

# Patch Finder Plus (PFplus): A web server for extracting and displaying positive electrostatic patches on protein surfaces

Shula Shazman<sup>1</sup>, Gershon Celniker<sup>1</sup>, Omer Haber<sup>1</sup>, Fabian Glaser<sup>2</sup> and Yael Mandel-Gutfreund<sup>1,\*</sup>

<sup>1</sup>Department of Biology and <sup>2</sup>The Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering, Technion - Israel Institute of Technology Haifa 32000, Israel

Received February 15, 2007; Revised April 30, 2007; Accepted May 1, 2007

## ABSTRACT

**Positively charged electrostatic patches on protein surfaces are usually indicative of nucleic acid binding interfaces. Interestingly, many proteins which are not involved in nucleic acid binding possess large positive patches on their surface as well. In some cases, the positive patches on the protein are related to other functional properties of the protein family. PatchFinderPlus (PFplus) <http://pfp.technion.ac.il> is a web-based tool for extracting and displaying continuous electrostatic positive patches on protein surfaces. The input required for PFplus is either a four letter PDB code or a protein coordinate file in PDB format, provided by the user. PFplus computes the continuum electrostatics potential and extracts the largest positive patch for each protein chain in the PDB file. The server provides an output file in PDB format including a list of the patch residues. In addition, the largest positive patch is displayed on the server by a graphical viewer (Jmol), using a simple color coding.**

## INTRODUCTION

The surface of a protein is the region where the protein interacts with other molecules such as other proteins, nucleic acids, membrane receptors and small ligands. The electrostatic potential is a fundamental property of the protein surface, playing a central role in recognition of other macromolecules (1). When the first 3D structures of protein–DNA complexes were solved it was noticed that charges are distributed asymmetrically on the protein surface, creating a patch of positive charges which complements the negative charge of the DNA (2). It was further suggested that charge complementarity is one of the first steps of recognition between proteins and DNA

(3,4). Indeed, large patches of positive charges have been suggested to be characteristic of protein–nucleic acid interfaces (5–8). Recently, several methods have been developed for automatic prediction of DNA-binding proteins based on the existence of large positive patches on the protein surface (6,7,8–11). In addition to nucleic acid binding, other essential protein functions could be dependent on the presence of large patches of positive charges on the protein surface (12). Among these are proteins which bind negatively charged membranes and receptor-binding ligands. Furthermore, although protein–protein interactions are usually known to be stabilized by a net neutral charge, different studies have revealed that positive and negative patches are commonly involved in protein–protein interfaces (13,14).

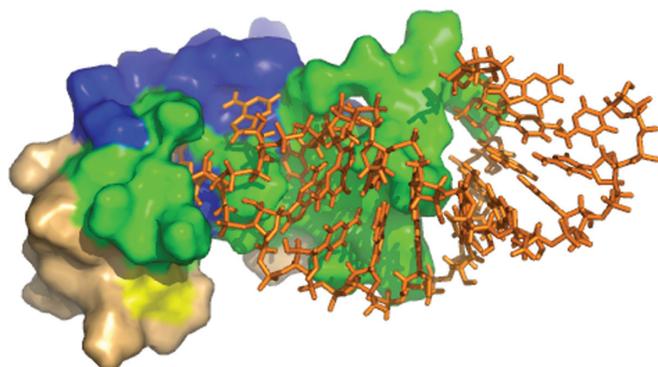
In general, positively charged surfaces can be detected by visualizing the electrostatic properties of the protein surface with graphical programs such as GRASP or GRASS (15,16). The first program for graphical representation and analysis of surface properties of macromolecules was the GRASP software (Graphical Representation and Analysis of Surface Properties), developed by Nicholls *et al.* (17). Among the features which are calculated and displayed by GRASP are the electrostatic potential and surface accessibility. The electrostatic potential displayed by GRASP is calculated using the finite-difference Poisson–Boltzmann equation (FDPB). A newer version of the GRASP program, GRASP2, was published more recently (16). In addition, the GRASS web server was developed (15) to exploit many of the features calculated by the previous programs on a simple interface over the World Wide Web.

These programs, however, are not designed to capture isolated patches on the protein surface. In order to specifically detect continuous regions on the protein surface that could be indicative of the protein function, we have previously developed an in-house program named PatchFinder (6), similar approaches have been developed later by several groups (7,8,18). PatchFinder was designed

\*To whom correspondence should be addressed. Tel: 972 4 8293958; Fax: 972 4 8225153; Email: yaelmg@tx.technion.ac.il

to calculate the Poisson–Boltzmann electrostatic potential of the protein and to construct the largest continuous positive patch on the protein surface. Earlier, we have shown that there is a high overlap between the largest electrostatic positive patches on protein surfaces and the DNA-binding interfaces (6). Figure 1 and Supplementary Figure S1 demonstrate the overlap (colored in green) between the largest positive patch, calculated with the PatchFinder algorithm (blue) and the real nucleic acid-binding interface (yellow) extracted from six selected co-crystal structures of DNA and RNA-binding proteins. Furthermore, the percent overlap between the patch and the interface for a randomly selected set of DNA–protein complexes is given in Supplementary Table S1 (the average percent-overlap for the random set was 75%). As shown in the figures, though the percent overlap varies between the different structures in all cases the calculated patch coincides with the binding site of the nucleic acids.

Recently, Ahmad and Sarai (18) have developed the Qgrid web server which identifies charge and hydrophobic clusters in proteins (<http://www.netasa.org/qgrid/index.html>). The Qgrid program calculates the distribution of charge and hydrophobic regions throughout the protein and applies a hierarchical clustering algorithm for clustering the atoms based on their charge. The output of Qgrid is a tree diagram of all grid points from which the user can interpret the different charge and hydrophobic clusters within the whole protein and the relationship between the different clusters. Here we describe a new web server, PatchFinderPlus (PFplus), for extracting electrostatic patches on the protein surface. Different than the Qgrid algorithm, PFplus is designed to map only the largest continuous positive patch on the protein surface. Furthermore, the PFplus algorithm searches for adjacent grid points above a given cutoff, and thus does not require any heuristic calculations, such as clustering. Our server provides a graphical output of the surface patch which presumably corresponds to the region on the protein surface involved in interaction with other molecules.



**Figure 1.** Overlap between the largest positive patch of the HIV-1 nucleocapsid protein (calculated with the PatchFinder algorithm) and the experimentally defined RNA-interface (1a1t). The overlap is shown in green, while the calculated patch that did not overlap with the real interface is colored in blue. Yellow represents the interface which was missed by the patch calculation. Overall, a high degree of overlap is observed between the largest positive patch calculated with the PFplus server and the actual RNA-binding interface calculated as described in Stawiski *et al.* (6).

## METHODS

The PFplus algorithm automatically assigns surface-positive patches by looking for adjacent points on the protein surface that meet a given electrostatic potential cutoff. The algorithm is built of five major steps:

- (i) Calculating the electrostatic potential of the protein on a 3D grid, using the Poisson–Boltzmann equation.
- (ii) Defining the grid points that fall closest to the protein surface while emitting all non-surface points.
- (iii) Extracting all 3D patches of adjacent grid points which meet the defined cutoff.
- (iv) Choosing the largest positive patch for each protein chain.
- (v) Assigning the protein residues related to the patch.

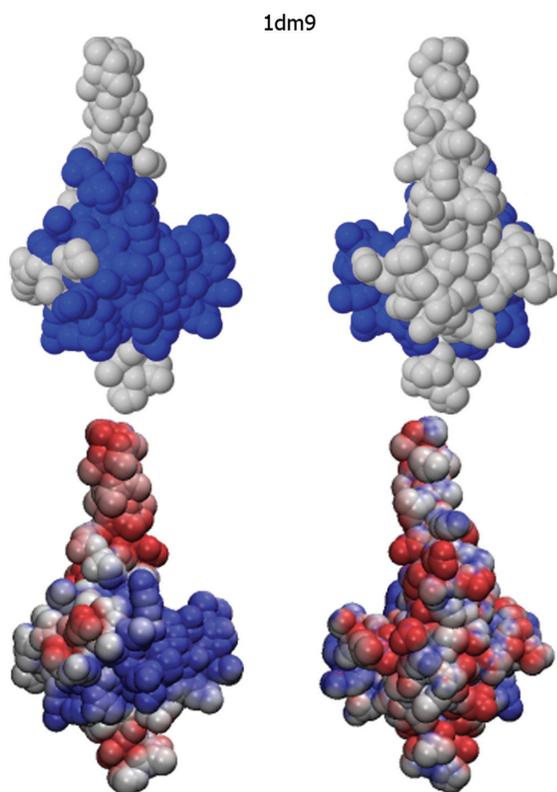
In the first step, the electrostatic potential is computed based on the Poisson–Boltzmann equation using the University of Houston Brownian Dynamics (UHBD) software package (19). In order to calculate the electrostatic potential, hydrogen atoms are added to the structures with the program HBPLUS (20). The input parameters for the UHBD program are given in Table 1. After applying the UHBD software, each grid point associated with the protein surface is assigned an electrostatic potential value. In the current version of PFplus, the input parameters used for calculating the electrostatics potential are kept fixed.

In the second step, the PatchFinder algorithm searches for continuous electrostatic patches on the protein surface. For this purpose, we first define all protein surface points using the open source DMS program, downloaded from the computer graphics laboratory at UCSF, <http://www.cgl.ucsf.edu/Resources/index.html>. The DMS program applies Richards' model (21) for obtaining the surface accessibility by rolling a ball of radius  $r$  along the van der Waals surface of the molecule. In the next step, we select the grid points, calculated by the UHBD program, which fall closest to the protein surface. At this stage all non-surface points on the grid are ignored. PFplus then searches for adjacent surface points on the 3D grid which meet a potential cutoff of  $2kT/e$ . In order to select the largest positive patch, the PFplus program calculates the size of each continuous surface patch, based on the number of grid points included within the patch. Subsequently, the patches are sorted by size. In the current version of the program only the largest positive

**Table 1.** Parameters used for electrostatics calculations

Parameter	Default values
Parameter set	OPLS
Probe radii	1.4 Å
Stern layer	2.0 Å
Temperature	298 K
Ionic strength	150 mM
Dielectric constant protein/solvent	2.0/78.0
Grid dimension	65 × 65 × 65
Grid spacing	1.5 Å

patch is provided to the user. In cases in which there is no surface grid point that meet the electrostatic potential cutoff ( $2kT/e$ ) no patches will be detected. Finally, the protein residues related to the grid points are selected and displayed. It is important to note that a protein residue is included within the patch if at least one of its atoms falls within the continuous patch of grid points. In Figure 2, the largest positive patch extracted from the PFplus server for a novel RNA-binding domain of the Hsp15 protein (PDB code 1dm9) is shown from two view points (top left and right) in comparison to the original electrostatic potentials calculated with the UHBD program (with the same input parameters as used in PFplus) and visualized with the VMD software <http://www.ks.uiuc.edu/Research/vmd/> (bottom left and right). The figure clearly shows that the PFplus algorithm captures a continuous positive region on the molecular surface. Supplementary Figure S2 provides four other examples comparing the positive patch obtained from the PFplus server to the electrostatic potential mapped onto the same molecule surface. The examples shown in Supplementary Figure S2 include two nucleic acid-binding proteins (The TATA-binding protein and the Nuclear



**Figure 2.** A comparison between the largest positive patch on the surface of the hsp15 protein (1dm9) calculated with the PFplus server (top left and right) using the  $2kT/e$  cutoff and the Poisson–Boltzmann electrostatic potentials (bottom left and right) calculated with the UHBD software (using the same parameters as applied in the PFplus server). For better representation the surface of the proteins is shown from two view points rotated  $180^\circ$  (left and right). As demonstrated the largest positive patch calculated by the PFplus server (top left) clearly corresponds to the continuous positively charged region on the molecular surface (bottom left).

Autoantigen SP100-B, Supplementary Figure S2 A and B, respectively) and two proteins which do not bind nucleic acid (Supplementary Figure S2 C and D). Among the non-nucleic acid-binding is the cytochrome P450terp (1cpt) in which the large positive patch is known, to be functional, possibly involved in protein–protein interaction (22). In addition, we show a relative small patch found in the barstar protein (1bta) which does not have a known functional role. As clearly demonstrated, both the small and large patches extracted with the PFplus algorithm correspond to the positive charged region on the molecular surface illustrated in the bottom panel. It is important to note that two of the examples shown in Supplementary Figure S1 (B and C) are protein structures that were solved by NMR, demonstrating that the PFplus method is applicable for low-resolution structures.

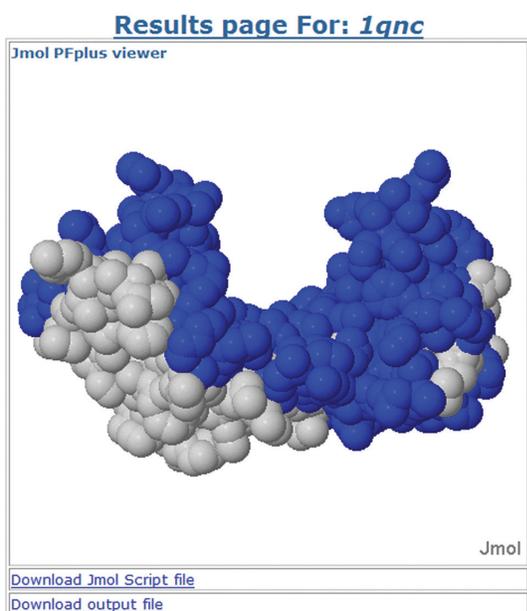
### PFplus INPUT AND OUTPUT

To extract the largest positive patch on a protein surface, the user should provide a PDB code of a protein or a protein complex. Alternatively, a user can upload a protein coordinate file in PDB format. The only mandatory fields are the ATOM records. All non-protein chains of the input file are ignored for patch calculation. The PFplus program displays the largest positive patch (defined by the number of grid points) without assigning a significance value.

The PFplus algorithm does not require a minimal resolution for calculating the largest positive patch, as long as all amino acid atoms are provided (models including only CA atoms are not accepted). Thus, the PFplus algorithm can be applied to structures solved by both X-ray crystallography and NMR, as illustrated in Supplementary Figure S2 B and C. As a default, the patch calculations for NMR structures are applied to the first model. It is important to note that the calculated electrostatic potential can fluctuate when changing input parameter, such as the grid spacing. An example of the largest positive patch calculated on the surface of the TATA-binding protein (TBP) from *Arabidopsis thaliana* is shown in Figure 3 and Supplementary Figure S3 (1qnc). Supplementary Figure S3 demonstrates the effect of the input parameters on the extraction of the largest positive patch, as exemplified on the TBP structure. A detailed output of the residues included in each patch is also provided in Supplementary Table S2. Notably, when using the  $2kT/e$  cutoff for the electrostatic potential (the default of the PFplus server) the patch calculation does not seem to strongly depend on the grid parameter. Nevertheless, when applying a higher cutoff for including a grid point within the continuous patch the definition of the patch seem to be much more sensitive to changes in the electrostatic potential parameters. This is probably due to the sparse data of grid points which extend the  $3kT/e$  cutoff.

Currently, the PFplus server provides the user with three different types of output. Shortly after submission

## The Patch Finder Plus - PFplus Server



**Figure 3.** PFplus output for the TATA binding protein from *Arabidopsis thaliana* (PDB code 1qnc). The residues of the largest positive patch are colored in blue. As illustrated in both cases, the saddle-shaped DNA-binding domain of the TATA-binding protein coincides with the largest positive patch on the protein surface.

(depending on the protein size and the number of chains) a Jmol molecular viewer window will appear displaying a representative protein chain (usually the first protein chain in the PDB file) in a spacefill full atom representation. All atoms affiliated to the largest positive patch are colored in blue (as illustrated in Figure 3). In addition, a Jmol script for displaying the patch residues is available for downloading, as well as an output file (in PDB format) that includes the patch residues extracted for each protein chain provided originally in the input file. Preceding the atomic coordinates of each protein chain, a PDB header is supplied providing the residue numbers of the all residues included in the largest positive patch.

### WORK IN PROGRESS

The PFplus web server is designed to calculate and display the largest positive patch on any given protein surface. We are currently integrating several more features which will be implemented in the next version of the PFplus web server. In the current version the cutoff for including a surface point in the continuous positive patch is predefined as  $2kT/e$ . This cutoff was chosen based on our comprehensive analysis of nucleic acid interfaces in protein–DNA complexes(6). In addition, the PFplus algorithm uses a set of default parameters for calculating the Boltzmann equations. In the next PFplus version, the user will be able to change default parameters within a given range specified by the program.

Generally the PatchFinder algorithm can support calculations of any continuous electrostatic patch on a protein surface, either a positive or a negative patch. In addition, the PatchFinder algorithm can provide a ranked list of surface patches for any given protein sequence. In the next version of PFplus, users will be able to define the number of surface patches required.

### CONCLUSIONS

The presence of positive patches on protein surfaces can be indicative of protein functions such as nucleic acid binding, membrane binding and others. Previous studies have suggested that the largest positive patch on the protein surface overlaps to a high extent with the nucleic acid-binding interface. We have developed an online web server for calculating and displaying the largest positive patch on protein surfaces. It is important to note that the PFplus server is not intended for predicting specific binding sites, however, it can suggest the location of binding interfaces on the protein surface. The input of the server is a PDB code or an input file in PDB format. The largest positive patch is provided as an output, both visually, displayed on a Jmol viewer, and as a text file (in PDB format) including a list of all patch residues.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

### ACKNOWLEDGEMENTS

We would like to thank Ora Furman Schueler, Hilda David-Eden for helpful suggestions and Martin Akerman for testing the server. This work was funded by the Mallat Family Foundation Granted to Y.M.G. S.S. is supported by the Israeli Science Foundation ISF grant number 923/05. Funding to pay the Open Access publication charges for this article was provided by Israeli Science Foundation ISF grant number 923/05.

*Conflict of interest statement.* None declared.

### REFERENCES

- Honig,B. and Nicholls,A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144–1149.
- Ohlendorf,D.H. and Matthew,J.B. (1985) Electrostatics and flexibility in protein-DNA interactions. *Adv. Biophys.*, **20**, 137–151.
- Mandel-Gutfreund,Y., Schueler,O. and Margalit,H. (1995) Comprehensive analysis of hydrogen bonds in regulatory protein DNA-complexes: in search of common principles. *J. Mol. Biol.*, **253**, 370–382.
- Mandel-Gutfreund,Y., Margalit,H., Jernigan,R.L. and Zhurkin,V.B. (1998) A role for CH...O interactions in protein-DNA recognition. *J. Mol. Biol.*, **277**, 1129–1140.
- Honig,B., Sharp,K. and Gilson,M. (1989) Electrostatic interactions in proteins. *Prog. Clin. Biol. Res.*, **289**, 65–74.
- Stawiski,E.W., Gregoret,L.M. and Mandel-Gutfreund,Y. (2003) Annotating nucleic acid-binding function based on protein structure. *J. Mol. Biol.*, **326**, 1065–1079.
- Bhardwaj,N., Langlois,R.E., Zhao,G. and Lu,H. (2005) Kernel-based machine learning protocol for predicting DNA-binding proteins. *Nucleic Acids Res.*, **33**, 6486–6493.

8. Jones, S., Shanahan, H.P., Berman, H.M. and Thornton, J.M. (2003) Using electrostatic potentials to predict DNA-binding sites on DNA-binding proteins. *Nucleic Acids Res.*, **31**, 7189–7198.
9. Shanahan, H.P., Garcia, M.A., Jones, S. and Thornton, J.M. (2004) Identifying DNA-binding proteins using structural motifs and the electrostatic potential. *Nucleic Acids Res.*, **32**, 4732–4741.
10. Tsuchiya, Y., Kinoshita, K. and Nakamura, H. (2004) Structure-based prediction of DNA-binding sites on proteins using the empirical preference of electrostatic potential and the shape of molecular surfaces. *Proteins*, **55**, 885–894.
11. Tsuchiya, Y., Kinoshita, K. and Nakamura, H. (2005) PreDs: a server for predicting dsDNA-binding site on protein molecular surfaces. *Bioinformatics*, **21**, 1721–1723.
12. Jones, S. and Thornton, J.M. (1997) Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.*, **272**, 133–143.
13. Sheinerman, F.B., Norel, R. and Honig, B. (2000) Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biol.*, **10**, 153–159.
14. Bradford, J.R., Needham, C.J., Bulpitt, A.J. and Westhead, D.R. (2006) Insights into protein-protein interfaces using a Bayesian network prediction method. *J. Mol. Biol.*, **362**, 365–386.
15. Nayal, M., Hitz, B.C. and Honig, B. (1999) GRASS: a server for the graphical representation and analysis of structures. *Protein Sci.*, **8**, 676–679.
16. Petrey, D. and Honig, B. (2003) GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Methods Enzymol.*, **374**, 492–509.
17. Nicholls, A., Bharadawaj, R. and Honig, B. (1993) GRASP: Graphical representation and analysis of surface properties. *Biophys. J.*, **64**, A166.
18. Ahmad, S. and Sarai, A. (2004) Qgrid: clustering tool for detecting charged and hydrophobic regions in proteins. *Nucleic Acids Res.*, **32**, W104–W107.
19. Davis, M.E., Mudura, J.D., Luty, B.A. and McCammon, J.A. (1991) Electrostatics and diffusion of molecules in solution. *Comp. Phys. Commun.*, **62**, 187–197.
20. McDonald, I.K. and Thornton, J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.
21. Richards, F.M. (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.*, **82**, 1–14.
22. Stayton, P.S. and Sligar, S.G. (1990) The cytochrome P-450cam binding surface as defined by site-directed mutagenesis and electrostatic modeling. *Biochemistry*, **29**, 7381–7386.