specifically optimised for these cases. HSPred reaches an overall precision and recall respectively of 61% and 69%, which roughly corresponds to a 10% improvement. An implementation of the described method is available as a web server at http://bioinf.cs.ucl.ac.uk/hspred. It is free to non-commercial users.

Citation: Lise S, Buchan D, Pontil M, Jones DT (2011) Predictions of Hot Spot Residues at Protein-Protein Interfaces Using Support Vector Machines. PLoS ONE 6(2): e16774. doi:10.1371/journal.pone.0016774

Editor: Collin Stultz, Massachusetts Institute of Technology, United States of America

Received October 23, 2010; **Accepted** December 30, 2010; **Published** February 28, 2011

Copyright: © 2011 Lise et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grant reference BB/E017452/1, from the Biotechnology and Biological Sciences Research Council (BBSRC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: d.jones@cs.ucl.ac.uk

Introduction

Alanine scanning mutagenesis is a powerful experimental methodology for investigating the structural and energetic characteristics of protein complexes [1]. Individual amino-acids are systematically mutated to alanine and changes in free energy of binding ($\Delta\Delta G$) measured. As alanine amino acids do not have a side-chain beyond the β-carbon, this procedure in effect tests the importance of individual side-chain groups for complex formation, providing a map of the so-called functional epitope. Results from a number of experiments indicate that only a small subset of contact residues contribute significantly to the binding free energy. These residues have been termed 'hot spots' and if mutated they can disrupt the interaction. For the majority of interface residues instead, the effect of an alanine mutation is minimal [2].

Hot spots are typically defined as those residues for which $\Delta\Delta G \geq 2$ kcal/mol. In recent years, several computational approaches have been developed to identify them at protein-protein interfaces [3–16]. Accurate predictive models provide a valuable complement to experimental studies and add to our understanding of the factors that influence affinity and specificity in protein-protein interfaces. In addition, they can have important applications in the field of drug discovery. A number of recent studies have been successful in developing (drug-like) small molecules that bind at hot spots and inhibit complex formation [17]. Reliable hot spots predictions could therefore represent the first step in rational drug design projects [18].

In a previous work, we presented a machine learning strategy to identify hot spot residues in protein-protein interfaces, given the structure of the complex [12]. We considered the basic energetic terms that contribute to hot spot interactions, i.e. van der Waals potentials, solvation energy, hydrogen bonds and Coulomb electrostatics, and treated them as input features of a Support Vector Machine (SVM) classifier. We found that the method could predict hot spots with overall good accuracy, comparing favourably to other available approaches. However, by grouping mutations according to the amino acid type, we observed that in some cases the SVM model did not perform too well, for example on predictions involving arginine or glutamic acid residues.

In this paper, we report the development of HSPred, a hot spot prediction method that aims to overcome the limitations highlighted above. For this purpose, we have integrated the original approach with two additional SVM classifiers, specifically built for mutations involving Arg and Glu residues. The two additional models are trained on the same data set as the 'general' model but are biased to perform well on Arg and Glu due to a different choice of input features. Employing a strict cross-validation scheme, we show that this strategy leads to a significant improvement over the previous version of the method. We further validate the results by applying HSPred to an external test case, which is not part of the original data set.

Results and Discussion

The problem we have investigated is the prediction of hot spot residues at a protein-protein interface using a machine learning approach. As input variables, we have considered basic energy terms (van der Waals, hydrogen bond, electrostatic and desolvation potentials) calculated from the complex structure. We have distinguished contributions from different structural regions in the complex, leading to 3 distinct types of interactions: side-chain inter-molecular, environment inter-molecular and side-chain

intra-molecular (see Figure 1). To each of them, we have associated 4 input features, corresponding to the energy terms above. In total therefore there are 12 input features but some of them have not been included in our models because scarcely informative (see Materials and Methods for more details). Support Vector Machines (SVMs) have then be used to learn from a training set to classify residues as hot spots ($\Delta\Delta G \geq 2$ kcal/mol) or non hot spots ($\Delta\Delta G < 2$ kcal/mol).

We have built a classifier, SVM_X, based on the following 7 features: van der Waals, hydrogen bond and solvation side-chain inter-molecular energies; van der Waals, hydrogen bond and solvation environment inter-molecular energies; van der Waals side-chain intra-molecular energy. A summary of the results is reported in Table 1 according to various performance measures. The precision P is the fraction of true hot spots among the set of residues predicted to be hot spots; the recall R is the fraction of correctly identified hot spots relative to all those present in the data set; the $F1$ score is a weighted average of the precision and recall; the Matthews Correlation Coefficient (MCC) is a commonly used measure of the quality of binary classifications (see Methods section for more details). SVM_X is very similar in its design and performance to the model described in [12]. With respect to the latter, SVM_X does not rely on any electrostatic term but it includes the van der Waals side-chain intra-molecular energy. We report in Table 2 the weight of each energy term in the linear scoring function.

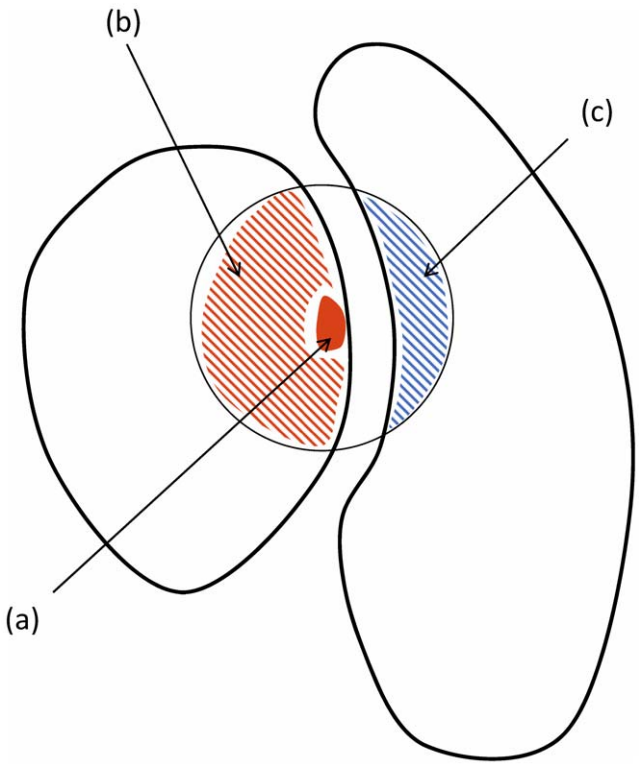


Figure 1. Schematic overview of protein structural regions which define the different energy contributions. The red filled area, (a), corresponds to side-chain atoms of the mutated residue; the red and blue striped regions, (b) and (c) respectively, correspond to atoms within 10A of the C_β of the mutated residue. We distinguish 3 types of interactions: *side-chain inter-molecular* between (a) and (c), *environment inter-molecular* between (b) and (c), *side-chain intra-molecular* between (a) and (b).
doi:10.1371/journal.pone.0016774.g001

Table 1. Summary of results.

Model	Precision	Recall	F1 score	MCC
SVM_X	0.54 ± 0.02	0.64 ± 0.04	0.59 ± 0.02	0.45 ± 0.02
HSPred	0.61 ± 0.02	0.69 ± 0.04	0.65 ± 0.02	0.54 ± 0.02

Cross-validated estimates of performances for SVM_X and HSPred. MCC is the Matthews correlation coefficient (see Methods section for definition of the various performance measures).
doi:10.1371/journal.pone.0016774.t001

We have analysed the SVM_X predictions by grouping mutations according to the amino acid type. In Figure 2(a) we report the results for the most frequent amino acids in the database. SVM_X has a good accuracy over most of amino acid types and is not biased toward some specific amino acid property (e.g. hydrophobic or charged residues). At the same time, however, it does not perform so well on mutations involving Arg and Glu. To tackle this problem, we have developed two additional classifiers, respectively SVM_E and SVM_R, specifically optimised for these amino acids. SVM_E and SVM_R have been trained using the whole data set but differ from SVM_X for the choice of input features and the associated weights (see Table 2).

As can be seen in Figure 2(b), SVM_E and SVM_R achieve significantly improved results on Glu and Arg predictions. A further confirmation of the improvement comes from analysing the correlation coefficients r between the classifiers output scores and the observed $\Delta\Delta G$ values. For Glu residues, r increases from $r = 0.37$ for SVM_X to $r = 0.60$ for SVM_E; for Arg, r increases from $r = 0.40$ for SVM_X to $r = 0.58$ for SVM_R. This suggests that SVM_E and SVM_R are indeed more effective than SVM_X in describing mutations involving Glu and Arg residues, respectively, and that the observed improvement is genuine and not due to over-fitting.

Table 2. Weight of energy terms in the scoring functions.

Feature (energy term)	SVM_X	SVM_E	SVM_R
Side-chain inter-molecular			
van der Waals	0.25 ± 0.03	–	0.79 ± 0.04
hydrogen bond	0.16 ± 0.04	0.63 ± 0.04	–
electrostatics	–	–	–
desolvation	0.21 ± 0.03	–	–
Environment inter-molecular			
van der Waals	0.13 ± 0.02	–	–
hydrogen bond	0.18 ± 0.03	0.69 ± 0.03	0.50 ± 0.04
electrostatics	–	–	–
desolvation	0.10 ± 0.01	–	–
Side-chain intra-molecular			
van der Waals	0.26 ± 0.06	0.60 ± 0.04	0.49 ± 0.04
hydrogen bond	–	–	–
electrostatics	–	–	0.47 ± 0.06
desolvation	–	–	–
Threshold	0.43 ± 0.05	0.54 ± 0.07	0.32 ± 0.07

We report the absolute value of the weight associated to each feature in the scoring functions, together with the threshold that defines the decision boundary. Energy terms which are not included in the scoring function are denoted with the – symbol.
doi:10.1371/journal.pone.0016774.t002

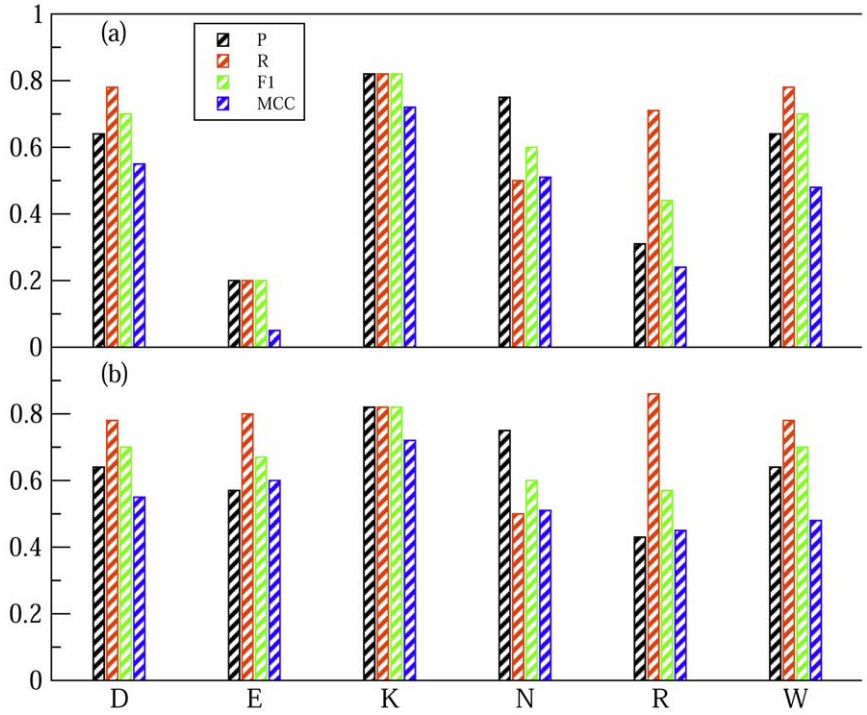


Figure 2. Predictions results for different amino acids. Only the most frequent amino acid in the database are reported. In (a) are the results for SVM_X, in (b) for HSPred, which includes SVM_E and SVM_R.
doi:10.1371/journal.pone.0016774.g002

We have combined SVM_X, SVM_E and SVM_R into a unique classifier, HSPred. SVM_E and SVM_R act respectively on Glu and Arg amino acids, SVM9 on all other amino acid types. We report a summary of the results for HSPred in Table 1. HSPred performs significantly better than SVM_X, reflecting the inclusion of SVM_R and SVM_E. As can be seen from Figure 2(b), predictions on Arg and Glu are roughly as accurate as for the other residues. HSPred therefore successfully overcomes the major limitations of the previously proposed method [12]. Most notable is the improvement on Glu predictions.

To further validate HSPred, we have applied it to the protein-protein complex Ras/RalGDS (PDB code: 1LFD). The Ras/RalGDS complex is not homologous to any of the complexes in the original data set and it can then be regarded as an independent external test case. Experimental $\Delta\Delta G$ values are available in [19], from which we have taken the data corresponding to 16 interface alanine mutations (7 on Ras and 9 on RalGDS). HSPred correctly identifies 6 hot spot (true positives) and 8 non hot spot residues (true negatives). However, 2 residues are wrongly predicted as hot spots (false positives). The predictions are illustrated in Figure 3. These results are in line with the cross-validated estimates in Table 1 and confirm the accuracy of HSPred.

We have implemented HSPred as a fully automatic web server, available at http://bioinf.cs.ucl.ac.uk/hspred. As input it requires a PDB formatted file containing the structure of the protein-protein complex. The user needs to define the interface to analyse by specifying the chain identifiers for each protein on either side of the interface. The output consists of two components: (i) a Jmol applet to visualise and explore the predictions using the protein structures and (ii) a table listing HSPred scores for each interface amino acid. The output page for an illustrative example is reported in Figure 4. The complex tested is Interleukin 4 (IL-4) bound to its receptor α chain (IL-4Rα) (PDB code: 1IAR). Alanine mutational data from experiments are available for this complex [20,21]. Out of 27 interface mutations, HSPred predicts 7 true positives, 14 true negatives, 4 false positive and 2 false negatives. These results further validate the predictive accuracy of HSPred.

To conclude, in this paper we have described HSPred, an accurate and reliable computational method to predict hot spot residues at protein-protein interfaces, given the structure of a complex. HSPred is available as a web server and it is free for non-commercial users. We believe that HSPred predictions will be useful in guiding biomedical experiments. In particular, we are currently testing its capacity to identify druggable binding sites at protein-protein interfaces [22].

Materials and Methods

Data sets

In our study, we have used the same data set as in [12]. It consists of 20 protein complex structures for which alanine mutational data are available. Only protein-protein interactions involving an extended interface are included (we have therefore ignored protein-peptide complexes). Following previous publications [23], we define hot spots as those alanine mutations for which $\Delta\Delta G \geq 2$ kcal/mol ($\Delta\Delta G$ is the change in binding free energy). Only mutations occurring at the complex interface are retained. In total the data set comprises 349 mutations, of which 81 correspond to hot spots. For cross-validation purposes, we have grouped homologous complexes and formed 16 non-homologous clusters. Accordingly, we have implemented a 16-fold cross-validation strategy. A detailed description of the data set, individual mutations and clustering criteria can be found in [12].

In addition, we have applied HSPred to the Ras/RalGDS protein-protein complex (PDB code: 1LFD) for which experimental $\Delta\Delta G$ values are available [19]. From the original reference, we

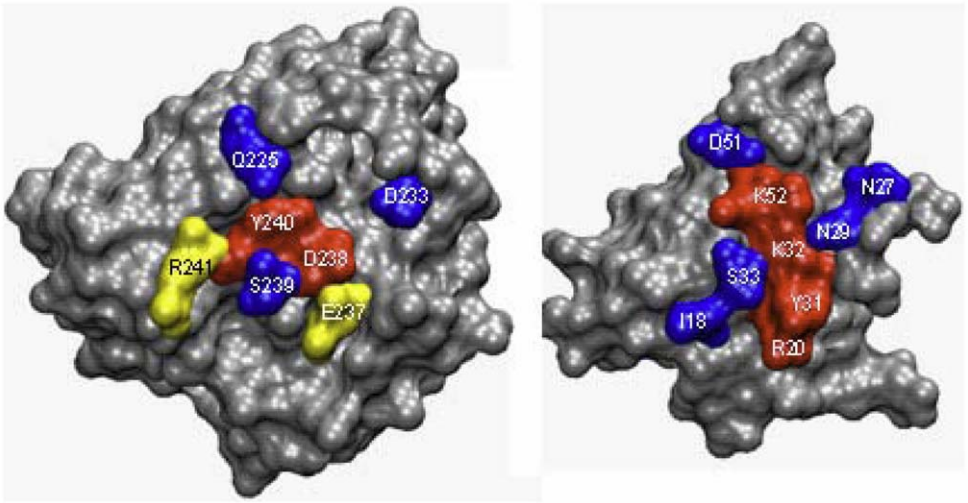


Figure 3. Ras/RalGDS complex. Mapping of HSPred predictions onto the the complex (PDB code: 1LFD). The monomers have been rotated to display the interface. Red residues are correctly predicted hot spots (true positives); blue residues are correctly predicted non hot spots (true negatives); yellow residues are non hot spots erroneously predicts as hot spots (false positives).
doi:10.1371/journal.pone.0016774.g003

have taken the data corresponding to 16 interface alanine mutations. As the Ras/RalGDS complex is not homologous to any of the complexes in the original data set, it can be regarded as an independent external test case. A similar data set had been used previously in [13] for validation purposes.

As a further illustrative example we have applied HSPred to Interleukin-4 (IL-4) bound to its receptor α chain (IL-4Rα) (PDB code: 1IAR). Experimental $\Delta\Delta G$ values are available for this complex too [20,21]. The IL-4/IL-4Rα complex is likely a remote homologue of the complex between human growth hormone (hGH) and its binding protein (hGHbp), which is part of our training set (PDB code: 1A22). IL-4 and hGH share only 8% sequence identity by optimal structural alignment but belong to the same homologous superfamily group (H-level) according to the CATH database [24]. Similarly, the sequence identity between IL-4Rα and hGHbp is only 14% but structural similarity suggests a homology relationship. It has however been pointed out that the IL-4/IL-4Rα complex differs in several important functional and structural aspects from the hGH/hGHbp complex [20,21,25]. It could therefore in effect be regarded as an additional independent test case.

Input features

As input features for the Support Vector Machines we have used basic energy terms that have been found to be important for the stability of protein complexes. These are van der Waals potential, hydrogen bonds, Coulomb electrostatics and desolvation energy. We distinguish contributions from 3 different structural regions (schematised in Figure 1):

- *Side-chain inter-molecular energies*: interaction energies between side-chain atoms of the mutated residue and atoms in the partner protein (respectively atoms in the red filled area and blue striped area in Figure 1).
- *Environment inter-molecular energies*: interaction energies between atoms in the two proteins that are within 10A of the C_β of the mutated residue (respectively atoms in the red striped area and blue striped area in Figure 1). We do not include the contribution from the mutated side-chain in this term.
- *Side-chain intra-molecular energies*: interaction energies between side-chain atoms of the mutated residue and other atoms in the same protein (respectively atoms in the red filled area and red striped area in Figure 1).

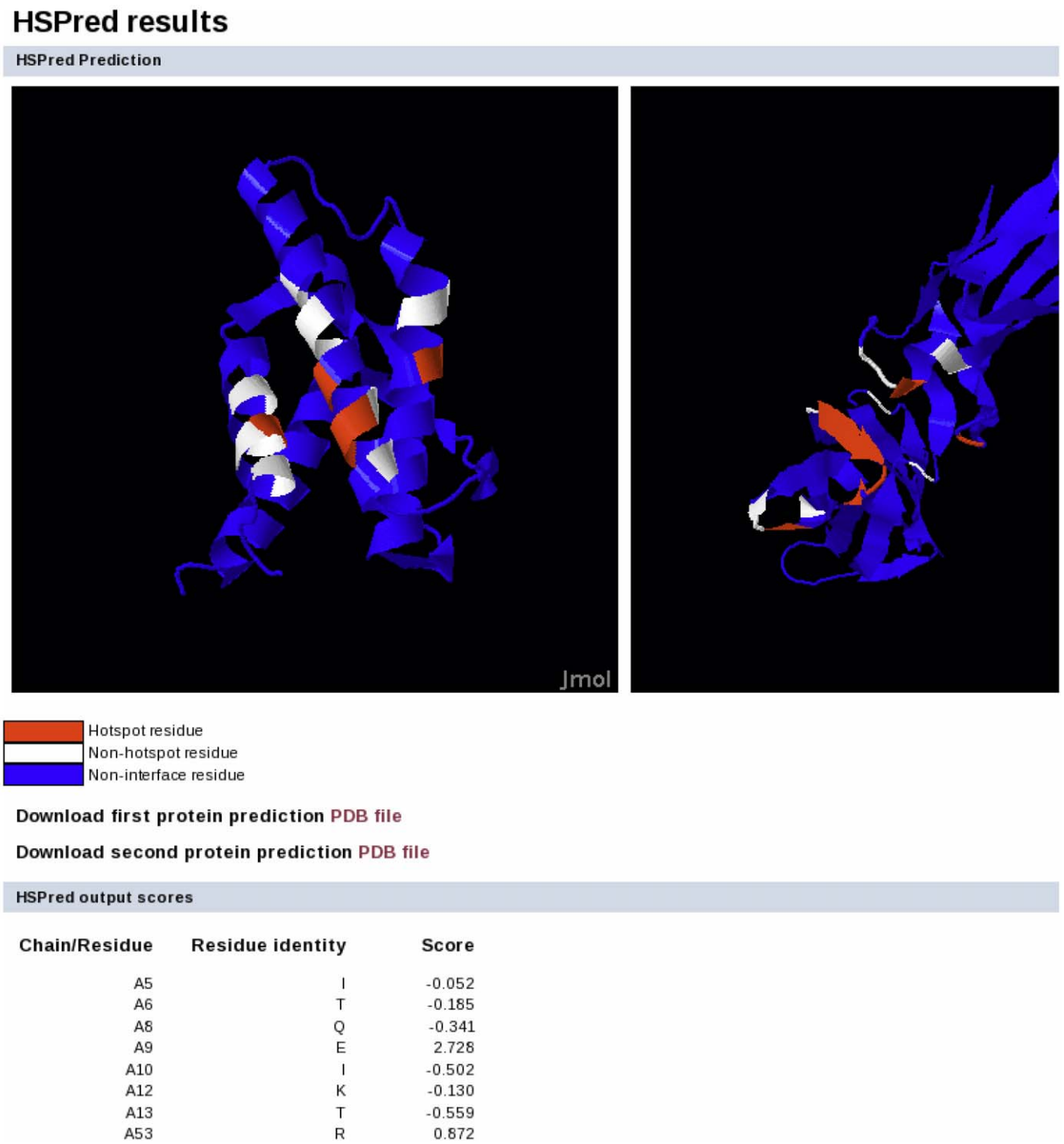
In total therefore there are 12 input features (4×3), although not all of them have been used to build our SVM models (we discuss our feature selection below). A detailed description of how energy components are calculated from the PDB structures is reported in [12].

Support Vector Machines models

We have used the program package SVM^{light} [26], which is available at the website http://svmlight.joachims.org/. As in [12], we have opted for a linear kernel and implemented a nested-loop cross-validation scheme. The latter consists of two nested cross-validation loops: an outer one for testing, an inner one for choosing hyper-parameters. In the inner cycle, the hyper-parameters are optimised by applying a grid search and the model performance is assessed by means of the F1 score. The nested-loop cross-validation scheme allows also to estimate statistical errors on performance measures (see [12] for details).

Models construction and feature selection. We have analysed the correlation coefficients r between energy features and the observed $\Delta\Delta G$ values (see Table 3). We have then built a 'baseline' model, SVM$_X$, including only the 7 features for which $r\geq0.2$. These are: van der Waals, hydrogen bond and solvation side-chain inter-molecular energies, van der Waals, hydrogen bond and solvation environment inter-molecular energies, and van der Waals side-chain intra-molecular energy. Note that the values of the correlation coefficients do not vary sensibly in the 16 different training sets, implying that this choice of features is robust.

We have analysed the predictions of SVM$_X$ by grouping mutations according to the amino acid type. In particular we have focused on the most frequent amino acids in our data set, i.e. those occurring more than 20 times with at least 5 hot spot examples. The list comprises the following 7 amino acid types: Arg, Asn, Asp, Glu, Lys, Trp and Tyr. We observe a good performance for all amino acids except Arg and Glu for which $F1<0.5$ (see Figure 2).



Chain/Residue	Residue identity	Score
A5	I	-0.052
A6	T	-0.185
A8	Q	-0.341
A9	E	2.728
A10	I	-0.502
A12	K	-0.130
A13	T	-0.559
A53	R	0.872

Figure 4. Sample output for the HSPred server. Screenshot of the results page for the IL-4/IL-4Rα complex (PDB code: 1IAR). On top, predictions are visualised using a Jmol applet. On the left is IL-4 (chain A), on the right IL-4Rα (chain B). Predicted hot spots are in red, non hot spots in white. Residues not part of the interface are in blue. Below, predictions scores for each interface residues (excluding Pro and Gly amino acids) are reported (note that only the first few residues are displayed here). Scores greater than zero corresponds to predicted hot spots.
doi:10.1371/journal.pone.0016774.g004

To overcome these limitations, we have built two separate SVM classifiers, SVM_R and SVM_E, for mutations involving respectively Arg and Glu.

In theory, one could use the amino acid identity as input feature or build a model using only, e.g., Glu mutations. In practice, at present this is not feasible as there are not enough mutational data. We have reasoned instead that SVM_E and SVM_R should not be completely different from SVM_X, rather they should differ only marginally from the latter. In this spirit, we have trained several different but related models. All models are trained using the whole data set (comprising therefore mutations from all amino acid types) but each of them corresponds to a different choice of input features. Within this ensemble of classifiers we have selected those that best perform on Arg and Glu.

Table 3. Correlation of energy terms with observed $\Delta\Delta G$ values.

Feature (energy term)	r
Side-chain inter-molecular	
van der Waals	**0.49**
hydrogen bond	**0.38**
electrostatics	0.01
desolvation	**0.45**
Environment inter-molecular	
van der Waals	**0.32**
hydrogen bond	**0.28**
electrostatics	0.04
desolvation	**0.32**
Side-chain intra-molecular	
van der Waals	**0.26**
hydrogen bond	0.09
electrostatics	0.12
desolvation	0.19

We report the absolute values of the correlation coefficients r between energy features and the observed $\Delta\Delta G$ (values greater than 0.2 are in bold).
doi:10.1371/journal.pone.0016774.t003

Our strategy has been to bias the classifiers to perform well on Arg and Glu by selecting a specific subset of features. This reflects the observation that some energy features appear to be more important for some amino acids than for others, i.e. for some amino acid they correlate better with the observed $\Delta\Delta G$s. Note that the hyper-parameters in each of the models in the ensemble are optimised over all the mutations in the training set. The identity of the amino acid of interest enters only when selecting the best model within the ensemble. We find this to be a robust strategy, i.e. it is not too sensitive to small modifications in the training set.

Given the starting 12 features, there is a huge number of possible combinations that can be selected and it is clearly not feasible to test them all. To simplify the problem, we have considered only combinations with 3 or 4 features, taken from the 7 features used for SVM_X. We have further constrained the selection by excluding pairs of highly correlated features, i.e. features for which $r>0.6$, because they would be redundant. For example, only one term between the van der Waals and solvation side-chain inter-molecular energies can be included. Similarly only one term among the 3 environment energies can be chosen. With these constraints, there are a total of 23 different feature combinations (6 combinations having 4 features and 17 having 3 features). We have built a classifier for each of them and then selected the one performing best on, e.g., Glu. In the case of Arg, the intra-molecular coulomb term appears to be also important (correlation coefficient with observed $\Delta\Delta G$ $r=0.4$). We have therefore tested additional 23 combinations which are obtained by adding the intra-molecular coulomb term to the set above.

It is important to underline that when assessing the results for SVM_E and SVM_R by cross-validation, the choice of the best model (feature combination) is performed within the inner loop of the nested-loop cross-validation scheme (i.e. using the training set only), similarly to the choice of hyper-parameters. This ensures that the optimal feature combination for either Arg or Glu is selected without ever considering the performance on the test set. It is worth noting that for both Arg and Glu the feature combination that gives the best results is consistent in the 16 different training sets. For example for Glu the optimal feature combination is always hydrogen bond side-chain inter-molecular, hydrogen bond environment and van der Waals side-chain intra-molecular. It is also worth mentioning that Glu and Arg can be singled out based on the performance of SVM_X in the training sets, therefore complying to the cross-validation scheme. We have not explicitly stated it above to keep the discussion as simple as possible.

Measures of prediction performance

We primarily assess the prediction performances of our method using the F1 score. Let TP, FP, FN refer to the number of true positives, false positives and false negative respectively. Precision (P, also called specificity) and recall (R, also called sensitivity) are defined as

$$P=\frac{TP}{TP+FP} \quad R=\frac{TP}{TP+FN} \tag{1}$$

The F1 score is the harmonic mean of precision and recall

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

We also calculate the Matthew's correlation coefficient (MCC) given by

$$MCC=\frac{TP\times TN-FP\times FN}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}} \tag{3}$$

where TN is the number of true negative and TP, FP and FN are as above.

Acknowledgments

SL, DB, MP and DTJ acknowledge financial support from the BBSRC.

Author Contributions

Conceived and designed the experiments: SL MP DTJ. Performed the experiments: SL. Analyzed the data: SL. Contributed reagents/materials/analysis tools: DB. Wrote the paper: SL DB DTJ.

References

1. Cunningham BC, Wells JA (1989) High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. Science 244: 1081–1085.
2. Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots–a review of the protein-protein interface determinant amino-acid residues. Proteins 68: 803–812.
3. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A 99: 14116–14121.
4. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320: 369–387.
5. Gao Y, Wang R, Lai L (2004) Structure-based method for analyzing protein-protein interfaces. J Mol Model 10: 44–54.
6. Li L, Zhao B, Cui Z, Gan J, Sakharkar MK, et al. (2006) Identification of hot spot residues at protein-protein interface. Bioinformation 1: 121–126.
7. Ofran Y, Rost B (2007) Protein-protein interaction hotspots carved into sequences. PLoS Comput Biol 3: e119–e119.
8. Darnell SJ, Page D, Mitchell JC (2007) An automated decision-tree approach to predicting protein interaction hot spots. Proteins 68: 813–823.
9. Grosdidier S, Fernandez-Recio J (2008) Identification of hot-spot residues in protein-protein interactions by computational docking. BMC Bioinformatics 9: 447–447.
10. Benedix A, Becker CM, de Groot BL, Caflisch A, Böckmann RA (2009) Predicting free energy changes using structural ensembles. Nat Methods 6: 3–4.

11. Cho Ki, Kim D, Lee D (2009) A feature-based approach to modeling protein-protein interaction hot spots. Nucleic Acids Res 37: 2672–2687.
12. Lise S, Archambeau C, Pontil M, Jones DT (2009) Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. BMC Bioinformatics 10: 365–365.
13. Krüger DM, Gohlke H (2010) DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. Nucleic Acids Res 38 Suppl: W480–W486.
14. Tuncbag N, Keskin O, Gursoy A (2010) HotPoint: hot spot prediction server for protein interfaces. Nucleic Acids Res 38 Suppl: W402–W406.
15. Meireles LMC, Dömling AS, Camacho CJ (2010) ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. Nucleic Acids Res 38 Suppl: W407–W411.
16. Xia JF, Zhao XM, Song J, Huang DS (2010) APIS: accurate prediction of hot spots in protein interfaces by combining protrusion index with solvent accessibility. BMC Bioinformatics 11: 174–174.
17. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450: 1001–1009.
18. González-Ruiz D, Gohlke H (2006) Targeting protein-protein interactions with small molecules: Challenges and perspectives for computational binding epitope detection and ligand finding. Curr Med Chem 13: 2607–2625.
19. Kiel C, Serrano L, Herrmann C (2004) A detailed thermodynamic analysis of ras/effector complex interfaces. J Mol Biol 340: 1039–1058.
20. Wang Y, Shen BJ, Sebald W (1997) A mixed-charge pair in human interleukin 4 dominates high-affinity interaction with the receptor alpha chain. Proc Natl Acad Sci U S A 94: 1657–1662.
21. Zhang JL, Simeonowa I, Wang Y, Sebald W (2002) The high-affinity interaction of human IL-4 and the receptor alpha chain is constituted by two independent binding clusters. J Mol Biol 315: 399–407.
22. Lise S, Jones DT. Predicting druggable binding sites at protein-protein interfaces by computational alanine scanning. In preparation.
23. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280: 1–9.
24. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, et al. (2009) The CATH classification revisited–architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic Acids Res 37: D310–D314.
25. Hage T, Sebald W, Reinemer P (1999) Crystal structure of the interleukin-4/receptor alpha chain complex reveals a mosaic binding interface. Cell 97: 271–281.
26. Joachims T (1999) Making large-scale svm learning practical. In: Schölkopf B, Burges C, Smola AJ, eds. Advances in Kernel Methods - Support Vector Learning, MIT Press.