PatchSearch: a web server for off-target protein identification

Julien Rey, Inès Rasolohery, Pierre Tufféry, Frédéric Guyon, Gautier Moroy

Published online 27 May 2019 *Nucleic Acids Research, 2019, Vol. 47, Web Server issue* ***W365–W372***
doi: 10.1093/nar/gkz478

PatchSearch: a web server for off-target protein identification

Julien Rey[1,2], Inès Rasolohery[1], Pierre Tufféry[1,2], Frédéric Guyon[1,*] and Gautier Moroy[1,*]

[1]Université Paris Diderot, Sorbonne Paris Cité, INSERM UMRS-973, Molécules Thérapeutiques *in silico* (MTi), F-75205 Paris, France and [2]Ressource Parisienne en Bioinformatique Structurale (RPBS), Paris, France

Received February 22, 2019; Revised April 26, 2019; Editorial Decision May 05, 2019; Accepted May 21, 2019

ABSTRACT

The large number of proteins found in the human body implies that a drug may interact with many proteins, called off-target proteins, besides its intended target. The PatchSearch web server provides an automated workflow that allows users to identify structurally conserved binding sites at the protein surfaces in a set of user-supplied protein structures. Thus, this web server may help to detect potential off-target protein. It takes as input a protein complexed with a ligand and identifies within user-defined or predefined collections of protein structures, those having a binding site compatible with this ligand in terms of geometry and physicochemical properties. It is based on a non-sequential local alignment of the patch over the entire protein surface. Then the PatchSearch web server proposes a ligand binding mode for the potential off-target, as well as an estimated affinity calculated by the Vinardo scoring function. This novel tool is able to efficiently detects potential interactions of ligands with distant off-target proteins. Furthermore, by facilitating the discovery of unexpected off-targets, PatchSearch could contribute to the repurposing of existing drugs. The server is freely available at http://bioserv.rpbs.univ-paris-diderot.fr/services/PatchSearch.

INTRODUCTION

During the drug discovery process, binding sites comparison can assist in the identification of interactions of drugs with undesired targets (off-targets) and the understanding of adverse effects. Binding site comparison is also helpful for drug repositioning and ligand selectivity optimization. Consequently, different approaches have been developed for this purpose and include ligand-based and structure-based approaches (1,2).

When based on the knowledge of the structure, off-target binding site identification faces the issue of structural plasticity, which hampers the identification of undesired binding partners. Different strategies have been considered and are mostly based on the fact that similar structures or regions of structure accessible to the solvent can be expected to bind similar ligands. Alignment-free methods perform an overall comparison of global properties and characteristics of binding sites such as shape, surface descriptors and physicochemical residue properties combined with atom types (3–5), Patch-Surfer (6,7), PocketMatch (8), PocketFeature (9). On the other hand, sequence order-independent alignments of residues or atoms are in general far more difficult to compute than alignment-free comparisons, but these methods allow for the identification of atoms or residues involved in the binding with a ligand. These methods are based on geometric hashing: TESS (10), SitesBase (11), SiteEngine (12) and I2I-SiteEngine (13), MultiBind (14,15) and PCalign (16), or on the Hungarian algorithm eMatchSite (17). A new approach based on deep learning has been recently published to compare binding site (18). Many methods also compute sequence order-independent alignment by searching for cliques in product graphs (19). The Bron–Kerbosh algorithm is the most efficient algorithm to search for all maximal cliques (20). For this purpose, it is widely used, in particular in computational chemistry (21) and is recognized as being one of the most efficient in practice (22). Many improved variants have since been described and more efficient algorithms for finding a maximum clique ex-

*To whom correspondence should be addressed. Tel: +331 57278385; Fax: +331 57278372; Email: gautier.moroy@univ-paris-diderot.fr
Correspondence may also be addressed to Frédéric Guyon. Tel: +331 44493073; Fax: +331 43065019; Email: frederic.guyon@univ-paris-diderot.fr
Present addresses:
Julien Rey, Université de Paris, Biologie Fonctionnelle et Adaptative CNRS UMR 8251, Computational Modeling of Protein-Ligand Interactions INSERM U1133, F-75205 Paris, France.
Pierre Tufféry, Université de Paris, Biologie Fonctionnelle et Adaptative CNRS UMR 8251, Computational Modeling of Protein-Ligand Interactions INSERM U1133, F-75205 Paris, France.
Gautier Moroy, Université de Paris, Biologie Fonctionnelle et Adaptative CNRS UMR 8251, Computational Modeling of Protein-Ligand Interactions INSERM U1133, F-75205 Paris, France.
Frédéric Guyon, Université de Paris, Biologie Intégrée du Globule Rouge, UMR_S1134, BIGR, INSERM, F-75015, Paris, France.

ists (22,23). However, the Bron–Kerbosh algorithm provides a mean to explore all maximal cliques and therefore all possible matchings. The first methods developing this strategy have been applied to protein structure comparisons since early 90's (24,25), and more recently, clique algorithms have been used in CavBase (26) and eF-site (27), SuMo (28). PocketMatch, SiteEngine, eF-site, MultiBind and ProBis (29) are available as web servers (Supplementary Table S1). Most of the above approaches compare or align binding sites only. ProBis web server is the only one able to search for a binding site on the entire surface of proteins based on local structural alignments. ProBis web server requires a query structure of a protein–ligand complex. The user can select a query binding site which is compared to entries in the non-redundant PDB (nr-PDB) or to a user-supplied list of PDB identifiers.

Molecular docking approaches can be also used to identify protein target of a ligand and consequently help the detection of off-target protein. Thus, IdTarget web server was developed to predict possible binding targets of a small chemical molecule via a divide-and conquer docking approach (30). It requires an input ligand file for the target screening. The user can choose to perform the search of potential binding targets among two predefined datasets of PDB identifiers or a user-supplied list of PDB identifiers.

Recently, PatchSearch (31) was developed to search for structurally conserved binding sites on the entire surface of a protein in order to help for the detection of potential off-target protein. It uses a quasi-clique approach which avoids a too stringent distance conservation between atoms and hence takes into account flexibility of binding sites. A quasi-clique is a dense subgraph. Our approach is similar to those used for dense subgraph or community detection in graph clustering (32–34). Cliques in correspondence graph involves the conservation of all internal distances between protein and patch surfaces. Based on Euclidean distance matrix properties, a well-chosen set of conserved internal distances is sufficient to ensure that all distances are equals or almost equals. Hence the complete node connection condition can be relaxed. In PatchSearch, a quasi-clique is a correspondence subgraph including a clique with added nodes ensuring a spatial similarity. It is computed with a greedy algorithm which starts from cliques and at each step add nodes with at least four connections with the current quasi-clique and maximizing a 3D similarity score. The quasi-clique approach compared to a classical clique technique allows for fast detection of larger and new patches on protein surfaces which could potentially provide new yet unpredicted off-target binding sites.

PatchSearch has been validated against a number of widely used datasets (35–40) (Supplementary Table S2) with a large diversity of ligands. First, PatchSearch ability to recognize proteins binding the same ligand has been assessed on four datasets (Kahraman, Homogeneous, Gunasekaran and Milletti). Second, it was applied to three drugs (ursodeoxycholic acid, prazosin and naproxen) from Drugs/sc-PDB dataset and to three drugs (sunitinib, imatinib and sorafenib) used in cancer therapy from the Multiple Target Ligand Database. The aim of these experiments was to prove that PatchSearch manages to identify true off-target proteins.

The main motivation of the present development of the PatchSearch web server is to provide the user with an automated workflow (Figure 1) to identify among a selection of proteins, which of them have a region sharing structural similarities with a ligand binding site. Through an energy minimization by steepest descent algorithm, the ligand conformation is locally optimized and scored within the binding site by smina program (41) using Vinardo scoring function (42).

MATERIALS AND METHODS

Patches and surfaces

A patch is a small piece of protein surface given by a set of solvent accessible atoms. Here, a patch is constituted by the solvent accessible atoms that have a distance smaller than 5Å to the ligand. Each atom of the patch is assigned a label corresponding to its type: N, O, S, C and Ca (for carbons α). Aromatic rings are replaced by one or two pseudo-atoms corresponding to one centroid for phenylalanine or tyrosine and two centroids for tryptophan). In this study, we are only interested in patch representing binding sites with ligands. The solvent accessibility is calculated using the program NACCESS (Hubbard, S.J. & Thornton, J.M. (1993), http://www.bioinf.manchester.ac.uk/naccess/) and all atoms with relative accessibility over 1% are conserved. This very small threshold allows to extract the largest surface involved in the interaction with the ligand.

For residues having at least one atom exposed, solvent accessible atoms and Cα coordinates are stored for the structural similarities calculations.

Two patches are considered to be similar if it exists a matching associating equivalent atoms. The largest match is computed such that all inner distances between patch atoms are preserved in target surfaces within a given tolerance. This is obtained by computing a sequence order-independent alignment, searching for cliques in product graphs (19) using the Bron–Kerbosh algorithm (20). Yet, this clique strategy, though accurate, presents some drawbacks: comparisons of proteins with several thousand atoms can lead to some very large product graphs, and hence to a large amount of running time. In addition, a too stringent distance conservation between atom does not take into account flexibility of binding sites. Therefore, in PatchSearch a quasi-clique strategy is employed: a quasi-clique is constructed from a small core clique found with the Bron–Kerbosh algorithm with a distance tolerance of 1.2Å, which is enriched using a greedy algorithm with less stringent distance tolerance set by default to 3Å. The greedy algorithm optimizes the relative Binet–Cauchy (BC) structural scores between patch and matched atoms by merging cliques and adding atom matchings.

The BC score is a geometric correlation score (43). The relative BC (rBC) score, is the BC score weighted by the percentage of retrieved atoms (N_{match}) relatively to the number of atoms of the patch (N_{patch}): $rBC = \frac{N_{\text{match}}}{N_{\text{patch}}} \times BC$

Rescoring step

For each targeted protein, the probable binding sites identified by PatchSearch are ranked according to the rBC score.

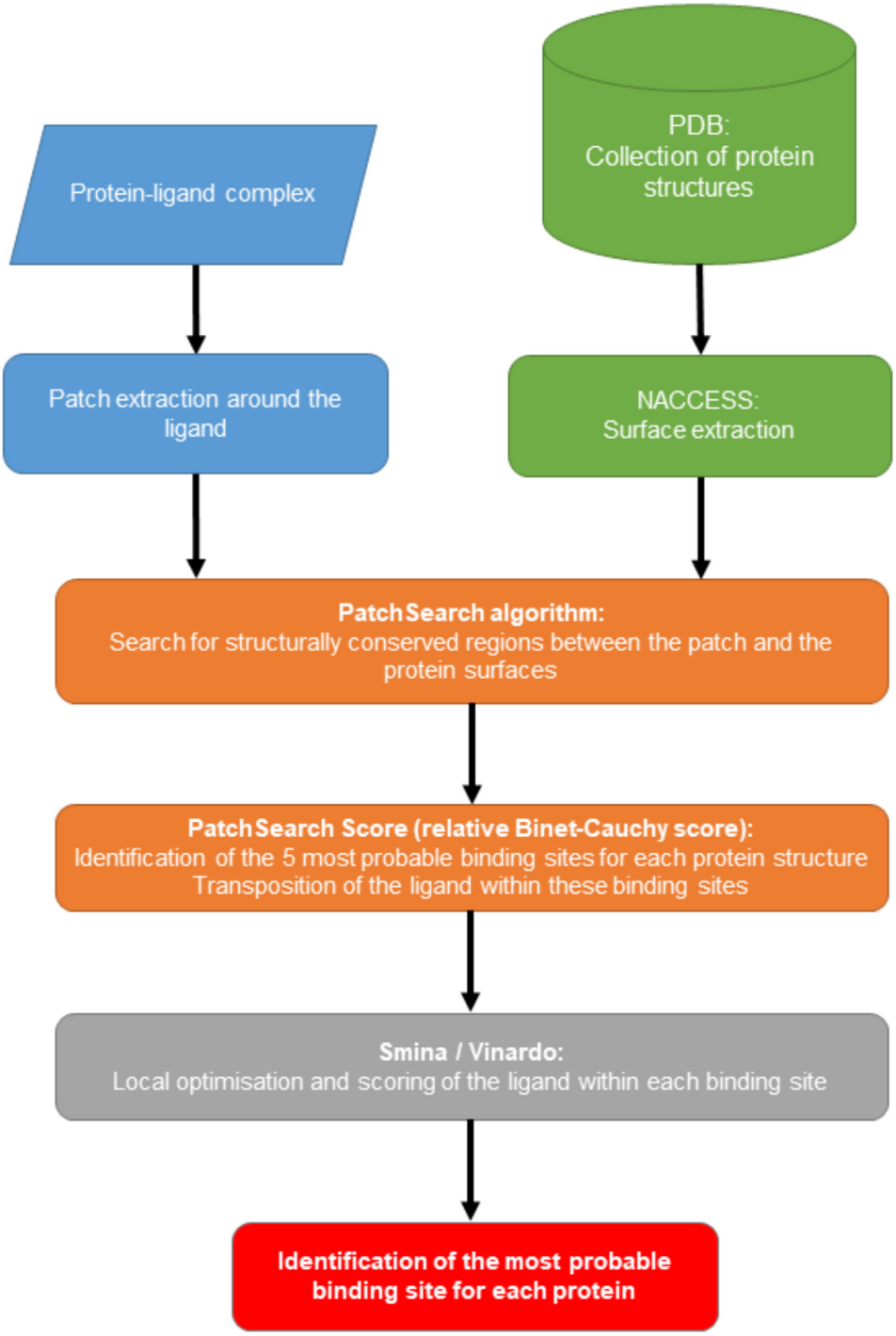


Figure 1. PatchSearch web server flowchart.

The five binding sites with the highest rBC score are kept. The ligand is transposed within the retained binding sites by rotating and translating ligand coordinates according to the alignment computed by PatchSearch between the atoms of the query patch and the matching atoms of the probable binding site. Using Vinardo scoring function, Smina program performs, through an energy minimization, a local optimization of the position and the conformation of the ligand within the probable binding sites. Vinardo scoring function takes into account steric attractions, steric repulsions, Lennard-Jones potentials, electrostatic interactions, hydrophobic interactions, non-hydrophobic interactions and non-directional hydrogen bonds. It has been recently shown to improve the scoring and ranking performances in docking experiments (42). During this step, the protein is rigid, the ligand is fully flexible and a maximum of 100 iterations of steepest descent algorithm are required. The lowest resulting energy among the five minimized mod-

els indicates which binding site can be considered as the most probable binding site at the protein surface.

WEB SERVER

The PatchSearch web server allows a local non-sequential searching for similar regions, called patches, on the entire protein surfaces, without any knowledge on binding site localizations or preliminary binding site detections. The service is fully embedded in the Mobyle framework (44). It embeds simple yet powerful data management features that allow the user to reproduce analyses. It gives to the users the possibility to create registered accounts, which allows user data and jobs to be maintained and managed across multiple work sessions and therefore to reuse and share data and results.

Input

There are two main inputs.

First, the user must provide a PDB identifier containing at least one protein–ligand complex. The ligand, for which similar binding sites have to be detected, can be selected in the list of non-protein molecule in the PDB file. The selected ligand, the patch around the ligand and the protein surface can be visualized by NGL viewer (45). If the user knows the residue number and the chain of the ligand of interest, he can directly fill the fields in the form.

Second, the user must define a collection of structures in which the search will be performed. This can be done either on the form of a collection of proteins defined as a list of PDB identifiers provided by the user. At present, this list cannot exceed 5000 proteins. Alternatively, five predefined lists are available. They correspond to subsets of the PDB with a minimal sequence length of 50 residues. These subsets are created using the Advanced Search Interface in the Protein Data Bank server. Three lists correspond to human proteins at different sequence identities (30%: 7922 structures; 50%: 9118 structures; 70%: 9836 structures), a list of 14 191 eukaryotic proteins with 30% of sequence identity and a list of 14 650 prokaryotic proteins with 30% of sequence identity.

Output

The server returns patches similar to the query identified among the collection of proteins. Pairings between the query atoms and the atoms in the targeted protein surface are detailed and patches are scored by the relative Binet–Cauchy score. Similar patches are output to a table with the PDB identifier, the number of patch atoms, alignment length (the number of retrieved atoms), the RMSD between the query patch and the retrieved patch, the relative Binet–Cauchy score and docking values resulting from the Smina rescoring calculations.

All the alignment results, as well the estimated affinity and the binding mode of the ligand for each targeted protein are downloadable in separated files.

Patch visualization

An interactive page allows to browse the retrieved patches. The residues forming the retrieved patch are in 'lines' representation with C atoms in green, the ligand is in 'sticks' representation with C atoms in cyan and the targeted protein in white 'cartoon' representation. The best solution for each targeted proteins are ranked according to the affinity value. For the visualization, we use NGL viewer (45) which takes advantage of WebGL capability of modern browsers for molecular graphics (Figure 2).

Execution times

Depending of the number of patch atoms and the protein size, average run times of searching one patch against an entire protein surface are <0.5 s (Supplementary Table S3). For the bigger predefined list of PDB containing 14 650 protein structures, the typical run times for a large patch formed by 100 atoms are of 1.5 h, but this depends on server load.

CASE STUDY

Distant off-targets

To illustrate PatchSearch effectiveness, we present an example of the identification of distant off-targets from the TOUGH-M1 dataset (46). The TOUGH-M1 dataset contains proteins with dissimilar global sequences and structures. The structural dissimilarity was measured by the TM-score (47). In this dataset, off-targets share no significant sequence (sequence identity < 30.0%) or fold similarity (TM-score < 0.4). This dataset is divided in two lists: a 'positive' list with proteins able to bind a chemically similar ligands and a 'negative' list in which the proteins interact with dissimilar ligands. Considering as input the binding site around Adenosine DiPhosphate (ADP) molecule in a myosin structure (PDB ID: 1lkx), we used the server to rank 20 proteins in 'positive' list and 20 proteins in 'negative' list for ADP. The patch around ADP was extracted from the 1lkx structure and compared to the entire surface of these 40 proteins. Note that all positive and negative proteins have very low sequence identities with the query —<23.0% and TM-scores <0.4. The results are reported in table 1 (Table 1).

The positive proteins have good affinity scores, with values less than $-10.0\ k_{\mathrm{cal/mol}}$, indicating that a site on the surface has a good affinity for ADP. Therefore, the positive proteins are correctly detected as off-targets. For all negative proteins, except for 4nym, the binding affinity scores are poorer around $-6\ k_{\mathrm{cal/mol}}$.

Results obtained for 4nym structure are interesting, because this protein is supposed to be unable to bind ADP. 4nym corresponds to the structure of GTPase HRas protein, which is involved in the activation of the Ras protein signal transduction. GTPase HRas protein is able to bind to Guanosine TriPhosphate (GTP) or Guanosine DiPhosphate (GDP). In this structure, GTPase HRas protein has been co-crystallized with Phosphoaminophosphonic Acid Guanylate Ester (ligand identifier for the PDB: GNP), a non-hydrolyzable analog of GTP, and the N-[1-(1H-indol-3-ylmethyl)piperidin-4-yl]-L-tryptophanamide (ligand identifier for the PDB: RND), a small molecule, altering experimentally the GTPase HRas activity. Logically, 4nym structure is considered as a negative protein for the binding site comparison between ADP

A

patchsearch.table.html

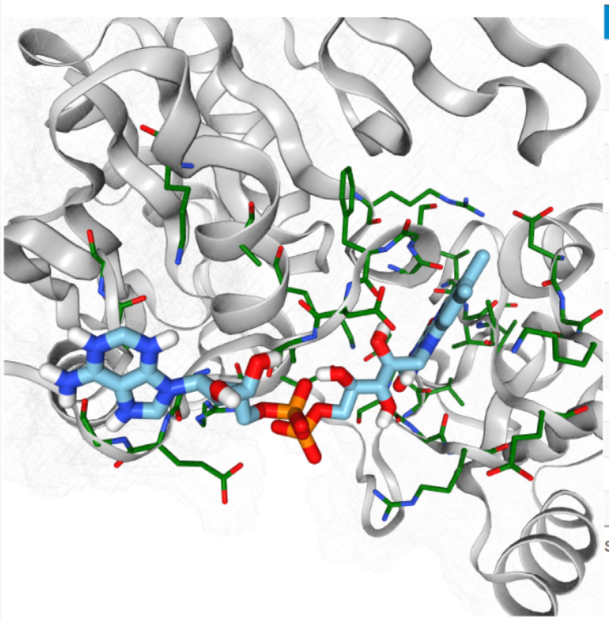
Id	Ch.	Protein	Organism	Cov.	RMSD	Score	Aff.
1efp	A	PROTEIN (ELECTRON TRANSFER FLAVOPRO...	PARACOCCUS DENITRIFICANS	60	2.309	0.335	-9.077
4h6r	A	PROLINE DEHYDROGENASE	DEINOCOCCUS RADIODURANS	46	2.225	0.267	-8.369
1o96	A	ELECTRON TRANSFERRING FLAVOPROTEIN ...	METHYLOPHILUS METHYLOTROPHUS	40	2.188	0.230	-8.275
4k1x	A	NADPH:FERREDOXIN REDUCTASE	RHODOBACTER CAPSULATUS	55	2.090	0.311	-8.027
3fst	A	5,10-METHYLENETETRAHYDROFOLATE REDU...	ESCHERICHIA COLI K-12	65	2.086	0.390	-7.911
3sf6	A	GLUTARYL-COA DEHYDROGENASE	MYCOBACTERIUM SMEGMATIS	54	2.150	0.302	-7.274
3of4	A	NITROREDUCTASE	IDIOMARINA LOIHIENSIS	47	2.063	0.266	-7.128
4u9u	A	N/A	N/A	44	2.028	0.247	-7.026
3ewk	A	SENSOR PROTEIN	METHYLOCOCCUS CAPSULATUS	48	1.815	0.276	-6.844
4pyt	A	UDP-N-ACETYLENOLPYRUVOYLGLUCOSAMINE...	UNIDENTIFIED	63	2.131	0.366	-6.835
3ahr	A	ERO1-LIKE PROTEIN ALPHA	HOMO SAPIENS	54	2.154	0.292	-6.781
1ep2	A	DIHYDROOROTATE DEHYDROGENASE B (PYR...	LACTOCOCCUS LACTIS	49	2.095	0.287	-6.665
2a1u	A	ELECTRON TRANSFER FLAVOPROTEIN ALPH...	HOMO SAPIENS	58	2.192	0.337	-6.605
2qtl	A	METHIONINE SYNTHASE REDUCTASE	HOMO SAPIENS	58	2.068	0.327	-6.584
3t58	A	SULFHYDRYL OXIDASE 1	MUS MUSCULUS	43	1.988	0.255	-6.481
2bvf	A	6-HYDROXY-D-NICOTINE OXIDASE	ARTHROBACTER NICOTINOVORANS	50	2.021	0.290	-6.432
2ok8	A	PUTATIVE FERREDOXIN--NADP REDUCTASE	PLASMODIUM FALCIPARUM 3D7	47	2.184	0.255	-6.405
3apy	A	METHYLENETETRAHYDROFOLATE REDUCTASE	THERMUS THERMOPHILUS	44	2.149	0.247	-6.386
4pwc	A	POLLEN ALLERGEN PHL P 4.0202	PHLEUM PRATENSE	34	2.178	0.170	-6.310
1jb9	A	FERREDOXIN-NADP REDUCTASE	ZEA MAYS	49	2.136	0.262	-6.257

Copy CSV Excel

Previous 1 2 3 4 Next

B

patchsearch.vis.html



Id	Aff.
1efp	-9.077
4h6r	-8.369
1o96	-8.275
4k1x	-8.027
3fst	-7.911
3sf6	-7.274
3of4	-7.128
4u9u	-7.026
3ewk	-6.844
4pyt	-6.835
3ahr	-6.781
1ep2	-6.665
2a1u	-6.605
2qtl	-6.584
3t58	-6.481

Showing 1 to 15 of 78 entries

Previous 1 2 3 4 5 6 Next

Atom pairs

Query: 55:C.O 54:C.CA 54:C.O 54:C.OE1 51:C.O 51:C.CA 21:C.CA 21:C.CG1 22:C.CA 23:C.N 26:C.CB 45:C.CA 47:C.N 46:C.CG 47:C.CA 47:C.CG2 47:C.OG1 50:C.CA 51:C.CG 51:C.N 51:C.OE1 55:C.CA 54:C.C 54:C.CB 55:C.CB 57:C.N 57:C.CA 57:C.O 57:C.CG2 57:C.CD1 58:C.CA 59:C.C 58:C.CG2 59:C.OD2 60:C.C 120:C.O 144:C.CA

Figure 2. Visualization of hits in the PatchSearch web server. (**A**) results table of proteins with patches similar to a query patch. (**B**) NGL viewer displaying the retrieved patch for 1efp structure. The protein is in white cartoon, ligand in thick sticks with C atoms in cyan and the residues of the retrieved patch are represented in lines with C atoms in green.

Table 1. Example of PatchSearch results: identification of distant off-targets for ADP

PDB ID	Sequence identity (%)	Structure similarity (TM-score)	TOUGH1-dataset list	Estimated Affinity ($k_{cal/mol}$)
1fqj	20.5	0.398	positive	−12.934
1in6	20.0	0.333	positive	−11.977
2o8b	22.1	0.270	positive	−11.923
1shz	18.5	0.386	positive	−11.853
4zkd	21.3	0.298	positive	−11.537
2fna	22.0	0.341	positive	−11.510
4nym	18.1	0.282	negative	−11.390
4d25	21.2	0.327	positive	−11.387
3fwy	23.5	0.365	positive	−11.361
1d2e	21.1	0.284	positive	−11.130
5bn3	21.6	0.271	positive	−10.829
1tq6	19.9	0.329	positive	−10.809
3iev	23.2	0.352	positive	−10.692
3u5z	21.3	0.299	positive	−10.616
1kk3	20.4	0.297	positive	−10.532
4kxf	19.3	0.298	positive	−10.491
1dg1	16.3	0.295	positive	−10.444
4djt	22.3	0.376	positive	−10.378
1sxj	22.9	0.307	positive	−10.368
3r7w	20.6	0.326	positive	−10.290
4yj1	15.1	0.279	positive	−10.271
2ozr	13.4	0.346	negative	−6.907
3d3h	22.3	0.355	negative	−6.861
1tve	18.8	0.339	negative	−6.780
5ai9	20.4	0.278	negative	−6.572
2vax	19.6	0.340	negative	−6.484
4rzm	18.1	0.339	negative	−6.425
1dsy	21.7	0.372	negative	−6.416
5hes	17.9	0.322	negative	−6.349
1y0g	19.0	0.343	negative	−6.328
2rjp	16.9	0.335	negative	−6.319
2f9a	20.6	0.294	negative	−6.298
3n0t	19.3	0.348	negative	−6.223
1kr1	18.4	0.370	negative	−6.214
1fbo	19.4	0.297	negative	−6.152
3njj	22.1	0.371	negative	−6.139
3od2	19.3	0.362	negative	−6.123
4s3r	20.8	0.283	negative	−6.059
2zjf	20.5	0.350	negative	−6.039
1lqy	19.6	0.361	negative	−5.923

The patch was extracted around ADP in the 1lkx myosin structure. Structural similarities between the patch and the entire surface of 40 PDB structures known to be able to interact, i.e. 'positive', or not, 'negative', with ADP or similar ligands, according to the TOUGH1-dataset. The sequence identity and the structure similarity were calculated between 1lkx against the each 40 PDB structures. The estimated affinity reported is the best score, computed by Smina program with Vinardo scoring function, between ADP and the five potential sites with the highest structural similarities according to the rBC score.

and RND, because these molecules are chemically dissimilar. However, the PatchSearch web server searches structural similarities onto the entire surface of the protein. The favorable computed binding affinity is basically due to the fact that PatchSearch has retrieved the GTP/GDP binding site of GTPase HRas protein (Figure 3). This unexpected result can be explained by the high chemical similarities between ADP and GDP.

It reinforces strengthens the interest of taking into account the entire surface protein, not only comparing predetermined binding sites for the finding of new off-targets.

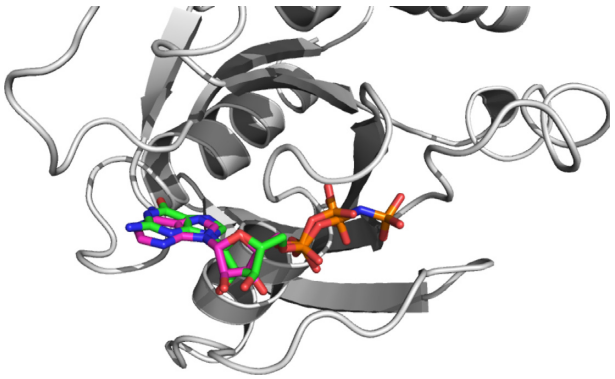


Figure 3. The position of ADP, in sticks with C atoms in magenta, proposed by the PatchSearch web server and the experimental position of GNP, in sticks with C atoms in green, in the 4nym structure.

DISCUSSION AND CONCLUSION

Based on a new algorithmic approach, PatchSearch web server allows fast structural comparisons between a binding site and the entire protein surfaces of a user-supplied collection of protein structures. PatchSearch recognizes structural

similarities based on geometry and physicochemical conditions by the matching of equivalent atoms. PatchSearch is able to perform the search on the entire protein surface to identify structural similarities without *a priori* knowledge about the binding sites. The use of quasi-cliques approach allows the detection of structurally flexible binding sites and the detection of similar binding sites having local structural distortions. In addition, PatchSearch web server is fast, the structural similarity calculations on ∼15 000 PDB structures are done in <2.0 h. To this respect, PatchSearch improves clearly over other methods dedicated only to the binding sites comparisons.

The PatchSearch web server benefits from a user-friendly submission and visualization interface. The transposed ligand within the potential binding site can be downloaded in PDB format for offline in-depth analysis. Consequently, the use of PatchSearch web server may be a preliminary step for the discovering of new interactions and hence is a valuable tool for predicting adverse effects, for helping in drug repositioning studies or for modifying a drug in a way that maintains binding to the intended target, but reduces binding to undesired proteins.

PatchSearch will often return a surprisingly high number of hits, in most cases with unknown status. However, it is important to keep in mind that the unwanted ligand interaction on a protein surface might show no effect on the protein behavior, especially if the potential binding site is not implicated in the protein biological activity, like enzymatic reactions or the interactions with partners. The potential off-targets have to be ascertained by further computational or experimental analysis.

Future directions are to extend PatchSearch to larger binding sites, such as protein–protein or protein–peptide binding sites, that have gain in recent years increasing interests.

DATA AVAILABILITY

The server is freely available via a user-friendly web interface at: http://bioserv.rpbs.univ-paris-diderot.fr/services/PatchSearch/

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Agence Nationale de la Recherche ANR-IA-2011-IFB [ANR-11-INSB-0013]. Funding for open access charge: Université Paris Diderot; INSERM [U1133].
Conflict of interest statement. None declared.

REFERENCES

1. Vulpetti,A., Kalliokoski,T. and Milletti,F. (2012) Chemogenomics in drug discovery: computational methods based on the comparison of binding sites. *Future Med. Chem.*, **4**, 1971–1979.
2. Jalencas,X. and Mestres,J. (2013) Identification of similar binding sites to detect distant polypharmacology. *Mol. Inform.*, **32**, 976–990.
3. Ritchie,D.W. and Kemp,G.J.L. (1999) Fast computation, rotation and comparison of low resolution spherical harmonic molecular surfaces. *J. Comput. Chem.*, **20**, 383–395.
4. Morris,R.J., Najmanovich,R.J., Kahraman,A. and Thornton,J.M. (2005) Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics*, **21**, 2347–2355.
5. Hoffmann,B., Zaslavskiy,M., Vert,J.-P. and Stoven,V. (2010) A new protein binding pocket similarity measure based on comparison of clouds of atoms in 3D: application to ligand prediction. *BMC Bioinformatics*, **11**, 99.
6. Sael,L. and Kihara,D. (2010) Binding ligand prediction for proteins using partial matching of local surface patches. *Int. J. Mol. Sci.*, **11**, 5009–5026.
7. Sael,L. and Kihara,D. (2012) Detecting local ligand-binding site similarity in nonhomologous proteins by surface patch comparison. *Proteins*, **80**, 1177–1195.
8. Yeturu,K. and Chandra,N. (2008) PocketMatch: a new algorithm to compare binding sites in protein structures. *BMC Bioinformatics*, **9**, 543.
9. Liu,T. and Altman,R.B. (2011) Using multiple microenvironments to find similar ligand-binding sites: application to kinase inhibitor binding. *PLoS Comput. Biol.*, **7**, e1002326.
10. Wallace,A.C., Borkakoti,N. and Thornton,J.M. (1997) TESS: a geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases. Application to enzyme active sites. *Protein Sci.*, **6**, 2308–2323.
11. Gold,N.D. and Jackson,R.M. (2006) Fold independent structural comparisons of protein-ligand binding sites for exploring functional relationships. *J. Mol. Biol.*, **355**, 1112–1124.
12. Shulman-Peleg,A., Mintz,S., Nussinov,R. and Wolfson,H.J. (2004) Protein-protein interfaces: recognition of similar spatial and chemical organizations. In: *Algorithms in Bioinformatics: 4th International Workshop, WABI 2004, Bergen, Norway, 2004*. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, Vol. **3240**, pp. 194–205.
13. Shulman-Peleg,A., Nussinov,R. and Wolfson,H.J. (2004) Recognition of functional sites in protein structures. *J. Mol. Biol.*, **339**, 607–633.
14. Shatsky,M., Shulman-Peleg,A., Nussinov,R. and Wolfson,H.J. (2006) The multiple common point set problem and its application to molecule binding pattern detection. *J. Comput. Biol.*, **13**, 407–428.
15. Shulman-Peleg,A., Shatsky,M., Nussinov,R. and Wolfson,H.J. (2008) MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. *Nucleic Acids Res.*, **36**, W260–W264.
16. Cheng,S., Zhang,Y. and Brooks,C.L. (2015) PCalign: a method to quantify physicochemical similarity of protein-protein interfaces. *BMC Bioinformatics*, **16**, 33.
17. Brylinski,M. (2014) eMatchSite: sequence order-independent structure alignments of ligand binding pockets in protein models. *PLoS Comput. Biol.*, **10**, e1003829.
18. Pu,L., Govindaraj,R.G., Lemoine,J.M., Wu,H.-C. and Brylinski,M. (2019) DeepDrug3D: classification of ligand-binding pockets in proteins with a convolutional neural network. *PLoS Comput. Biol.*, **15**, e1006718.
19. Ullmann,J.R. (1976) An algorithm for subgraph isomorphism. *J. Assoc. Comput. Mach.*, **23**, 31–42.
20. Bron,C. and Kerbosch,J. (1973) Algorithm 457: finding all cliques of an undirected graph. *Commun. ACM*, **26**, 48–50.
21. Gardiner,E.J., Willett,P. and Artymiuk,P.J. (2000) Graph-theoretic techniques for macromolecular docking. *J. Chem. Inf. Comput. Sci.*, **40**, 273–279.
22. Cazals,F. and Karande,C. (2008) A note on the problem of reporting maximal cliques. *Theor. Comput. Sci.*, **407**, 564–568.
23. Konc,J. and Janezic,D. (2007) An improved branch and bound algorithm for the maximum clique problem. *MATCH Commun. Math. Comput. Chem.*, **58**, 569–590.
24. Grindley,H.M., Artymiuk,P.J., Rice,D.W. and Willett,P. (1993) Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol.*, **229**, 707–721.
25. Gardiner,E.J., Artymiuk,P.J. and Willett,P. (1997) Clique-detection algorithms for matching three-dimensional molecular structures. *J. Mol. Graph. Model.*, **15**, 245–253.
26. Schmitt,S., Kuhn,D. and Klebe,G. (2002) A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.*, **323**, 387–406.

27. Kinoshita,K., Furui,J. and Nakamura,H. (2002) Identification of protein functions from a molecular surface database, eF-site. *J. Struct. Funct. Genomics*, **2**, 9–22.
28. Jambon,M., Imberty,A., Deléage,G. and Geourjon,C. (2003) A new bioinformatic approach to detect common 3D sites in protein structures. *Proteins*, **52**, 137–145.
29. Konc,J. and Janezic,D. (2010) ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics*, **26**, 1160–1168.
30. Wang,J.C., Chu,P.Y., Chen,C.M. and Lin,J.H. (2012) idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach. *Nucleic Acids Res.*, **40**, W393–W399.
31. Rasolohery,I., Moroy,G. and Guyon,F. (2017) PatchSearch: a fast computational method for off-target detection. *J. Chem. Inf. Model.*, **57**, 769–777.
32. Girvan,M. and Newman,M.E. (2002) Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7821–7826.
33. Clauset,A. (2005) Finding local community structure in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **72**, 026132.
34. Blondel,V.D., Guillaume,J., Lambiotte,R. and Lefebvre,E. (2008) Fast unfolding of communities in large networks. *J Stat Mech*, P10008.
35. Kahraman,A., Morris,R.J., Laskowski,R.A. and Thornton,J.M. (2007) Shape variation in protein binding pockets and their ligands. *J. Mol. Biol.*, **368**, 283–301.
36. Hoffmann,B., Zaslavskiy,M., Vert,J.P. and Stoven,V. (2010) A new protein binding pocket similarity measure based on comparison of clouds of atoms in 3D: application to ligand prediction. *BMC Bioinformatics*, **11**, 99.
37. Gunasekaran,K and Nussinov,R. (2007) How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.*, **365**, 257–273.
38. Chen,C., He,Y., Wu,J. and Zhou,J. (2015) Creation of a free, Internet-accessible database: the multiple target ligand database. *J. Cheminform.*, **7**, 14.
39. Milletti,F. and Vulpetti,A. (2010) Predicting polypharmacology by binding site similarity: from kinases to the protein universe. *J. Chem. Inf. Model.*, **50**, 1418–1431.
40. Schomburg,K.T. and Rarey,M. (2014) Benchmark data sets for structure-based computational target prediction. *J. Chem. Inf. Model.*, **54**, 2261–2274
41. Koes,D.R., Baumgartner,M.P. and Camacho,C.J. (2013) Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *J. Chem. Inf. Model.*, **53**, 1893–1904.
42. Quiroga,R. and Villarreal,M.A. (2016) Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PLoS One*, **11**, e0155183.
43. Guyon,F. and Tufféry,P. (2014) Fast protein fragment similarity scoring using a Binet-Cauchy kernel. *Bioinformatics*, **30**, 784–791.
44. Néron,B., Ménager,H., Maufrais,C., Joly,N., Maupetit,J., Letort,S., Carrere,S., Tuffery,P. and Letondal,C. (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics*, **25**, 3005–3011.
45. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.
46. Govindaraj,R.G. and Brylinski,M. (2018) Comparative assessment of strategies to identify similar ligand-binding pockets in proteins. *BMC Bioinformatics*, **19**, 91.
47. Zhang,Y. and Skolnick,J. (2004) Scoring function for automated assessment of protein structure template quality. *Proteins*, **57**, 702–710.