

HotPoint: hot spot prediction server for protein interfaces

Nurcan Tuncbag, Ozlem Keskin* and Attila Gursoy*

Center for Computational Biology and Bioinformatics, and College of Engineering, Koc University, Rumelifeneri Yolu, 34450 Sariyer Istanbul, Turkey

Received February 12, 2010; Revised April 13, 2010; Accepted April 17, 2010

ABSTRACT

The energy distribution along the protein–protein interface is not homogenous; certain residues contribute more to the binding free energy, called ‘hot spots’. Here, we present a web server, HotPoint, which predicts hot spots in protein interfaces using an empirical model. The empirical model incorporates a few simple rules consisting of occlusion from solvent and total knowledge-based pair potentials of residues. The prediction model is computationally efficient and achieves high accuracy of 70%. The input to the HotPoint server is a protein complex and two chain identifiers that form an interface. The server provides the hot spot prediction results, a table of residue properties and an interactive 3D visualization of the complex with hot spots highlighted. Results are also downloadable as text files. This web server can be used for analysis of any protein–protein interface which can be utilized by researchers working on binding sites characterization and rational design of small molecules for protein interactions. HotPoint is accessible at <http://prism.ccbb.ku.edu.tr/hotpoint>.

INTRODUCTION

Most molecular and cellular processes are controlled by protein–protein interactions. Proteins interact through interfaces. The energy distribution along the interface region is not homogenous; certain residues contribute more to the binding free energy, called ‘hot spots’ (1–3). Hot spots form tightly packed regions in protein interfaces (4). Presence of hot spots is important as a target to disrupt malfunctioning association of proteins by therapeutic molecules and for rational design of highly specific protein complexes (5,6). Experimentally, a hot spot can be found by evaluating the change in binding free energy

upon mutating it to an alanine. For a limited number of protein complexes, alanine mutations are available and this information is deposited in databases (7,8). Highly efficient computational methods emerged to identify hot spots under the occurrence of limitations in experimental information. Although there is not a strict rule to identify hot spots, combination of several physical and chemical features of residues gives successful results. Several groups developed energy based methods (9–12), learning based methods (13–19) and molecular dynamics based methods (20–22) to predict hot spot residues computationally. Some of these methods work as servers such as Robetta (10,11) and KFC server (23). Robetta server (10,11) performs computational alanine scanning based on estimating energy (including van der Waals, H-bonds) at atomic level for a given complex and outputs changes in the binding free energy values for each residue in the interface. KFC server (23) predicts hot spots for a given complex using a machine learning approach which considers the shape specificity and surrounding structural features of the residues. Server output is the confidence scores and the predictions. Results can be visualized by an interactive viewer.

Here, we present HotPoint web server, which provides a user-friendly interface to run the method developed by Tuncbag *et al.* (19) for online prediction of hot spots in protein interfaces. Our aim is to provide an efficient server at a single location for analysis of any protein–protein interface which can be utilized by researchers interested in protein binding sites. The method principally considers the solvent accessibility and the total contact potential of the interface residues. The output tabulates the interface residues with the highlighted hot spots and their features. Additionally, it provides an interactive 3D visualization of the submitted protein–protein interface with the predicted hot spots for observing their localization. Distinct features of HotPoint from existing servers (Robetta and KFC server) are the improved efficiency and accuracy. The calculation of solvent accessibility and pair potentials of

*To whom correspondence should be addressed. Tel: +90 212 338 1720; Fax: +90 212 338 1548; Email: agursoy@ku.edu.tr
Correspondence may also be addressed to Ozlem Keskin. Tel: +90 212 338 1538; Fax: +90 212 338 1548; Email: okeskin@ku.edu.tr

residues are faster than atomic level computations performed by Robetta, and the prediction accuracy is higher than both Robetta and KFC server.

THE HOTPOINT METHOD

HotPoint is based on a few simple rules consisting of solvent accessibility and energetic contribution of residues. The thresholds of the model are adjusted according to a data set composed of 150 experimentally alanine mutated residues of which 58 residues are hot spots and 92 residues are non-hot spots. The interface residues, whose mutations change the binding free energy at least 2.0 kcal/mol, are considered as experimental hot spots. If the mutation results in a change <0.4 kcal/mol, that residue is labeled as experimental non-hot spot. The independent test set is derived from Binding Interface Database (BID) (7), composed of 112 residues (of which 54 residues are hot spots and 58 residues are non-hot spots). The predictive performance of this method is assessed using accuracy (the ratio of number of correctly predicted residues to number of all predicted residues), precision (the ratio of number of correctly classified hot spot residues to the number of all residues classified as hot spots), recall (the proportion of number of correctly classified hot spot residues to the number of all hot spot residues), specificity (the proportion of number of correctly predicted non-hot spot residues to the number of all non-hot spot residues) and F1 score (the balance between precision and recall).

Several empirical and machine learning methods are trained and tested using several features [relative accessible surface area (ASA) in complex state, relative change in ASA upon complexation, conservation, amino acid propensity and total contact potential]. At the end, the best performance is achieved by an empirical model based on relative accessibility in complex state and total pair potentials. According to this model, if an individual interface residue is buried (its relative ASA in complex state is $\leq 20\%$) and its total contact potential is ≥ 18.0 , this residue is flagged as a hot spot; else, it is flagged as a non-hot spot. The thresholds of the model (20% and 18.0) are inferred from training set. This model demonstrates an accuracy of 0.70, a precision of 0.73, a recall of 0.59 and a specificity of 0.79 on the independent test set, which exceeds the performance of existing approaches [such as, KFC (15), KFCA (15), ISIS (18), Robetta (11)] and machine learning approaches (such as, SVM, BayesNet, decision tree, etc.) on the same test set. The details of the data sets, methodology and an exhaustive comparison with other approaches are available in Tuncbag *et al.* (19).

HOTPOINT WEB SERVER

The HotPoint web server is available at <http://prism.cccb.ku.edu.tr/hotpoint>. Server interface is coded in PHP. The code to predict hot spots is written in Python.

Input

Input data is the protein structure in PDB formatted coordinate file, two chain identifiers forming the interface and the interface definition. User can either run the server with default distance thresholds to extract interface residues or can change the interface definition by submitting a distance threshold. There are two options to submit a structure file. User can enter the four letter PDB code of a protein which is directly downloaded from the ftp site of PDB. The second option is uploading a structure file that is in the PDB format. HotPoint requires two chain identifiers which confine to a protein interface. Server does not work for PDB files containing only one chain and returns an error. For NMR structures, it uses the first model in the prediction and gives results for the first model. HotPoint is specific to protein-protein interfaces; chains corresponding to DNA structures return a warning in the web server.

When there is not enough input data, the server informs the users of what is missing. The HotPoint web server is free and open to all users and there are no login requirements.

Extraction of computational hot spots

When a protein structure with its chain identifiers is submitted, HotPoint server starts the calculation of three consecutive steps:

- (1) Extraction of interface residues: a protein interface is defined as a set of amino acids which represents a region that links two protein chains by non-covalent interactions. According to the default interface definition in the server, if the distance between any two atoms belonging to two residues, one from each chain, is less than the sum of their van der Waals radii plus a 0.5 Å tolerance, these two residues are defined as interacting. Users can change this definition by submitting a distance threshold.
- (2) Calculation of the features: Residue solvent accessibilities are calculated using Naccess (24). The residue accessibilities in complex state and in monomer state are converted into relative accessibilities by dividing them to maximum accessibility of that residue. Knowledge-based solvent mediated inter-residue potentials are taken from Keskin *et al.* (25). In the contact potential matrix, there are 210 distinct contact potentials between all possible pairs of 20 amino acids in RT unit (R , universal gas constant; T , absolute temperature). To calculate the total contact potential of a residue in the interface, server extracts the neighbors of that residue whose side chain center of mass are closer than the cutoff (7.0 Å). Another constraint for neighbor extraction is that they should not be close neighbor in sequence ($|i-j| \geq 4$ where i and j are residue numbers). Total contact potential of the residue is defined as the absolute of the sum of the contact potentials with its neighbors (19).
- (3) Prediction based on empirical model: Finally, the empirical model [presented in Tuncbag *et al.* (19)]

is applied on the residue to determine whether it is a computational hot spot or not. If the relative accessibility of an individual interface residue is $\leq 20\%$ and its total contact potential is ≥ 18.0 , it is labeled as hot spot (19).

Output

During the processing, the server informs users about the steps it is performing. The output of the server is a table consisting of the interface residues with their features (Figure 1). The interface residues are tabulated with chain names, one-letter residue names, residue numbers, their relative ASA in complex, relative ASA in monomer and total pair potentials. In the last column of the table, the prediction is presented as H (hot spot) or NH (non-hot spot). Background of the predicted hot spots is highlighted with red color. The prediction results as a text file and interface residue coordinates in PDB file format are also downloadable by the user. In this way, the results can be visualized in any visualization tool. Besides the

downloadable files, overall complex, the interface residues and hot spots can be visualized interactively using the Jmol (26) applet window in the HotPoint server.

An independent case study: Interleukin-2 and its receptor complex

Interleukin-2 (IL-2) is a cytokine immune system signaling molecule. IL-2 gets functional when it associates with the IL-2 receptor. To find the residues necessary for binding, several residues (K35, R38, M39, T41, F42, K43, F44, Y45, E62, P65, V69 and L72) on IL-2 are mutated to alanine. Among these residues, F42, Y45 and E62 reduce binding affinity of IL-2 to its receptor >100 -folds. Further, small inhibitor molecule SP4206 also targets these hot spots of the receptor (27). HotPoint predicts all three experimental hot spots (F42, Y45 and E62) correctly for the IL-2/IL-2 receptor complex (PDB code: 1z92, chain A is IL-2 and chain B is IL-2 receptor). According to our interface definition, M39 cannot be found in the interface residues. So, for the

Online calculation in progress...

It may take up to 3 minutes to complete prediction process. Please be patient...

Retrieving PDB file 1ycs from the RCSB Protein Data Bank...
Interface Name: 1ycsAB
Running NACCESS...
Calculating Pair Potentials and Predicting Hot Residues...

Calculation finished!

Interface file in PDB format: [1ycsAB.pdb](#)
Hot Spot Prediction result file: [1ycsAB.output](#)

Legend for Jmol visualization:

Interface residues on chain A Interface residues on chain B
Predicted Hot spots in interface (Chain A) Predicted Hot spots in interface (Chain B)

Click for an interactive 3D model (Opens in new window): [Jmol Visualization](#)

Prediction Results						
Residue Number	Residue Name	Chain	RelCompASA	RelMonomerASA	Potential	Prediction
178	H	A	26.41	81.22	12.72	NH
179	H	A	11.06	21.00	28.12	H
181	R	A	51.37	75.59	6.75	NH
182	C	A	15.42	44.93	9.48	NH
183	S	A	80.50	104.88	4.48	NH
241	S	A	13.51	54.58	8.00	NH
243	M	A	27.77	84.33	25.95	NH
247	N	A	3.22	46.15	17.93	NH
248	R	A	13.33	79.39	27.47	H
249	R	A	1.36	15.50	21.12	H
280	R	A	34.54	56.21	8.46	NH
422	M	B	33.17	45.71	13.41	NH
423	T	B	0.00	0.24	15.51	NH
424	Y	B	21.54	58.82	19.03	NH
425	S	B	27.63	74.71	6.39	NH
427	M	B	21.15	63.93	24.97	NH
469	Y	B	8.29	13.72	30.49	H
472	Q	B	37.96	82.31	18.23	NH

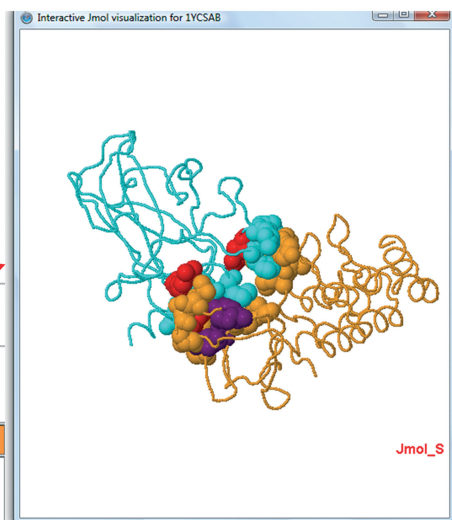


Figure 1. The output page of HotPoint for the p53 DNA binding domain/53BP2 protein complex (pdb:1ycs, chain A and B). Interface residues of this complex are shown in the table with hot spot predictions: (1) the coordinates of interface residues can be downloaded; (2) hot spot prediction results are also downloadable; and (3) interface with predicted hot spots can be visualized by Jmol.

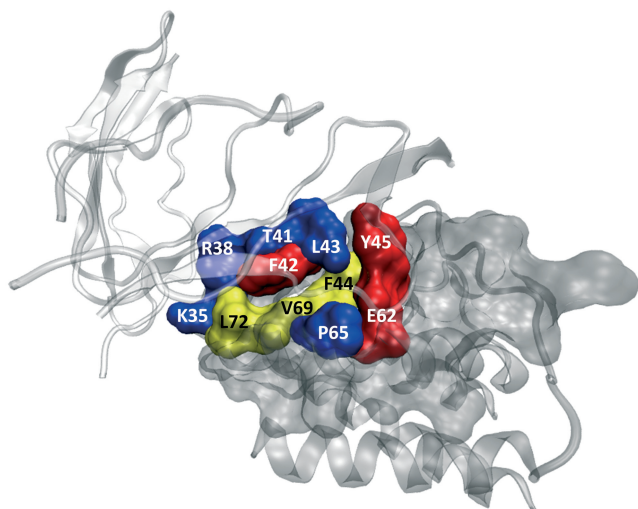


Figure 2. IL-2/IL-2 receptor complex. The PDB code for this complex is 1z92. The red-colored residues are correctly predicted hot spots. The blue-colored ones are correctly predicted non-hot spots. The yellow-colored residues represent non-hot spot residues that are incorrectly predicted as hot spots.

remaining eight residues, HotPoint labels five residues (K35, R38, T41, K43 and P65) as non-hot spot, correctly. However, three residues come as false positives (F44, V69 and L72) from HotPoint prediction. As a result, 8 out of 11 alanine mutations are correctly predicted. This protein complex is independent from the training and test sets. The predictions are illustrated in Figure 2 in 3D using the output files obtained from HotPoint.

CONCLUSIONS

A small subset of residues in protein interfaces comprises a large portion of binding free energy, namely hot spots. We present HotPoint server to determine computational hot spots in protein interfaces based on solvent accessibility and pair potentials which allows online calculation for all protein interfaces within practical running times. Further, the model outperforms other existing approaches. It tabulates residue level features and prediction results for a given protein complex which are also downloadable. We hope that with its simple architecture and visualization tool, HotPoint would be useful both for the experimentalists and computational scientist working on protein recognition, modeling of protein complexes and drug design.

ACKNOWLEDGEMENTS

We thank to Abdullah M. Turan for his help in developing the web service.

FUNDING

TUBITAK (Research Grant No 109T343 and 109E207). TUBITAK fellowship (to N.T.). Funding for open access charge: Koc University.

Conflict of interest statement. None declared.

REFERENCES

- Bogan, A.A. and Thorn, K.S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, **280**, 1–9.
- Clackson, T. and Wells, J.A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383–386.
- Wells, J.A. (1991) Systematic mutational analyses of protein-protein interfaces. *Meth. Enzymol.*, **202**, 390–411.
- Keskin, O., Ma, B. and Nussinov, R. (2005) Hot regions in protein-protein interactions: the organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol.*, **345**, 1281–1294.
- Keskin, O., Gursoy, A., Ma, B. and Nussinov, R. (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.*, **108**, 1225–1244.
- Keskin, O., Tuncbag, N. and Gursoy, A. (2008) Characterization and prediction of protein interfaces to infer protein-protein interaction networks. *Curr. Pharm. Biotechnol.*, **9**, 67–76.
- Fischer, T.B., Arunachalam, K.V., Bailey, D., Mangual, V., Bakhr, S., Russo, R., Huang, D., Paczkowski, M., Lalchandani, V., Ramachandra, C. *et al.* (2003) The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. *Bioinformatics*, **19**, 1453–1454.
- Thorn, K.S. and Bogan, A.A. (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics*, **17**, 284–285.
- Guerois, R., Nielsen, J.E. and 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.
- Kortemme, T. and Baker, D. (2002) A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl Acad. Sci. USA*, **99**, 14116–14121.
- Kortemme, T., Kim, D.E. and Baker, D. (2004) Computational alanine scanning of protein-protein interfaces. *Sci STKE*, **2004**, pl2.
- Guharoy, M. and Chakrabarti, P. (2009) Empirical estimation of the energetic contribution of individual interface residues in structures of protein-protein complexes. *J. Comput. Aided Mol. Des.*, **23**, 645–654.
- Assi, S.A., Tanaka, T., Rabbitts, T.H. and Fernandez-Fuentes, N. (2009) PCRPI: Presaging Critical Residues in Protein interfaces, a new computational tool to chart hot spots in protein interfaces. *Nucleic Acids Res.*, **38**, e86.
- Cho, K.I., Kim, D. and Lee, D. (2009) A feature-based approach to modeling protein-protein interaction hot spots. *Nucleic Acids Res.*, **37**, 2672–2687.
- Darnell, S.J., Page, D. and Mitchell, J.C. (2007) An automated decision-tree approach to predicting protein interaction hot spots. *Proteins*, **68**, 813–823.
- Guney, E., Tuncbag, N., Keskin, O. and Gursoy, A. (2008) HotSprint: database of computational hot spots in protein interfaces. *Nucleic Acids Res.*, **36**, D662–D666.
- Lise, S., Archambeau, C., Pontil, M. and Jones, D.T. (2009) Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. *BMC Bioinformatics*, **10**, 365.
- Ofran, Y. and Rost, B. (2007) Protein-protein interaction hotspots carved into sequences. *PLoS Comput. Biol.*, **3**, e119.
- Tuncbag, N., Gursoy, A. and Keskin, O. (2009) Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics*, **25**, 1513–1520.
- Gonzalez-Ruiz, D. and 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.
- Huo, S., Massova, I. and Kollman, P.A. (2002) Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.*, **23**, 15–27.
- Rajamani, D., Thiel, S., Vajda, S. and Camacho, C.J. (2004) Anchor residues in protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **101**, 11287–11292.

23. Darnell,S.J., LeGault,L. and Mitchell,J.C. (2008) KFC Server: interactive forecasting of protein interaction hot spots. *Nucleic Acids Res.*, **36**, W265–W269.
24. Hubbard,S.J. and Thornton,J.M. (1993) *Department of Biochemistry and Molecular Biology*. London, University College.
25. Keskin,O., Bahar,I., Badretdinov,A.Y., Ptitsyn,O.B. and Jernigan,R.L. (1998) Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. *Protein Sci.*, **7**, 2578–2586.
26. Herraiez,A. (2006) Biomolecules in the computer: Jmol to the rescue. *Biochem Mol Biol Educ*, **34**, 255–261.
27. Thanos,C.D., DeLano,W.L. and Wells,J.A. (2006) Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc. Natl Acad. Sci. USA*, **103**, 15422–15427.