HotSpot Wizard: a web server for identification of hot spots in protein engineering

Antonín Pavelka, Eva Chovancová and Jiří Damborsky*

Loschmidt Laboratories, Institute of Experimental Biology and National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5/A4, 625 00 Brno, Czech Republic

Received February 18, 2009; Revised April 18, 2009; Accepted May 2, 2009

ABSTRACT

HotSpot Wizard is a web server for automatic identification of ‘hot spots’ for engineering of substrate specificity, activity or enantioselectivity of enzymes and for annotation of protein structures. The web server implements the protein engineering protocol, which targets evolutionarily variable amino acid positions located in the active site or lining the access tunnels. The ‘hot spots’ for mutagenesis are selected through the integration of structural, functional and evolutionary information obtained from: (i) the databases RCSB PDB, UniProt, PDQSWS, Catalytic Site Atlas and nr NCBI and (ii) the tools CASTp, CAVER, BLAST, CD-HIT, MUSCLE and Rate4Site. The protein structure and e-mail address are the only obligatory inputs for the calculation. In the output, HotSpot Wizard lists annotated residues ordered by estimated mutability. The results of the analysis are mapped on the enzyme structure and visualized in the web browser using Jmol. The HotSpot Wizard server should be useful for protein engineers interested in exploring the structure of their favourite protein and for the design of mutations in site-directed mutagenesis and focused directed evolution experiments. HotSpot Wizard is available at http://loschmidt.chemi.muni.cz/hotspotwizard/.

INTRODUCTION

Protein engineering represents a powerful approach for production of useful enzymes (1,2). Directed evolution and rational protein design are two distinct technologies currently used in protein engineering. While directed evolution randomly modifies the residues throughout the entire protein structure, the rational protein design and focused directed evolution targets only selected residues, called ‘hot spots’. Modification of ‘hot spots’ often leads to desired changes in protein properties and significantly reduces the number of mutants that must be biochemically characterized to discover an interesting variant (3,4).

The functional amino acid residues that mediate the substrate binding, transition-state stabilization or product release, i.e. the residues located in the active sites and lining the access tunnels, are frequently selected as ‘hot spots’ for modification of enzyme catalytic properties (4–12). However, some functional residues may be indispensable for catalysis and their replacement leads to complete loss of enzyme function. Information about the evolutionary conservation of individual residues can serve as a useful indicator for their mutability. Mutagenesis targeting the functional residues located in highly variable positions is an effective strategy for altering catalytic properties with reduced risk of losing a catalytic activity (9).

This strategy is implemented in HotSpot Wizard, which automatically identifies the functional residues for engineering of catalytic properties and estimates their mutability. For this purpose, HotSpot Wizard integrates several bioinformatics databases and computational tools. Structural analyses are conducted to identify the residues that potentially come into contact with the substrates or products. The mutability of individual amino acid residues is derived from their conservation level. Higher mutability implies a better chance that the amino acid replacement will not compromise the enzyme function. HotSpot Wizard assigns all functional residues with high mutability as ‘hot spots’. HotSpot Wizard results, including the annotated structure, sequence and summary tables of ‘hot spots’, functional residues, pockets and tunnels, can be interactively visualized in the web browser or downloaded to a local computer. For individual residues, information about their mutability, structural location, functional role and annotations are provided.

A variety of computational tools has been developed to assist rational design of proteins and recently reviewed (13). Novelty of HotSpot Wizard lies in a tight integration of functional, evolutionary and structural information. This integration enables automation of the
proteins. The HotSpot Wizard is designed to identify amino acid residues indispensable for enzymatic function and to collect available information about mutagenesis experiments and naturally occurring variants. The tool is accessible to the scientific community and can be used by researchers to gain insights into the structure and function of proteins.

METHODS

The HotSpot Wizard employs a flowchart depicted in Figure 1. The input to the tool can be either a PDB code or a protein structure in PDB format. The output is an annotated structure, with hotspots for mutagenesis selected through the integration of structural, functional, and evolutionary information obtained from bioinformatics databases.

The method begins with a PDB file either downloaded from the RCSB PDB or provided by the user. The PDBSWS database is used to identify UniProt entries corresponding to individual PDB chains. If the structure is provided by the user, the UniProt entries are identified through a BLAST sequence search. The UniProt entries are retrieved using the WSDbfetch web service. Residue annotations are extracted from the UniProt fields 'Active site', 'Binding site', 'Mutagenesis' and 'Natural variant' to determine amino acid residues indispensable for enzymatic function and to collect available information about mutagenesis experiments and naturally occurring variants. PDBSWS is then used for the second time to map the indexes of extracted UniProt annotations to the corresponding PDB residues.

CASTp is used to find all pockets present in the query structure and to select active site pockets based on the volume of individual pockets. For each identified active site, a center of mass for the catalytic residues is computed, where unitary weight is assigned to each residue. In the next step, another point is calculated where (i) a sphere circumscribed around this point intersects no atom in a protein structure and (ii) the distance between this point and the center is minimal, ensuring its positioning within the space occupied by the substrate molecule. The CAVER calculation is then performed to identify tunnels connecting this point with the outside solvent.

In the final step, the BLAST search against the non-redundant database at NCBI is performed to gather protein sequences similar to the query. Sequences are clustered using CD-HIT and representatives of the clusters are aligned using MUSCLE. These tools were chosen for their ability to process large datasets within a short time period with required accuracy. The multiple sequence alignment is used for estimating site-specific conservation scores. The evolutionary rates of individual amino acid positions are estimated by the empirical Bayesian method as implemented in the Rate4Site program. The evolutionary rates are then converted to the ConSurf conservation scale, which is used to assign mutability to individual residues of the query enzyme.

DESCRIPTION OF THE SERVER

Input

HotSpot Wizard requires a structure of the query protein, provided as a PDB code or PDB file, and an e-mail address as the only obligatory inputs. Additionally, a number of optional settings are available to control the calculation. The user may specify protein chains of interest. By default, the application attempts to extract information about the biologically relevant molecule from the PDB file. If this information is not available, the first
chain of the PDB file is used for the calculation. Another option enables the selection of the resource for annotation of the catalytic residues or the manual input of the catalytic residues. Two parameters can be set for the calculation of tunnels: (i) the minimal tunnel radius specifying the minimal tunnel width (default 1.4 Å) and (ii) the minimal starting radius specifying the required free space around the tunnel starting point (default 1.6 Å). The calculation of evolutionary conservation is, by default, conducted on the dataset obtained by the BLAST search. The appropriate setting of the \(E\)-value (default \(1E^{-12}\)) enables the user to perform the analysis on enzymes with desired similarities or function. The user can also change the maximum number of sequences to be analyzed (default 50) and thus potentially increase the accuracy of the evolutionary rate estimation. However, demand on calculation time grows with increasing number of sequences, and it is currently not recommended to analyze more than 100 sequences at once. Results of the conservation analysis are stored on the server for 3 months. If a new job with identical parameters is submitted during this period, the precalculated results are used instead of conducting the conservation analysis. The usage of precalculated results may be disabled. Users can also upload their own multiple sequence alignment and the phylogenetic tree. Users interested only in the structural annotation or some specific analysis may skip the conservation analysis as it is the most time-consuming part of the HotSpot Wizard calculation.

**Output**

The ‘Job’ panel appears when the calculation is finished and enables navigation through the obtained results. The results may be accessed either by an interactive web application called the ‘Results Browser’ (Figure 2) or downloaded to a local computer. Moreover, the ‘Job’ panel provides links to the external databases and servers with further information about the query protein.

**Residue annotations.** For each residue, information about its mutability, potential function (catalytic, located in the active site or access tunnel), available mutagenesis data, existing sequence variants and amino acids occurring at the corresponding position of the sequence alignment is provided. The mutability of individual sites, which estimates the tolerance of a given position to amino acid replacement, is derived from evolutionary conservation. The mutability scale ranges from 1 (low mutability) to 9 (high mutability). For visualization purposes, mutability is converted to a color scale ranging from blue (low mutability) to red (high mutability). The color scale is additionally extended by the violet, pink and ivory colors to indicate catalytic residues, residues with unreliable assignment and residues with missing information about evolutionary conservation, respectively. It is important to note that low mutability of a particular position does ‘not’ necessarily mean that the corresponding residue is immutable. However, the residues at such positions are often essential for maintaining structural or functional properties of an enzyme and should not be mutated without careful consideration of the potential effects of these mutations. To prevent mutagenesis at the sites indispensable for a function, catalytic residues are assigned as immutable.

**Summary tables.** The main output of the HotSpot Wizard application is summarized in the ‘Mutagenesis Hot Spots’ table, listing the residues assigned as ‘hot spots’, based on their high mutability (6–9) and structural location. ‘Hot spots’ are ordered by their mutability, starting with the most mutable ones, and accompanied by the residue annotations. For some proteins, only very few ‘hot spots’ are identified by HotSpot Wizard, because functional residues of the protein are highly conserved. The users are encouraged to use the ‘Functional residues’ table to see annotations for all active site and tunnel residues identified in the query structure. However, selecting the targets for mutagenesis from this table increases a risk of producing inactive mutants, especially when the positions with low mutability are targeted. The ‘Functional residues’ table is also useful for users who want to select their own ‘hot spots’ based on other criteria than mutability. The ‘All Residues’ and ‘All Residues by Mutability’ tables provide information about each individual residue of the query enzyme.

**Annotated structure.** HotSpot Wizard results are mapped on the enzyme structure and are visualized directly in the web browser using the Jmol applet (http://jmol.sourceforge.net/). Besides the enzyme structure, ligands and identified tunnels may be displayed as well. The default visualization highlights the ‘hot spots’ and the functional residues. The residue annotations may be invoked by positioning the mouse cursor over a given residue in the Jmol applet. Users can select predefined groups of residues and change their coloring and visualization styles using the ‘Controls’ panel. The Jmol applet is interactively interconnected with all other sections of the ‘Results Browser’ also enabling selection from the ‘Sequence’ panel and the summary tables.

**Annotated sequence.** The panel ‘Sequence’ displays the annotated sequence of the query enzyme. Its functionality comprises selections of residues and displaying the residue annotations. The coloring of the sequence is synchronized with the structure and the summary tables.

**Results for download.** The results may be downloaded to a local computer all in one package or as individual files. The package includes the summary tables in a text format for direct reading or software parsing, the original outputs obtained from individual servers and databases, and the Python script enabling visualization and analysis of the HotSpot Wizard results in PyMOL (http://pymol.sourceforge.net/).

**Limitations**

HotSpot Wizard processes information from databases and computational tools in a highly automated manner. A specification of the query protein structure is the only required input information. Tight integration of individual components in the HotSpot Wizard workflow may result in the propagation of incorrect results from one tool to the remaining components of the cascade. (i) Currently, the most critical step of the protocol is the...
assignment of the catalytic residues. Information about the catalytic residues is used to avoid mutagenesis at the residues critical for the enzyme function, but also for the assignment of the active site pocket and the starting point for the calculation of tunnels. If catalytic residues are not found in databases either for the query sequence or for homologous sequences, users must specify this information manually. Otherwise, the pocket with the largest volume will be assumed to contain the active site and no tunnels will be computed. (ii) The identification of the active site pocket is another critical step of the computation. Surface residues can be incorrectly assigned as part of the active site pocket. This sometimes happens at the interface of two subunits forming a very large pocket. Exclusion of one of the chains is a simple solution to this problem. Miss-assignment of the residues of the active site pocket happens also for proteins carrying large open depressions at the protein surface, which has to be recognized by the user, and these residues should not be considered for mutagenesis experiments. (iii) Finally, the sequence databases might not contain a sufficient number of homologous sequences for some of the proteins and their mutability thus cannot be calculated. In such cases, HotSpot Wizard lists the functional residues without mutability scores.

EXAMPLES

The ‘hot spot’ residues identified computationally by HotSpot Wizard were compared with the mutations extracted from the Protein Mutant Database (29) and the primary literature for four different proteins: (i) haloalkane dehalogenase, (ii) phosphotriesterase, (iii) 1,3-1,4-β-D-glucan 4-glucanohydrolase and (iv) β-lactamase.

**Haloalkane dehalogenase**

Haloalkane dehalogenase DhaA from *Rhodococcus* sp. catalyzes hydrolytic dehalogenation of a broad range of halogenated aliphatic hydrocarbons. The structure of the
DhaA enzyme was specified by the PDB code 1BN6 (30). In further text, the numbering from the PDB file is used, while 11 residues need to be subtracted to obtain the numbering used in the literature. HotSpot Wizard identified 17 reliable ‘hot spots’ (mutability range 6–9) lining 1 active site pocket and 2 tunnels: F142, R144, I146, P147, E151, W152, F155, A156, T159, V183, K186, C187, E234, H241, V256, P259, E262 (Figure 3). Kretz et al. performed saturation mutagenesis of the entire dhaA gene and constructed its mutability map (31). While one third of the residues of DhaA could not be replaced without compromising enzyme activity according to the experiments of Kretz et al., all but one ‘hot spots’ identified by HotSpot Wizard, were able to accommodate a wide range of mutations. The only exception is the residue F142 located just next to the catalytic residue E141. It is well known that the mutations introduced to the regions closest to the catalytic residues are tricky as they can disrupt the geometry of the transition state and lead to loss of enzyme activity. In another project, error-prone PCR was used for the introduction of random mutations to the DhaA structure by Gray et al. (32) and Bosma et al. (33). The biochemical screening of 10,000 variants against toxic environmental pollutant 1,2,3-trichloropropane (TCP) provided two double-point mutants C187Y + Y284F (33) and G14D+C187F (32), respectively, showing 3.5 times and 4 times higher activities with TCP than the wild-type enzyme. These two variants obtained independently in the directed evolution experiments carry the mutation in the position C187. HotSpot Wizard identified the position C187 as the ‘hot spot’ due to its location at the tunnel opening and high mutability. Pavlova et al. (manuscript submitted) used rational design and selected five positions lining the access tunnels: I146, W152, C187, V256 and L257. The mutant of Bosma et al. (C187Y and Y284F) and the triple mutant (W152F, C187Y and Y284F) were used as templates for focused directed evolution, subjecting the positions I146, V256 and L257 to a simultaneous saturation mutagenesis. The screening of 5000 variants against TCP provided 25 unique mutants with higher activities towards TCP; the best mutant showing 32 times higher activity than the wild type. Out of the five positions rationally selected in this study, four positions were also identified as ‘hot spots’ by HotSpot Wizard.

**Phosphotriesterase**

Phosphotriesterase from *Pseudomonas diminuta* catalyzes the hydrolysis of a variety of organophosphates, including the widely used insecticide paraaxon and the chemical warfare agent sarin. The crystal structure of the phosphotriesterase homodimer, PDB code 1I0D (34), was analyzed and nine reliably assigned ‘hot spots’ were identified: G60, L136, R139, S205, D235, A270, L271, L272, F306. The data from mutagenesis experiments were found for 4 of these residues: G60, L136, L271 and F306. Gopal et al. (35) rationally designed eight mutants, including L136Y, L271Y, F306A and F306Y, intending to enhance activity towards a chemical warfare agent VX [O-ethyl S-(2-diisopropyl aminoethyl) methylphosphonothioate] and five other organophosphorus compounds. L136Y was the only mutation under study that leads to an increased relative activity towards VX. Five mutants, including L136Y and F306Y, displayed increased activity towards demeton-S-methyl. In another study, improvements of the catalysis of insecticide DFP up to one order of magnitude were obtained with the mutants F306Y and F306H (36). Systematic alanine-scanning of phosphotriesterase binding pocket, including residues G60, L271 and F306, was carried out to elucidate structural determinants of stereoselectivity and overall reactivity of this enzyme with different substrates (37). Substitution G60A lead to a dramatic enhancement of the chiral preference for the S<sub>r</sub>-enantiomers, increasing the enantioselectivity for several chiral substrates, e.g. from 21 to >100 for p-nitrophenyl ethyl phenyl phosphate (6). Substitutions in L271 and F306 did not improve stereoselectivity, while the catalytic rates were decreased relative to those of the wild-type enzyme for most of the tested substrates (37).

**1,3,1,4-β-D-Glucan 4-glucanohydrolase**

1,3-1,4-β-D-Glucanase from *Bacillus licheniformis* is an endoglucosidase that cleaves β-glucans containing mixed β-1,3- and β-1,4-linkages. The structure of 1,3-1,4-β-glucanase was
specified by the PDB code 1GBG (38). In further text, the numbering from the PDB file is used, while 29 residues needs to be added to obtain the numbering used in the literature. HotSpot Wizard identified 5 reliable ‘hot spots’: Y24, N28, M29, T69, V189. Three of them, Y24, N28 and M29, are described in the experimental study of Pons et al. (39). Pons et al. conducted alanine-scanning mutagenesis of the major loop (residues 22–38) of 1,3-1,4-beta-glucanase to investigate its role in substrate binding. The substitution Y24 by alanine resulted in a 3.4-fold increase in $k_{cat}/K_m$ towards the substrate 4-methylumbelliferyl 3-O-M29, 4-O-Y24, N28, M29, T69, V189. Three of them, Y24, N28 and M29, were systematically randomized to identify the determinants of catalysis for the important antibiotic cephalosporin L293P. Compared with the wild-type beta-lactamase, the catalytic efficiencies of the L293P mutant were increased 27- and 11-fold for cefepime and ceftazidime, respectively. In the next step, 14 additional substitutions were introduced to the position L293 by site-directed mutagenesis. More than half of the mutants conferred increased resistances towards cefepime and ceftazidime.

**Validation**

The effectivity of mutagenesis targeting highly variable positions was assessed by comparing HotSpot Wizard results with the data obtained from the systematic mutagenesis studies: (i) 3315 mutants of lactose repressor (43), (ii) 1930 mutants of T4 lysozyme (44), (iii) 676 mutants of barnase (45) and (iv) 364 mutants of HIV reverse transcriptase (29,46). The individual mutagenesis studies differed in the activity assays and in the classification of the effects. Therefore, these data sets are not directly comparable and had to be evaluated independently. For each data set, those mutations which lead to the protein variants with very low or no activity, were assigned as deleterious. The proportion of deleterious mutations in the positions assigned as highly mutable by HotSpot Wizard, i.e. with the mutability grades from 6 to 9, were compared with proportion of deleterious mutations in the entire protein structure. Calculated ratio of deleterious mutations in the mutable positions versus in the entire protein structure were: (i) 2.7% versus 25.3% for lactose repressor, (ii) 1% versus 9.1% for T4 lysozyme, (iii) 0.4% versus 4.8% for barnase and (iv) 3.9% versus 30.7% for HIV reverse transcriptase. These results demonstrate that mutagenesis targeting the ‘hot spot’ positions identified by HotSpot Wizard provides significantly higher proportion of viable variants than blind mutagenesis. Detailed information on all presented examples can be downloaded from http://loschmidt.chemi.muni.cz/hotspotwizard/data.

**Conclusions and Outlook**

The development of the HotSpot Wizard server was motivated by a growing demand for comprehensive tools assisting protein engineers with the rational design of enzymes. The primary use of HotSpot Wizard is the identification of ‘hot spots’ for the site-directed mutagenesis or focused directed evolution experiments. Mutagenesis targeting the ‘hot spots’ found by HotSpot Wizard should increase the yield of active mutants with altered catalytic properties. Alternatively, HotSpot Wizard can be used for the annotation of protein structures. HotSpot Wizard automates a protein engineering protocol by the integration of several bioinformatics databases and computational tools and saves the user’s time. Minimal demands on input information make this web server potentially useful for the users with no prior knowledge of structural or bioinformatics analyses. We are currently integrating additional features to HotSpot Wizard, e.g. a visualization of the residue solvent accessibility and volumes of pockets. We also aim to improve the estimation of mutability by including additional factors into the calculation of the prioritization function.

**Acknowledgements**

We would like to express our thanks to the authors of the tools used by HotSpot Wizard and to Peter Lisak (Brno University of Technology) for the LLWS utility enabling a simple HTTP access to software installed on Linux. Access to the METACentrum supercomputing facilities provided under the research intent MSM638391720 is highly appreciated.

**Funding**

Czech Ministry of Education, Youth and Sport [LC06010 to E.C.H., MSM0021622412 to J.D.]; the Grant Agency of the Czech Republic [201/07/0927 to J.D.]; and the Grant Agency of the Czech Academy of Sciences [IAA401630901 to J.D.]. Funding for open access charge: LC06010 and MSM0021622412.

Conflict of interest statement. None declared.
REFERENCES


