Journal of Molecular Graphics and Modelling 29 (2011) 895–905

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

dxTuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics

Martin Raunest, Christian Kandt*

Computational Structural Biology, Department of Life Science Informatics, B-IT, Life & Medical Sciences (LIMES) Center, University of Bonn, Dahlmannstr 2, 53113 Bonn, Germany

ARTICLE INFO

Article history:
Received 17 November 2010
Received in revised form 4 February 2011
Accepted 14 February 2011
Available online 24 February 2011

Keywords:
Cavity analysis
Protein structure
Molecular dynamics simulation
Residence probability
Pore profile

ABSTRACT

Empty space in a protein structure can provide valuable insight into protein properties such as internal hydration, structure stabilization, substrate translocation, storage compartments or binding sites. This information can be visualized by means of cavity analysis. Numerous tools are available depicting cavities directly or identifying lining residues. So far, all available techniques base on a single conformation neglecting any form of protein and cavity dynamics. Here we report a novel, grid-based cavity detection method that uses protein and solvent residence probabilities derived from molecular dynamics simulations to identify (I) internal cavities, (II) tunnels or (III) clefts on the protein surface. Driven by a graphical user interface, output can be exported in PDB format where cavities are described as individually selectable groups of adjacent voxels representing regions of high solvent residence probability. Cavities can be analyzed in terms of solvent density, cavity volume and cross-sectional area along a principal axis. To assess dxTuber performance we performed test runs on a set of six example proteins representing the three main classes of protein cavities and compared our findings to results obtained with SURFNET, CAVER and PyMol.

1. Introduction

The successful determination of a protein's high-resolution three-dimensional structure is a landmark on the way to understanding its function. Next to the actual atomic XYZ coordinates, 3D structures often contain additional information that can yield further insights into the protein in question. For example, empty space in a protein structure can provide valuable insight into protein properties such as internal hydration, structure stabilization, substrate translocation, storage compartments or substrate binding sites [1,2]. This information can be visualized by means of cavity analysis. Over the years numerous cavity detection tools have been developed including [3–17] that depict cavities either directly [3–6,8,10–17] or indirectly by identifying lining residues [9] or filling a cavity with water molecules [7]. The main strategies used in these geometry-based algorithms [1] can be grouped into four categories plus combinations of these.

One approach employs the rolling probe sphere method [18–21] where molecular surfaces are calculated combining different probe radii and surface types in order to (a) distinguish external from internal residues and (b) detect protein cavities. Examples of this approach include, GRASP [11], 3V [16] and PyMol since version 1.3 [4]. Specifically aiming at pores and tunnel-like cavities, another strategy calculates pathways connecting protein interior and exterior. Most tools employing this approach begin with a user-defined starting point inside the protein from where an exit pathway is calculated while simultaneously recording a profile of the pore dimensions. Expressed as radius or diameter of a tunnel-filling sphere at given way points, such pore profiles allow a detailed characterization of the channel in question. Examples of this approach include HOLE [15], its successor CAVER [14] and MoleAxis [17]. On the other hand, CAST [9] and the algorithm introduced by Maeda and coworkers [10] resort to alpha shape theory, Voronoi diagrams and Delaunay tessellation [22–24] to identify protein cavities. Whereas the CAST algorithm is intended for general cavity detection, the program by Maeda et al. specifically focuses on interprotein cavities. The fourth category of tools describes cavities as entities filled with (a) probe spheres of parameterizable size as in SURFNET [8] and HOLLOW [6], (b) water molecules as in VOIDOO [7] or (c) voxels like in the algorithm introduced by Exner and coworkers [5] or dxTuber in this publication.

Examples where combinations of these main strategies have been implemented include the CAVER successor MOLE [13], PoreWalker [12] and CHUNNEL [3]. Whereas MOLE and PoreWalker employ the path finding approach in conjunction with Voronoi diagrams and alpha shape theory, CHUNNEL combines the path finding strategy with probe-sphere based molecular surfaces. However, unlike other path finding programs, CHUNNEL calculates multiple pathways leading into the protein without requiring a user-defined starting point.

* Corresponding author. Tel.: +49 228 2699 324; fax: +49 228 2699 341.
E-mail address: kandt@bit.uni-bonn.de (C. Kandt).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.02.003

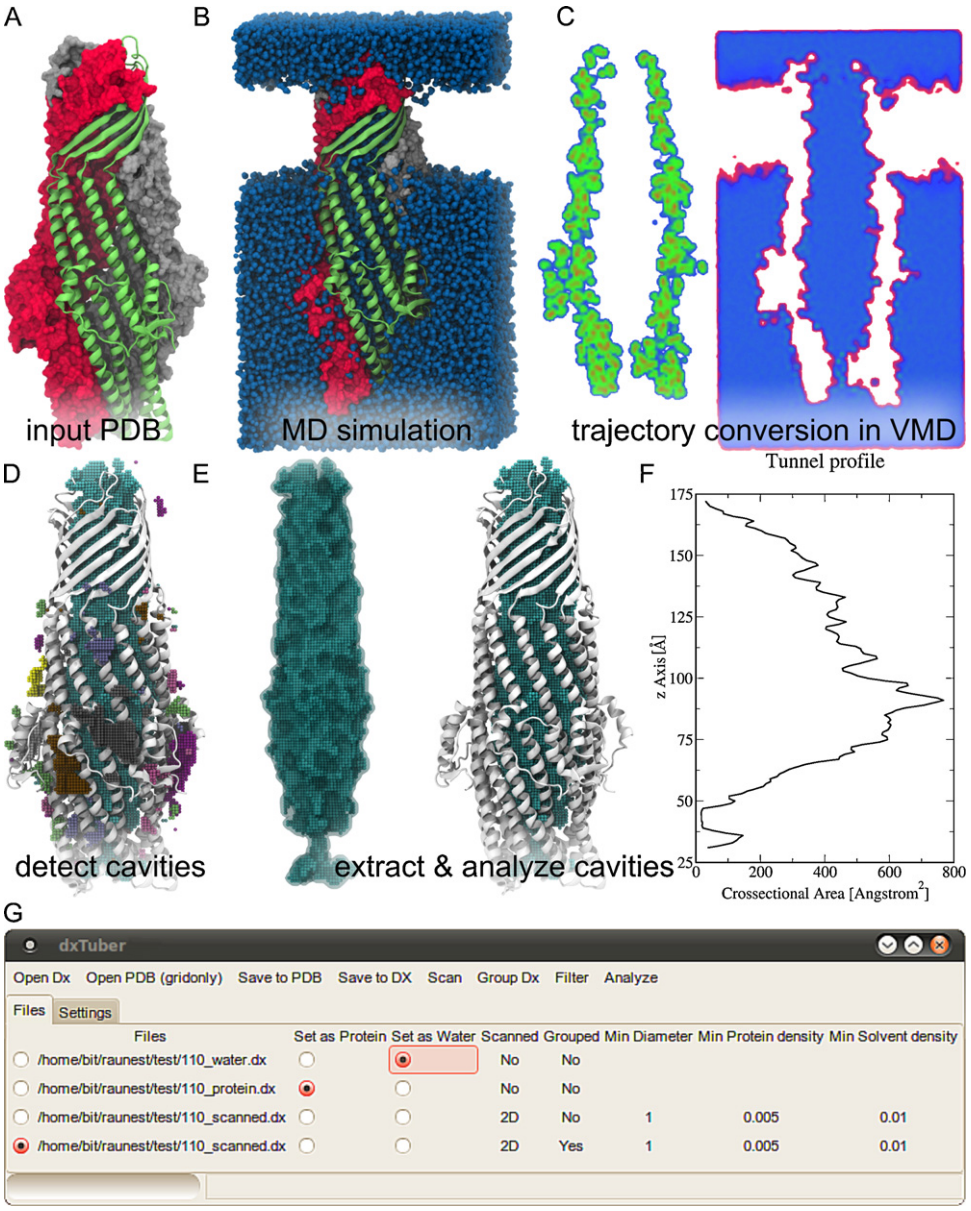

Fig. 1. dxTuber work flow. Starting from an experimentally determined protein structure (A) a molecular dynamics simulation is performed to sample protein and solvent dynamics (B). Trajectories are converted to voxel-based residence probabilities using the VolMap function in VMD [25] (C). The resulting time-averaged and mass-weighted OpenDX protein and solvent densities constitute the dxTuber input files used to separate inside from outside solvent densities to determine cavities, channels or surface clefts (D). Cavities can be exported as OpenDX or PDB files where each cavity is stored as an individually selectable group of voxels (E). Further analysis can be done computing cavity volume or profiles of cross sectional area along a principal axis (F). dxTuber is controlled via a graphical user interface (G).

All available techniques of cavity detection published to date have in common that they detect cavities based on a single protein conformation, neglecting any form of protein and cavity dynamics. Here we introduce a novel cavity detection method that uses protein and solvent residence probabilities derived from molecular dynamics (MD) simulations to detect cavities, tunnels and surface clefts. Using the VolMap plugin in VMD 1.8.7 [25], solvent and protein trajectories are converted to a voxel representation of OpenDX mass-weighed spatial density maps which serve as input files for our algorithm. The general work flow of the analysis procedure is summarized in Fig. 1. Driven by a graphical user interface, dxTuber then separates protein-internal from protein-external solvent voxels, describing cavities as groups of adjacent voxels that represent protein-internal regions of high solvent residence probability. To this end three different search algorithms have been implemented aiming at the detection of three major types of protein cavities: (I) isolated protein internal cavities with no connection to the protein exterior – both intra- and inter-molecular, (II) tunnel-like cavities with at least one connection to the protein exterior and (III) clefts on the protein surface.

dxTuber results can be exported in PDB format where each cavity is an individually selectable object of coherent voxels written as pseudo atoms. For each voxel solvent residence probability information is stored as averaged and mass-weighed solvent density encoded as formal B-factors. Cavities and solvent density can be readily visualized and processed using common molecular viewers. Cavities can also be analyzed in terms of volume and profiles of cross-sectional area along a principal axis. The profiles are exportable in ASCII file format, and can be processed by common spreadsheet applications. dxTuber is available free of charge for academic use at http://www.csb.bit.uni-bonn.de/downloads.html.

Table 1
Test proteins.

Protein	PDB	Cavity type
Photoactive yellow protein PYP [26]	2ZOH	Small intra-molecular cavity
Molybdenum storage protein MoSto [27]	2OGX	Large inter-molecular cavity
Tolerance to colicins TolC [28]	1EK9	Wide continuous tunnel
M2 proton channel [29]	2RLF	Narrow continuous/discontinuous tunnel(s)
BtuF periplasmic binding protein [30]	1N4D	Large ligand binding cleft
GTPase H-Ras P21 [31]	3K8Y	Small ligand binding cleft, partially occluded

To assess dxTuber performance we performed test runs on a set of six example proteins [26–31] representing each of the three cavity classes by a large and small variant (Table 1) and compared our findings to results obtained with SURFNET, CAVER and PyMol. While an in-depth comparison of all cavity detection tools would certainly be desirable, such a review is beyond the scope of this paper. In our benchmark tests we therefore restricted ourselves to these three cavity detection applications which we selected based on their wide spreading in the community and their frequent usage in our lab.

2. Materials and methods

2.1. *Creating DX input files*

2.1.1. *MD simulations*

To generate a sample of solvent and protein dynamics, MD simulations are carried out with the protein of interest solvated in, for example, water (Fig. 1A). The simulation time required depends on the diffusion velocity of the solvent and whether protein atoms are position-restrained. The usage of position restraints allows controlling the amount of protein dynamics to be probed. If the cavity structure close to the protein X-ray conformation is of interest, protein dynamics can be reduced or disabled using position restraints during the MD. Without position restraints, full protein and cavity dynamics will be considered. For water as solvent, simulation times of 100 ps where found sufficient to sample cavities, clefts and tunnels when position restraints of 1000 kJ mol^{-1} nm^{-2} were used to keep the protein close to its X-ray conformation. Due to extended equilibration times of protein and environment, unbiased MD requires longer simulation time, depending on the system of interest.

2.1.2. *VolMap options*

VolMap offers different options to convert MD trajectories into OpenDX density maps. Whereas “max” and “min” assign to each voxel the highest or lowest number of atoms detected at this position throughout the trajectory, “avg” yields a time-averaged value of the number of atoms per voxel (Fig. 2A). For the use with dxTuber we applied two combinations of settings: (a) “average” for both water and protein and (b) “average” for water and “minimum” for protein atoms (Table 2). The latter setting filters out regions of high protein flexibility, yielding the maximum volume accessible by water and thus the largest possible extension of a given cavity. To eliminate protein drift in the simulation box, trajectories should be aligned to a reference structure prior to VolMapping.

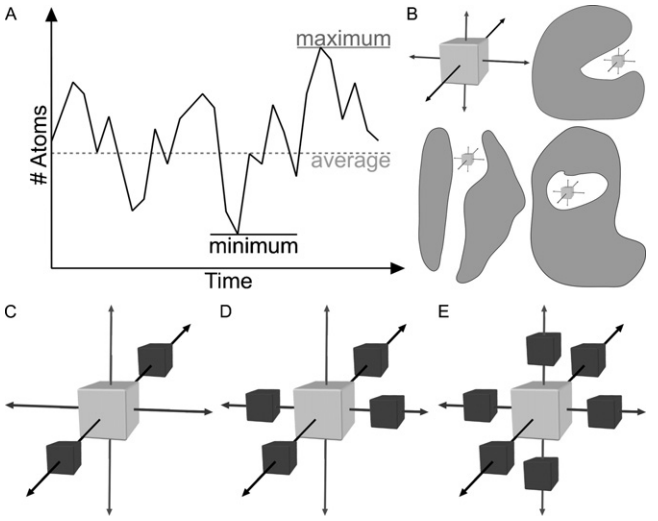

Fig. 2. To convert an MD trajectory to a voxel representation VMD can calculate the average, minimum or maximum atom density per voxel (A). Depending on the type of cavity to be detected (B), dxTuber offers three search algorithms using different criteria to decide whether a solvent voxel (grey) is considered protein-internal (C–E). Aiming at surface clefts, 1D scanning requires at least one protein voxel in *each* direction of *one* principal axis (C). Intended for tunnel-like cavities, 2D scanning requires at least one protein voxel (dark grey) in *each* direction of *two* principal axes (D), whereas 3D scanning expects at least one protein voxel in *each* direction of *three* principal axes before a solvent voxel is regarded protein-internal.

2.2. *Cavity detection*

2.2.1. *Scanning density voxels*

To separate protein-interior solvent densities from exterior ones (Fig. 2), three search algorithms have been implemented in dxTuber, each aiming at the detection of a different type of cavity (Fig. 2B). Whereas 3D scanning is intended for isolated buried cavities in the protein interior, and 2D scanning primarily aims for channel-like cavities that have a connection to the protein exterior, 1D scanning is recommended for clefts on the protein surface. In each method every solvent voxel – with a density exceeding 0.01 atoms/Å^3 – is tested if any protein voxels are located in both directions of the three principal axes. Voxel by voxel the algorithm proceeds away from the tested solvent voxel in posi-

Table 2
Molecular dynamics simulation, VolMap and dxTuber settings.

Protein	Simulation time	Used for cavity detection	Restraint on protein atoms	figure*	dxTuber settings	VolMap settings Protein|solvent densities	Filter
PYP	100 ps|10 ns	All|all	Yes|no	4	3D	Avg|avg	None
MoSto	100 ps	All	Yes	4	3D	Avg|avg	None
TolC	20 ns	All	Yes	5	2D	Avg|avg	None
TolC	21 ns	Last 10 ns	No	5	2D	Avg|avg	None
M2 (His37 deprotonated)	20 ns	Last 10 ns	No	6	2D	Avg and min|avg	None
M2 (His37 protonated)	20 ns	Last 10 ns	No	6	2D	Avg and Min|avg	None
BtuF	100 ps	All	Yes	7	1D (*Z*-axis)	Avg|avg	Neighbor 6
H-Ras P21	100 ps	All	Yes	7	2D	Avg|avg	None

tive and negative direction of the *X*, *Y* and *Z* principal axes until either a protein voxel is found or the end of the simulation box is reached. Depending on the scanning criterion used, different conditions of protein voxel distributions need to be fulfilled before a solvent voxel is considered protein-internal. For 1D scanning at least one protein voxel in *each* direction of *one* principal axis is required (Fig. 2C). For 2D scanning at least one protein voxel in *each* direction of *two* principal axes is necessary (Fig. 2D), whereas 3D scanning expects at least one protein voxel in *each* direction of all *three* principal axes before a solvent voxel is classified as protein-internal (Fig. 2E). 2D scanning is the default method and recommended for an initial scan. Protein voxels are identified based on a minimum density cut-off of 0.005 protein atoms/Å^3. This is the standard setting that we found to give good results; however, protein and solvent density thresholds are freely adjustable by the user.

After scanning, results may contain false positives, i.e. incorrectly assigned internal solvent voxels (ISVs) on the protein surface. As long as these are not connected to true positives, dxTuber can separate them via subsequent grouping. However, if false positives are connected to true positives, a more restrictive scanning method should be applied. In terms of restrictiveness dxTuber's scanning algorithms are ranked 3D > 2D > 1D. Beyond that, results can be post-processed by filtering.

2.2.2. Grouping cavities and analysis

Detected ISVs can be summarized into groups of voxels representing protein cavities. To that end dxTuber loops over the coordinate system of the ISVs starting in *Z* direction followed by *Y* and then *X* (Fig. 3A). The first ISV detected will become group one. Subsequent ISVs are tested based on their already visited neighborhood to decide whether they are part of a known group or constitute a new one (Fig. 3B). If ISVs belong to different groups that are later found to nevertheless have common neighbors, all respective voxels will be merged into a single group that is assigned the smaller (=earlier) group identification number (Fig. 3C). Once completed, ISVs have been clustered into individual cavities (Figs. 1D and 3B).

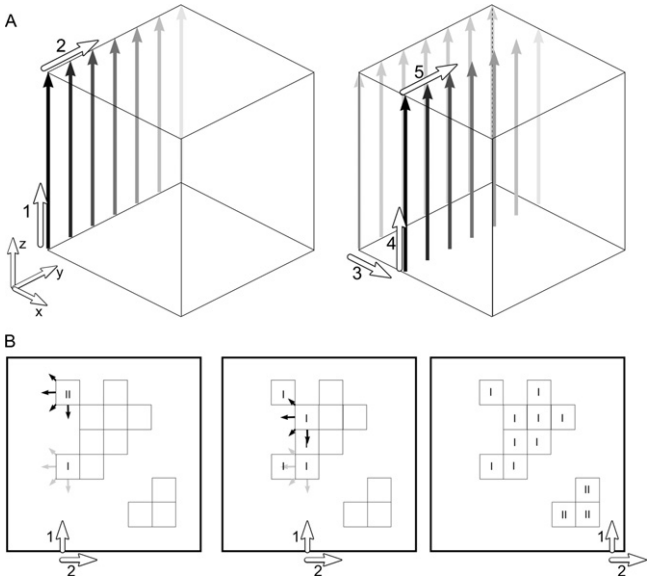

Fig. 3. To group protein-internal solvent voxels (ISVs) into cavities, dxTuber loops over the coordinates space spanned by the ISVs starting in *Z* direction, followed by *Y* and then *X* (A). The first ISV detected will become group one. Subsequent ISVs are tested based on their already visited neighborhood to decide whether they are part of a known group or constitute a new one (B, left). If ISVs belong to different groups that are later found to nevertheless have common neighbors (B, middle), all respective voxels will be merged into a single group that is assigned the smaller (=earlier) group identification number (B, right).

Once ISVs have been grouped into cavities, dxTuber can derive the cross-sectional area of a detected cavity along a selectable principal axis (*X*, *Y* or *Z*). To that end dxTuber loops over the chosen axis, counting in each step the number of ISVs in the plane perpendicular to the search axis. That way cross-sectional profiles can be generated which are exported as ASCII text files (Fig. 1F).

2.2.3. Filtering

Next to grouping or employing a more restrictive scanning algorithm, dxTuber results can be filtered to reduce the amount of false positives. In particular if more restrictive scanning is not successful due to incomplete cavity detection or if false positive ISVs are still connected to true positives subsequent filtering is recommended.

By applying the *neighbor filter* each ISV is re-evaluated based on its 26 neighbor voxels and will be deleted if the amount of neighbor ISVs is smaller than a user-defined minimum. The default for this minimum threshold is ten. The number of neighboring ISVs for the filter option should be set carefully, since filtering always leads to a loss of data as true positives not fulfilling the minimum neighbor criterion are also deleted. Therefore the threshold should be set to a low value (<10) to keep data loss to a minimum. The filter works best if two cavities of interest are connected by a thin line of voxels.

2.3. Example applications

2.3.1. Test proteins

Each of the three main classes of protein cavities is represented by two proteins, each featuring a large and small variant of a given cavity type. Whereas isolated internal cavities are represented by 2ZOH photoactive yellow protein PYP after chromophore removal [26] and the 2OGX molybdenum/tungsten storage protein MoSto [27], 1EK9 TolC [28] and the 2RLF Influenza M2 proton channel [29] serve as examples for tunnel-like cavities. Clefts on the protein surface are represented by the 1N4D periplasmic vitamin B12 binding protein BtuF [30] and the 3K8Y GTPase H-Ras P21 [31].

2.3.2. Simulation settings

Molecular dynamics simulations were performed using the GROMACS 4.0.3 package [32,33] and GROMOS96-53a6 force field [34]. In the simulations, all bond lengths were constrained by LINCS so that an integration time step of 2 fs could be chosen [35]. Systems were simulated at a temperature of 310 K, maintained separately for protein, lipids (where present) and water by a Berendsen thermostat [36] with a time constant of $\tau_T = 0.1$ ps. Pressure coupling was done using a Berendsen barostat [36] employing 1 bar reference pressure and a time constant of 4 ps. Whereas isotropic pressure coupling was employed for the water-soluble proteins PYP, MoSto, H-Ras P21 and BtuF, semiisotropic pressure coupling was used for the membrane proteins TolC and M2 to allow for bilayer fluctuations in the membrane plane. Electrostatic interactions were calculated using particle mesh Ewald (PME) summation [37,38] and twin range cut-offs of 1.0 nm and 1.4 nm were applied for computing the van-der-Waals interactions.

TolC and M2 were inserted in pre-equilibrated 9.6 nm × 9.6 nm and 6.8 nm × 6.8 nm palmitoyloleoyl-phosphatidylethanolamine (POPE) bilayer patches based on [39] using INFLATEGRO [40]. All systems were solvated with simple point charge water molecules [41]. Standard protonation states were assumed for titratable residues except for M2 which was simulated with His37 in both protonated and deprotonated form. An overview of simulation times and position restraints applied in each system is given in Table 2.

Table 3
SURFNET search parameters.

Protein	Grid size [Å]	Minimum probe radius [Å]	Maximum probe radius [Å]
PYP	0.8	1.5	5.5
MoSto	2	2.5	10
TolC	2	3	10
M2 (His37 deprotonated)	0.8	1.5	5.5
M2 (His37 protonated)	0.8	1.6	5.5
BtuF	0.8	3	10
H-Ras P21	0.8	1	10

Table 4
CAVER search parameters.

Protein	Start coordinates [distance to reference atoms]			Exit pathways
TolC	Thr163-CG2 chain A: 11 Å, B: 10.3 Å, C: 9.9 Å			2
M2 (His37 deprotonated)	Trp19-HZ2 chain A: 6.6 Å, B: 4.6 Å, C: 4 Å			5
	His15-CG chain A: 4.5 Å, B: 2.3 Å, C: 12.3 Å			2
M2 (His37 protonated)	Trp19-CG chain A: 5.8 Å, B: 7.5 Å, D: 3.5 Å			11
BtuF	Gln154-NE2: 1.7 Å Phe222-O: 2.9 Å Glu223-O: 2 Å			3
H-Ras P21		1st coordinate	2nd coordinate	1 from 1st coordinate
	GppNHp-PB:	2.0 Å	2.1 Å	2 from 2nd coordinate
	GppNHp-PG:	1.1 Å	1.0 Å	
	GppNHp-N3B:	1.6 Å	1.2 Å	

2.3.3. dxTuber, SURFNET, CAVER, and PyMol settings

All dxTuber analyses were run on *average* protein and water densities calculated at 1 Å^3 resolution via VolMap in VMD 1.8.7. For M2, additional *minimum* protein densities were computed whereas for PYP and TolC additional density map resolutions of 0.5 and 2.0 Å^3 were calculated to assess how voxel resolution affects calculation speed. dxTuber search algorithms and filters used are detailed in Table 2. SURFNET [8] has three main user-adjustable parameters affecting the cavity search: grid size, minimum gap sphere radius and maximum gap sphere radius. To keep the number of false positives to a minimum these parameters were optimized for each protein (Table 3). Main cavities were isolated using the SURFNET tool MASK. As CAVER [42] is designed to find a path that leads out of the protein interior, cavity analysis requires a set of starting coordinates inside the protein as well as the number of how many different exit pathways to look for. Both parameters were optimized for each protein to describe the cavities of interest as accurate as possible (Table 4). For better comparison with dxTuber, CAVER's radius-based tunnel profiles were converted to cross-sectional area profiles. Implemented since version 1.3, cavity detection in PyMol [4] is based on the MSMS molecular surface algorithm [43,44] and controlled via the search parameters cavity detection radius and cavity detection cut-off. The settings used for each protein are listed in Table 5. All cavity calculations were performed using a single Intel® Xeon® E5410 2.33 GHz CPU on a 4 GB RAM DELL Precision T5400 workstation running 32 bit Ubuntu Linux 10.04.

Table 5
PyMOL search parameters.

Protein	Detection radius [Å]	Detection cutoff radius [Å]
Pyp	7	3
MoSto	6	4
TolC	20	7
M2 (His37 deprotonated)	7	4
M2 (His37 protonated)	8	4
BtuF	20	4
H-Ras P21	20	3

All results obtained with CAVER and SURFNET and PyMol are based on the reference X-ray structures, except for M2: representative for the channel's open and closed state, simulation snapshots at 18.7 ns (open state) and 13.9 ns (closed state) were used. The conformations were selected based on their minimum Cα-RMSD to the simulation average structures calculated over the last 10 ns of the trajectories. All molecular illustrations were created using VMD [25] or PyMol [4].

3. Results

We developed a novel cavity detection method that uses protein and solvent residence probabilities derived from molecular dynamics simulations to detect isolated internal cavities, tunnels and surface clefts. To assess dxTuber performance we carried out test runs on six example proteins and compared our findings to results obtained with SURFNET, CAVER and PyMol.

3.1. Isolated internal cavities

Specifically designed to find a path out of the protein interior CAVER does not support this type of cavities and has therefore been excluded from this section. PyMol does not provide any quantitative volume information on detected cavities.

The molybdenum storage protein (MoSto) [27] features an inter-molecular cavity, large enough to accommodate up to 100 molybdenum (Mo) or tungsten (W) atoms [27]. Whereas SURFNET, dxTuber and PyMol were all able to detect the cavity, there are differences in cavity shape and volume. In SURFNET the cavity is represented having a smooth shape and a volume of 10720 Å^3 (Fig. 4B), whereas in dxTuber a volume of 9223 Å^3 is calculated and the cavity has an overall less regular shape revealing more structural detail (Fig. 4A). PyMol gives a detailed view of the storage cavity but also finds false positives on the protein surface which are partly connected to the cavity. In this example restrained protein dynamics were used for the dxTuber analysis, yielding a Cα-RMSD of 0.3 Å in the simulation end conformation.

Photoactive yellow protein (PYP) is a 125 residue signaling blue-light photo receptor [26] whose covalently bound chromophore

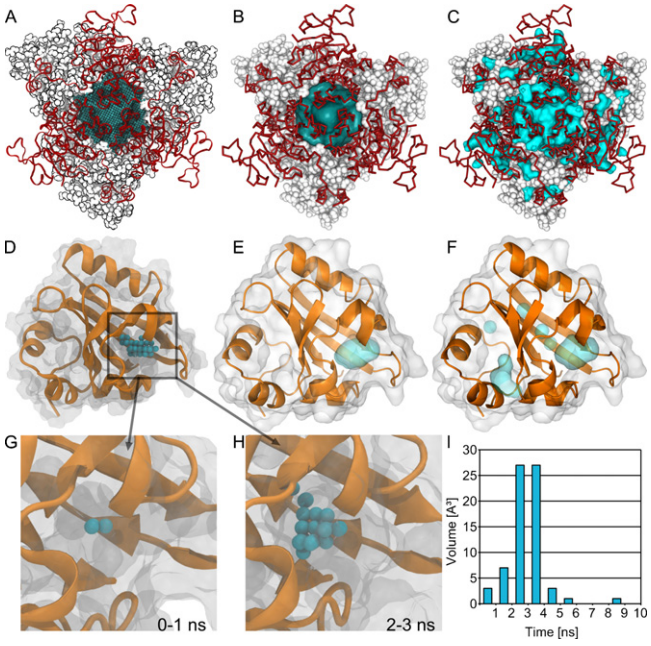

Fig. 4. Isolated internal cavities with no connection to the protein exterior are represented by the molybdenum/tungsten storage protein MoSto (A–C) and photoactive yellow protein PYP (D–I). Whereas MoSto serves an example for a large intermolecular cavity, PYP after chromophore removal represents small intra-molecular cavities. Each protein's main cavity is depicted as detected by dxTuber with (A and D) and without position restraints on the protein atoms (G and H), SURFNET (B and E) and PyMol (C and F). As cavity separation is not implemented in PyMol, (C) and (F) also show secondary cavities and surface clefts. When PYP dynamics are unrestrained, the main cavity fluctuates in size and volume and disappears after 6 ns (G–I).

(para-coumaric acid, pCA) was removed to create a cavity representative for small intra-molecular cavities. For dxTuber analysis both restrained (Fig. 4D) and unbiased protein dynamics were investigated (Fig. 4G–I) yielding Cα-RMSDs of 3.5 Å and 0.3 Å in the end conformations. For the unbiased simulation, cavity volume was monitored throughout the trajectory using time windows of 1 ns (Fig. 4I). dxTuber (Fig. 4D), SURFNET (Fig. 4E), and PyMol (Fig. 4F) find the cavity, and whereas PyMol additionally detects five independent cavities with the main cavity of interest being similar to SURFNET. In dxTuber the cavity is smaller (50 Å^3 volume versus 152 Å^3 as reported by SURFNET) and also more extended by 2 Å towards the center of the protein. The innermost section of the dxTuber cavity coincides with a secondary cavity detected in PyMOL that is absent in SURFNET. Without position restraints on the protein, the cavity fluctuates in size and volume (Fig. 4G and H) and disappears after 6 ns (Fig. 4I).

3.2. Tunnel-like cavities

In this section the calculation of cross-sectional profiles along the membrane normal was restricted to CAVER and dxTuber, as SURFNET and PyMol do not include this feature.

The outer membrane protein TolC functions as a major efflux duct in the outer membrane of *Escherichia coli* and at a length of 140 Å and 20 Å in diameter, the TolC channel is large enough to accommodate and transport entire proteins [45]. dxTuber analysis was performed using both restrained (Fig. 5A) and unrestrained protein dynamics (Fig. 5B) yielding end Cα-RMSDs of 0.3 Å and 2.5 Å respectively. While all four applications detect the main cavity, the results differ in size and shape (Fig. 5A–E). In SURFNET the channel ends 10 Å before TolC's lower (periplasmic) exit (Fig. 5C), whereas in dxTuber, the cavity is detected completely only when restrained protein dynamics are applied (Fig. 5A). When unrestrained protein dynamics are used the channel cavity ends 15 Å before the upper (extracellular) exit (Fig. 5A). In CAVER the channel is narrowest of all five cases (Fig. 5D) and requires two tunnels set in the CAVER input parameters to be detected completely. If only a single tunnel is used, a similar effect as with SURFNET is seen with the lower 10 Å of the cavity missing. PyMol (Fig. 5E) describes the main cavity in a similar way as SURNET, however the tunnel is shorter in the upper (extracellular) exit region but longer towards the lower (periplasmic) exit (approximately half way between the SURFNET and CAVER findings). Additionally PyMol generates false positives on the protein surface at the TolC equatorial region.

Compared to CAVER (Fig. 5G) the channel's profile of cross-sectional area is larger in dxTuber (Fig. 5F), exhibiting a continuous progression with a maximum peak at 85 Å. The CAVER profile is discontinuous displaying abrupt jumps and a maximum peak at 100 Å. While the dxTuber profiles are similar to each other, the ones obtained with restrained protein dynamics are larger than those derived from unrestrained MD (Fig. 5F).

With a trans-membrane channel 30 Å long and only a few water molecules wide, the pH-gated Influenza proton channel M2 [29] serves as an example for a narrow tunnel-like cavity. M2 switches between two conformations where extracellular and intracellular water phases are either connected (protonated His37, open state) or separated (deprotonated His37, closed state) (Fig. 6). Whereas closed state M2 was captured in the NMR structure, a transition to the open state could be observed within 10 ns MD simulation when His37 was considered in protonated form. No position restraints were applied throughout the simulations. The M2 end conformations exhibit a total Cα-RMSD of 5.3 Å over all four

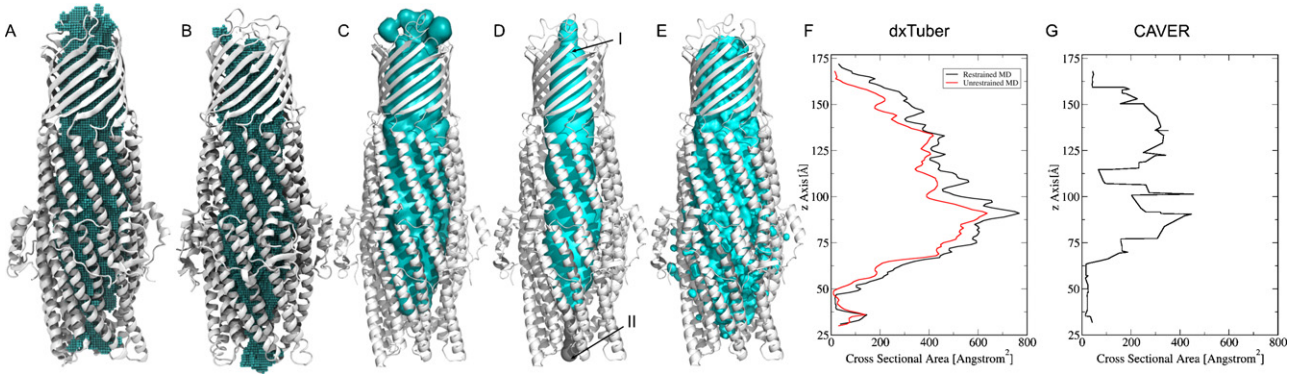

Fig. 5. The bacterial efflux duct TolC is an example of a large tunnel-like cavity whose main channel is shown as depicted by dxTuber with (A and F, black tunnel profile) and without position restraints on the protein atoms (B and F, red tunnel profile), SURFNET (C), CAVER (D and G) and PyMol (E). In CAVER two exit tunnels had to be calculated to represent the TolC channel completely. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

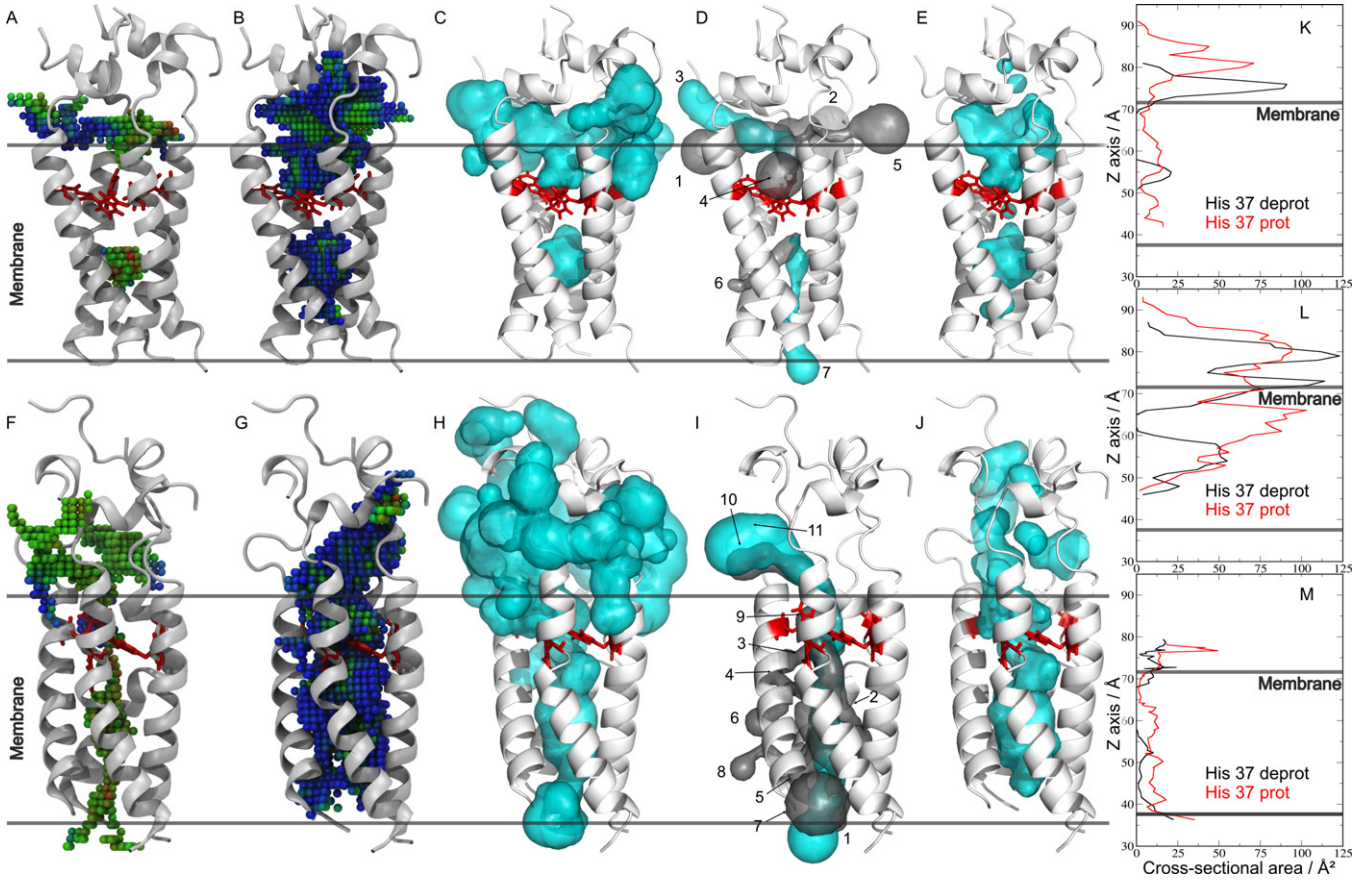

Fig. 6. The pH-gated Influenza M2 proton channel serves as an example of a narrow tunnel-like cavity. Controlled by His37, the M2 tunnel either occurs as two separate half channels when His37 is deprotonated (A–E) or as single channel when His37 is protonated (F–J). The M2 channel is depicted as detected by dxTuber showing average (A, F and K) and maximum cavity extension (B, G and L) colored by solvent density (blue = low density, red = high density), SURFNET (C and H), CAVER (D, I, M) and PyMol (E, J). Whereas closed state M2 required the calculation of seven CAVER exit tunnels (D), eleven exit tunnels were necessary for the open state (I). CAVER tunnels used for the profiles of cross-sectional area appear cyan.

monomers when His37 was protonated, and 4.4 Å when His37 was deprotonated.

All four applications identify separated cavities for closed state M2 – 2 in dxTuber (Fig. 6A and B), SURFNET (Fig. 6C) and CAVER (Fig. 6D) and 4 in PyMol (Fig. 6E) – and a continuous tunnel cavity for open state M2 (Fig. 6F–J). Results differ in (a) channel narrowness – ranging from dxTuber using average protein density (Fig. 6A and F) and CAVER (Fig. 6D and I) up to dxTuber run on minimum protein density (Fig. 6B and G) – and (b) the amount of false positives, i.e. cavity sections extruding to the protein exterior. This effect is most pronounced in SURFNET (Fig. 6C and H) and, to a lesser degree, in CAVER (Fig. 6D and J) which required seven tunnels for the closed state and eleven tunnels for open state M2 to find the cavities of interest.

Apart from the exit regions the tunnel profiles are similar in CAVER (Fig. 6M) and dxTuber run on average protein density (Fig. 6K). When dxTuber is run on protein minimum density, the channel is wider, the difference between open and closed state is more pronounced and the bottle neck region in closed state M2 is 5 Å shorter (Fig. 6L).

3.3. Surface clefts

In the periplasmic vitamin B12 binding protein BtuF [30] substrate binding occurs in a wedge-shaped, un-occluded binding cleft that is approximately 20 Å wide and 15 Å deep when the protein is in its closed state [46]. The cleft is detected by dxTuber (Fig. 7A), SURFNET (Fig. 7B) and partially by CAVER (Fig. 7C). PyMol

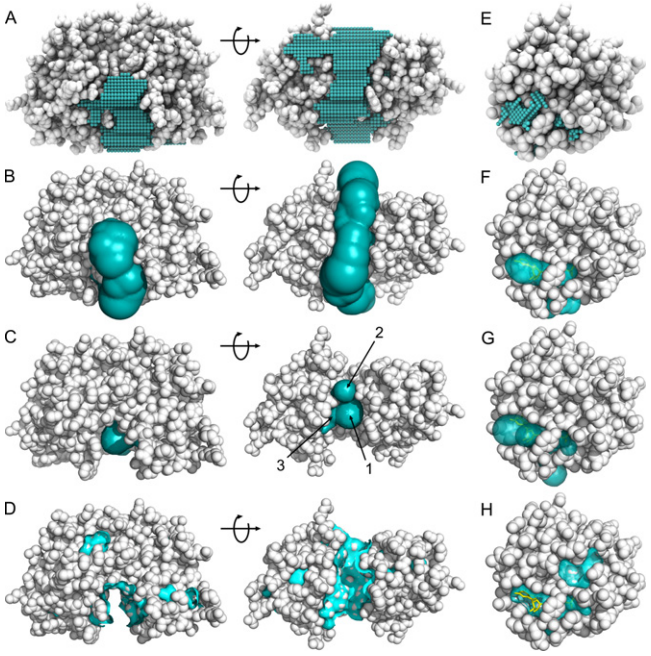

Fig. 7. The large un-occluded vitamin B12 binding cleft in BtuF (A–D) and the small partially occluded GTP binding niche in H-Ras p21 (E–H) serve as examples for surface clefts. Clefts are shown as detected by dxTuber (A and E), SURFNET (B and F), CAVER (C and G) and PyMol (D and H). Three CAVER tunnels were calculated for BtuF (C) and four for H-Ras p21 (H).

only detects the borders of the binding pocket; while additionally producing false positives lining the main binding cleft (Fig. 7D). Although dxTuber produces the most detailed rendering of the binding cleft, it incorrectly includes adjacent grooves in the surface of the N-terminal and C-terminal domains. In SURFNET the cleft appears hose-like protruding by up to 9 Å on both sides of the binding cleft whereas in CAVER, only a fraction of the cleft is detected calculating 3 exit pathways. Restrained protein dynamics were employed during the dxTuber MD leading to an end Cα-RMSD of 0.4 Å.

The partially occluded GTP binding niche of the activated signal transduction protein H-Ras P21 [31] concludes our set of example applications. All four tools detect the entire binding cleft (Fig. 7E–H), with SURFNET (Fig. 7F) and CAVER (Fig. 7G) yielding almost identical results of hose-shaped cavities whereas dxTuber and PyMol show more structural detail of the binding cleft (Fig. 7E and H). Three tunnels were computed in the CAVER analysis and restrained protein dynamics were employed for dxTuber resulting in an end Cα-RMSD of 0.3 Å.

3.4. dxTuber calculation speed

Representing the smallest and the largest of our test proteins, the PYP and TolC analyses were additionally carried out at 0.5 and 2.0 $Å^3$ voxel size to assess how voxel resolution influences the calculation speed. At 0.5 $Å^3$ the PYP analysis required 3:53 min for trajectory conversion in VMD and 5:42 min for cavity detection in dxTuber. At 1.0 $Å^3$ resolution the analysis took 0:29 (VMD) plus 0:19 min (dxTuber) and 0:08 plus 0.02 min at 2.0 $Å^3$ voxel size. For TolC, the analysis was complete after 24:54 min in VMD plus 130:05 min in dxTuber when a voxel size of 0.5 $Å^3$ was used. At 1.0 $Å^3$ resolution the analysis required 3:03 min plus 6:17 min and 0:35 min plus 0:18 when a voxel size of 2.0 $Å^3$ was selected.

4. Discussion

Proteins are in a continuous state of motion that can easily exceed mere thermal fluctuations. To take into account the element of conformational dynamics which has so far been neglected by other cavity detection tools, we have developed a novel method employing protein and solvent residence probabilities obtained from molecular dynamics simulations to detect and analyze cavities, tunnels and surface clefts. After comparing dxTuber's performance to SURFNET, CAVER and PyMol we proceed discussing our approach's methodological characteristics in context of the other three methods. We conclude proposing general proceeding recommendations for cavity detection.

4.1. Performance and methodological characteristics

Each of the three main classes of protein cavities presents particular challenges to cavity detection and the six example proteins we used as representatives might raise the question whether the number of test cases is sufficient. While it is naturally desirable to have as many test cases as possible, the six proteins used in this study are nevertheless likely to give a reasonable representation of the palette of protein cavities. Given that each class is exemplified by a small and a large variant, representing extreme sizes of each cavity type, it is reasonable to assume a tool's performance in these extreme cases is likely transferable to the range of intermediate cases in between. Future applications to other proteins of different sizes will show whether if this assumption is valid.

Whereas isolated cavities with no contact to the protein exterior are detected equally well by SURFNET, PyMol and dxTuber (Fig. 4), the CAVER algorithm does not support this type of cavity but performs expectedly better with tunnel-like cavities as long as multiple pathways are calculated (Figs. 5 and 6). PyMol and dxTuber also yield comparably good results here, whereas SURFNET did not detect the entire tunnel in case of TolC (Fig. 6) or produced a large amount of false positives on the protein surface in case of M2 (Fig. 7). Whereas H-Ras p21's small and partially occluded GTP binding cleft was detected by all four applications (Fig. 7E–H), the large, un-occluded vitamin B12 binding cleft in BtuF (Fig. 7A–D) was only detected by SURFNET and dxTuber although in both cases the detected cavity exceeds the actual binding cleft.

4.1.1. SURFNET

Each strategy to cavity detection bears its own limitations. Using gap spheres [8] placed on a 3D grid, SURFNET results depend on the orientation of the protein of interest. Common to all grid-based algorithms, this limitation should be kept in mind particularly when comparing different conformations of the same protein – such as open and closed state M2 (Fig. 7). To minimize this artifact, structures should be aligned to the same reference conformation prior to cavity analysis. When larger cavities like the storage compartment in MoSto (Fig. 4) are of interest, SURFNET's hard-coded limit of 120,000 initial gap spheres must be compensated at the cost of a lower spatial resolution by making the search grid coarser. In case of MoSto we additionally increased the minimum gap sphere radius to ensure the entire cavity is detected completely. The overall smoothness of cavity shape, the general lack of fine structural detail as well as the protruding BtuF ligand binding cleft (Fig. 7B) are another consequence of SURFNET's gap sphere technique that is also seen with other probe-sphere employing methods like CAVER or PyMol. Although the SURFNET standard settings are a good starting point, grid size, minimum and maximum gap sphere radius should be optimized for each protein to achieve the best results (Table 3). Often this requires a trade-off between completeness of cavity detection and the amount of false positives. For example, although a complete detection of the entire TolC channel (Fig. 5C) is possible using a smaller maximum gap sphere radius, the channel cavity then merges with the clefts on the outer protein surface into one inseparable entity under these conditions. On the other hand, for the narrow channel in M2 (Fig. 6C and H) a larger maximum gap sphere radius eliminates not only the false positives on the outer surface but also the tunnel cavity itself. Any SURFNET analysis is restricted to the 50 largest cavities and although initially handled as one object, individual cavities can be manually separated into individual isosurfaces using the MASK script of the SURFNET package. Although SURFNET calculates cavity volume, pore profiles cannot be computed.

4.1.2. CAVER

Specifically designed to find the shortest way out of the protein interior along a 3D grid from a user-defined starting point, the CAVER algorithm [14] is not applicable to class I cavities (Fig. 4). Whereas the large, un-occluded surface cleft in BtuF was only detected partially (Fig. 7C), the smaller and partially occluded GTP binding niche in H-Ras p21 was rendered correctly (Fig. 7G). However, like with the TolC (Fig. 5D) and M2 channels (Fig. 6D and I), this requires calculating several exit tunnels and/or using multiple starting points (Table 4). This is a consequence of (a) the 3D grid constructed from the user-defined starting point and (b) the cost function CAVER employs to calculate the exit pathway. Aimed at maximizing the tunnel radius as quickly as possible, the cost function favors an increase of the tunnel radius. In each CAVER run several exit tunnels are calculated and the actual output tunnel is selected based on the amount of favored steps of radius increase and disfavored steps of radius decrease. As seen with TolC and M2, this is problematic with bottleneck regions. One way to sample and output this disfavored regions is by placing the starting point into the bottleneck and always calculate more than one exit tun-

nel. While two tunnels were sufficient for TolC (Fig. 5D), eleven output tunnels had to be calculated for open state M2 (Fig. 6I). CAVER tunnel profiles approximate cross-sectional area by the radius of a tunnel-filling probe sphere. Common to all tools using that approach, this leads to inaccuracies with tunnel cavities whose cross-sectional profile is not circular [10]. Furthermore, in case of TolC we observed jumps in the CAVER pore profile (Fig. 5G) suggesting the presence of a bottleneck when there is none. For reasons we do not understand yet, the CAVER path here temporarily deviates from the tunnel center and approaches the channel wall where only the placement of small probe spheres is sterically possible.

4.1.3. PyMol

Cavity detection is a recent addition to the PyMol molecular viewer [4] and is based on the MSMS molecular surface algorithm [43,44]. Cavities cannot be further analyzed nor separated into individually selectable objects, and the large BtuF surface cleft was not detected (Fig. 7D). Nevertheless, fast and robust results were yielded for all other example proteins (Figs. 4C and F, 5E, and 6E and J) when the search parameters cavity detection radius and cavity detection cut-off were manually optimized for each protein (Table 5).

4.1.4. dxTuber

dxTuber is the first cavity detection tool that fully takes into account protein flexibility. This is done on the grounds of protein and solvent residence probability distributions derived from molecular dynamics simulations using solvent molecules to probe for cavities. Achieving a more realistic and accurate representation of protein and cavity structure dxTuber allows to study cavities within the framework of the dynamic protein.

Whereas statistics data on cavities can be difficult to obtain in other methods, the dxTuber approach offers an easy way of doing that. For each cavity its average and minimum or maximum extension can be calculated (Figs. 2A and 6A, B, F, and G) based on (a) how protein and solvent trajectories were converted into residence probability distributions or (b) which residence probability cut-off is used for cavity display using the averaged and mass-weighed solvent density per voxel encoded as formal B-factor in the exported PDB. When multiple trajectory time windows are chosen for analysis (Fig. 4G–I) cavity dynamics can also be computed over time, monitoring profiles of cross-sectional area or cavity volume as in Fig. 4I.

As the choice which molecules are used for cavity probing is user-defined, dxTuber offers a high degree of application flexibility. For example, whereas assigning a prevalent solvent like water allows a more general cavity search, less widespread molecules such as bound ligands can be used for high-precision mapping of substrate binding sites or to determine preferred ligand–protein interaction sites. Recent MD studies where the dxTuber approach could be useful include the mapping of protein-internal water distributions [47–49], monitoring the opening state of transport proteins [50,51], investigating protein-internal gas diffusion pathways [52,53] or exploring ligand binding sites [54].

Unlike many other cavity detection tools specialized in one or two cavity classes, dxTuber is capable of reliably detecting all three main types of cavities where – apart from initial choices of protein orientation, probing solvent and scanning algorithm – no fine tuning of search parameters is required. Controlled by a graphical user interface, dxTuber also offers a high degree of accessibility and usability. A common problem in cavity detection is the occurrence of false positives – i.e. additional cavities and surface grooves not in the current focus of analysis – and the necessity to separate these from the actual cavities of interest for later visualization or analysis. As detected cavities can be exported in PDB format where cavities are stored as individually selectable groups of voxels sharing the same atom name, dxTuber offers an easy and efficient way to identify and extract cavities of interest.

Relying on molecular dynamics simulation trajectories to probe protein cavities, dxTuber's major limitation is the amount of effort necessary to perform a cavity analysis. Next to simulation length and system size, the total time required for a dxTuber analysis is determined by the voxel resolution used in trajectory conversion and cavity detection. Whereas for a small system like PYP analysis times range from seconds to minutes using voxel resolution of 2.0 $Å^3$, 1.0 $Å^3$ and 0.5 $Å^3$, a large system like TolC requires seconds to hours depending on voxel sizes applied. While a coarser resolution might be beneficial for a quick assessment of large cavity structures, in our test runs we did not find the increased resolution of 0.5 $Å^3$ leading to considerable improvements over 1.0 $Å^3$ resolution which we recommend as standard setting.

Another limitation is that per definition only those cavities will be detected that actually contain solvent molecules designated for cavity probing. Whereas using water as the only solvent in our test runs was unproblematic, additional solvent types might become necessary to probe for, for example, hydrophobic compartments in a protein, micelle or a lipid bilayer. Pore profiles in dxTuber are not computed via the common tunnel-filling probe sphere approach used in [12,13,15,42]. Instead the cross-sectional area is calculated as an area sum of protein-internal solvent voxels sharing the same Z coordinate. While this leads to a higher accuracy and detail in tunnel profiles (Figs. 5F and 6K and L) independent of a channel's cross-sectional shape, the approach is restricted to approximately linear tunnels running parallel to a principal axis. A potential way to overcome this limitation could be employing a slicing plane through the cavity voxels that is aligned perpendicular to the local tunnel normal. On the slicing plane local area, diameter or circumference of the tunnel could then be measured. Furthermore the exit regions of channels or surface clefts should be interpreted carefully as they might be represented incompletely due to the grid approach's dependence on the initial orientation of protein and cavity. When such cavities are of interest (Fig. 7A) and need to be represented completely, we recommend orienting the protein in such a way that the cavity is aligned to the X, Y or Z axes. Although dxTuber generated a complete rendition of the BtuF substrate binding cleft that way (Fig. 7A), smaller adjacent surface clefts were also detected that could not be separated or eliminated automatically. This is a limitation of the employed 1D scanning algorithm which is currently the only way to represent large surface clefts completely. Since all cavities can be exported in PDB format, we suggest selecting and removing such artifact voxels manually. On the other hand, the different lengths of TolC tunnel profiles derived from restrained (Fig. 5F, black profile) or unrestrained protein dynamics (Fig. 5F, red profile) are not a limitation of the algorithm but instead result from a partial closure of the protein in the unrestrained MD simulation.

Similarly, the fluctuating and eventually disappearing cavity dxTuber detects in unrestrained PYP (Fig. 4G–I) reflects an adaption of the protein to deleted chromophore and an expulsion of initially cavity-internal water molecules. On the other hand, when PYP is position-restrained (Fig. 4D) the cavity is stable but smaller than in SURFNET (Fig. 4E) or PyMOL(Fig. 4F). When protein position restraints are applied during the MD, one should keep in mind that force constants of 1000 kJ mol^{-1} nm^{-2} still allow modest protein fluctuation – as reflected by Ca-RMSDs between 0.3 and 0.4 Å – that can limit water diffusion as seen with the MoSto or PYP dxTuber cavity volumes (Fig. 4A and D) which are 1497 $Å^3$ and 102 $Å^3$ smaller than their SURFNET counterparts (Fig. 4B and E). To completely eliminate any form of protein motion we recommend a force constant of 100,000 kJ mol^{-1} nm^{-2}.

4.2. General proceeding recommendations

For a first impression of a protein with unfamiliar cavity structure, we found PyMol a good starting point, yielding fast and robust results for the majority of cavity classes while requiring little effort in optimizing cavity search parameters. For a more in-depth analysis of a protein with familiar cavity structure there are two scenarios to be considered.

When molecular dynamics simulation data are not available, SURFNET gives robust results for isolated internal cavities or surface clefts whereas CAVER is a reasonable choice for tunnel-like cavities. In both cases each method's individual limitations should be considered as discussed above. It is also beneficial to perform each analysis in combination with an initial PyMol search. That way useful information is gained regarding (a) cavity types and choice of search algorithm, (b) cavity orientation and shape which determine the placement of starting points, grid size as well as minimum and maximum gap sphere radius and (c) validation of results and search parameters.

When molecular dynamics simulation data are available and conformational dynamics are to be taken into account, dxTuber is currently the only cavity detection tool available yielding excellent results for all cavity classes bearing in mind the specific limitations discussed above. Again, a prior PyMol analysis can provide useful information regarding the choice of search algorithm or protein and cavity orientation, but is also useful to inspect a cavity's hydration state prior to MD simulation, and as a means for general cross-checking. To explore cavity structure near the X-ray conformation, protein atoms should be position-restrained (Figs. 4A and D, 5A, and 7A and E) whereas unbiased MD is recommended to investigate protein and cavity dynamics (Figs. 4G–I, 5B, and 6A and F). Depending on whether a cavity's average, minimum or maximum extension is of interest, trajectory to voxel conversion in VMD should be carried out using the respective setting for the calculation of protein and water densities.

5. Conclusions

We introduce a novel cavity detection method that fully takes into account cavity dynamics using protein and solvent residence probabilities derived from molecular dynamics simulations. Once MD trajectories have been converted to a voxel representation of mass-weighed spatial density maps using VMD, dxTuber identifies cavities as groups of coherent voxels representing protein-internal regions of high solvent density. Three different search algorithms have been implemented that yielded excellent results on six test proteins representing the three main classes of protein cavities: isolated internal cavities, tunnels and surface clefts. dxTuber results can be exported in PDB format where each cavity is an individually selectable object of coherent voxels written as pseudo atoms, each holding the residence probability information in form of averaged and mass-weighed solvent density encoded as formal B-factors. Cavities can be (a) readily visualized and processed using common molecular viewers and (b) further analyzed in terms of volume and profiles of cross-sectional area along a principal axis. dxTuber is available free of charge for academic use at http://www.csb.bit.uni-bonn.de/downloads.html.

Acknowledgements

We thank Dr. Frank Wennmohs, Nadine Fischer, Thomas Schmidt and Lars Lüdicke for helpful comments and extensive beta testing. This work was financially supported by the Ministerium für Innovation, Wissenschaft, Forschung und Technologie des Landes Nordrhein-Westfalen. ChK is a junior research group leader funded by the NRW Rückkehrerprogramm.

References

[1] P. Schmidtke, C. Souaille, F. Estienne, N. Baurin, R.T. Kroemer, Large-scale comparison of four binding site detection algorithms, J. Chem. Inf. Model. 50 (2010) 2191–2200.
[2] A. Volkamer, A. Griewel, T. Grombacher, M. Rarey, Analyzing the topology of active sites: on the prediction of pockets and subpockets, J. Chem. Inf. Model. 50 (2010) 2041–2052.
[3] R.G. Coleman, K.A. Sharp, Finding and characterizing tunnels in macromolecules with application to ion channels and pores, Biophys. J. 96 (2009) 632–645.
[4] W.L. DeLano, The PyMOL Molecular Graphics System, 2002.
[5] T. Exner, M. Keil, G. Moeckel, J. Brickmann, Identification of substrate channels and protein cavities, J. Mol. Model. 4 (1998) 340–343.
[6] B.K. Ho, F. Gruswitz, W. Hollo, generating accurate representations of channel and interior surfaces in molecular structures, BMC Struct. Biol. 8 (2008) 49.
[7] G.J. Kleywegt, T.A. Jones, Detection, delineation, measurement and display of cavities in macromolecular structures, Acta Crystallogr. D: Biol. Crystallogr. 50 (1994) 178–185.
[8] R.A. Laskowski, SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions, J. Mol. Graph. 13 (1995) 323–330, 307–308.
[9] J. Liang, H. Edelsbrunner, C. Woodward, Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design, Protein Sci. 7 (1998) 1884–1897.
[10] M.H. Maeda, K. Kinoshita, Development of new indices to evaluate protein–protein interfaces: assembling space volume, assembling space distance, and global shape descriptor, J. Mol. Graph. Model. 27 (2009) 706–711.
[11] A. Nicholls, R. Bharadwaj, B. Honig, GRASP graphical representation and analysis of surface properties, Biophys. J. (1993) 64.
[12] M. Pellegrini-Calace, T. Maiwald, J.M. Thornton, PoreWalker:, a novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure, PLoS Comput. Biol. 5 (2009) e1000440.
[13] M. Petrek, P.n. KosinovÃ, J. Koca, M. Otyepka, Mole: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels, Structure 15 (2007) 1357–1363.
[14] M. Petrek, M. Otyepka, P. BanÃs, P.n. KosinovÃ, J. Koca, J. DamborskÂ, CAVER: a new tool to explore routes from protein clefts, pockets and cavities, BMC Bioinformatics 7 (2006) 316.
[15] O.S. Smart, J.G. Neduvelil, X. Wang, B.A. Wallace, M.S. Sansom, Hole: a program for the analysis of the pore dimensions of ion channel structural models, J. Mol. Graph. 14 (354–60) (1996) 76.
[16] N.R. Voss, M. Gerstein, 3V: cavity, channel and cleft volume calculator and extractor, Nucleic Acids Res. 38 (Suppl.) (2010) W555–W562.
[17] E. Yaffe, D. Fishelovitch, H.J. Wolfson, D. Halperin, R. Nussinov, MolAxis: efficient and accurate identification of channels in macromolecules, Proteins 73 (2008) 72–86.
[18] M.L. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, Science 221 (1983) 709–713.
[19] M.L. Connolly, Analytical molecular-surface calculation, J. Appl. Crystallogr. 16 (1983) 548–558.
[20] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–400.
[21] F.M. Richards, Areas, volumes, packing and protein structure, Annu. Rev. Biophys. Bioeng. 6 (1977) 151–176.
[22] H. Edelsbrunner, E.P Mucke, 3-Dimensional alpha-shapes, ACM Trans. Graph. 13 (1994) 43–72.
[23] A. Poupon, Voronoi and Voronoi-related tessellations in studies of protein structure and interaction, Curr. Opin. Struct. Biol. 14 (2004) 233–241.
[24] F.M. Richards, The interpretation of protein structures: total volume, group volume distributions and packing density, J. Mol. Biol. 82 (1974) 1–14.
[25] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (33–38) (1996) 27–28.
[26] J. Vreede, J. Juraszek, P.G. Bolhuis, Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 2397–2402.
[27] J.r. Schemberg, K. Schneider, U. Demmer, E. Warkentin, A. MÃžller, U. Ermler, Towards biological supramolecular chemistry: a variety of pocket-templated, individual metal oxide cluster nucleations in the cavity of a mo/w-storage protein, Angew. Chem. Int. Ed. Engl. 46 (2007) 2408–2413.
[28] V. Koronakis, A. Sharff, E. Koronakis, B. Luisi, C. Hughes, Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export, Nature 405 (2000) 914–919.
[29] J.R. Schnell, J.J. Chou, Structure and mechanism of the M2 proton channel of influenza A virus, Nature 451 (2008) 591–595.
[30] N.K. Karpowich, H.H. Huang, P.C. Smith, J.F. Hunt, Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding, J. Biol. Chem. 278 (2003) 8429–8434.
[31] G. Buhrman, G. Holzapfel, S. Fetics, C. Mattos, Allosteric modulation of Ras positions Q61 for a direct role in catalysis. Proc Natl Acad Sci U S A 107 (2010) 4931–4936.

[32] H.J.C. Berendsen, D. Vanderspoel, R. Vandrunen, Gromacs – a message-passing parallel molecular-dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.
[33] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, J. Chem. Theory Comput. 4 (2008) 435–447.
[34] C Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, J. Comput. Chem. 25 (2004) 1656–1676.
[35] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, Lincs: a linear constraint solver for molecular simulations, J. Comput. Chem. 18 (1997) 1463–1472.
[36] H.J.C. Berendsen, J.P.M. Postma, W.F. Vangunsteren, A. Dinola, J.R. Haak, Molecular-dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[37] T. Darden, D. York, L. Pedersen, Particle mesh Ewald – an $N\cdot\mathrm{Log}(N)$ method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[38] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.
[39] D.P. Tieleman, H.J. Berendsen, A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer, Biophys. J. 74 (1998) 2786–2801.
[40] C. Kandt, W.L. Ash, D.P. Tieleman, Setting up and running molecular dynamics simulations of membrane proteins, Methods 41 (2007) 475–488.
[41] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans, Interaction models for water in relation to protein hydration, Int. Forces (1981) 331–342.
[42] M. Petrek, M. Otyepka, P. Banas, P. Kosinova, J. Koca, J. Damborsky, Caver: a new tool to explore routes from protein clefts, pockets and cavities, BMC Bioinformatics 7 (2006) 316.
[43] M.F. Sanner, A.J. Olson, J.C. Spehner, Fast and robust computation of molecular surfaces, in: Proceedings of the 11th Annual ACM Symposium on Computational Geometry, Vancouver, British Columbia, Canada, 1995, pp. 406–407.
[44] M.F. Sanner, A.J. Olson, J.C. Spehner, Reduced surface: an efficient way to compute molecular surfaces, Biopolymers 38 (1996) 305–320.
[45] L. Federici, D. Du, F. Walas, H. Matsumura, J. Fernandez-Recio, K.S. McKeegan, et al., The crystal structure of the outer membrane protein VceC from the bacterial pathogen Vibrio cholerae at 1.8 A resolution, J. Biol. Chem. 280 (2005) 15307–15314.
[46] C. Kandt, Z. Xu, D.P. Tieleman, Opening and closing motions in the periplasmic vitamin B12 binding protein BtuF, Biochemistry 45 (2006) 13284–13292.
[47] C. Kandt, K. Gerwert, J. Schlitter, Water dynamics simulation as a tool for probing proton transfer pathways in a heptahelical membrane protein, Proteins 58 (2005) 528–537.
[48] C. Kandt, J. Schlitter, K. Gerwert, Dynamics of water molecules in the bacteriorhodopsin trimer in explicit lipid/water environment, Biophys. J. 86 (2004) 705–717.
[49] H. Yin, G. Feng, G.M. Clore, G. Hummer, J.C. Rasaiah, Water in the polar and nonpolar cavities of the protein interleukin-1beta, J. Phys. Chem. B. 114 (2010) 16290–16297.
[50] C. Kandt, D.P. Tieleman, Holo-BtuF stabilizes the open conformation of the vitamin B12 ABC transporter BtuCD, Proteins 78 (2010) 738–753.
[51] J. Sonne, C. Kandt, G.H. Peters, F.Y. Hansen, M.O. Jensen, D.P. Tieleman, Simulation of the coupling between nucleotide binding and transmembrane domains in the ATP binding cassette transporter BtuCD, Biophys. J. 92 (2007) 2727–2734.
[52] R. Baron, C. Riley, P. Chenprakhon, K. Thotsaporn, R.T. Winter, A. Alfieri, et al., Multiple pathways guide oxygen diffusion into flavoenzyme active sites, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 10603–10608.
[53] Y. Wang, S.A. Shaikh, E. Tajkhorshid, Exploring transmembrane diffusion pathways with molecular dynamics, Physiology (Bethesda) 25 (2010) 142–154.
[54] J. Li, R. Shi, C. Yang, X. Zhu, Exploration of the binding of benzimidazole-biphenyl derivatives to hemoglobin using docking and molecular dynamics simulation, Int. J. Biol. Macromol. 48 (2011) 20–26.