Biotechnology Advances xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Software tools for identification, visualization and analysis of protein tunnels and channels

Jan Brezovsky [a], Eva Chovancova [a], Artur Gora [a], Antonin Pavelka [a,b,c], Lada Biedermannova [a,1], Jiri Damborsky [a,c,*]

[a] Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic
[b] Human Computer Interaction Laboratory, Faculty of Informatics, Masaryk University, Botanicka 68a, 602 00 Brno, Czech Republic
[c] International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic

ARTICLE INFO

Article history:
Received 13 July 2011
Received in revised form 17 January 2012
Accepted 2 February 2012
Available online xxxx

Keywords:
Tunnel
Channel
Pore
Pathway
Ligand binding
Ligand transport
Substrate entry
Product release
Protein design
Protein engineering

ABSTRACT

Protein structures contain highly complex systems of voids, making up specific features such as surface clefts or grooves, pockets, protrusions, cavities, pores or channels, and tunnels. Many of them are essential for the migration of solvents, ions and small molecules through proteins, and their binding to the functional sites. Analysis of these structural features is very important for understanding of structure–function relationships, for the design of potential inhibitors or proteins with improved functional properties. Here we critically review existing software tools specialized in rapid identification, visualization, analysis and design of protein tunnels and channels. The strengths and weaknesses of individual tools are reported together with examples of their applications for the analysis and engineering of various biological systems. This review can assist users with selecting a proper software tool for study of their biological problem as well as highlighting possible avenues for further development of existing tools. Development of novel descriptors representing not only geometry, but also electrostatics, hydrophobicity or dynamics, is needed for reliable identification of biologically relevant tunnels and channels.

Contents

* Corresponding author at: Loschmidt Laboratories, Department of Experimental Biology, Faculty of Science, Masaryk University, Kamenice 5/A4, 625 00 Brno, Czech Republic. Tel.: +420 549493467; fax: +420 549496302.
E-mail address: jiri@chemi.muni.cz (J. Damborsky).
[1] Present address: Laboratory of Biomolecular Recognition, Institute of Biotechnology AS CR, Videnska 1083, 142 20 Prague 4, Czech Republic.

0734-9750/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.biotechadv.2012.02.002

1. Introduction

The three dimensional structure of a protein is composed of tightly packed atoms (Richards, 1977). The packing is not perfect, and gives rise to the highly complex system of voids containing many specific features such as surface clefts or grooves, pockets, protrusions, cavities, channels or pores, and tunnels (Fig. 1). These features play an essential role in many biological processes, since they represent sites of interactions with ligands, ions, solvent molecules and other proteins or nucleic acids (Janin and Chothia, 1990; Rhodes and Burley, 2000; Zvelebil and Thornton, 1993). Channels through transmembrane proteins enable the selective transport of ions, small molecules and even large macromolecules across biological membranes (Bezanilla, 2008; Gold et al., 2007). The channel selectivity is determined by its geometry and composition of amino acid residues (Gouaux and Mackinnon, 2005). Channel dysfunctions have been related to various human diseases (Niemeyer et al., 2001). Protein cavities, clefts, grooves and pockets provide a suitable micro-environment for biochemical reactions. In many enzymes, the cavities are buried deeply inside the protein core, shielded from water molecules and accessible from the bulk solvent via tunnels (Petrek et al., 2006). The enzyme specificity is determined not only by protein-substrate interactions in the active site, but is also influenced by the selectivity of these access tunnels (Chaloupková et al., 2003). Similarly to the channels, the selectivity of the tunnels is governed by the size, shape, and physico-chemical properties of the residues lining it.

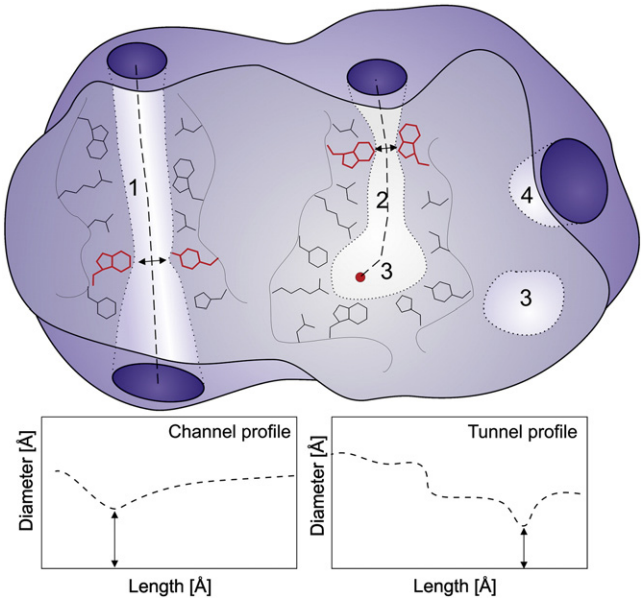

Fig. 1. Schematic representation of a protein with a channel/pore (1), a tunnel (2), a buried cavity (3) and a surface cleft/groove (4). Void interior is light purple, channel and tunnel entrance is dark purple, channel- and tunnel-lining residues are in black and bottleneck residues are in red. Arrows indicate a bottleneck; channel and tunnel profiles are shown at the bottom and correspond to the dashed line. The red dot in the buried cavity represents the starting point for the tunnel calculation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The importance of pockets and cavities for protein function has been recognized for a long time, therefore, many tools for their identification and analysis have been developed, e.g., POCKET (Levitt and Banaszak, 1992), VOIDOO (Kleywegt and Jones, 1994), SURFNET (Laskowski, 1995), LIGSITE (Hendlich et al., 1997), CAST (Liang et al., 1998), PASS (Brady and Stouten, 2000), Q-SiteFinder (Laurie and Jackson, 2005), LigandFit (Venkatachalam et al., 2003), CASTp (Binkowski et al., 2003), SITEHOUND (Ghersi and Sanchez, 2009), SplitPocket (Tseng et al., 2009), Fpocket (Le Guilloux et al., 2009), and POCASA (Yu et al., 2010). A detailed discussion of the strengths, weaknesses and usability of some of these tools can be found in the review of Laurie and Jackson (Laurie and Jackson, 2006). These tools were designed to find functional pockets on the protein surface or cavities inside the protein, often overlooking tunnels connecting these cavities with the surrounding solvent. However, to study the ligand migration through the tunnels and channels, the molecular dynamics (MD) simulations can be recommended as a method of choice. Readers interested in the comparison of various MD simulation methods for the investigation of transport processes in proteins should refer to the recent review of Arroyo-Mañez and co-workers (Arroyo-Mañez et al., 2011). Setup, computation and analysis of such MD simulations usually require a lot of time and expertise. Therefore, several approximate methods for a direct study of tunnel and channel geometry have been recently developed (Coleman and Sharp, 2009; Ho and Gruswitz, 2008; Pellegrini-Calace et al., 2009; Petrek et al., 2006; Petrek et al., 2007; Smart et al., 1996; Voss and Gerstein, 2010; Yaffe et al., 2008a), representing another possibility to analyze ligand pathways in proteins. This review aims to critically compare these geometry-based methods, point out advantages and pitfalls in their usage, provide users with guidance on how to select appropriate tools for a particular problem and outline their possible practical applications.

2. Software tools for analysis of protein tunnels and channels

The current tools for analysis of protein tunnels and channels (Table 1) can be divided according to their aim to: (i) study of tunnels connecting buried cavities with bulk solvents—CAVER (Petrek et al., 2006), MOLE (Petrek et al., 2007) and MOLAXIS (Yaffe et al., 2008a); (ii) study of channels or pores penetrating throughout the proteins—HOLE (Smart et al., 1996), MOLAXIS (Yaffe et al., 2008a), CHUNNEL (Coleman and Sharp, 2009) and POREWALKER (Pellegrini-Calace et al., 2009); and (iii) analysis of overall voids in proteins—HOLLOW (Ho and Gruswitz, 2008) and 3V (Voss and Gerstein, 2010).

2.1. Software tools specialized in analysis of protein tunnels

These tools calculate tunnels connecting occluded protein cavities with the surrounding bulk solvent. Since finding a pathway from the protein surface to the internal cavity is not an easy task, all current tools are designed to solve the inverse definition of this problem—tracing the pathway from the internal cavity to the protein surface (Petrek et al., 2006, 2007; Yaffe et al., 2008a). All these methods employ Voronoi diagram (Fig. 2A) to describe a skeleton of voids within the protein structure. In this way, all possible pathways connecting the starting point in the internal cavity with the bulk solvent are

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

Table 1
Comparison of the software tools for analysis of tunnels and channels in proteins.

Software		CAVER	MOLE	MOLAXIS	HOLE	CHUNNEL	POREWALKER	HOLLOW	3V
Structural feature	Tunnels	+	+	+	−	−	−	+[c]	+[c]
	Channels	+[a]	+[a]	+[b]	+	+	+	+[c]	+[c]
	Overall voids	−	−	−	−	−	−	+	+
Operating system	Linux	+	+	+	+	+	+[d]	+	+
	Windows	+	+	+[d]	−	−	+[d]	+	+
	Mac OS X	+	+	+[d]	−	−	+[d]	+	+
Availability	Web server	+[e]	+	+	−	−	+	−	+
	PYMOL plugin	+	+	−	−	−	−	−	−
	VMD plugin	−	−	+	−	−	−	−	−
	Binary file	+	+	+	+	−	−	−	−
	Source code	−	−	−	+[f]	+	−	+[g]	+
Graphical interface		+	+	+	−	−	+	−	+
Multiple pathways analysis		+	+	+	−	+	−	−	+
Input structure	Single	+	+	+	+	+	+	+	+
	Multiple	+	+	+	+	−	−	−	−
Essential outputs	Pathway profile	+	+	+	+	+	+	−	−
	Pathway residues	+	+	+	−	+	+	−	−
	Bottleneck radius	+	+	+	+	+	+	−	−
Visualization	PYMOL	+	+	−	+	+	+	+	+
	VMD	+	+	+	+	−	+	+	−
	CHIMERA	−	−	−	+	−	+	+	+
	Web applet	+	+	+	−	−	+	−	−

[a] channels can be analyzed by the manual combination of two tunnels; [b] MOLAXIS provides a separate module for channel analysis performing an automatic concatenation of two tunnels, [c] channels and tunnels are identified as an integral part of overall void analysis, [d] web server, [e] beta version as Java Web Start application, [f] source code upon request, [g] source code in interpreted scripting language which can be run without compilation.

described (Aurenhammer, 1991). Subsequently, the Dijkstra's algorithm (Dijkstra, 1959) is used to search this skeleton for optimal tunnels, based on the criteria defined by a cost function.

2.1.1. CAVER

CAVER was the first software tool designed for finding tunnels connecting buried protein cavities with the protein surface, and is still being developed (Petrek et al., 2006). Original CAVER 1.0 performed the search for voids inside the protein structure using the memory and CPU demanding finite grid limiting the calculation to smaller systems (Petrek et al., 2007). The current version of CAVER 3.0 employs Voronoi diagrams, constructed for approximated structure (Fig. 2B), in which individual atoms are replaced by 4, 8, 12 or 20 spheres with a radius of the smallest atom of a given protein (Chovancova et al., unpublished results). Due to this approximation, the Voronoi diagram can be then constructed correctly from spheres of the fixed size. Subsequently, all Voronoi edges narrower than the user-defined minimal width are pruned from the diagram to speed up a tunnel calculation. The user-defined starting point is optimized to prevent its collision with protein atoms. CAVER employs two types of boundary points for tunnel calculations. First, the outer protein surface is derived by rolling the probe of the user-specified radius around the protein, and second, the inner surface of protein is defined by rolling another probe with the radius equal to the user-defined minimal width of tunnels. This second probe cannot depart from the outer protein surface for more than a specified distance. Paths

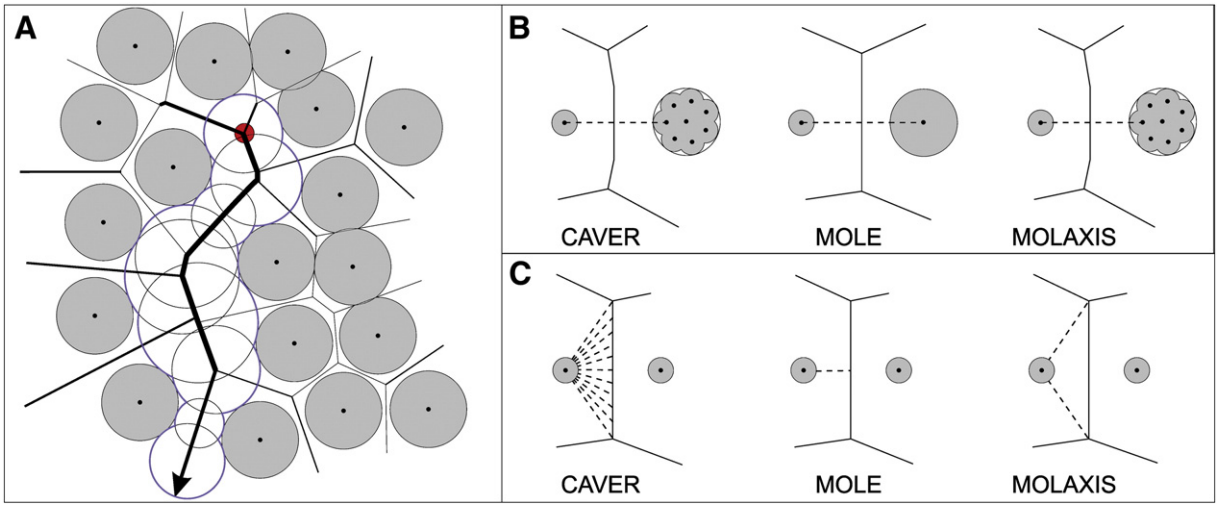

Fig. 2. Algorithms implemented in the tools specialized in analysis of protein tunnels. A) schematic representation of idealized Voronoi diagram: atoms are represented by gray circles, edges by lines, tunnel starting point by a red dot, tunnel exit by arrow, tunnel surface by a blue contour. Thickness of edges corresponds to their width—the thicker edges have lower cost and higher probability of serving as a route). B) Construction of Voronoi edge between two atoms of different size. CAVER and MOLAXIS software approximate large atoms by a set of spheres, each with a radius of the smallest atom of a given protein. The edge is positioned in the middle of two centers of atoms of the same size in an equal distance from the atoms surfaces. In MOLE, the edge is positioned in the middle of two atom centers, not equally distant from the atoms surfaces. C) Determination of edge width. CAVER integrates the distances from the surface of the closest atom to the points regularly sampled on the edge. In MOLE, the distance is measured from the surface of the closest atom to the edge. MOLAXIS averages two distances from the surface of the closest atom to both vertices of the edge.

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

are then identified in the two-step search. In the first step, the best pathways leading from the starting point to all reachable points on the inner surface are identified. All these pathways are then optimally extended up to the protein surface in the second step. Ranking of the tunnels is based on the cost function, which evaluates each Voronoi edge according to its length and distances of points, evenly sampled along a given edge, to their respective closest atom (Fig. 2C). The cost function has an adjustable parameter to control the optimal balance between the width and length of tunnels. All tunnels found in a given protein structure are clustered based on their relative proximity in space by average linkage hierarchical clustering, and for each cluster the best tunnel is retained. CAVER is implemented as a PYMOL (Schrödinger, LLC., USA, http://www.pymol.org/) plugin, stand-alone multiplatform application in Java and as CAVER VIEWER graphical interface (currently beta-version). All implementations require a structure of the target protein in PDB format, a definition of the starting point and specification of the minimal width of tunnels. In the output, plugin and stand-alone version provide information about tunnel parameters, data for plotting tunnel profiles, list of surrounding atoms and residues, and either direct visualization of identified tunnels, as in the case of the plugin and CAVER VIEWER, or PDB files and scripts for visualization of tunnels in PYMOL and VMD (Humphrey et al., 1996). CAVER VIEWER explores Java Web Start technology to deliver visualization capabilities designed for the tunnel analysis, and enables an automatic identification of the starting point from the Catalytic Site Atlas database (Porter et al., 2004). The release version of CAVER VIEWER will also process trajectories from AMBER (Pearlman et al., 1995), GROMACS (Berendsen, 1995) and CHARMM (Brooks et al., 1983) as the input, and will provide a module for a simplified analysis of the protein channels. The stand-alone application is customized for the analysis of MD simulations and besides the essential outputs, it also provides a summary of parameters for individual tunnel clusters.

2.1.2. MOLE

MOLE was the first tool employing Voronoi diagrams for identification of tunnels (Petrek et al., 2007). The Voronoi diagram is constructed without consideration to the variability in radii of individual protein atoms (Fig. 2B). Before the tunnels are searched, the position of the starting point is optimized to prevent its collision with protein atoms, and the convex hull of the protein is calculated to define boundary points for tunnel searching. The MOLE cost function evaluates each edge in the Voronoi diagram by its width, i.e. the distance from the edge to the closest atom (Fig. 2C), and its length. The search is then performed for tunnels with the minimal overall cost. When the optimal tunnel is identified, the cost of searching the same part of the skeleton is increased allowing to locate other tunnels. Each identified tunnel is compared with all other tunnels identified in a given structure. Based on their similarity in the sets of tunnel-lining atoms, the tunnel is either recognized as a new type of a tunnel or assigned as similar to another tunnel and discarded. The MOLE algorithm is implemented as a stand-alone application for Windows, Linux and Mac OS X, plugin for popular visualization program PYMOL and web server. All these implementations require a structure of the target protein in the PDB format, definition of the starting point and specification of the number of tunnels to be found. In the output, MOLE produces information about tunnel parameters, a tunnel profile, a list of surrounding atoms, and PDB files and scripts for visualization of tunnels in PYMOL or VMD. The stand-alone application additionally allows analysis of multiple structures obtained from MD simulations, supplied as AMBER trajectory and topology files. The MOLE ONLINE server provides automatic identification of the starting point, either from coordinates of catalytic residues retrieved from the Catalytic Site Atlas database or as a centre of selected pocket identified by CASTp (Binkowski et al., 2003). The server also provides a user-friendly overview of results, visualization of tunnels and surrounding residues in Jmol (Herráez, 2006), plotted profiles of tunnels and matrix, and dendrogram indicating the similarity of individual tunnels.

2.1.3. MOLAXIS

MOLAXIS was designed for identification of both the tunnels and channels (Yaffe et al., 2008a). The algorithm allows an efficient search for potentially relevant pathways. In the first step, MOLAXIS replaces all atoms in the protein under scrutiny with a collection of spheres, each with a radius equal to the radius of the smallest atom in a given protein (Fig. 2B). The protein atoms are approximated in way guaranteeing the maximal difference between the van der Waals (VDW) surface of the original protein atoms and that of the approximated atoms is smaller than user-defined resolution enabling correct construction of Voronoi diagram. MOLAXIS search for wide and relatively short tunnels using the cost function, which evaluates the length of each Voronoi edge, and the average distance between the two edge vertices and their closest atom (Fig. 2C). In the MOLAXIS module for the analysis of tunnels, the starting point is placed either in the user-specified sphere or in the center of the largest cavity. The tunnel calculation is stopped at points of a bounding sphere, i.e., the sphere of the user-defined radius centered at the starting point. All pathways from the starting point to the bounding sphere are calculated at once, and the tunnels are ranked according to their total cost. Very similar tunnels are discarded using the user-defined forking threshold. The tunnel is reported only if it splits from another tunnel before reaching the forking threshold, i.e., the user-defined distance from the starting point. Otherwise, it is considered as a duplicate of the original tunnel and discarded. The MOLAXIS algorithm is implemented as a stand-alone application for the Linux platform, web server (Yaffe et al., 2008b) and VMD plugin. All implementations require a PDB format structure of the target protein, automatic or manual definition of the starting point and specification of the bounding sphere. The stand-alone application additionally allows defining the starting point as a center of protein mass. The outputs of the stand-alone application are the text file for plotting of the tunnel profile, and data for visualization of tunnels via the VMD plugin. The plugin also allows identification of tunnels in multiple structures. The web server provides list of all found tunnels together with their bottleneck radii, surrounding residues, and split distances from the starting point. Identified tunnels are visualized in Jmol.

2.1.4. Critical comparison

The three tools available for analysis of protein tunnels differ mainly in the: (i) description of the protein by the Voronoi diagram, (ii) definition of the boundary between the protein surface and bulk solvent, (iii) calculation of the cost of individual tunnels, (iv) handling of the multiple tunnels, (v) ability to analyze MD simulations and (vi) provided outputs (Table 2). The analysis of channels by these tools is possible, but less convenient in the case of MOLE and CAVER, since it requires manual assembly of two identified tunnels. To overcome this inconvenience, MOLAXIS provides a special module for analysis of channels.

2.1.4.1. Voronoi diagram. The MOLE algorithm considers the errors caused by using the un-weighted Voronoi diagram on atoms of different size as negligible (Fig. 2B). While this error is small for most of the heavy atoms frequently occurring in biomolecules (Petrek et al., 2007), the error increases up to 2 Å if hydrogen atoms are included in the calculations (Yaffe et al., 2008a). One way to avoid this casual error in MOLE is to remove hydrogen atoms prior to the calculation. The second way adopted by, CAVER and MOLAXIS is to approximate each atom in the protein by a set of spheres with a radius of the smallest atom (Fig. 2B). The error of these tools is thus limited to a small constant, known in advance. The main drawback of this

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

Table 2
Characteristics of software tools specializing in protein tunnels.

Software	CAVER	MOLE	MOLAXIS
Input	PDB file; AMBER, GROMACS, CHARMM trajectory	PDB file, AMBER trajectory	PDB file
Atom approximation	approximation of atoms by up to 20 small spheres	no approximation - atoms of dissimilar radii are used	approximation of atoms by small spheres to user-defined resolution
Starting point			
Manual	coordinates, list of residues and/or atoms	coordinates, list of residues	coordinates
Automatic	residues from Catalytic Site Atlas	residues from Catalytic Site Atlas, center of selected cavity from CASTp	center of protein mass, center of largest cavity
Protein surface	rolling probe	convex hull	bounding sphere
Cost function	$C(edge) = \int_0^{L(edge)} \frac{1}{d(l)^n} dl$ - L represents the *edge* length - $d_p(l)$ is the distance of the point p to the closest atom. The point p is in the distance l from the fixed vertex of *edge*. - n is the user defined real number determining the balance of width and length	$C(edge) = \frac{L(edge)}{d(edge)^2+\varepsilon}$ - L represents the *edge* length, - d is the distance from the *edge* to the closest atom - ε is a small number to avoid division by zero	$C(edge) = \frac{L(edge)}{d(edge)^2}$ - L represents the *edge* length - d is the average distance of the two vertices forming the *edge* to their respective closest atoms or the user-defined distance, if it is smaller than calculated average distance
Multiple tunnels	tunnels are clustered based on their proximity in space, the best tunnel from each cluster is retained	tunnels with similar tunnel-lining atoms are discarded	similar tunnels are discarded based on the user-defined forking distance
Analysis of MD simulations	tunnel clustering, time evolution of tunnels through the entire MD simulation	tunnel clustering, time evolution of tunnels through the entire MD simulation	tunnel characteristics in individual MD snapshots
Outputs	tunnel bottleneck, tunnel length, tunnel curvature, tunnel profile, tunnel-lining residues, files for visualization in PYMOL and VMD	tunnel bottleneck, tunnel length, tunnel profile, tunnel-lining residues, PDB file for tunnel visualization, similarity matrix	tunnel bottleneck, tunnel profile, tunnel-lining residues, file for visualization in VMD
WWW	caver.cz	mole.chemi.muni.cz	bioinfo3d.cs.tau.ac.il/MolAxis

approximation is longer calculation times due to the significant expansion in the number of atoms employed in the diagram construction.

2.1.4.2. Surface boundary. The estimation of protein surface by the rolling probe used in CAVER is the most detailed, but the user has to select the appropriate probe size, which will not enter the inner voids of a protein as no tunnel would be found. The MOLAXIS approximation of the protein-solvent boundary by a sphere of large radius does not reflect the complex shape of the protein surface.

2.1.4.3. Cost function. All three tools identify tunnels based on very similar cost functions. Whilst tunnel cost is evaluated similarly in all three tools, they differ in their definition of edge width (Fig. 2C). In CAVER, the cost of traveling through a particular path reflects the shape of the path well thanks to the regular sampling of points on edges. Furthermore, users can adjust the cost function to strengthen or weaken the importance of the width over the length, based on their needs. MOLE considers only the narrowest site of the edge as its width. MOLAXIS assumes that the whole edge has a constant width, equal to the average distance from its two ends to the closest atom, thus systematically overestimating the edge width. MOLAXIS does not distinguish between widths above the user-defined threshold, and thus only optimizes the length of the tunnel in such parts of the diagram, i.e., the parts where edges are wider than the given threshold. Neither MOLAXIS nor MOLE allows shifting of a predefined balance between length and width.

2.1.4.4. Multiple tunnels. Both MOLAXIS and CAVER aim to identify all tunnels simultaneously, and there is therefore no need to specify a number of searched tunnels prior to calculation, like in MOLE. After the tunnels are found, MOLE and CAVER cluster them to reduce their number, whilst searching algorithm of MOLAXIS removes any tunnel which is forking from already identified tunnels, and is beyond the forking threshold.

2.1.4.5. MD simulations. The analysis of MD simulations can be performed by each of these three tools. MOLE and CAVER use the clustering to provide information about time evolution of individual tunnels. MOLAXIS does not employ clustering, and thus analyzes the tunnels from individual snapshots separately.

2.1.4.6. Outputs. The calculated tunnels can be opened in several visualization programs, with the exception of MOLAXIS, which does not provide output for tunnel visualization outside VMD plug-in and its own web server. Tunnel characteristics provided by all these tools are similar. CAVER additionally reports the curvature of tunnels, while MOLE produces the similarity matrix of tunnels. MOLAXIS does not report the tunnel lengths.

It is important to underline that none of the three discussed tools are suitable for analysis of irregular tunnels or the extra voids enfolding tunnels since they by definition can only detect tunnels with circular cross-sections (Fig. 3).

2.2. Software tools specialized in analysis of protein channels

In contrast to the tools specializing in tunnel analysis, the tools focusing on channels all employ very different algorithms (Fig. 4). HOLE and MOLAXIS need information about the approximate position and direction of the channel, while CHUNNEL and POREWALKER identify channels without any prior input from the user.

2.2.1. HOLE

HOLE is a pioneering program for the discovery of holes inside the macromolecules (Smart et al., 1996). HOLE was developed to analyze a dimension of ion channels using the Metropolis Monte Carlo simulated annealing procedure (Kirkpatrick et al., 1983; Metropolis et al., 1953). The channel is sliced into parallel planes perpendicular to the user-defined vector, which approximates the general direction of the channel. The search starts on a plane, containing the initial

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

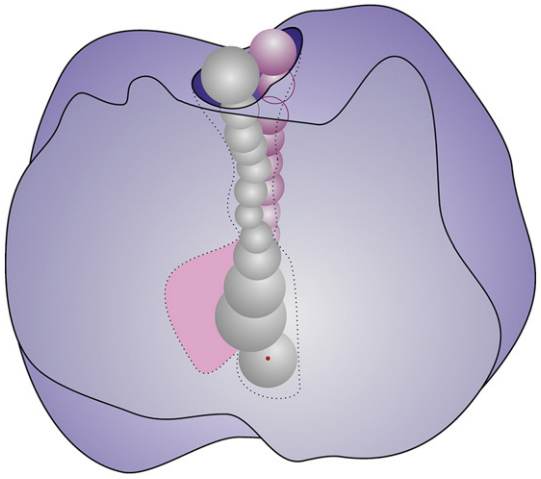

Fig. 3. Schematic representation of limitation common to the software specialized in analysis of protein tunnels due to the approximation of an irregular shape of tunnel by the spheres. The tools will identify two spherical tunnels instead of one true asymmetric tunnel, and the part of the void surrounding the tunnel remains undetected (pink). The shape of internal void is shown by the dotted line, tunnels identified by the tools are shown by spheres, and the tunnel entrance is shown in dark blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

point located inside the channel, which is selected by the user. The position of the point is randomly adjusted by the Monte Carlo procedure to maximize the radius of the sphere centered at this point which does not intersect the VDW surface of surrounding atoms. When the maximal sphere is found, its center is transferred along the channel vector into the next plane, and the whole process is repeated. The algorithm can be viewed as a process of squeezing a flexible sphere through the center of the channel. To correctly describe an anisotropic nature of many larger protein channels, the spherical probe can be replaced by a capsule, and then the search for the capsule with maximal area on a given plane is conducted. The HOLE program is available as a suite of stand-alone applications for Linux operating system while the HOLE web server is under development (currently in the beta testing). In the output, it provides numerical data for plotting of the channel profile, centerlines of channel and a set of spheres colored according to the number of water molecules fitting the spheres for display in various visualization tool. Data can also be used to estimate the channel conductivity. Additionally, the latest version (HOLE 2.2) enables an automatic estimation of the channel vector and the initial point based on the channel symmetry. This version can also be applied for the analysis of MD simulations. A set of PDB files or a CHARMM trajectory is accepted as the input. Visualization of the results in VMD is facilitated by a script.

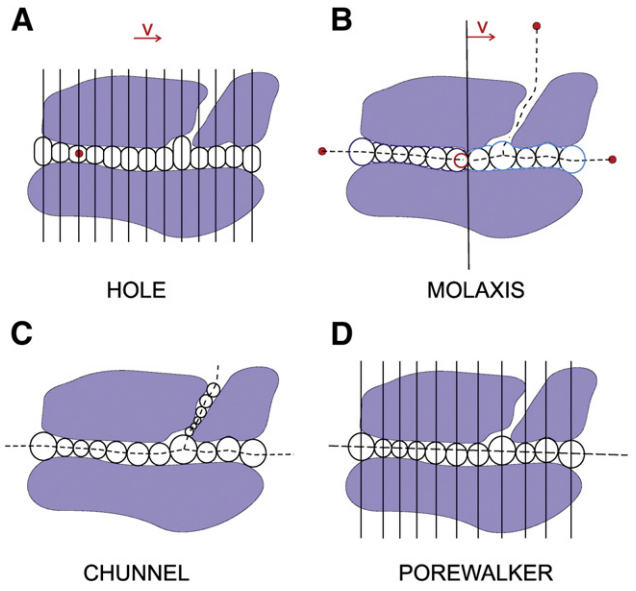

Fig. 4. Algorithms employed by the tools specialized in channel analysis. A) HOLE divides the protein into regular slices perpendicular to the channel vector provided by the user (red arrow). The capsule with the largest area fitting the channel is found in each slice in both directions, starting from the user-specified point (red dot). B) MOLAXIS searches the Voronoi diagram for pathways from all starting points (red dots) outside the protein. User provides a vector (red arrow), allowing separation of the space into two parts by the plane normal to this vector. The pairs of pathways coming from different sides of the plane are combined to create channels. The best channel going through the user defined sphere (in red) is reported. C) CHUNNEL Travel-Depth and Travel-Out distances, in combination with geometrical and topological analysis of a protein, are used to automatically discover all (even branching) channels leading through the protein. D) POREWALKER identifies channel axis by heuristic iterative approach. The protein is then divided into slices perpendicular to the axis, and the largest spheres fitting the channel are identified. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2.2. MOLAXIS

The MOLAXIS mode for traversing protein channels employs a similar algorithm to the tunnel mode discussed in the chapter 2.1.3 (Yaffe et al., 2008a). The starting points are automatically placed on the bounding sphere. The paths traversing the channel are then created by connecting two paths coming from different sides of a plane perpendicular to the user-defined vector. Finally, the best paths crossing the user-defined sphere localized in the channel are reported. In the output of the MOLAXIS channel-module, the channel is divided into several segments. For each segment, the distance from the start of the channel, radius of the channel segment and its surrounding residues are listed. The channel geometry is visualized by Jmol. The stand-alone version also provides the channel profile and the data for visualization of the channel via the VMD plugin.

2.2.3. CHUNNEL

CHUNNEL is based on a multiple source shortest paths algorithm for measurement of the depths of cavities and clefts, which is modified to allow unsupervised analysis of protein channels (Coleman and Sharp, 2009). It employs the Travel-Depth distance (Coleman and Sharp, 2006), quantifying the shortest protein-avoiding pathway from the convex hull of the protein to all points on the proteins surface as well as the Travel-Out distance (Coleman and Sharp, 2009). The Travel-Out distance starts at the molecular surface, and is propagated outwards into the solvent. The maxima of Travel-Out distance are localized in the centers of the channel representing its approximate shape skeleton. This can be used to travel the path along the center of the channel by adjusting the algorithm to start searching at the maximum of both the Travel-Depth and Travel-Out distances, and following the channel skeleton of Travel-Out distance with decreasing Travel-Depth distance in two opposite directions. The CHUNNEL tool performs geometrical and topological analyses of the proteins molecular surface to calculate the number of all channels present in the protein structure, to ensure the correct positioning of the starting point in the channel, as well as to define opposite directions in the channels. By the procedure of loop identification and regularization, the loop running around the narrowest part of the channel is found, and a plug is constructed inside this loop to separate the channel in two parts. When the channel is identified, it is masked and the procedure is repeated to identify another channel, until all channels are found and plugged. From each plug, the two half-pathways out of the channels center are calculated using the branch-and-bound algorithm. All pathways coming out of each plug are combined into a tree, which is then processed to provide all possible non-cycling paths connecting all combinations of points. The final step is the verification that the found pathways actually traverse through the channel to avoid falsely reported channels. CHUNNEL is

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

implemented as a suite of programs written in FORTRAN and Python compatible with Linux and Mac OS X. In the input, users have to provide a structure of the target protein in the PDB format. The output for all identified channels comprises: the channel length, minimum channel radius, first minimum channel radii from each of the channel ends, and maximum radius between these two minima, channel curvature, list of channel lining residues as well as the channel entrance, exit and plug it passes through. CHUNNEL also provides scripts for visualization of the results in PYMOL.

2.2.4. POREWALKER

POREWALKER is the second tool designed for automatic detection and analysis of channels, more specifically channels passing through transmembrane proteins (Pellegrini-Calace et al., 2009). It applies heuristic and iterative algorithms, which aim to identify channel-lining residues, and use them to define the channel axis and center. Initially, the POREWALKER algorithm defines the channel axis as the average of the axes of transmembrane secondary structures, and the channel center as a center of a protein mass. The channel axis is then used to find channel-lining residues as those belonging to solvent accessible residues which are both close to the axis and oriented towards it. The new channel center is in turn redefined as the center of a newly detected channel-lining residue. The axis is then translated to pass through this new center. This whole procedure is repeated as long as additional channel-lining residues can be identified. The channel axis is then refined by the iterative slice-based approach in which all protein atoms are placed on a grid, and sliced along the channel axis in 1 Å thick layers. In each layer, the local channel center is initially assigned to the average position of atoms of putative channel-lining residues belonging to the given layer. The sphere is constructed at the center, and then expanded while avoiding clashes with protein atoms. The sphere center is systematically shifted from the initial position through the layer to maximize its radius. The center of the largest sphere is considered the new local channel center for the layer. When all local centers for all layers are identified, the channel axis is derived from the position of the four last centers at each end of the channel. The refinement of the channel axis, and identification of the local channel centers are repeated until the angle between the current and new channel axis is less than 0.5 degrees. Finally, the channel defined by the refined channel axis, and refined set of local channel centers is analyzed to describe the composition and geometry of the channel. POREWALKER is implemented as a web server requiring the PDB file of the protein as the only input. The outputs of the server comprise the list of channel-lining atoms and residues, a simplification of the channel shape by a set of truncated cones and cylinders describing trends in channel segments with constant and varying diameters, the profile of the channel, and a description of the channel regularity by a series of straight or wiggly lines. The server also prepares figures of the channel showing positions of all local channel centers from several viewpoints, and a slideshow of channel cross-sections along the channel. The server provides a downloadable PDB file with the channel centers at 3 Å slices, indicated by water molecules, and channel-lining residues, indicated by their B-factors.

2.2.5. Critical comparison

The tools specialized in channel identification are, by definition, not convenient for identification of tunnels leading from occluded protein cavities, with the exception of the MOLAXIS module for tunnels. The main difference between the tools lies in their ability to: (i) analyze multiple channels, (ii) perform automatic search, (iii) describe the channel anisotropy, and (iv) provide detailed outputs of the analysis (Table 3).

2.2.5.1. Multiple channels. With the exception of CHUNNEL, all tools analyze only one channel per protein. If the analysis of different channels of the protein is desired, the identified channel has to be manually masked, e.g. by mutation or introduction of a small molecule, to allow the identification of secondary channels.

2.2.5.2. Automatic analysis. CHUNNEL and POREWALKER calculations can run fully unsupervised. HOLE can perform a crude estimation of channel direction, and location based on the channel symmetry, but

Table 3
Characteristics of software tools specialized on analysis of protein channels.

Software	HOLE	MOLAXIS	CHUNNEL	POREWALKER
Input	PDB file, CHARMM trajectory	PDB file	PDB file	PDB file
Channel center				
Manual	coordinates	coordinates	–	–
Automatic	center of C_α atoms	–	from analysis of molecular surface	center of protein mass initially, iteratively improved as center of mass of channel-lining residues
Channel axis				
Manual	vector	vector	–	–
Automatic	along *X*-, *Y*- or *Z*-axis	–	from analysis of molecular surface	secondary structure vectors initially, iteratively improved from local centers of channel
Multiple channels	only one channel detected	only one channel detected	all non-cycling channels including branches	only one channel detected
Channel geometry	spheres, capsules	spheres	spheres	spheres
Outputs	channel-lining atoms, channel profile, center lines of channel, set of spheres colored by number of fitting water molecules for various visualization tools	distance of channel segment from channel start, segment bottlenecks, residues surrounding segments, file for visualization in VMD, channel profile	channel length, minimal channel bottleneck, first bottlenecks from each channel end, maximum radius between bottlenecks, channel curvature, channel-lining residues, channel entrance, bottleneck and exit residues, file for visualization in PYMOL	channel-lining atoms and residues, descriptors of changes in channel diameter and regularity, channel profile, figures of all local channel centers, slideshow of channel cross-sections, PDB file with channel centers indicated by water molecules
WWW	hole.biop.ox.ac.uk/hole	bioinfo3d.cs.tau.ac.il/MolAxis	crystal.med.upenn.edu/travel_distance.html	www.ebi.ac.uk/thornton-srv/software/PoreWalker

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

due to the non-deterministic nature of its algorithm, the proper setting of the minimal number of evaluations in Monte Carlo minimization is needed to find the pathway close to its optimum. MOLAXIS needs full specification of the channel direction and location by the user.

2.2.5.3. Channel anisotropy. HOLE is the only tool able to handle irregularity in channel dimensions as it performs the search for capsule of maximal area, while the other tools provide correct information only on the spherical parts of channels similarly to the in Section 2.1. On the other hand, the search by HOLE is performed separately in each plane along the channel, considering only pathway width, which can produce a discontinuous pathway, deviating from the ideal channel axis.

2.2.5.4. Outputs. The information provided by all tools is similar with the distinction of POREWALKER, which provides novel simplified descriptors of the channel shape and geometry suitable for a large-scale comparison of various channels. CHUNNEL and MOLAXIS produce scripts for detailed visualization of results in PYMOL and VMD, respectively, HOLE and POREWALKER allow import of their results as modified PDB files into many standard visualization programs.

2.3. Tools specialized in tunnels and channels as a part of overall protein voids

These tools aim to accurately represent overall voids in a particular protein structure, including channels, tunnels, surface gorges, and occluded pockets or closed cavities. Both HOLLOW and 3V employ the method of rolling probe (Fig. 5A) to define the exterior envelope of the protein as well as the voids inside the protein (Fig. 5B). The rolling probe method is a grid-based approach, in which a probe of a given radius is placed on all grid points, where it does not overlap with a protein atom, enabling the close circumscription of the protein surface.

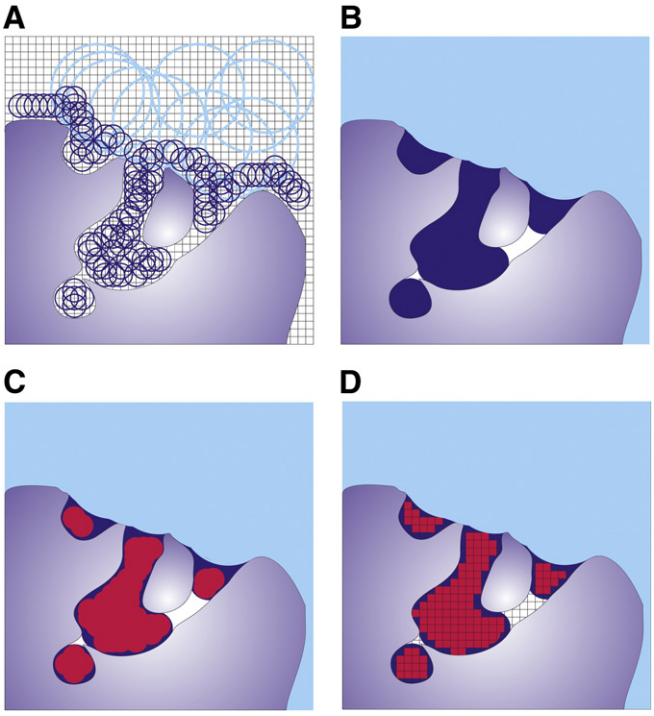

Fig. 5. Algorithms embedded in the tools specialized in analysis of protein voids. A) Rolling probe method. The surface and inner voids of the protein are defined by the placement of probes (large for surrounding and small for voids) in each point of the grid. B) The identified volumes are divided into the protein surrounding (light blue), and internal voids (dark blue), while undetected internal volumes are white. C) HOLLOW represents identified voids using the dummy atoms fitting these voids (red spheres). D) 3V represents detected internal void by voxels (red squares). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3.1. HOLLOW

HOLLOW represents a protein interior (Ho and Gruswitz, 2008) by a filling composed of dummy atoms (Fig. 5C). The algorithm uses two probes of different radii. The protein envelope is defined by the larger probe, which is rolled over the surface atoms of the protein with the solvent-accessible surface area larger than 9 $Å^2$. Consequently, the smaller probes (dummy atoms) are placed on the rectangular grid surrounding the protein molecule, and the probes overlapping with any protein atom or located outside of the protein envelope are removed. Both removals of dummy atoms beyond the envelope, and a calculation of solvent accessible surface areas of protein atoms are provided by the Sharke–Rupey dot-density method (Shrake and Rupley, 1973). The set of the dummy atoms, remaining between the proteins surface and its envelope, represents the volume of the overall void in the protein. The number of dummy atoms is selectively reduced, while simultaneously maintaining the total volume of the void. HOLLOW is implemented as a platform-independent command-line utility written in Python. This implementation allows a customization of the grid spacing, probe radii and dummy atom type. A user can also specify spherical or cylindrical constrains to focus void calculation on a selected protein region, allowing its more detailed analysis. The output is the PDB file containing a reduced set of dummy atoms, representing the overall volume of the void, which can be visualized, and further analyzed in many standard visualization programs.

2.3.2. 3V

3V is an abbreviation for the Voss Volume Voxelator. The tool automatically extracts and analyzes all internal volumes in protein molecules (Voss and Gerstein, 2010). 3V employs a method based on two probes, resulting in two types of volumes: a solvent-excluded volume and shell volume. A solvent-excluded volume is obtained by the smaller probe; while a shell volume is determined using the larger probe. The subtraction of these two volumes is a volume of the overall void space in the protein. Identified void volumes are represented by voxels, i.e., small cubic portions of the grid, belonging to the void in the protein (Fig. 5D). 3V is implemented as a web server, and its source code is provided as C libraries, allowing compilation of the program for various platforms. The server provides six modules of which three are relevant for pathway identification. The Channel Find module allows the identification of individual continuous voids, fulfilling particular cutoff criteria. These criteria can be the minimal volume of the void in $Å^3$, minimal volume of the void as a percentage of the protein shell volume, or the number of separate voids to be detected. The Channel Extract module provides only one void, which contains the user-defined coordinate point. The Exit Tunnel Extract module is tailored for investigation of the exit tunnels of the 50S ribosomal subunit. In the input, the user defines radii of the smaller and larger probes, selects from four predefined grid resolutions, and specifies cutoff criteria or void location. The output of the modules comprises the volume and surface area of each void, void sphericity, effective radius of the void, center of its mass and its reduced center. For each individual void, pictures showing the void from several angles are provided as well as density map files in Medical Research Council (MRC) format enabling visualization of the results in PYMOL or CHIMERA (Pettersen et al., 2004).

2.3.3. Critical comparison

Unlike majority of the tools in Sections 2.1. and 2.2., HOLLOW and 3V tools can describe the geometry of nearly any type of void. The detailed visualization of voids demands long computational time, which can be reduced by using a lower grid resolution or larger probes, but only at the expense of accuracy. The obvious disadvantage is the

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

need for specification of proper radii for both probes. Basically, the smaller probe has to be small enough to fit the internal voids, while the larger probe should not, otherwise incorrect voids will be found. The larger probe should still be small enough to closely delineate the protein surface, preventing the detection of many irrelevant surface patches. If the appropriate radii of probes are not known *a priori*, an iterative trial-and-error approach should be applied. The HOLLOW and 3V tools vary in: (i) void representation, (ii) adjustability of grid resolution, (iii) focus on the particular void regions, and (iv) range of provided outputs (Table 4).

2.3.3.1. Void representation. HOLLOW and 3V differ in the way of presenting and storing detected internal voids. HOLLOW uses the dummy atoms, which allow analysis of results in many molecular visualization tools. The output of 3V is stored as voxels, which is not suitable for further analysis nor commonly supported by visualization tools.

2.3.3.2. Grid resolution. The option for setting grid resolution is very important for the preparation of highly detailed visualization of voids as well as for optimization of calculation speed useful for larger proteins. While HOLLOW provides the user with the full control over this setting, the 3V server only allows selection from four predefined resolutions.

2.3.3.3. Void focusing. The results produced by HOLLOW and 3V provide limited information on tunnels and channels characteristics. To enable a more detailed study of these features, the identified voids have to be separated from each other. In HOLLOW, the user-defined constrains are used to restrict the void search to a predefined location, e.g., selected tunnel. The drawback of this approach is that input of the approximate location of the tunnel is required prior to calculation. It is advisable to first perform a crude search on the whole protein to obtain a rough idea about the tunnel location. Since the HOLLOW output uses dummy atoms, it is possible to easily modify reported voids manually. The conceptually different approach is employed in 3V, which offers several filtering tools for the selection of individual voids. However, a tool allowing the extraction of part of the void from the continuous void is not available.

Table 4
Characteristics of tools specializing in tunnels and channels as part of overall voids in proteins.

Software	HOLLOW	3V
Input	PDB file, CHARMM trajectory	PDB file
Grid resolution	user-defined	2.0 Å, 1.0 Å, 0.75 Å or 0.5 Å
Protein surface	rolling of the larger probe over protein atoms with a solvent-accessible surface area > 9Å^2	rolling of the larger probe over all protein atoms
Void representation	dummy atoms	voxels
Void focusing	spherical or cylindrical constrains defined by users,	filtering by total or relative minimal volume of void,
	manual removal of dummy atoms	number of voids or location of voids
Output	PDB file containing dummy atoms	pictures of void from several directions, volume and surface area of void, void sphericity, effective radius of the void, center of void mass, reduced center of void mass, Medical Research Council (MRC) files with stored voxels
WWW	hollow.sourceforge.net	3vee.molmovdb.org

3. Practical applications of software tools

The software tools introduced in this review can be used for identification, visualization and analysis of tunnels and channels in protein structures. Whilst tunnel and channel visualization still represents the most common application area of these tools (O'Donoghue et al., 2010; Pervushin et al., 2009; Sharma et al., 2010; Shimada et al., 2010; Sidhu et al., 2010; Fig. 6), there is a growing number of examples demonstrating their utility for detailed mechanistic understanding of protein function or rational design of proteins with improved properties. Since the number of these examples has been growing steadily in recent years, only selected examples will be provided in the following section.

3.1. Applications for the analysis of biological systems

The analysis of tunnels by software tools can provide unique insight into a role of individual tunnels for a reaction mechanism of various enzymes. The investigation of the main tunnel in norcoclaurine synthase from *Thalictrum flavum* helped to improve understanding of mechanisms involved in the benzylisoquinoline alkaloid biosynthetic pathway (Ilari et al., 2009). The main tunnel is composed of side-chains of tyrosine, lysine, aspartic and glutamic acid residues, and provides the structural determinants governing the stereospecificity of the Pictet–Spengler cyclization.

The role of three different tunnels connecting the bulk solvent with the internal active site of the heterotetrameric sarcosine oxidase from *Corynebacterium* sp. U-96 was characterized in detail (Moriguchi et al., 2010). Two of these tunnels were found in the inter-subunit space, and presented main pathways for the substrate access, while the third tunnel could possibly be used by an iminium intermediate.

Similarly, the accessibilities of individual tunnels in the membrane-intrinsic protein-cofactor complex photosystem II from the cyanobacterium *Thermosynechococcus elongatus* were analyzed pinpointing four tunnels, which could serve for water and oxygen transport, and a previously unknown plastoquinone-transfer tunnel (Guskov et al., 2009). Existence of this second plastoquinone-transfer tunnel suggested three possible plastoquinol-plastoquinone exchange mechanisms: (i) both ligands proceed via both tunnels in an alternate fashion, (ii) plastoquinone enters exclusively via the first tunnel while plastoquinol exits only through the second tunnel, and (iii) transport occurs only through the second tunnel.

To examine the dynamical nature of tunnels, the tunnel-intersecting residues of mammalian CYP450 isoforms were analyzed. The results of this analysis enabled the definition of gating models. Twenty-eight two-state gating models and thirty-seven singlet gating-residue models were devised. The interaction patterns between the residues forming these gates suggest the presence of a regulatory network enabling coordinated gating for multiple tunnels present in CYP450 isoforms (Zawaira et al., 2011).

The anticipated role of the tunnels can be confirmed by experimental mutagenesis of selected tunnel residues. In this way, the importance of a tunnel in MoFe nitrogenase from *Azotobacter vinelandii* for substrate and product transport was tested by mutagenesis of four tunnel residues (Barney et al., 2009). The mutants showed a significant increase in Michaelis-Menten constants K_m while maintaining their maximum rates V_{max}, verifying the functional role of this tunnel. Similarly, the putative hydrophobic H_2 tunnel in *Clostridium acetobutylicum* FeFe hydrogenase was identified, and its relevance for substrate transport was investigated by mutagenesis (Lautier et al., 2011). Eight constructed mutants showed only minimal effects on the enzymes catalytic properties, suggesting that H_2 diffuse either via temporary voids or through a water-filled tunnel, which has not been previously considered as a transport pathway for hydrophobic ligands.

Information derived by reviewed tools can also be applied in drug development, as illustrated on study of Aquaglyceroporin

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

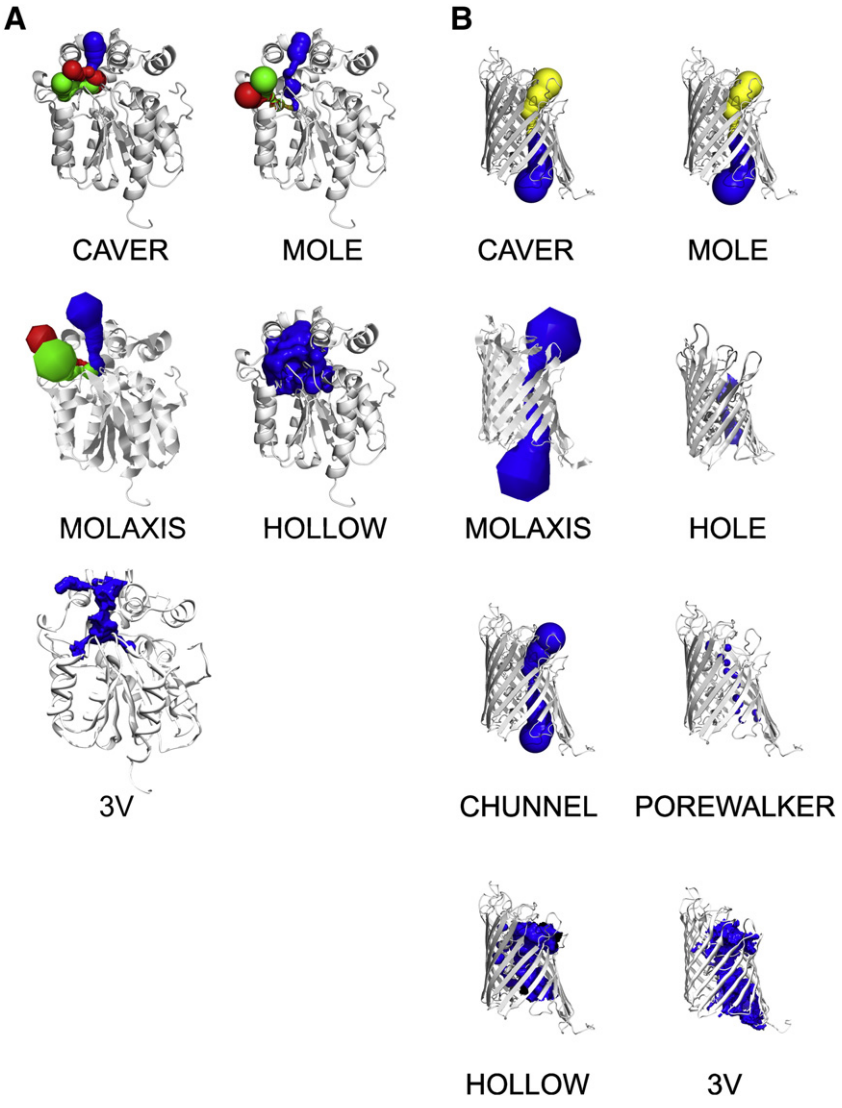

Fig. 6. Visualization of tunnels and channels calculated by the reviewed tools. A) Tunnels of haloalkane dehalogenase DhaA (PDB ID: 1CQW). Protein structure is in gray cartoon, the first scoring tunnel is in blue, the second in green and the third in red. B) Channel through porin OprP (PDB ID: 2O4V). Protein structure is in gray cartoon, and the channel is in blue. In case of CAVER and MOLE, the two best tunnels of opposite direction are depicted. The first scoring tunnel is in blue, and the best scoring tunnel of opposite direction is in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

LmAQP1, which is responsible for the transport of antimony drugs into the cell of human parasite *Leishmania major* (Mukhopadhyay et al., 2011). The analysis of this channel revealed that substitution of Ala163 located near the channel entrance can reduce channel permeability towards the antimonite while maintaining the water and solute conductivity essential for its biological function. This study thus showed that substitutions in position 163 of LmAQP1 may be important for drug resistance, opening new possibilities for structure-based drug development.

3.2. Applications for the design of biological systems

In some cases, the information obtained from the analysis of protein tunnels and channels has already been employed to design proteins with novel functional properties.

The engineering of the tunnel in human cytochrome P450 2B6 resulted into mutants with increased catalytic activity towards a chemotherapeutic agent cyclophosphamide (Nguyen et al., 2008), which is activated due to conversion by this cytochrome. When the most active mutant was expressed in the cancer cells, it switched the resistant cancer cells into cells sensitive toward cyclophosphamide allowing for cancer treatment via gene-directed enzyme prodrug therapy (Waxman et al., 1999).

The detailed study of access tunnels of haloalkane dehalogenase DhaA from *Rhodococcus rhodochrous* led to the identification of the key tunnel residues. These residues were targeted by focused directed evolution. Using this procedure, the mutant enzyme DhaA31, with up to 32-fold higher catalytic activity towards the anthropogenic pollutant 1,2,3-trichloropropane, was obtained (Pavlova et al., 2009). The improvement in catalytic activity was explained by decreased accessibility of the active site for water molecules, promoting the formation of the reactive complex.

Site-saturation mutagenesis of tunnel residues located near the active site entrance of cellobiose phosphorylase form *Cellulomonas uda* was performed, yielding mutant enzymes with significantly broader substrate specificity (De Groeve et al., 2010). These engineered cellobiose phosphorylase variants can be useful for biosynthesis of industrial surfactants β-cellobiosides, cellulase inhibitors or prebiotic sugars.

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

4. Conclusions and future prospects

Investigation of small molecules, ions and solvent pathways in proteins is becoming the standard procedure for protein characterization and design. Several tools with conceptually different algorithms are currently available for this purpose. We have overviewed algorithms implemented in these tools, their advantages, limitations and requirements for their successful applications. Since none of these tools can guarantee the identification of tunnels and channels of all very different shapes, the recommended practice should be to combine several tools, and complement obtained results with visualization of the proteins internal surface using a probe with a radius smaller than the tunnel or channel bottleneck.

While the use of current tools for the analysis of protein tunnels and channels is established, their application for the design of modified proteins is limited. Improved tools should provide more robust automatic identification of initial conditions, which are currently supplemented by users. The protein engineers would also benefit from the tools enabling the selection of appropriate mutations to obtain transport pathways with desired characteristics, e.g. with specific bottleneck size, hydrophobicity or dynamics.

Detailed information on individual pathways, beyond their geometry, e.g., electrostatics, hydrophobicity, dynamics or druggability, would be very useful. Molecular descriptors resembling the shape and regularity parameters of POREWALKER should be implemented to allow condensed description of various pathways. Such descriptors could be used for large-scale comparisons of different pathways. Existing and newly developed descriptors should ultimately lead to distinguishing functionally relevant pathways from biologically irrelevant voids.

Acknowledgments

This work was supported by the European Regional Development Fund (CZ.1.05/2.1.00/01.0001 to E.C., J.B. and J.D.), the Grant Agency of the Czech Republic (P202/10/1435 to A.P., P503/12/0572 to J.D.) and the Grant Agency of the Czech Academy of Sciences (IAA401630901 to J.B. and J.D.). MetaCentrum is acknowledged for providing access to computing facilities, supported by the Ministry of Education of the Czech Republic (LM2010005). The work of A.G. was supported by SoMoPro programme No. SIGA762 funded by the European Community within the 7th FP under grant agreement No. 229603, and co-financed by the South Moravian Region.

References

Arroyo-Mañez P, Bikiel DE, Boechi L, Capece L, Di Lella S, Estrin DA, et al. Protein dynamics and ligand migration interplay as studied by computer simulation. Biochim Biophys Acta 2011;1814:1054–64.
Aurenhammer F. Voronoi diagrams—a survey of a fundamental geometric data structure. Comput Surv 1991;23:345–405.
Barney BM, Yurth MG, Dos Santos PC, Dean DR, Seefeldt LC. A substrate channel in the nitrogenase MoFe protein. J Biol Inorg Chem 2009;14:1015–22.
Berendsen H. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun 1995;91:43–56.
Bezanilla F. How membrane proteins sense voltage. Nat Rev Mol Cell Biol 2008;9:323–32.
Binkowski TA, Naghibzadeh S, Liang J. CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res 2003;31:3352–5.
Brady Jr GP, Stouten PF. Fast prediction and visualization of protein binding pockets with PASS. J Comput Aided Mol Des 2000;14:383–401.
Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
Chaloupková R, Sýkorová J, Prokop Z, Jesenská A, Monincová M, Pavlová M, et al. Modification of activity and specificity of haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 by engineering of its entrance tunnel. J Biol Chem 2003;278:52622–8.
Coleman RG, Sharp KA. Travel depth, a new shape descriptor for macromolecules: application to ligand binding. J Mol Biol 2006;362:441–58.
Coleman RG, Sharp KA. Finding and characterizing tunnels in macromolecules with application to ion channels and pores. Biophys J 2009;96:632–45.
De Groeve MRM, Remmery L, Van Hoorebeke A, Stout J, Desmet T, Savvides SN, et al. Construction of cellobiose phosphorylase variants with broadened acceptor specificity towards anomerically substituted glucosides. Biotechnol Bioeng 2010;107:413–20.
Dijkstra EW. A note on two problems in connexion with graphs. Numer Math 1959;1:269–71.
Ghersi D, Sanchez R. EasyMIFS and SiteHound: a toolkit for the identification of ligand-binding sites in protein structures. Bioinformatics 2009;25:3185–6.
Gold VAM, Duong F, Collinson I. Structure and function of the bacterial Sec translocon. Mol Membr Biol 2007;24:387–94.
Gouaux E, Mackinnon R. Principles of selective ion transport in channels and pumps. Science 2005;310:1461–5.
Guskov A, Kern J, Gabdulkhakov A, Broser M, Zouni A, Saenger W. Cyanobacterial photosystem II at 2.9-A resolution and the role of quinones, lipids, channels and chloride. Nat Struct Mol Biol 2009;16:334–42.
Hendlich M, Rippmann F, Barnickel G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. J Mol Graph Model 1997;15:359–63. (389).
Herráez A. Biomolecules in the computer: Jmol to the rescue. Biochem Mol Biol Educ 2006;34:255–61.
Ho BK, Gruswitz F. HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. BMC Struct Biol 2008;8:49.
Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–8.
Ilari A, Franceschini S, Bonamore A, Arenghi F, Botta B, Macone A, et al. Structural basis of enzymatic (S)-norcoclaurine biosynthesis. J Biol Chem 2009;284:897–904.
Janin J, Chothia C. The structure of protein–protein recognition sites. J Biol Chem 1990;265:16027–30.
Kirkpatrick S, Gelatt Jr CD, Vecchi MP. Optimization by simulated annealing. Science 1983;220:671–80.
Kleywegt GJ, Jones TA. Detection, delineation, measurement and display of cavities in macromolecular structures. Acta Crystallogr D Biol Crystallogr 1994;50:178–85.
Laskowski RA. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J Mol Graph 1995;13(323–30):307–8.
Laurie ATR, Jackson RM. Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. Bioinformatics 2005;21:1908–16.
Laurie ATR, Jackson RM. Methods for the prediction of protein–ligand binding sites for structure-based drug design and virtual ligand screening. Curr Protein Pept Sci 2006;7:395–406.
Lautier T, Ezanno P, Baffert C, Fourmond V, Cournac L, Fontecilla-Camps JC, et al. The quest for a functional substrate access tunnel in FeFe hydrogenase. Faraday Discuss 2011;148:385–407.
Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics 2009;10:168.
Levitt DG, Banaszak LJ. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. J Mol Graph 1992;10:229–34.
Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. Protein Sci 1998;7:1884–97.
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys 1953;21:1087.
Moriguchi T, Ida K, Hikima T, Ueno G, Yamamoto M, Suzuki H. Channeling and conformational changes in the heterotetrameric sarcosine oxidase from *Corynebacterium* sp. U-96. J Biochem 2010;148:491–505.
Mukhopadhyay R, Mandal G, Atluri VSR, Figarella K, Uzcategui NL, Zhou Y, et al. The role of alanine 163 in solute permeability of *Leishmania major* aquaglyceroporin LmAQP1. Mol Biochem Parasitol 2011;175:83–90.
Nguyen T-A, Tychopoulos M, Bichat F, Zimmermann C, Flinois J-P, Diry M, et al. Improvement of cyclophosphamide activation by CYP2B6 mutants: from in silico to ex vivo. Mol Pharmacol 2008;73:1122–33.
Niemeyer BA, Mery L, Zawar C, Suckow A, Monje F, Pardo LA, et al. Ion channels in health and disease. 83rd Boehringer Ingelheim fonds international titisee conference. EMBO Rep 2001;2:568–73.
O'Donoghue S, Goodsell D, Frangakis A, Jossinet F, Laskowski R, Nilges M, et al. Visualization of macromolecular structures. Nat Methods 2010;7:S42–55.
Pavlova M, Klvana M, Prokop Z, Chaloupkova R, Banas P, Otyepka M, et al. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. Nat Chem Biol 2009;5:727–33.
Pearlman D, Case D, Caldwell J, Ross W, Cheatham T, Debolt S, et al. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 1995;91:1-41.
Pellegrini-Calace M, Maiwald T, Thornton JM. PoreWalker: a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. PLoS Comput Biol 2009;5:e1000440.
Pervushin K, Tan E, Parthasarathy K, Lin X, Jiang FL, Yu D, et al. Structure and inhibition of the SARS coronavirus envelope protein ion channel. PLoS Pathog 2009;5:e1000511.
Petrek M, Otyepka M, Banás P, Kosinová P, Koca J, Damborský J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. BMC Bioinformatics 2006;7:316.
Petrek M, Kosinová P, Koca J, Otyepka M. MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. Structure 2007;15:1357–63.
Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002

Porter CT, Bartlett GJ, Thornton JM. The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. Nucleic Acids Res 2004;32:D129–33.
Rhodes D, Burley SK. Protein–nucleic acid interactions. Curr Opin Struct Biol 2000;10:75–7.
Richards FM. Areas, volumes, packing and protein structure. Annu Rev Biophys Bioeng 1977;6:151–76.
Sharma M, Yi M, Dong H, Qin H, Peterson E, Busath DD, et al. Insight into the mechanism of the influenza A proton channel from a structure in a lipid bilayer. Science 2010;330:509–12.
Shimada A, Ishikawa H, Nakagawa N, Kuramitsu S, Masui R. The first crystal structure of an archaeal metallo-beta-lactamase superfamily protein; ST1585 from *Sulfolobus tokodaii*. Proteins 2010;78:2399–402.
Shrake A, Rupley JA. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. J Mol Biol 1973;79:351–71.
Sidhu RS, Lee JY, Yuan C, Smith WL. Comparison of cyclooxygenase-1 crystal structures: cross-talk between monomers comprising cyclooxygenase-1 homodimers. Biochemistry 2010;49:7069–79.
Smart OS, Neduvelil JG, Wang X, Wallace BA, Sansom MS. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. J Mol Graph 1996;14:354–60.
Tseng YY, Dupree C, Chen ZJ, Li W-H. SplitPocket: identification of protein functional surfaces and characterization of their spatial patterns. Nucleic Acids Res 2009;37:W384–9.
Venkatachalam CM, Jiang X, Oldfield T, Waldman M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J Mol Graph Model 2003;21:289–307.
Voss NR, Gerstein M. 3V: cavity, channel and cleft volume calculator and extractor. Nucleic Acids Res 2010;38:W555–62.
Waxman DJ, Chen L, Hecht JE, Jounaidi Y. Cytochrome P450-based cancer gene therapy: recent advances and future prospects. Drug Metab Rev 1999;31:503–22.
Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R. MolAxis: efficient and accurate identification of channels in macromolecules. Proteins 2008a;73:72–86.
Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov R. MolAxis: a server for identification of channels in macromolecules. Nucleic Acids Res 2008b;36:W210–5.
Yu J, Zhou Y, Tanaka I, Yao M. Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics 2010;26:46–52.
Zawaira A, Coulson L, Gallotta M, Karimanzira O, Blackburn J. On the deduction and analysis of singlet and two-state gating-models from the static structures of mammalian CYP450. J Struct Biol 2011;173:282–93.
Zvelebil MJ, Thornton JM. Peptide–protein interactions: an overview. Q Rev Biophys 1993;26:333–63.

Please cite this article as: Brezovsky J, et al, Software tools for identification, visualization and analysis of protein tunnels and channels, Biotechnol Adv (2012), doi:10.1016/j.biotechadv.2012.02.002