MolAxis: a server for identification of channels in macromolecules

Eitan Yaffe¹, Dan Fishelovitch², Haim J. Wolfson¹, Dan Halperin^{1,}* and Ruth Nussinov^{2,3}

¹School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, ²Department of Human Genetics, Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel and ³SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, NCI – Frederick, Bldg 469, Rm 151, Frederick, MD 21702, USA

Received February 11, 2008; Revised April 3, 2008; Accepted April 10, 2008

ABSTRACT

MolAxis is a freely available, easy-to-use web server for identification of channels that connect buried cavities to the outside of macromolecules and for transmembrane (TM) channels in proteins. Biological channels are essential for physiological processes such as electrolyte and metabolite transport across membranes and enzyme catalysis, and can play a role in substrate specificity. Motivated by the importance of channel identification in macromolecules, we developed the MolAxis server. MolAxis implements state-of-the-art, accurate computational-geometry techniques that reduce the dimensions of the channel finding problem, rendering the algorithm extremely efficient. Given a protein or nucleic acid structure in the PDB format, the server outputs all possible channels that connect buried cavities to the outside of the protein or points to the main channel in TM proteins. For each channel, the gating residues and the narrowest radius termed 'bottleneck' are also given along with a full list of the lining residues and the channel surface in a 3D graphical representation. The users can manipulate advanced parameters and direct the channel search according to their needs. MolAxis is available as a web server or as a stand-alone program at [http://bioinfo3d.cs.tau.ac.il/MolAxis.](http://bioinfo3d.cs.tau.ac.il/MolAxis)

INTRODUCTION

Channels in proteins are putative sites for binding, conducting ions, small molecules, nucleic acids, peptides and water. In enzymes, substrate specificity is defined not only by interactions at the binding site but also by the selectivity of the substrate by the pathways leading to the active site. This selectivity is a function of the pathway

geometry and chemical properties. Predicting and better understanding of pathways in macromolecules are critical issues in biology and chemistry, and in rational drug design. The major goal in channel finding is to detect all possible distinct channels, their dimensions including their length, narrowest radius (termed bottleneck) and gating residues taking into account the global geometry of the macromolecule. Most of the earlier algorithms (1–8) were developed to find functional sites such as binding sites on the surface of a protein or in buried areas of proteins. However, the *pathway* to the active site was neglected. Voids and chamber finding algorithms are rather limited in their ability to detect the pathways of ligands to the active site and in finding transmembrane (TM) channels. The first program for the search of holes inside macromolecules is HOLE (9). HOLE finds and displays the pore dimensions of ion channels but is limited to TM channels and does not find channels emanating from cavities. The first program designed to explore routes from protein clefts and cavities is CAVER (10). In CAVER the protein is mapped onto a 3D grid. Each cell is weighted such that the lowest weighted cells are surrounded by empty space. The search algorithm detects the lowest weighted cells and finds the lowest cost paths from a user-specified starting point to the surface of the protein. A recently developed tool, called MOLE (11), replaces the large number of grid vertices of CAVER by smaller number of vertices, which are located on the Voronoi diagram of the centers of the atoms. This renders the channel search more efficient. Medek et al. (8) implemented an algorithm based on the computation of the Voronoi diagram of the atom centers. Their algorithm is similar to the MOLE algorithm, yet it is not accessible via a web server. To the best of our knowledge the only available servers for channel identification in macromolecules are those of CAVER and MOLE. The MolAxis algorithm was found to be considerably more efficient than the CAVER algorithm with running times differences of several orders of magnitude. On the other hand, MOLE and MolAxis

*To whom correspondence should be addressed. Tel: $+972-3-6406478$; Fax: $+972-3-6405387$; Email: danha@post.tau.ac.il Correspondence may also be addressed to Ruth Nussinov. Tel: 301-846-5579; Fax: 301-846-5598; Email: ruthn@ncifcrf.gov

2008 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License [\(http://creativecommons.org/licenses/](http://creativecommons.org/licenses/) by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

File	Source Point	Resolution (A)	Split Distance (A)	Bounding Sphere Radius Max Radius (A)	(A)	Include Hetatoms	Results
1mwvA.pdb	$-8.01 - 61.76 -$ 34.99	0.5	$\mathbf{1}$	50	6	HEM	Linux Windows
Channel	Bottleneck Radius (A)	1.51	Bottleneck Amino Acids 141: A 1500: A 324: A	Split (A) 0.00			
Channel_1							
Channel_2		0.65	108: A 1500: A	0.00			
⊽ $ \nabla$ ⊽ Channel 3		0.54	239: A 240: A 388: A	0.00			
⊽ Channel_4 ⊽ Channel 5	0.62 0.65		254: A 391: A	0.71 0.95			

Figure 1. The MolAxis chamber mode output page results. The results page contains a header table, channels table (left) and a Jmol viewer (right). The channel table presents the data relating to all detected channels emanating from the active site of a bacterial catalase–peroxidase. Each row in the table contains information about a channel including the bottleneck radius, bottleneck residues along with their chain ID and split radii of each channel relative to the starting point. The enzyme is colored gray and represented by ribbons. The active site heme is colored orange and represented as space-fills. Each channel is colored with respect to the channel table colors.

manifest similar running times; however, they differ in their performance: MOLE outputs a partially redundant list of channels that emanate from a chamber. It attempts to solve the redundancy problem by clustering of channels. In addition, for TM proteins MOLE outputs several channels, some of which are not biologically relevant. In contrast, MolAxis (12) permits the user to conduct searches for channels emanating from voids and to detect TM channels using a single server. All detected channels are geometrically distinct with no need for clustering analysis. Since biological systems could have different topologies, the server provides several intuitive optional parameters that allow more flexibility in running MolAxis. This ability to adapt the channel search to a specific topology renders it more efficient and sensitive and yields a higher quality output. In both search types, the MolAxis server enables the user to better control the output by changing the parameters of the channel search, adapting the channel detection to a given topology. The MolAxis server was tested on a diverse dataset of enzyme structures including several Cytochrome P450 isoforms, haloalkane-dehalogenase, trans-aldolase, catalase–peroxidase (Figure 1), hydrolase, lipase and many more enzymes with buried active sites as well as on ion channels, transporters and receptors (Supplementary data). MolAxis can run on proteins as well as on nucleic acid structures. MolAxis is very efficient and can handle large structural files such as the large ribosomal subunit (Supplementary data). It has a simple

BETWEEN MODEL CARDS

interface and uses the Jmol visualization tool (www.jmol. org). All tests were carried out on a Pentium IV 3.0 GHz machine with 1 GB of RAM running a LINUX native operating system and the running times span between 5 and 30 seconds in most cases depending on the initial structure dimensions and topology.

Here, we describe the web server for channel identification in macromolecules. Details related to the algorithm theory and its application to macromolecular structures are provided in the server 'About MolAxis' page and elsewhere (12).

MolAxis: A CHANNEL FINDING ALGORITHM BASED ON COMPUTATIONAL-GEOMETRY **TECHNIQUES**

Computational-geometry techniques were developed to represent and analyze molecular structures. The α -shapes theory (13) was used to describe the topological and geometrical features of molecules, including measuring the surface area and volume of pockets (14). However, it is rather difficult to directly use the α -complex of a molecule to describe features of channels such as its spine or diameter. MolAxis is based on the α -shapes theory and on a geometric concept called the medial axis (MA). The MA of a general surface is the collection of 3D points that have more than one closest point on the surface. Here, the surface is the van der Waals surface of a molecule. We represent molecular channels using corridors. A corridor is

a probable route taken by a small molecule passing through a channel. MolAxis uses a novel algorithm that allows fast identification of corridors in the complement of the molecule. It approximates a useful subset of the MA of the complement of the molecule. We convert a 3D problem to a 2D problem, which improves dramatically the performance of the algorithm. MolAxis can automatically compute a source point in the center of the main void with a high success rate or allow a user-specified source point. A complete description of the theory and algorithm behind the MolAxis server can be found in the pdf files at the bottom of the 'About MolAxis' and the 'web server' page. In brief, the MolAxis algorithm operates in four steps: (i) representation of the input molecule using a collection of fixed-sized balls; (ii) based on the Voronoi diagram of the centers of the fixed-sized balls, we construct an approximation of the MA of the complement of the molecule; (iii) computation of a minimal weight tree from a user-specified point or from an automatically computed starting point, using the Dijkstra's shortest-path algorithm (15). We avoid reporting duplicate channels by using a split distance parameter that bounds the distance to the least common ancestor of the channel thereby not reporting previously detected channels and (iv) scoring the constructed corridors according to the total weight of the edges in the graph. The output includes a ranked list of the protein channels according to the 'flux score', that weighs in the length and width of the channel and favors channels that are relatively short and wide. In TM channels, the source point is defined to be at infinity. In that case, the reported channel is a concatenation of two paths in the corridor tree, which pass through a user-defined sphere and reach the bounding sphere of the entire molecule from two opposite directions.

MolAxis: INPUT, OUTPUT AND USER INTERFACE

MolAxis is a free web server with a user-friendly interface, which contains seven pages: (i) *about page* that describes the algorithm behind MolAxis with a complete theoretical background at the bottom of the page; (ii) web server page that is the main page for submitting queries. When submission is complete the user is redirected to a *results* page. This page is automatically refreshed every 5 seconds until the calculation completes; (iii) *download page* that contains a downloadable, stand-alone version for Linux operating system; (iv) help page that describes the parameters that can be modified by the user in the web server as well as in the stand-alone version and the output created by MolAxis; (v) FAQ page that presents questions and answers; (vi) links page that contains links to related web sites and (vii) *tutorial page* that describes how to run the MolAxis stand-alone version on several examples.

INPUT

MolAxis distinguishes between (i) channels that emanate from an inner chamber and (ii) TM channels. Therefore,

there are two separate forms for the two channel types: chamber and transmembrane. Both forms require a single input structure in a Protein Data Bank (PDB) format (16). The PDB file can be uploaded to the server by using the browse button or can be automatically retrieved by the server from the PDB by entering a PDB code and chain ID. An optional e-mail address field exists and if filled, the link to the results page is sent to the specified address. Otherwise the link appears at the bottom of the results page.

Channels out of chambers input form

In this channel type, the PDB file is the only required input. An optional Source type parameter that sets the starting point (Auto void mode–default) can be changed to a user-defined starting point. In this case, Cartesian coordinates need to be added in the User source sphere field for the starting point, following a positive value for the radius of the source sphere (default $1 \AA$).

Advanced optional parameters

There are ten advanced optional fields in the channel search main form: (i) Resolution: the resolution of the channel search (default 0.5 Å); (ii) Split distance: above that distance pathways splitting are ignored (default 4\AA); (iii) Bounding sphere radius: maximal channel length (default 30 Å). This parameter is the distance from the starting point to the last point along the search; (iv) End channel at radius: limits the display of channel spheres to a radius under this value (default 4 Å); (v) *Include hetatoms* is a list of all hetero atoms that are to be included in the search. By default all hetero atoms are ignored; (vi) Include hydrogens: use this field to take into account hydrogen atoms (default NO). This parameter is useful for analyzing molecular dynamics simulations snapshots; (vii) Mesh quality controls the visualization quality of the Jmol viewer (default Normal); (viii) Via sphere: Cartesian coordinates of the center of a sphere and its radius: when addressed, MolAxis reports channels that pass through this sphere. This reduces the search to a specified channel in multi channel systems; (ix) Radii table: MolAxis uses a default radii table. To use another set of atom radii upload a file with the browse button, in the format of the given template files and (x) Probe radius: the user can change the radius of the probe sphere of the channel search (default 1.4 Å).

TM channels input form

When the user chooses to search a TM channel he is redirected to a form similar to that of the chamber search. The advanced parameters are the same as in chamber channels with the exception of Split distance that is omitted and Via sphere that is an obligatory parameter. In addition, the user needs to fill in the Channel vector parameter which is a vector representing the general direction of the channel or the main axis of the channel. Given the input macromolecular topological feature diversity in both channel types (out of chambers and TM channels), we suggest inspection of the results and manually changing the parameters if needed.

OUTPUT

Channels out of chambers output form

In chamber channel type with multiple channels (default), the results page contains a header table (top), channels table (left) and a Jmol viewer applet window on the right (Figure 1). The header table presents all the parameter values assigned for the run and the last column (right) contains a link for downloading the results files in Linux or Windows format. The channel table presents the data for all detected channels. Each row in the table contains information of a channel including its bottleneck radius, bottleneck residues along with their chain ID and split radius of the channel relative to the starting point. By clicking on the channel checkbox in the channels table, the user can observe the channel surface in the Jmol viewer window. Upon clicking on the channel gating residues list identified by MolAxis, they will be shown as spheres in Jmol. The gating residues are located in the channel table and denoted as bottleneck amino acids. The user can hide the bottleneck amino acids by pressing the 'Hide residues' button. The user can also view the pathway tree of all the detected channels by clicking the checkbox 'Toggle all spines'.

There are five different file types (all are text files) that can be downloaded: (i) data.txt is a file with all the parameter values; (ii) [PDB code].stats jmol.txt is a file with running time statistics of a given run; (iii) [PDB_code].graph_pathway_X.txt is a file that contains the trajectory of the detected channel. The first column contains the channel radius along the channel path at the distance shown in the second column relative to the beginning of the channel path. The next columns contain information about the closest atoms that contact the channel. This information includes how many atoms contact the channel at the distance shown in the second column, their line numbers in the PDB file and to which residue they belong; (iv) [PDB code].graph pathway X all.txt is a file that contains the Cartesian coordinates of the MA of the channel along with the respective radii along the axis (fourth column) and (v) [PDB_code].graph_pathway_X_bottle.txt is a file that records the bottleneck radius, the distance of the center of the corresponding sphere from the beginning of the channel and information about the closest atoms that contact the bottleneck sphere including their line numbers and to which amino acid they belong.

Figure 1 presents all channels emanating from the heme containing active site of the bacterial catalase–peroxidase (KatG) of Burkholderia Pseudomallei (PDB code 1MWV:A). Bacterial KatGs are bi-functional enzymes that disintegrate hydrogen peroxide by a catalase activity $[2H_2O_2 \rightarrow 2H_2O + O_2]$ or by a peroxidase activity $[H_2O_2 + 2AH \rightarrow 2H_2O + 2A]$ (17). Their function is believed to reduce reactive oxygen species as a response to oxidative stress arising from metabolic generation or environmental factors. The top ranked channel found by MolAxis (colored blue in Figure 1) is the main access channel of catalase–peroxidases (18). Residues D141 and S324 that constitute the narrowest part of the main channel, were shown experimentally to be the gating residues in corresponding residues of other bacterial catalase–peroxidase isoforms (19,20).

TM channels output form

In TM mode the result page is very similar to that of chamber mode with the exception of the channels table. This table has four columns. The first column is the serial number of a given point along the TM channel. The second column is the distance of that point from the beginning of the TM channel. The third column is the radius value of the TM channel at the given distance of the second column and the fourth column contains the amino acids that border the sphere and can be viewed when clicked. The downloadable files are similar to those of the chamber mode: [PDB_code].graph_main.txt that is similar to [PDB_code].graph_pathway_X.txt, [PDB_code]. graph main all.txt that is similar to [PDB code]. graph_pathway_X_all.txt, [PDB_code].graph_main_ bottle.txt is similar to [PDB code].graph pathway X bottle.txt and a statistics file of running times.

Figure 2 shows the main TM channel of the nicotinic acetylcholine (ACh) receptor (PDB code 2BG9). This receptor is located at the synapse between nerve and muscle cells. When ACh binds the receptor, the receptor conformation changes, opening the channel. This allows positively charged ions to cross the membrane and initiate muscle constriction. The receptor has three major domains including an extra cellular domain that consists of the ACh binding pocket, a TM domain containing TM helices that traverse the cytoplasmatic membrane and a cytoplasmatic domain (21). We focus on the TM domain in Figure 2 that is believed to be located at the middle portion of the membrane spanning pore (21). (The channel surface is given in the Supplementary file.) Among the lining residues of the TM channel we detected six leucines (colored red in Figure 2) that gate the channel and are located around the middle of the TM domain. These leucines include L257:B, L265:B, L265:C, L273:C, L259:E and L267:E (Figure 2). Based on mutagenesis studies, it was suggested that these conserved leucines are involved in the gating of the ACh receptor (22,23) and MolAxis identifies these leucines.

CONCLUSIONS

The MolAxis server was found to be sensitive, accurate and very efficient in locating channels in macromolecules in a variety of biological systems. The server efficiently detects substrate and water channels leading from deep active sites to the protein surface, even if they are almost closed; and it further detects the main TM channel in TM proteins. It identifies distinct channels with no redundancy and can be applied to very large systems, including proteins and nucleic acids. The user can manipulate the channel search such that it will better fit the specified biological input. We hope that the server will be useful to the biological and chemical communities, assisting in the comprehension of gating mechanisms and substrate selectivity of channels as well as in drug design.

Accurate Identification of Channels in Macromolecules [About MolAxis] [Web Server] [Download] [Help] [FAQ] [Links] [Tutorial] Contact: ppdock@post.tau.ac.il

42	66.75	3.16	273:C	
43	69.52	3.79	259:D 267:E	
44	70.22	3.85		
45	70.44	3.80	255:D 265:B	
46	71.22	3.57		
47	72.37	3.28		
48	74.57	2.77		
49	74.68	2.77	269:C	
50	74.88	2.77	261:B	
51	75.96	2.89		
52	78.31	3.71	263:E	
53	78.86	3.76	258: B	
54	79.13	3.76	265:C	
55	80.30	3.82	257: B	
56	81.83	4.20		
57	82.71	4.09		
58	83.42	4.00	259: E	
59	84.37	3.77		
60	86.39	3.38	262.0	

Figure 2. The MolAxis TM mode output page results. The results page contains a header table (not shown), channels table (left) and a Jmol viewer (right). The columns from the left are: a serial number of a given point along the channel, the distance of that point from the beginning of the channel, the radius value of the TM channel at the specified distance and amino acids bordering the sphere, along with their chain ID, that can be viewed when clicking on them. This example shows the ACh receptor main TM channel. The TM domain of the receptor is represented as trace and colored gray. The main channel surface is colored blue and six conserved and gating leucines are colored red and represented as space-fills.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to Dina Schneidman-Duhovny for very helpful discussions on setting up the server. This work has been supported in part by the IST Programme of the EU as Shared-cost RTD (FET Open) Project under Contract No IST-006413 (ACS - Algorithms for Complex Shapes), by the Israel Science Foundation (grant no. 236/ 06) and by the Hermann Minkowski—Minerva Center for Geometry at Tel Aviv University. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under contract number N01-CO-12400. The content of this publication neither does necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations that imply endorsement by the U.S. Government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. Funding to pay the Open Access publication charges for this article was provided by SAIC-Frederick contract number N01-CO-12400.

Conflict of interest statement. None declared.

REFERENCES

1. Levitt,D.G. and Banaszak,L.J. (1992) POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. J. Mol. Graph., 10, 229–234.

- 2. Kleywegt,G.J. and Jones,T.A. (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. Acta. Crystallogr. D, 50, 178–185.
- 3. Laskowski,R.A. (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J. Mol. Graph., 13, 323-330.
- 4. Hendlich,M., Rippmann,F. and Barnickel,G. (1997) LIGSITE: automatic and efficient detection of potential small moleculebinding sites in proteins. J. Mol. Graph. Model., 15, 359–363.
- 5. Liang,J., Edelsbrunner,H. and Woodward,C. (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci., 7, 1884–1897.
- 6. Venkatachalam,C.M., Jiang,X., Oldfield,T. and Waldman,M. (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J. Mol. Graph. Model., 21, 289–307.
- 7. Laurie,A.T.R. and Jackson,R.M. (2005) Q-SiteFinder: an energybased method for the prediction of protein–ligand binding sites. Bioinformatics, 21, 1908-1916.
- 8. Medek,P., Benes,P. and Sochor,J. (2007) Computation of tunnels in protein molecules using Delaunay triangulation. J. WSCG07, 1, 107–114.
- 9. Smart,O.S., Goodfellow,J.M. and Wallace,B.A. (1993) The pore dimensions of gramicidin A. Biophys. J., 65, 2455-2460.
- 10. Petrek,M., Otyepka,M., Banas,P., Kosinova,P., Koca,J. et al. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinformatics, 7, 316–324.
- 11. Petrek, M., Kosinová, P., Koca, J. and Otyepka, M. (2007) MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. Structure, 15, 1357-1363.
- 12. Yaffe,E., Fishelovitch,D., Wolfson,H.J., Halperin,D. and Nussinov,R. (2008) MolAxis: efficient and accurate identification of channels in macromolecules. Proteins: Structure, Function, and Bioinformatics, doi:10.1002/prot.22052.
- 13. Edelsbrunner, H. and Mücke, E.P. (1994) Three-dimensional alpha shapes. ACM Trans. Graph., 13, 43-72.
- 14. Edelsbrunner,H., Facello,M.A. and Liang,J. (1998) On the definition and the construction of pockets in macromolecules. Discrete Appl. Math., 88, 83-102.
- 15. Dijkstra,E.W. (1959) A note on two problems in connexion with graphs. Numerische Math., 1, 269–271.
- 16. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank. A computer-based archival file for macromolecular structures. Eur. J. Biochem., 80, 319–324.
- 17. Donald,L.J., Krokhin,O.V., Duckworth,H.W., Wiseman,B., Deemagarn,T., Singh,R., Switala,J., Carpena,X., Fita,I. and Loewen,P.C. (2003) Characterization of the Catalase-Peroxidase KatG from Burkholderia pseudomallei by Mass Spectrometry. J. Biol. Chem., 278, 35687–35692.
- 18. Jakopitsch, C., Droghetti, E., Schmuckenschlager, F., Furtmüller, P.G., Smulevich,G. and Obinger,C. (2005) Role of the main access channel of catalase-peroxidase in catalysis. J. Biol. Chem., 280, 42411–42422.
- 19. Jakopitsch,C., Auer,M., Regelsberger,G., Jantschko,W., Furtmüller, P.G., Rüker, F. and Obinger, C. (2003) Distal site

aspartate is essential in the catalase activity of catalase-peroxidases. Biochemistry, 42, 5292–5300.

- 20. Yu,S., Girotto,S., Lee,C. and Magliozzo,R.S. (2003) Reduced affinity for isoniazid in the S315T mutant of mycobacterium tuberculosis KatG is a key factor in antibiotic resistance. J. Biol. Chem., 278, 14769–14775.
- 21. Miyazawa,A., Fujiyoshi,Y., Stowell,M. and Unwin,N. (1999) Nicotinic acetylcholine receptor at 4.6 Å resolution: transverse tunnels in the channel. J. Mol. Biol., 288, 765-786.
- 22. Unwin, N. (1993) Nicotinic acetylcholine receptor at 9 \AA resolution. J. Mol. Biol., 229, 1101–1124.
- 23. Filatov,G.N. and White,M.M. (1995) The role of conserved leucines in the M2 domain of the acetylcholine receptor in channel gating. Mol. Pharmacol., 48, 379–384.