

Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: *trj_cavity*

Teresa Paramo,[†] Alexandra East,[†] Diana Garzón,[†] Martin B. Ulmschneider,[‡] and Peter J. Bond^{*,†,§,||}

[†]Unilever Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[‡]Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, Maryland 21218, United States

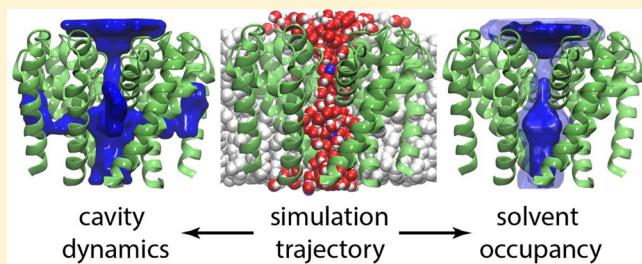
[§]Bioinformatics Institute (A*STAR), 30 Biopolis Str, #07-01 Matrix, Singapore 138671

^{||}Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, 117543 Singapore

Supporting Information

ABSTRACT: Protein cavities and tunnels are critical in determining phenomena such as ligand binding, molecular transport, and enzyme catalysis. Molecular dynamics (MD) simulations enable the exploration of the flexibility and conformational plasticity of protein cavities, extending the information available from static experimental structures relevant to, for example, drug design. Here, we present a new tool (*trj_cavity*) implemented within the GROMACS (www.gromacs.org) framework for the rapid identification and characterization of cavities detected within MD trajectories.

trj_cavity is optimized for usability and computational efficiency and is applicable to the time-dependent analysis of any cavity topology, and optional specialized descriptors can be used to characterize, for example, protein channels. Its novel grid-based algorithm performs an efficient neighbor search whose calculation time is linear with system size, and a comparison of performance with other widely used cavity analysis programs reveals an orders-of-magnitude improvement in the computational cost. To demonstrate its potential for revealing novel mechanistic insights, *trj_cavity* has been used to analyze long-time scale simulation trajectories for three diverse protein cavity systems. This has helped to reveal, respectively, the lipid binding mechanism in the deep hydrophobic cavity of a soluble mite-allergen protein, Der p 2; a means for shuttling carbohydrates between the surface-exposed substrate-binding and catalytic pockets of a multidomain, membrane-proximal pullulanase, PulA; and the structural basis for selectivity in the transmembrane pore of a voltage-gated sodium channel (NavMs), embedded within a lipid bilayer environment. *trj_cavity* is available for download under an open-source license (<http://sourceforge.net/projects/trjcavity>). A simplified, GROMACS-independent version may also be compiled.



INTRODUCTION

Protein cavities are empty, concave surfaces associated with the tertiary or quaternary structure of a protein. The archetypal cavity is buried within the protein core, far from the surface, but cavities may also lie on the protein exterior (forming pockets) or may connect different surface sites (to form a tunnel or channel). Cavities enable the creation of isolated, high affinity sites of interaction, facilitating specific binding of ligands or drugs. They can also help to create specialized microenvironments with unusual physicochemical features, which may facilitate processes such as enzyme catalysis.¹ Similarly, many biological enzymes carry out reactions using oxygen or other gas molecules with efficiencies that are many folds greater than those attained by artificial catalysts, with structural and dynamic adaptations proposed to facilitate uptake mechanisms and substrate reactivity.² The chemical and structural features that have evolved within proteins have afforded a means for high-affinity binding to a wide range of ligands. These include small signaling molecules, such as steroid hormones or neuro-

transmitters, which interact with nuclear receptors³ or nicotinic receptors,⁴ respectively. There are also many examples of larger molecules binding to cavities; for example, antigenic peptides bind to the polar surface grooves of MHC proteins,⁵ while a range of hydrophobic molecules containing fatty acid tails can associate with intracellular fatty acid-binding proteins (FABPs)⁶ or immunological lipid binding domains.^{7–10} Cavities often take the form of tunnels, particularly in membrane proteins, which encode 20–30% of many genomes¹¹ and represent major drug targets.¹² Such tunnels can be critical for facilitating and/or controlling metabolite distribution across membranes. In some cases, gating transitions between closed and open states may result from modulation of the tunnel connectivities, for example, due to allosteric changes in response to protein/ligand binding or electrical signals^{13–15} or via substrate diffusion assisted by fluctuating, flexible side chain motions.¹⁶

Received: December 19, 2013

Published: March 28, 2014



It is evident that the geometric and chemical properties of a cavity, pocket, or tunnel will play an essential role in ligand binding¹⁷ and molecular transport. More generally, the presence of cavities can directly impact protein stability.^{18–20} Thus, the identification and characterization of cavities is critical for understanding protein structure and function and to make predictions in the context of protein druggability and rational design.²¹ At present, several programs exist to characterize cavities. These use a range of algorithms and are often designed to identify particular types of cavities. One of the most popular strategies for cavity detection is the rolling probe sphere method, in which a spherical probe of defined radius is used to identify enclosed surfaces within proteins.²² This method has been implemented within programs such as GRASP²³ and more recently 3V.²⁴ The alpha shape theory²⁵ is another widely used strategy, featured in tools such as CAST,²⁶ for general cavity search, while CAVER²⁷ and MolAxis²⁸ focus in particular on the detection of tunnel-like cavities. The latter tools complement the alpha shape strategy with the use of a weighting function, which accelerates the search of paths from the interior of the protein to bulk water. Other geometry-based approaches represent protein voids as a set of overlapping spheres, such as in SURFNET,²⁹ HOLE,³⁰ and VOIDOO,³¹ or as empty voxels in a 3D grid.³²

Regardless of the strategy used, the common starting point for cavity analysis will be a protein structure, whose coordinates have been obtained from (or are based upon) X-ray crystallography or nuclear magnetic resonance (NMR) data. Unfortunately, the structure of proteins under *in vitro* conditions is not always representative of the physiological state, leading to uncertainties in, for example, side chain orientation, the conformation of unstructured loops, or the presence of water or ligand molecules^{33,34}—features which can all influence local features of pockets or cavities. Moreover, ligand binding often causes or results from conformational transitions that may not be evident from a single, static structure.³⁵ Molecular dynamics (MD) simulations provide a route to characterize the flexibility and conformational plasticity of proteins and are now routinely used alongside experimental approaches to relate protein structure at the atomic level to function and, where appropriate, to ligand binding,^{36–38} channel permeation,^{39,40} or gating.^{41–43} MD has proven useful in understanding large-scale conformational changes associated with ligand binding/release^{9,10,44} and in predicting likely binding modes,^{45,46} while there has been a resurgence in the use of MD-based alchemical approaches for accurate binding energy predictions.⁴⁷ Enhanced-statistics MD simulations enabled the first observation of complete, spontaneous diffusion of dioxygen from surface to active site of flavoenzymes and, combined with site-directed mutagenesis and rapid kinetics experiments, revealed how multiple, dynamic substrate diffusion paths converge toward a few key residues neighboring the reactive moiety of the flavin cofactor.¹⁶ Furthermore, recent simulation approaches enable the *ab initio* identification of potentially druggable protein cavities that are not evident in experimental structures.^{48–51} Consequently, some tools have begun to incorporate the knowledge gained from molecular simulations into a cavity-analysis workflow. For example, OMD⁵² uses an optimized MD protocol to identify solvent binding sites around the protein, improving the subsequent calculation of the static cavity volume. dxTuber⁵³ calculates an average cavity structure by using a spatial density map, obtained after the preprocessing of an MD trajectory with the VMD

volmap tool.⁵⁴ MolAxis,²⁸ Caver 3.0,⁵⁵ and MDpocket⁵⁶ have each added trajectory capabilities to their cavity search protocol via input of a series of protein data bank (PDB) structures. Finally, a genuine integration of traditional simulation-based analysis with cavity detection geared toward characterizing small molecule binding sites was introduced in PocketAnalyzer^{PCA},⁵⁷ which applies principal component analysis (PCA) and clustering to the output of a grid-based pocket detection method based on a variant of the LIGSITE algorithm.⁵⁸

As simulations reach more biologically informative time scales, facilitated by continuing improvements in algorithms and hardware,^{36,59} the computational cost of postsimulation analysis and storage has become critical. To address this, we now present a new program: *trj_cavity*, a rapid, grid-based approach for generalized cavity search of compressed MD trajectories, which enables efficient identification and analysis of protein cavities, tunnels, and pockets. The tool features a novel algorithm which performs an extremely fast neighbor search, maintaining a calculation cost that is linear with the number of grid voxels used to represent the protein structure. The tool is implemented within the GROMACS⁶⁰ framework (www.gromacs.org), enabling the use of a range of file formats (Table 1) and standard GROMACS system options (Table 2),

Table 1. GROMACS⁶⁰ Compatible *trj_cavity* File Formats

input files	option	optional	format
structure file	-s	no	tpr, tpb, tpa, gro, g96, pdb
trajectory	-f	yes	xtc, trr, trj, gro, g96, pdb, cpt
index file	-n	yes	ndx
output files		optional	format
cavities structure file	-o	no	tpr, tpb, tpa, gro, g96, pdb
cavities trajectory	-ot	yes	xtc, trr, trj, gro, g96, pdb, cpt
cavities volume vs time	-ov	no	xvgr, xmgr
cavities structure and occupancy	-ostat	yes	pdb
overlapping solvent vs time	-os	yes	xvgr, xmgr
overlapping solvent occupancy	-osstat	yes	pdb
tunnel structure file	-ot	yes	tpr, tpb, tpa, gro, g96, pdb
tunnel trajectory	-ott	yes	xtc, trr, trj, gro, g96, pdb, cpt
tunnel cross-sectional area or radius (defined by -sector) versus axis (defined by -axis)	-ota	yes	xvgr, xmgr
tunnel bottleneck area vs time	-otb	yes	xvgr, xmgr

including atom index groups. *trj_cavity* generates a trajectory of discovered cavities, together with a corresponding quantification of time-dependent cavity volume. Average statistical properties such as grid occupancy and residency of user-specified atom groups such as solvent molecules, ions, or ligands may be output for subsequent graphical analysis. Similarly, tunnel characterization descriptors, including shape and radius profiles as a function of a defined axis, enable channels to be studied within, for example, simulations of membrane proteins.

To test its utility, *trj_cavity* has been used to study long-time scale simulation trajectories associated with three diverse systems in which cavity dynamics are key to function, for a soluble protein, a transmembrane (TM) protein, and a peripheral membrane protein. Each protein corresponds to a rather different cavity, namely, (i) the deep hydrophobic binding cavity of mite-allergen Der p 2 protein,⁶¹ which can

Table 2. *trj_cavity* User Options and Standard GROMACS⁶⁰ Options

option	type	default value	description
-dim	int	5	number of surrounding protein directions (max 6)
-spacing	real	1.4	grid spacing
-mode	enum	all	cavities to output: all or max
-min	real	50	minimum size of cavities (\AA^3)
-cutoff	real	0	search distance cutoff (\AA)
-seed	vector	0 0 0	coordinates of seed point (\AA)
-axis	enum	Z	direction along which to calculate tunnel: z, x, or y
-sector	enum	area	calculate sector as area or radius
-ff_radius	string	no	use force field van der Waals radius; can be calculated directly from a topology (tpr) file or directly from the force field if user provides the folder path (with the -ff_path option)
-ff_path	string	no	path to force field folder
-axis_value	real	0	output tunnel cross-sectional area at a defined coordinate (\AA) along tunnel axis.
standard options	type	value	description
-h	bool	no	print help info and quit
-version	bool	no	print version info and quit
-nice	int	19	set the nicelevel
-b	time	0	first frame (ps) to read from trajectory
-e	time	0	last frame (ps) to read from trajectory
-dt	time	0	only use frame when t MOD dt = first time (ps)
-xvg	enum	xmgrace	xvg plot formatting: xmgrace, xmgr, or none
-skip	int	1	only write every nr-th frame

bind bacterial lipid; (ii) the TM pore within the voltage-gated Na^+ channel pore from *Magnetococcus sp.* (NavMs);⁶² and (iii) the surface-exposed pocket of a catalytic site specialized for carbohydrate degradation within the large, multidomain pullulanase (PulA) from *K. oxytoca* (Francetic et al., submitted; PDB id 2YOC). Application of *trj_cavity* has provided new insights into the relationship between the conformational dynamics of each protein and their functional mechanisms. In the absence of further direct experimental data concerning the dynamic behavior of each cavity, the quality of the results for each test case was assessed by comparison with several other commonly used analysis tools, including MDpocket,⁵⁶ CAVER,⁵⁵ and, for the NavMs channel, HOLE.³⁰ This revealed significantly improved usability and computational performance across all systems studied.

METHODOLOGY

The objective of *trj_cavity* is to be a general-purpose, comprehensive and efficient tool for the identification and/or characterization of cavities within long time-scale biomolecular simulation trajectories. Grid-based methods often offer high performance rates since they allow spatially focused calculations rather than brute force inspection of all atom pairs, which result in a large number of operations due to the $O(n^2)$ relationship. Most grid-based cavity search methods tend to preprocess the protein structure in order to calculate certain geometrical properties, such as iterative protein density factors.^{29,53} This normally requires inspection of the protein grid voxels several times, increasing the order of magnitude of the cavity search

algorithm. The algorithm we present here has instead been designed to minimize the inspection of grid voxels, maintaining the computational cost as a linear function of the number of grid cells. The number of predefined options necessary to run the program has also been minimized, improving the usability. Finally, the output of the code takes the form of text data files, interpretable by any standard graphing program, and standard GROMACS-compatible structure files, including PDBs, enabling straightforward graphical analysis of cavity dynamics. An overview of the *trj_cavity* method is presented below and in Figures 1 and 2, with a summary of the program file formats and options presented in Tables 1 and 2.

Algorithm—Protein Representation. The first step in the algorithm (Figure 1A) involves the transformation of the molecular coordinates to a 3D grid of user-defined voxel size. Each atom is mapped onto the grid with its corresponding van der Waals radius, by default according to Bondi's compilation⁶³ or using a customized radii set provided by the user. However, the van der Waals radius of each atom may also be derived from the Lennard-Jones potential, by extracting the force field parameters used to run the simulation, via either a simulation run input file or from a specified GROMACS force field directory (-ff_radius and -ff_path options).

The voxel size is an optional parameter (-spacing option) and is critical to the subsequent search algorithm since it determines the expansion of each cavity through differently spanned pathways. While a reduction of the voxel size may increase the precision of the cavity detection, it will also increase the computational time required to process the structure. To overcome this limitation, *trj_cavity* offers the possibility to use index files to restrict the grid calculation to certain atom groups of interest (-n option). This also enables the program to produce results relative to specific sites rather than for the whole structure, avoiding the noise that a large number of cavities could potentially produce in the final output.

Algorithm—Cavity Search. Once atoms are positioned within the grid, the cavity search begins (Figures 1A and 2). By default, a complete scanning of the atom-mapped grid is performed. Alternatively, a seed point may be provided (-seed option), in which case *trj_cavity* scans for cavities within a default radius of 5 \AA around the seed point. This accelerates the cavity search since it is no longer necessary to navigate through the entire structure; this may also facilitate subsequent user analysis of the output. When analyzing trajectories with a seed point, the coordinates of cavities found in previous frames are used as seed points for subsequent frames, in order to minimize the effect of rigid-body protein dynamics.

The process of identification of cavities is incremental: once a voxel has been assigned as a part of a cavity, the algorithm searches for potential cavity members among all immediate voxel (ν) neighbors in the six positive/negative directions along the x , y , and z axes ($\nu_x^{\pm 1}, \nu_y^{\pm 1}, \nu_z^{\pm 1}$), proceeding in an iterative fashion that only terminates when potential neighboring cavity members in a certain number of directions are no longer found (Figures 1B and 2). A successful cavity member must be (i) an empty voxel; (ii) connected to at least one other cavity point; and (iii) surrounded by protein atoms in a certain number of directions. To determine the number of directions in which a voxel is surrounded by protein atoms (the “burial state”), a linear search is performed along the positive (b^+) and negative (b^-) directions of the three grid dimensions until a protein atom is found or the algorithm reaches the end of the grid (or a cutoff distance—see below), i.e.,

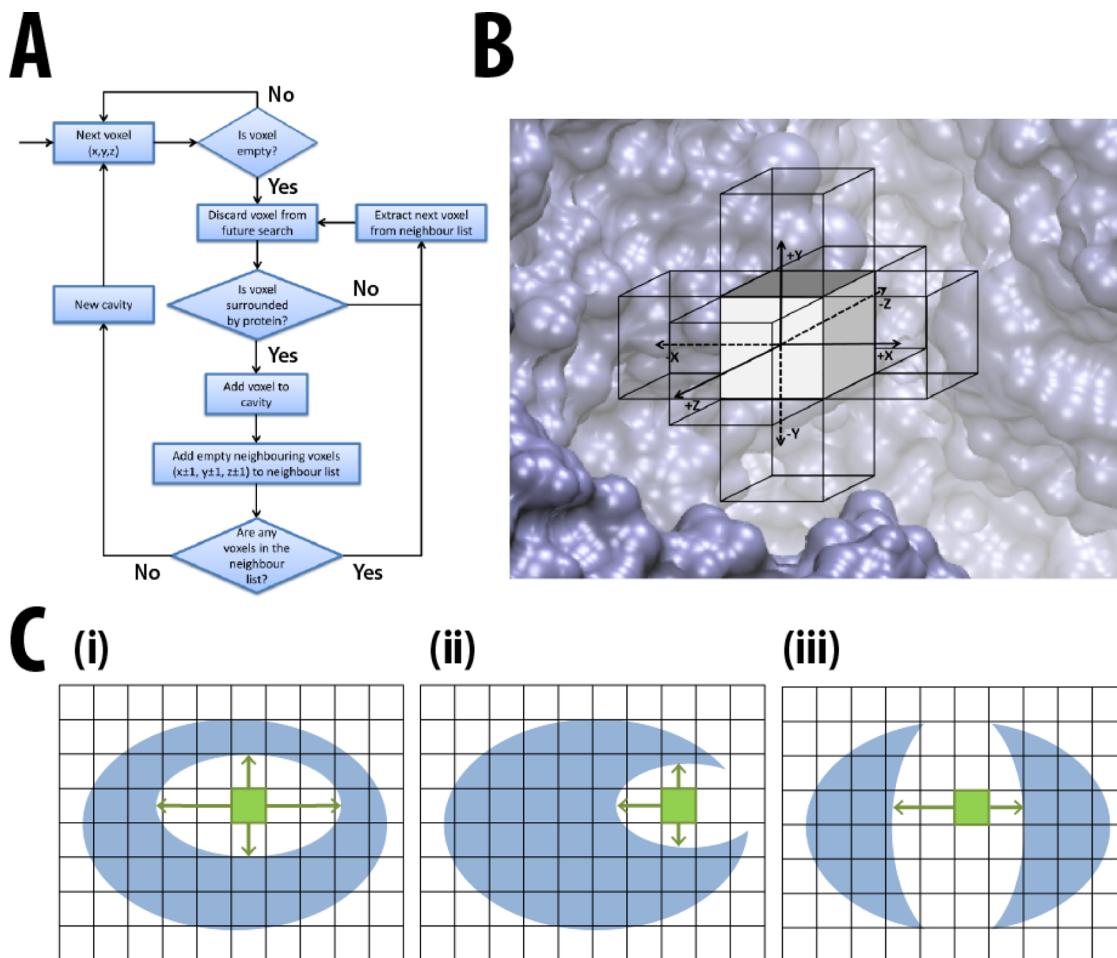


Figure 1. Overview of the *trj_cavity* program. (A) Workflow indicating the main features of the *trj_cavity* search algorithm. (B) The process of incremental identification of new cavity members; a hypothetical growing grid that surrounds a central cavity-member voxel is overlaid on a protein cavity surface. (C) Schematic 2D illustration of the relation between “burial state” and type of protein cavity. The number of directions (green arrows) surrounding a central voxel (green box) are enclosed by protein (blue) for (i) an internal protein cavity; (ii) a surface pocket; or (iii) a tunnel. Note that, in reality, a voxel may be surrounded by protein in up to six (rather than four) possible directions.

$$\text{“burial state”} = b_x^+ + b_x^- + b_y^+ + b_y^- + b_z^+ + b_z^-$$

$$\text{where } b_x^\pm, b_y^\pm, b_z^\pm = \begin{cases} 0 & \text{if no protein atoms found} \\ 1 & \text{if protein atoms found} \end{cases}$$

The “burial state” required to maintain the grid search depends upon the cavity type of interest (Figure 1C) and may be chosen by the user (*-dim* option). The number of directions must be equal to or higher than the *-dim* parameter to maintain the cavity search, and therefore the greater its value, the more restrictive the result (Figure 1C). For example, if the objective of the search is to find an internal cavity, any cavity candidates would be expected to be covered by protein in all six voxel directions (*-dim* 6, both positive and negative directions of *x*, *y* and *z* axes), whereas a tunnel might require four directions (*-dim* 4) and a protein pocket only two (*-dim* 2). To avoid potential problems derived from protein packing (i.e., areas with lower atom density), the algorithm uses voxels already identified as a part of a cavity to complete missing dimensions.

Finally, it should be noted that the cavity search in each dimension may be combined with a distance cutoff (*-cutoff* option), which sets a maximum distance around the currently inspected voxel up to which protein atoms may be found. This

may, for example, help to accurately define cavities in very large multidomain proteins, to avoid the inclusion of atoms from neighboring domains for a particular cavity.

Trajectory Analysis. Subject to provision of a suitable input simulation structure (*-s* option) and trajectory file (*-f* option), cavity analysis will subsequently be performed for each frame of the trajectory. Information about all the cavities discovered that meet the requested input criteria will be returned in the form of (i) a cavity structure file and corresponding trajectory; (ii) a text file containing cavity voxel volume versus time; and (iii) an optional PDB file containing the coordinates of all voxels where a cavity was detected, with the mean occupancy statistics of each voxel included within the B-factor column (*-ostat* option). In order to limit the analysis to the most significant cavities, the output may be restricted to the largest cavity found (*-mode max* option) or to all cavities found that are larger than a predefined minimum volume (*-min* option).

Since the output of cavity information is coordinate-based, a prior structural fit of the trajectory (for example, using the *trjconv* tool within GROMACS) to a sensible atom selection with regard to the cavities of interest is recommended (though it is unnecessary for the cavity search algorithm itself). Fitting of the trajectory may be required for two primary reasons. First,

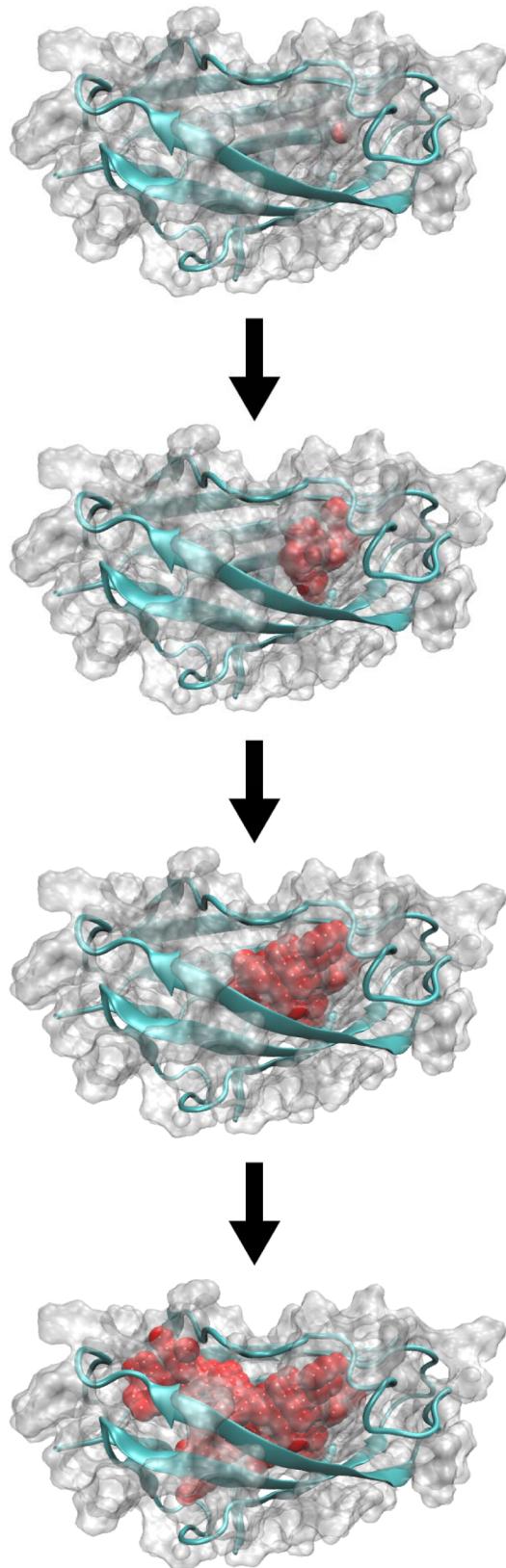


Figure 2. Evolution of the cavity described by the *trj_cavity* algorithm. The process of incremental identification of new cavity voxel members and growth of the resultant grid (red spheres) is shown for the Der p 2 cavity. The protein is shown in cyan cartoon format, with its external molecular surface highlighted in transparent gray.

to obtain useful quantitative descriptors for a specific cavity, it needs to be identified throughout the simulation. In the *trj_cavity* algorithm, a cavity identified from one frame to the next is considered to be the same one if some of its Cartesian coordinates overlap. If a cavity is relatively stable during simulation, then fitting to any reasonable set of atoms (such as all $C\alpha$ atoms of the protein or those of the region of interest in a multidomain protein) should be sufficient to maintain consistency in cavity detection. More critically, fitting is essential for calculation of the cavity grid statistics, and hence for accurate graphical representation of the cavity. This is because the output PDB generated for the cavity of interest contains the coordinates and mean occupancies of all detected free voxels. The choice of fitting group in this case is nontrivial and rather system-dependent, since if a protein is particularly complex or dynamic, the occupancy statistics may not be meaningful. A reasonable choice may, for example, be guided by the preliminary generation of static cavity maps using *trj_cavity*, enabling identification of surrounding protein atoms (or protein domain(s)) which may then be used to guide the trajectory fitting procedure followed by a complete cavity characterization.

Solvent (or Ligand) Statistics. The presence of solvent atoms (or any other group of atoms, such as ions or protein ligands, as defined in an optional index file) within each of the cavities discovered may also be monitored during the analysis of the trajectory. A quantification of the number of those atoms that overlap with cavities per time step is then performed (*-os* option) and, analogously to the cavity statistics output, a PDB file with the overlapping atom frequencies will be generated (*-osstat* option).

Tunnel Detection. A cavity is considered to be a tunnel in a pore or channel if any line of the protein-free grid points can be traced along a predefined alignment axis (*-axis* option) and reaches two extremes of the grid, i.e., both positive and negative directions of an inspected axis are required to be considered as a buried dimension:

$$\begin{aligned} \text{"burial state"} = & (b_x^+ + b_x^- - 1) \times 2 + (b_y^+ + b_y^- - 1) \\ & \times 2 + (b_z^+ + b_z^- - 1) \times 2 \end{aligned}$$

If a cavity is identified as a tunnel, the cross-sectional area or radius (*-sector area/radius* option) of the cavity will be calculated as a function of the predefined axis (*-ota* option), providing a straightforward way to analyze the mean pore profile. In order to reduce the dependency of the algorithm upon the voxel size and yield a more accurate estimate of the cross-sectional area, cavities defined as tunnels are refined by calculating any possible gaps between radial cavity voxels and their nearest protein atom. The bottleneck cross-sectional area (the minimum area along the tunnel axis) (*-otb* option) or, alternatively, the cross-sectional area of the tunnel at a defined location along the axis (*-axis_value* option) may also be output as a function of time, in order to help identify possible conformational changes in key regions of the tunnel during the simulation trajectory.

Simulation Details. The apo Der p 2 structure was based on the X-ray structure⁶¹ (PDB id 1KTJ) after ligand removal (two lysophosphatidylcholine molecules), while the PulA structure (Francetic et al., submitted; PDB id 2YOC) was modified to add 19 missing amino-terminal residues using Modeller.⁶⁴ The PulA membrane consisted of an equilibrated bilayer of 318 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoetha-

nolamine (POPE) molecules. PulA was manually placed near the surface of the membrane, guided by experimental data (Francetic et al., submitted). The open conformation of the NavMs pore was derived from the crystal structure⁶² (PDB id 4F4L) and inserted in a 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine (POPC) bilayer comprising 204 lipid molecules. The three systems were solvated with TIP3P water. Der p 2 was embedded in a periodic truncated octahedral box of unit cell dimensions ~70 Å, containing ~2,000 protein atoms and ~3,500 water molecules, and neutralized with a single sodium ion, leading to a total system size of ~12,000 atoms. PulA was placed in a cubic box of dimensions 100 Å × 100 Å × 190 Å, containing ~16,000 protein atoms, ~50,000 waters, ~40,000 lipid atoms, and ~300 ions (equivalent to a ~0.1 M NaCl solution), leading to a total system size of ~200,000 atoms. Finally, NavMs was placed in a cubic box of ~80 Å × 80 Å × 80 Å, containing ~6,000 protein atoms, ~8,000 waters, ~20,000 lipid atoms, and ~200 ions (equivalent to 0.5 salt solutions of either NaCl or KCl), leading to a total system size of ~50,000 atoms. In all systems, a minimum of 20 Å separated protein atoms from the periodic box edges.

The Der p 2 and PulA systems were simulated with GROMACS 4.5,⁶⁰ while the NavMs system was simulated using the DESMOND version implemented on the Anton supercomputer.⁶⁵ For all systems, the CHARMM22/CMAP all-atom force field^{66,67} was used to represent the protein and membrane lipid utilized CHARMM36.⁶⁸ Compatible ion parameters were taken from ref 69 for the Der p 2 and PulA systems and from ref 70 for the high-salt concentration NavMs system. The charge state of all ionizable residues was assigned assuming neutral pH. Simulations were performed in the *NpT* ensemble, at a temperature of 298 K (Der p 2) or 310 K (PulA, NavMs) and 1 atm pressure, employing the velocity-rescale thermostat⁷¹ and the Parrinello–Rahman barostat,⁷² coupling the system isotropically (Der p 2) or semi-isotropically (PulA, NavMs). The equations of motion were integrated using the leapfrog method⁷³ with a 2 fs time step, and LINCS was used to constrain bond lengths.⁷⁴ Nonbonded pairlists were generated every 5 (NavMs) or 10 (Der p2, PulA) steps using a distance cutoff of 14 Å. A cutoff of 12 Å was used for Lennard-Jones (excluding scaled 1–4) interactions, which were smoothly switched off between 10 and 12 Å; a dispersion correction was not employed for the membrane simulations. Electrostatic interactions were computed using the Particle-Mesh-Ewald algorithm⁷⁵ with a real-space cutoff of 10 Å (NavMs) or 12 Å (Der p2, PulA). A homogeneous external electric force field of 18.6 mV/nm was applied to the membrane normal (*z* axis) of the NavMs system, in order to reproduce the voltage-dependent properties of the pore. This equates to a 166 mV drop across the membrane. Production simulations were run for 100 ns (PulA, Der p 2) or 1000 ns (NavMs).

RESULTS

Case Study 1: Der p 2 (Large Internal Cavity). The MD-2 related lipid-recognition (ML) superfamily is composed of immunoglobulin-like proteins involved in the recognition and metabolism of lipids.⁷⁶ The most representative proteins of this family are MD-2 and MD-1, which play an essential role in innate immunity by recognizing pathogenic lipopolysaccharide (LPS).^{77,78} Other distantly related proteins have been classified as members of this family as well, such as the mite allergens Der f 2 and Der p 2, cholesterol-binding protein NPC2, and the

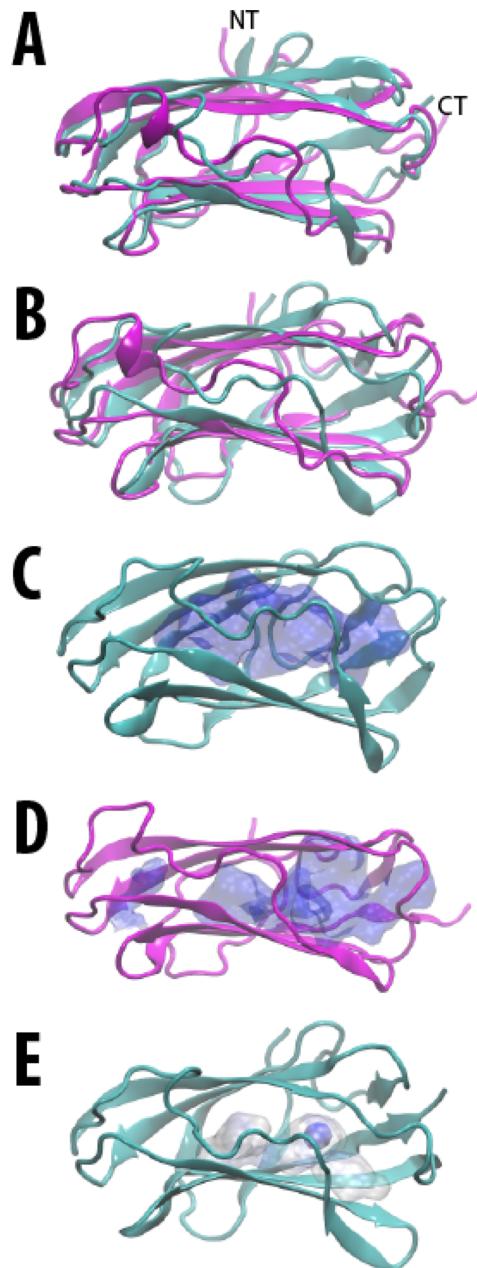


Figure 3. Flexible β -sheet “clamshell” enables changes in the internal cavity of Der p 2, as illustrated for the apo-state NMR structure, the ligand-bound X-ray structure, and simulation of the solvated X-ray structure following ligand removal. Structural alignments of (A) X-ray (cyan) vs NMR (magenta) structures and (B) final simulation snapshot (magenta) vs X-ray (cyan) structure. Internal cavity surfaces (blue) of (C) X-ray and (D) NMR structures. (E) Mean cavity surface over the final 20 ns of simulation, with probability of >50% (white) and >90% (blue). In all cases, protein is shown in schematic cartoons format, and the amino- (NT) and carboxy- (CT) termini are indicated in (A).

GM2 activator protein. These single-domain proteins are rich in β -strands, adopting a characteristic β -cup fold with a large hydrophobic cavity.^{23,61,79,80} Structures of Der p 2 in its apo⁸¹ and holo⁶¹ forms revealed a differently sized binding cavity (Figure 3), leading to the hypothesis, supported by some other structures of ML proteins,²³ that the ligand binding mechanism may involve transitions between open and closed states.

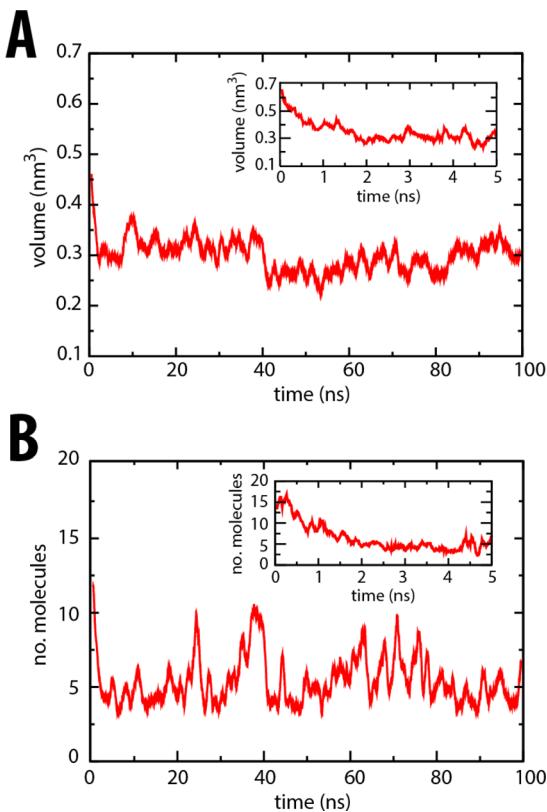


Figure 4. Adaptation of the Der p 2 hydrophobic cavity during simulation. Changes in (A) cavity volume and (B) cavity-resident solvent molecules, over 100 ns of simulation in the ligand-free apo state. An initial, rapid phase reflecting hydrophobic collapse is highlighted via the inset graphs. Cavity collapse and dehydration were coupled, with a correlation coefficient over 100 ns of 0.88 ± 0.02 (and similarly, 0.86 ± 0.04 during the first 5 ns).

To explore this hypothesized binding mechanism, simulations of the lipid-bound X-ray structure of Der p 2 were performed upon ligand removal, i.e., in an artificial apo state, and subsequently analyzed with *trj_cavity*. Like the rest of the ML protein family members, Der p 2 features a characteristic β -cup fold, with an additional loop covering the binding cavity opening. Given this protein architecture, the cavities were considered as internal, and the “burial state” was set appropriately (i.e., *-dim 6* option). The NMR apo structure (PDB id 1A9V) was also analyzed using the same parameters, so that results between both structures could be compared, with a default grid spacing of 1.4 Å in each case. Thus, the command line was as follows:

```
trj_cavity -s protein.pdb -f protein.xtc -dim 6
          -stat cavity_stats.pdb
```

As an aside, it is worth considering the variation of these input parameters. As described in Methodology, the burial state modifies the topology of the expected cavity; reducing its value makes the algorithm less restrictive and enables the cavity to expand through additional, “less buried” exit paths (Supporting Information Figure S1A). This is obvious for the Der p 2 crystal structure, where the estimated cavity volume incrementally increases from ~ 0.8 nm³ (*-dim 6*) to ~ 1.3 nm³ (*-dim 2*). At lower dimensionality, two additional, less buried exit paths were discovered (these were also discovered by other cavity-search

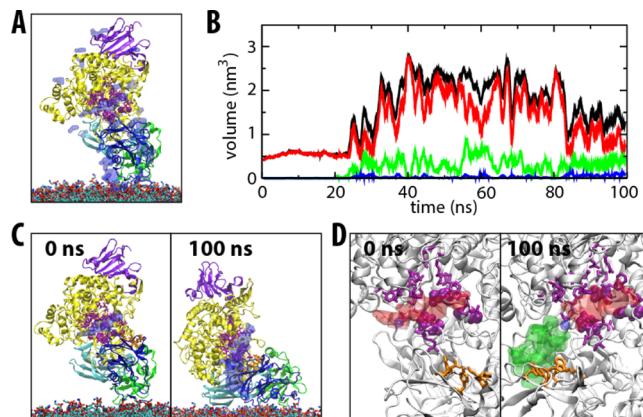


Figure 5. Cavity dynamics in PulA_{Kox}. (A) The surface of all cavities initially detected by *trj_cavity* are overlaid (transparent blue) on the X-ray structure, with possible sugar-binding (purple sticks format) and catalytic (spacefill format) residues highlighted. The protein is shown in cartoon format, colored according to the five domains: N1 (dark blue), N2 (green), N3 (cyan), A (yellow), and C (purple). Membrane lipids are shown in CPK sticks format. (B) Changes in volume of main cavities, merging into a single large cavity, over 100 ns simulation, which occurs following expansion of the flexible N1 domain after ~ 25 ns. (C) Evolution of the total internal cavity volume (transparent blue surface) at 0 and 100 ns, leading to connection of the catalytic and substrate binding sites with another potential sugar binding region (orange sticks format). (D) Close-up of the individual cavities at 0 and 100 ns, colored according to (B), with sugar-binding (purple sticks format) and catalytic (spacefill format) residues highlighted and the protein fold shown in gray.

programs, such as CAVER—see next section); while subjective, it is visually rather intuitive that these do not belong to the internal binding cavity and should be disregarded for accuracy (Supporting Information Figure S1A). Similarly, the precision of the cavity topology may be tuned by varying the grid spacing value (*-spacing* option), with the expansion of tunnels and pockets evident upon reduction of the grid size (Supporting Information Figure S1B). The use of a 1.4 Å grid spacing seems most appropriate, given its correspondence to the effective van der Waals radius of a water molecule, and again, its correctness is visually intuitive for Der p 2 and matches well with the presence of overlapping ligand-bound crystallographic density in the X-ray structure (Supporting Information Figure S1B). Nevertheless, in some cases, variation of the grid spacing may be useful, such as for characterization of small, intricate internal binding pockets, or when there is interest in the binding of variably sized ligands. Finally, an alternative means of explicitly defining (or excluding) exit paths, or particular regions of a cavity, would entail defining a group of atoms for only part of the protein of interest within an index file (*-n* option). In this way, it is possible to “force” the premature termination of the Der p 2 cavity, irrespective of grid spacing or burial state (Supporting Information Figure S1C). Importantly, while the definition of a pocket or cavity can often be subjective, it is essential that the same parameters be chosen consistently when comparing across multiple simulations or related systems.

Using *trj_cavity*, two cavities were identified for the experimental (NMR) structure of Der p 2: a larger cavity was found buried between the two β -sheets (~ 0.49 nm³) and a smaller one (~ 0.06 nm³) was located close to the mouth of the cavity (Figure 3). Interestingly, the number and location of the cavities was similar to those measured using GRASP⁸² for the

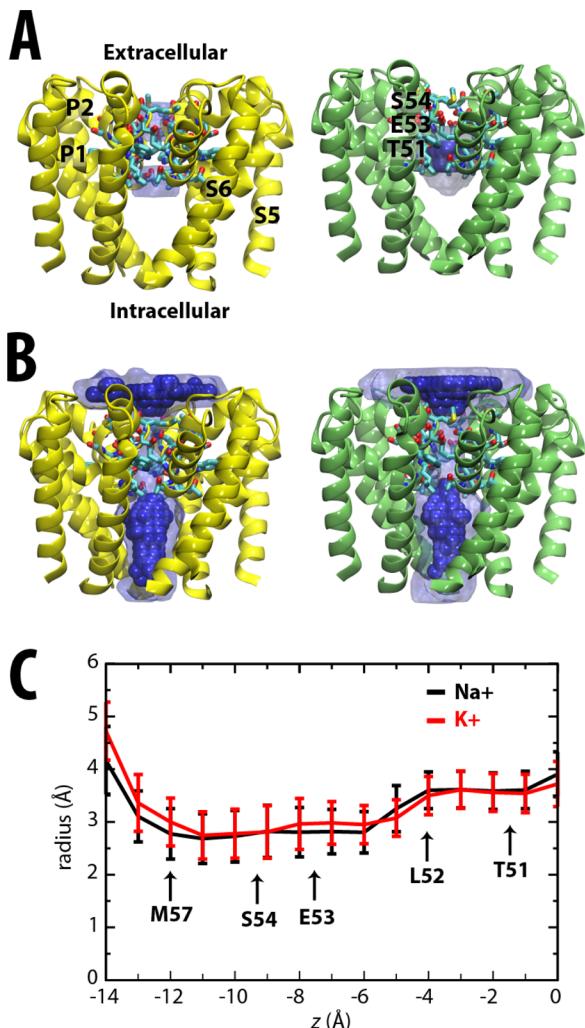


Figure 6. Characterization of the NavMs channel during 1000 ns of simulation. (A) Mean ion occupancy for potassium (left, yellow protein) and sodium (right, green protein) systems. Protein is shown in cartoon format; helix domains, orientation (left), and key pore residues (right) are labeled. Occupancy frequencies are colored solid blue (>50%) or transparent blue (>25%). (B) Mean water occupancy for potassium (left) and sodium (right) systems, colored according to (A). (C) Average NavMs pore radius as a function of the pore axis (z) observed during simulations in the presence of sodium (black line) or potassium (red line). The locations of key residues within the pore are indicated. Negative z coordinates represent the extracellular side of NavMs in (A) and (B).

NMR structure of the closely related Der f 2 protein.²³ The largest cavity was also present in the crystal structure but was nearly double its size ($\sim 0.80 \text{ nm}^3$), due to a wider separation of the β -sheets, presumably as a result of ligand accommodation (Figure 3). On the other hand, simulation following ligand removal had a dramatic effect upon the volume of the main Der p 2 cavity, which rapidly decreased to $\sim 0.30 \text{ nm}^3$. This was due to hydrophobic collapse of the cavity and was accompanied by simultaneous expulsion of the water molecules placed within the protein during the equilibration step (Figure 4). This cavity collapse was coupled to protein conformational changes during which the two β -sheets closed up, adopting a separation similar to the apo NMR structure (Figure 3).

The observed conformational plasticity of Der p 2, and the apparent convergence of the protein fold and its enclosed cavity

structure to the apo NMR state supports the hypothesis that the existence of closed ligand free and open ligand bound states may be a common feature of members of the ML protein family.²³ Indeed, recent simulations of the endotoxin coreceptor MD-2¹⁰ lend further support to this notion and suggest that transitions between the closed and open states occur via “clamshell-like” motions. These findings have a potential therapeutic impact, since many ML family members are implicated in a range of allergic and inflammatory responses.⁸³

Case Study 2: Pullulanase (Surface Pocket). Pullulanase (PulA) is a special type of glucanase which hydrolyses the α -1,6-glucosidic linkage of α -glucans, such as amylopectin and pullulan. PulA has an increasing importance in the glucose manufacturing industry due to its efficient debranching of starch.⁸⁴ This cell-surface anchored lipoprotein is present in a variety of microorganisms,⁸⁵ sharing four highly conserved domains corresponding to common substrate binding sites in amylases.^{86,87} PulA is folded in the periplasm and secreted to the bacterial surface via the Type II secretion system (T2SS). PulA is the only known substrate of the T2SS of *K. oxytoca* and therefore has been used to study the molecular basis of secretion specificity. The resulting structure, resolved by X-ray crystallography at 2.9 \AA resolution, is similar to that of the pullulanase of *K. pneumoniae*,⁸⁸ which shares 88% sequence identity with PulA_{Kox} (Francetic et al., submitted; PDB id 2YOC).

PulA is composed of five domains (N1, N2, N3, A, and C), where N1 and N3 have been identified as carbohydrate-binding domains,⁸⁸ while the catalytic domain is buried within a large cavity formed by the A and C domains (Figure 5A). Each subunit of the PulA_{Kox} crystal structure presents a different orientation of the N1 domain with respect to the rigid core containing N2, N3, and C domains, in agreement with the poor density and high B-factors observed for this region in both PulA_{Kox} and PulA_{KPn} structures. The flexibility of this carbohydrate-binding domain suggests that N1 could play a role in carbohydrate presentation to the catalytic domains (Francetic et al., submitted; PDB id 2YOC).

To further investigate the conformational dynamics of PulA_{Kox} in its native environment and to gain insights into the role of membrane lipids in protein orientation and mobility, PulA was simulated for 100 ns on the surface of a POPE membrane. The resulting trajectory was then studied with *trj_cavity* in order to characterize the dynamics of the catalytic cavity. Given the large protein structure (1070 residues), an initial grid size of 2 \AA was used along with a 10 \AA cutoff to reduce processing time as well as to minimize “cavity tunnelling”. Since visual analysis of the X-ray structure did not reveal any obvious internal cavity, a preliminary inspection of the semiburied cavities was performed by requesting five protein-surrounded directions (i.e., *-dim 5* option). This preliminary analysis of all pockets within the protein helped to identify a large cavity of $\sim 0.5 \text{ nm}^3$ volume (Figure 5A) nearby to the catalytic residues His607, Asp677, and His682⁸⁹ within the A domain and overlapping with an area in which maltooligosaccharide was bound in the homologous PulA_{KPn} structure (Supporting Information Figure S2).⁸⁸ Since this protein site is of functional interest, we focused subsequent analysis by using a random coordinate within the large cavity as a seed point:

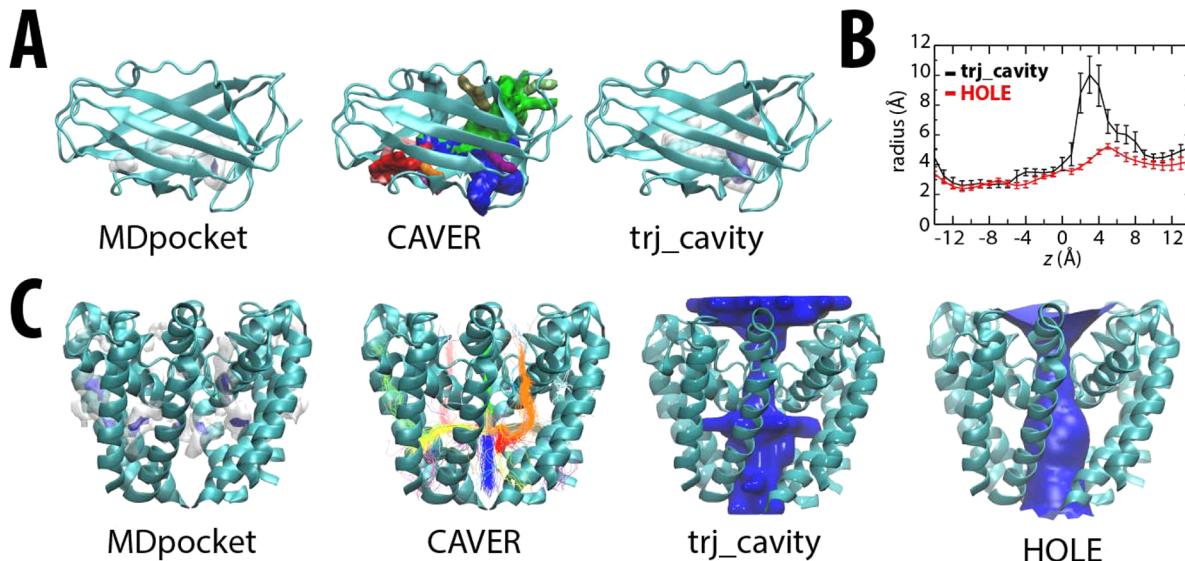


Figure 7. Comparison of *trj_cavity* output with those of other tools. (A) Graphical characterization of the mean cavity surface for ligand-free simulation of Der p 2 (cartoons format), with probability of >50% (white) and >90% (blue) shown for *trj_cavity* and MDpocket, and the individual tunnel clusters shown in different colors for CAVER. (B) Mean pore profiles for the NavMs simulation in the presence of sodium, calculated using *trj_cavity* (black line) or CAVER (red line), with standard deviations indicated. (C) Graphical characterization of the mean internal cavity surface(s) detected for the channel of NavMs (cartoons format) simulated in the presence of sodium. The cavities detected by MDpocket are shown with probability of >50% (white) and >90% (blue), while the individual tunnel clusters are shown in different colors for CAVER. The single transmembrane channels discovered by *trj_cavity* and HOLE are shown in blue surface representation.

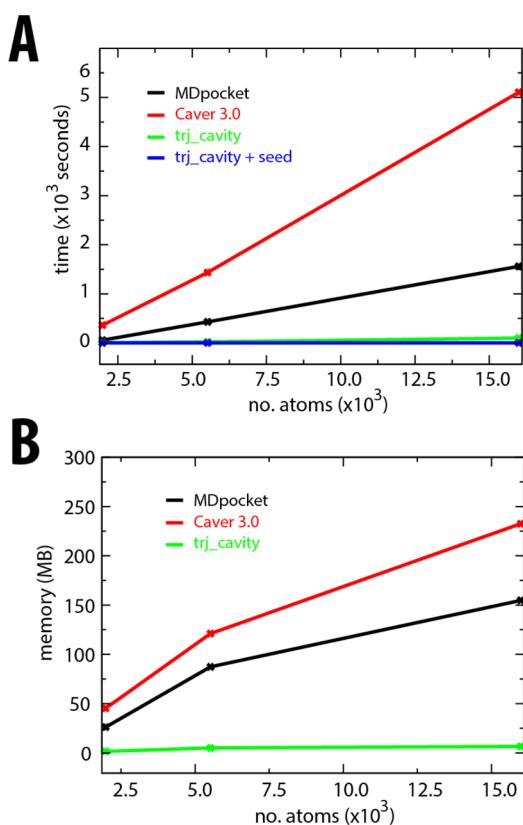


Figure 8. Performance of *trj_cavity* in comparison with other tools, tested for 100 frames of each system as described in the text. Total (A) CPU time and (B) memory usage for each tool is shown as a function of number of protein atoms.

```
trj_cavity -s protein.pdb -f protein.xtc -dim 5 -spacing 2
-cutoff 10 -seed 48.9 41.9 119.8
```

Analysis of the PulA trajectory revealed that the large catalytic cavity began to grow after ~25 ns as a result of motions of the flexible N1 domain relative to the rest of the protein (Figure 5B), fluctuating around a volume of ~1–2 nm³, as well as spawning some additional smaller pockets. The gradual growth of the cavity led to the formation of a contiguous surface (Figure 5C,D) that encompassed residues (Trp80, Trp95, Asp138, Tyr78, and Lys133) that were observed to interact with an additional maltose moiety more distant from the catalytic site in the PulA_{KPn} crystal structure (Supporting Information Figure S2). Thus, in the presence of the native membrane environment, the flexibility of the N1 domain enables formation of an expansive binding cavity surface that connects two known sugar binding sites to the catalytic core, providing a potential means for substrate shuttling and presentation.

Case Study 3: Voltage-Gated Sodium Channel (Transmembrane Tunnel). Voltage-gated ion channels are TM proteins essential for electrical signaling in neurons and other excitable cells.^{90,91} These channels are generally composed of several TM helices arranged around a central pore, where the narrowest part constitutes the selectivity filter (SF) which provides a means for the channel to restrict the passage to specific ions.^{92,93} The crystal structure of the *Magnetococcus sp.* voltage-gated sodium channel (NavMs) in the open conformation⁶² has been used as a basis for ion translocation properties,⁹⁴ via microsecond-time scale simulations of NavMs embedded within a POPC bilayer. Constant electric fields, corresponding to TM potentials of 39–665 mV, were applied in order to reproduce voltage-dependent properties of activation leading to the passage of ions through the channel.

Table 3. Performance for Der p 2 (1,981 atoms × 100 frames)

program	parameters	input size (MB)	output size (MB)	internal cavity found?	starting point?	calculation time ^a (s)
MDpocket	-v 1000	25.49	10.67	yes	no	12.40 (cavity search)
					yes	44.83 (descriptor calculation)
Caver3.0 <i>trj_cavity</i>	starting_point	25.49	19.76	yes	yes	361.58 (seed)
	-dim 6				no	7.06
					yes	0.24

^aMDpocket performs cavity characterization in two independent runs (cavity search followed by descriptor calculation); calculation times for both parts are provided. Calculation times for two possible modes of execution of *trj_cavity* (search and characterization of all possible cavities or guided search with seed point) are also provided.

Table 4. Performance for NavMs (5,520 atoms × 100 frames)

program	parameters	input size (MB)	output size (MB)	tunnel found?	starting point?	calculation time ^a (s)
MDpocket	-v 1000	125.29	29.41	no	no	330.80 (cavity search)
					yes	1279.84 (descriptor calculation)
Caver3.0 <i>trj_cavity</i>	starting_point	125.29	107.25	yes	yes	5104.64
	-dim 4				no	99.90
					yes	2.26

^aSee Table 1.

Table 5. Performance for PulA (15,990 atoms × 100 frames)

program	parameters	input size (MB)	output size (MB)	pocket found?	starting point?	calculation time ^a (s)
MDpocket	-v 1000	50.49	36.84	yes	no	57.06 (cavity search)
					yes	372.69 (descriptor calculation)
Caver3.0 <i>trj_cavity</i>	starting_point	50.49	70.57	no	yes	1436.20
	-dim 5				no	24.13
					yes	1.80

^aSee Table 1.

Here we analyze the 166 mV system, which has a flux of 37 ions μs^{-1} for sodium and 2 ions μs^{-1} for potassium.

Although channels are usually ion-specific, similarly sized ions may sometimes pass through their SFs. To test the ion specificity of NavMs, simulations of its open conformation were conducted in both 0.5 M NaCl and KCl salt solutions. The long time scale of these simulations permitted the observation of multiple, complete potassium ion passage events through the channel but revealed a lower conductance than for sodium ions.⁹⁴ In contrast with the closely related potassium channel family, K⁺ ions were observed to translocate in a partially hydrated form past the side chains of the consensus sequence TLESWS that face inside the pore within the NavMs SF. On the basis of X-ray structural analysis of the SF, the principal binding site was expected to be at the Glu53 side chain carboxylate groups of the consensus sequence, although the Thr51 hydroxyl group was also hypothesized to potentially bind to water molecules of a hydrated ion.⁹⁵

To examine the relationship between the conformational dynamics of the protein pore and ion transitions, *trj_cavity* was used to process the entire microsecond-long trajectories. Since a tunnel is expected to be surrounded by protein in at least four out of six dimensions, the enclosing dimensions used were four (i.e., -dim 4 option), with a default grid spacing of 1.4 Å. For this study, we only considered the largest of the cavities found (-mode max option), assuming this cavity would be the pore tunnel, and we supplied an index file for sodium or potassium ions in order to detect their overlap with the tunnel volume during each simulation:

```
trj_cavity -s protein+ions.pdb -f protein+ions.xtc -n ions.ndx
-dim 4 -osstat ion_stat.pdb -ota tunnel_profile.xvg
-sector radius
```

The measured mean occupancy of K⁺ ions was ~1.5 times lower than that of Na⁺ ions within the cavity, in agreement with the reported conductivities of each ion. Interestingly, the study of ion presence within the tunnel cavity revealed both similarities and differences in binding site affinities for the two types of ions. As previously hypothesized,⁹⁵ both hydrated Na⁺ and K⁺ ions were retained in the region enclosed by the Thr51 side chains of the TLE consensus sequence, due in part to polar interactions with the ion solvation shells (Figure 6). Sodium ions were found to be distributed throughout the core TLE consensus region of the SF for ≥50% of the simulation time, consistent with the pore selectivity (Figure 6). On the other hand, two maximal distributions (of 25% simulation time each) were observed for potassium ions, at the bottom of the SF near Thr51/Leu52, and at the top, above Glu53/Ser54 (Figure 6). The observed discontinuity in the K⁺ ion permeation pathway reflects the reduced conductivity of the channel in comparison with Na⁺; i.e., the SF region between Leu52 and Glu53 acts as a barrier to K⁺ passage.

Inspection of the protein pore radius as a function of pore axis revealed a very similar mean profile along the SF for both systems (Figure 6). In each case, the bottleneck of the pore was situated within a polar ring comprising the side chains of Leu52 and Glu53, with a mean radius of $\sim 2.6 \pm 0.4$ Å. Both Na⁺ and K⁺ ions move across the majority of the channel hexacoordinated by water molecules. However, as the solvation

shells of both ions (3.1 Å for Na⁺ and 3.4 Å for K⁺) exceeds the bottleneck radius, both need to lose some of those waters in order to translocate the pore. This is reflected in the reduced water density in this region (Figure 6). Following ion dehydration, the Glu53 side chain carboxylate groups help to compensate for the partial solvation shell. The smaller size (and higher effective charge density) of Na⁺ leads to more favorable interactions with the Glu53 ring, lowering the energy barrier necessary for the transition in comparison with K⁺, thus helping to determine the channel selectivity.

Assessment of Performance. The performance and relative quality of output of *trj_cavity* were compared with two popular dynamics-compatible cavity search programs, MDpocket⁵⁶ and CAVER 3.0.⁵⁵ It should be noted that the algorithms used in these programs are based on variations of the alpha shape theory, first implemented in CAST,²⁶ whose $O(n^2)$ relationship is computationally demanding. To improve performance, as is key for the analysis of large proteins and/or long-time scale trajectories, both CAVER and MDpocket rely on modified algorithms, which necessitate specialization for particular cavity types. The MDpocket cavity detection suite was based on a previously developed tool, fpocket,⁹⁶ which uses a combination of the alpha shape method²⁵ and Voronoi tessellation, and which is specifically designed to target ligand binding pockets. CAVER was also based on an open source program, qhull,⁹⁷ which implements another variation of the alpha shape method with Delaunay triangulation, combining its results with graph theory in order to find the shortest path to exit pathways. Therefore, CAVER is designed to identify tunnels or tunnel-like cavities. In contrast with the specializations of MDpocket and CAVER, the grid-based algorithm implemented within *trj_cavity* is not restrictive to the form of cavity analyzed.

Since each tool uses a different algorithm, default parameters were used in each case. CAVER 3.0 requires a seed point for the cavity calculation, and this was obtained from *trj_cavity*. The MDpocket cavity search algorithm does not require a predefined starting point but does require the coordinates of a specific cavity in order to calculate its descriptors. Therefore, to ensure a fair comparison, the initial cavity coordinates were again provided by *trj_cavity*. Only *trj_cavity* could characterize all cavity types; the tunnel-based characterization of CAVER was unsuitable for the single voluminous Der p 2 cavity (Figure 7A), while MDpocket is specialized for surface pockets and was not designed to detect the channel within NavMs (Figure 7C). Moreover, even MDpocket could not correctly detect or characterize the catalytic cavity that evolved during simulation of Pula, due to the high mobility of the protein domains which hampered the fitting process. Nevertheless, in the cases where comparisons could be made, reasonable agreement was found in the quantitative output descriptors. Thus, the average shape defined by MDpocket and *trj_cavity* of the collapsed internal hydrophobic cavity of Der p 2 was similar (Figure 7A), with a mean volume estimate of $0.4 \pm 0.2 \text{ nm}^3$ and $0.3 \pm 0.1 \text{ nm}^3$ for the two programs, respectively (Supporting Information Figure S3A); the small fluctuations in total volume estimated by *trj_cavity* reflect the structural stability of the protein fold (Supporting Information Figure S3B). We also attempted to compare the bottleneck radii within NavMs. CAVER generated a radically different output from *trj_cavity*, with several tunnel clusters covering the protein pore (Figure 7C); however, we selected one of the clusters from CAVER located within the selectivity filter, and its bottleneck radius of $2.3 \pm 0.1 \text{ \AA}$ was

comparable to the mean radius of $2.6 \pm 0.3 \text{ \AA}$ calculated by *trj_cavity* (Supporting Information Figure S3C) when bound at the equivalent channel-axis coordinate ($z = -11 \text{ \AA}$, at the center of the selectivity filter, Figure 6C).

To test the quality of the NavMs channel description in detail, we also compared the output with that of HOLE,³⁰ a popular tool for the analysis of the dimensions of transmembrane protein pores. Analysis of the pore radius as a function of the z -axis revealed similar mean profiles, particularly around the selectivity filter region (Figure 7B). However, a key difference may be observed within the profiles between ~ 2 to $\sim 8 \text{ \AA}$ along the z -axis, due to the presence of tunnels or “fenestrations” which connect the pore to the outer transmembrane surface of the protein and which have been proposed to act as entry portals for hydrophobic ligands or lipids.⁹⁴ As may be observed graphically in Figure 7C, *trj_cavity* (as well as CAVER) detects these fenestrations, in contrast to HOLE, leading to the divergence in pore radius by as much as $\sim 6 \text{ \AA}$. This is because the algorithm within HOLE finds the best route for a sphere (or spherocylinder) of variable radius to squeeze through the channel, which may occasionally limit accuracy when characterizing pores with complex topologies such as NavMs.

The computational performance of *trj_cavity* was measured in terms of total calculation time as a function of protein size, for 100 ns simulation trajectories with a 10 ps output rate (10,000 frames) for Der p 2 ($\sim 2,000$ atoms), NavMs ($\sim 5,500$ atoms), and Pula ($\sim 16,000$ atoms). The total CPU time (including user and system calls) over 100 frames for each system is reported in Figure 8. While each program exhibited an approximately linear relationship between system size and total calculation time, *trj_cavity* consistently outperformed the others (Figure 8, Tables 3–5), with a calculation cost of $\leq 18 \text{ ms atom}^{-1}$, or just $\leq 0.4 \text{ ms atom}^{-1}$ for cavity characterization when using a seed point. In contrast, speeds for cavity characterization of ~ 60 – 230 ms atom^{-1} were achieved for MDpocket, increasing to ~ 90 – 900 ms atom^{-1} in the case of CAVER. Thus, *trj_cavity* generates its cavity descriptors for a range of different protein types and cavity topologies, at a rate that is orders of magnitude faster compared to other tools. Similar speed-ups were observed when calculating cavities for a single frame; with a defined seed point, *trj_cavity* achieved speeds of $\leq 0.007 \text{ ms atom}^{-1}$, compared to $\leq 2 \text{ ms atom}^{-1}$ and $\leq 6 \text{ ms atom}^{-1}$ for MDpocket and Caver, respectively. In addition, for even a small protein such as Der p 2, the complete 10,000-frame trajectory corresponded to a disk space usage of $\sim 2.5 \text{ GB}$ for the input PDBs required for MDpocket and CAVER 3.0, whereas the equivalent compressed GROMACS trajectory occupied $\sim 70 \text{ MB}$. Similarly, the size of the output files provided by these tools can be significant (Figure 8, Tables 3–5) in comparison with *trj_cavity*.

CONCLUSIONS

As a result of increasing computer power and more efficient algorithms, MD simulations have started to reach biologically relevant time scales, toward the μs –ms regime and beyond. Consequently, the resultant trajectories require development of more efficient analysis tools, in terms of both calculation speed and disk storage. *trj_cavity* has been introduced as a novel tool for the all-purpose and efficient identification and characterization of protein cavities, pockets, and tunnels within long-time scale simulation trajectories. It has been designed to facilitate the analysis of protein cavities while optimizing both

speed and usability. Furthermore, the accuracy of the program can be effectively tuned by modifying just two parameters (protein surrounding dimensions and grid size), simplifying the use of this program and the subsequent interpretation of its results. Implemented within the GROMACS suite, it allows the rapid processing of compressed trajectories in GROMACS compatible formats, as well as providing the use of atom index groups. The results are also produced in standard, general file formats that allow direct inspection within a range of plotting and molecular graphics visualization tools.

To test the validity of its results, *trj_cavity* has been used to characterize the dynamic behavior of a variety of protein cavity types, yielding useful biological insights. In comparison with other common cavity and channel analysis programs, *trj_cavity* consistently demonstrated accurate cavity definition coupled to better computational performance. The straightforward calculation of geometrical descriptors such as the cavity volume and tunnel cross-sectional area/radius also enabled a systematic comparison of cavities and their dynamics across each trajectory. Only *trj_cavity* could characterize the complete range of cavity topologies tested, while reasonable agreement could be demonstrated when comparing with those tools which could be successfully applied to common systems in terms of both numerical and graphical descriptors. Thus, the generic nature of the cavity analysis algorithm and the relatively small number of input parameters required provides an “all-in-one” tool for generalized analysis of all forms of protein cavity.

ASSOCIATED CONTENT

Supporting Information

Additional figures demonstrating the dependence of input parameters upon cavity characterization by *trj_cavity*, quantitative cavity descriptors, and predicted sugar binding modes in PuLA. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*(P.J.B.) Mailing address: Bioinformatics Institute (A*STAR), 30 Biopolis Str, #07-01 Matrix, Singapore 138671. Tel.: +65 6478 8378. E-mail: peterjb@bii.a-star.edu.sg.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the Darwin Supercomputer of the University of Cambridge, the Swiss National Supercomputing Center, via DECI/PRACE-2IP, and the HECToR supercomputing service via the UK High-End Computing Consortium for Bimolecular Simulation/EPSRC. We thank Florian Roessler, Maite Ortiz-Suarez, and Rubben Torella for helpful discussions.

REFERENCES

- (1) Gutteridge, A.; Thornton, J. M. Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **2005**, *30*, 622–629.
- (2) Baron, R.; McCammon, J. A.; Mattevi, A. The oxygen-binding vs. oxygen-consuming paradigm in biocatalysis: structural biology and biomolecular simulation. *Curr. Opin Struct. Biol.* **2009**, *19*, 672–679.
- (3) Weatherman, R. V.; Fletterick, R. J.; Scanlan, T. S. Nuclear-receptor ligands and ligand-binding domains. *Annu. Rev. Biochem.* **1999**, *68*, 559–581.
- (4) Arias, H. R. Topology of ligand binding sites on the nicotinic acetylcholine receptor. *Brain Res. Brain Res. Rev.* **1997**, *25*, 133–191.
- (5) Schumacher, T. N.; Heemels, M. T.; Neefjes, J. J.; Kast, W. M.; Melief, C. J.; Ploegh, H. L. Direct binding of peptide to empty MHC class I molecules on intact cells and in vitro. *Cell* **1990**, *62*, 563–567.
- (6) Hertzel, A. V.; Bernlohr, D. A. The mammalian fatty acid-binding protein multigene family: molecular and genetic insights into function. *Trends Endocrinol. Metab.* **2000**, *11*, 175–180.
- (7) Brigl, M.; Brenner, M. B. CD1: antigen presentation and T cell function. *Annu. Rev. Immunol.* **2004**, *22*, 817–890.
- (8) Viriyakosol, S.; Tobias, P. S.; Kitchens, R. L.; Kirkland, T. N. MD-2 binds to bacterial lipopolysaccharide. *J. Biol. Chem.* **2001**, *276*, 38044–38051.
- (9) Garzon, D.; Anselmi, C.; Bond, P. J.; Faraldo-Gomez, J. D. Dynamics of the Antigen-binding Grooves in CD1 Proteins: Reversible Hydrophobic Collapse in the Lipid-Free State. *J. Biol. Chem.* **2013**, *288*, 19528–19536.
- (10) Paramo, T.; Piggot, T. J.; Bryant, C. E.; Bond, P. J. The structural basis for endotoxin-induced allosteric regulation of the TLR4 innate immune receptor. *J. Biol. Chem.* **2013**, *288*, 36215–36225.
- (11) Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580.
- (12) Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discovery* **2006**, *5*, 993–996.
- (13) Chang, Y. C.; Wu, W.; Zhang, J. L.; Huang, Y. Allosteric activation mechanism of the cys-loop receptors. *Acta Pharmacol. Sin.* **2009**, *30*, 663–672.
- (14) Traynelis, S. F.; Wollmuth, L. P.; McBain, C. J.; Menniti, F. S.; Vance, K. M.; Ogden, K. K.; Hansen, K. B.; Yuan, H.; Myers, S. J.; Dingledine, R. Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **2010**, *62*, 405–496.
- (15) Baenziger, J. E.; Corringer, P. J. 3D structure and allosteric modulation of the transmembrane domain of pentameric ligand-gated ion channels. *Neuropharmacology* **2011**, *60*, 116–125.
- (16) Baron, R.; Riley, C.; Chenprakhon, P.; Thotsaporn, K.; Winter, R. T.; Alfieri, A.; Forneris, F.; van Berkel, W. J.; Chaiyen, P.; Fraaije, M. W.; Mattevi, A.; McCammon, J. A. Multiple pathways guide oxygen diffusion into flavoenzyme active sites. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 10603–10608.
- (17) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (18) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **1992**, *255*, 178–183.
- (19) Ohmura, T.; Ueda, T.; Ootsuka, K.; Saito, M.; Imoto, T. Stabilization of hen egg white lysozyme by a cavity-filling mutation. *Protein Sci.* **2001**, *10*, 313–320.
- (20) Kadono, T.; Chatani, E.; Hayashi, R.; Moriyama, H.; Ueki, T. Minimization of cavity size ensures protein stability and folding: structures of Phe46-replaced bovine pancreatic RNase A. *Biochemistry* **2003**, *42*, 10651–10658.
- (21) Henrich, S.; Salo-Ahen, O. M.; Huang, B.; Rippmann, F. F.; Cruciani, G.; Wade, R. C. Computational approaches to identifying and characterizing protein binding sites for ligand design. *J. Mol. Recognit.* **2010**, *23*, 209–219.
- (22) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
- (23) Ichikawa, S.; Hatanaka, H.; Yuuki, T.; Iwamoto, N.; Kojima, S.; Nishiyama, C.; Ogura, K.; Okumura, Y.; Inagaki, F. Solution structure of Der f 2, the major mite allergen for atopic diseases. *J. Biol. Chem.* **1998**, *273*, 356–360.
- (24) Voss, N. R.; Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **2010**, *38*, W555–W562.
- (25) Edelsbrunner, H.; Mücke, E. P. Three-dimensional alpha shapes. *ACM Trans. Graphics* **1994**, *13*, 43–72.

- (26) 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*, 1884–1897.
- (27) Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics* **2006**, *7*, 316.
- (28) Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R. MolAxis: efficient and accurate identification of channels in macromolecules. *Proteins* **2008**, *73*, 72–86.
- (29) Laskowski, R. A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graphics* **1995**, *13* (323–330), 307–328.
- (30) Smart, O. S.; Neduvilil, J. G.; Wang, X.; Wallace, B. A.; Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graphics* **1996**, *14* (354–360), 376.
- (31) Kleywegt, G. J.; Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 178–185.
- (32) Exner, T.; Keil, M.; Moekel, G.; Brickmann, J. Identification of Substrate Channels and Protein Cavities. *J. Mol. Model.* **1998**, *4*, 340–343.
- (33) Bond, P. J.; Faraldo-Gomez, J. D.; Deol, S. S.; Sansom, M. S. Membrane protein dynamics and detergent interactions within a crystal: a simulation study of OmpA. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 9518–9523.
- (34) Bond, P. J.; Derrick, J. P.; Sansom, M. S. Membrane simulations of OpcA: gating in the loops? *Biophys. J.* **2007**, *92*, L23–25.
- (35) Mobley, D. L.; Dill, K. A. Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure* **2009**, *17*, 489–498.
- (36) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **2012**, *41*, 429–452.
- (37) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.
- (38) Hurst, D. P.; Grossfield, A.; Lynch, D. L.; Feller, S.; Romo, T. D.; Gawrisch, K.; Pitman, M. C.; Reggio, P. H. A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem.* **2010**, *285*, 17954–17964.
- (39) Khalili-Araghi, F.; Tajkhorshid, E.; Roux, B.; Schulten, K. Molecular dynamics investigation of the omega-current in the Kv1.2 voltage sensor domains. *Biophys. J.* **2012**, *102*, 258–267.
- (40) Fowler, P. W.; Beckstein, O.; Abad, E.; Sansom, M. S. Detailed Examination of a Single Conduction Event in a Potassium Channel. *J. Phys. Chem. Lett.* **2013**, *4*, 3104–3109.
- (41) Romo, T. D.; Grossfield, A.; Pitman, M. C. Concerted interconversion between ionic lock substates of the beta(2) adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* **2010**, *98*, 76–84.
- (42) Nury, H.; Poitevin, F.; Van Renterghem, C.; Changeux, J. P.; Corringer, P. J.; Delarue, M.; Baaden, M. One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 6275–6280.
- (43) Jensen, M. O.; Jogini, V.; Borhani, D. W.; Leffler, A. E.; Dror, R. O.; Shaw, D. E. Mechanism of voltage gating in potassium channels. *Science* **2012**, *336*, 229–233.
- (44) Bond, P. J.; Faraldo-Gomez, J. D. Molecular mechanism of selective recruitment of Syk kinases by the membrane antigen-receptor complex. *J. Biol. Chem.* **2011**, *286*, 25872–25881.
- (45) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **2011**, *133*, 9181–9183.
- (46) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13118–13123.
- (47) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
- (48) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879–1881.
- (49) Ivetac, A.; McCammon, J. A. Mapping the druggable allosteric space of G-protein coupled receptors: a fragment-based molecular dynamics approach. *Chem. Biol. Drug Des.* **2010**, *76*, 201–217.
- (50) Wise, J. G. Catalytic transitions in the human MDR1 P-glycoprotein drug binding sites. *Biochemistry* **2012**, *51*, 5125–5141.
- (51) Arnarez, C.; Mazat, J. P.; Elezgaray, J.; Marrink, S. J.; Periole, X. Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc1. *J. Am. Chem. Soc.* **2013**, *135*, 3112–3120.
- (52) Bhinge, A.; Chakrabarti, P.; Uthanumallian, K.; Bajaj, K.; Chakrabarty, K.; Varadarajan, R. Accurate detection of protein-ligand binding sites using molecular dynamics simulations. *Structure* **2004**, *12*, 1989–1999.
- (53) Raunest, M.; Kandt, C. dxTuber: detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J. Mol. Graphics Modell.* **2011**, *29*, 895–905.
- (54) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (33–38), 27–38.
- (55) Chovancova, E.; Pavalka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **2012**, *8*, e1002708.
- (56) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **2011**, *27*, 3276–3285.
- (57) Craig, I. R.; Pfleger, C.; Gohlke, H.; Essex, J. W.; Spiegel, K. Pocket-space maps to identify novel binding-site conformations in proteins. *J. Chem. Inf. Model.* **2011**, *51*, 2666–2679.
- (58) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graphics Modell.* **1997**, *15* (359–363), 389.
- (59) Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery through the computational microscope. *Structure* **2009**, *17*, 1295–1306.
- (60) Hess, B.; Kutzner, C.; Van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (61) Derewenda, U.; Li, J.; Derewenda, Z.; Dauter, Z.; Mueller, G. A.; Rule, G. S.; Benjamin, D. C. The crystal structure of a major dust mite allergen Der p 2, and its biological implications. *J. Mol. Biol.* **2002**, *318*, 189–197.
- (62) McCusker, E. C.; Bagneris, C.; Naylor, C. E.; Cole, A. R.; D'Avanzo, N.; Nichols, C. G.; Wallace, B. A. Structure of a bacterial voltage-gated sodium channel pore reveals mechanisms of opening and closing. *Nat. Commun.* **2012**, *3*, 1102.
- (63) Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441–451.
- (64) Fiser, A.; Sali, A. Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **2003**, *374*, 461–491.
- (65) Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. Presented at ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida, November 11–17, 2006.
- (66) MacKerell, A. D.; Bashford, D.; Bellott, Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical

- Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (67) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.
- (68) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
- (69) Beglov, D.; Roux, B. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* **1994**, *100*, 9050–9063.
- (70) Joung, I. S.; Cheatham, T. E., 3rd. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (71) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (72) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (73) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; McGraw-Hill Inc.: 1981.
- (74) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (75) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (76) Inohara, N.; Nunez, G. ML – a conserved domain involved in innate immunity and lipid metabolism. *Trends Biochem. Sci.* **2002**, *27*, 219–221.
- (77) Shimazu, R.; Akashi, S.; Ogata, H.; Nagai, Y.; Fukudome, K.; Miyake, K.; Kimoto, M. MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J. Exp. Med.* **1999**, *189*, 1777–1782.
- (78) Miyake, K.; Ogata, H.; Nagai, Y.; Akashi, S.; Kimoto, M. Innate recognition of lipopolysaccharide by Toll-like receptor 4/MD-2 and RP105/MD-1. *J. Endotoxin Res.* **2000**, *6*, 389–391.
- (79) Friedland, N.; Liou, H. L.; Lobel, P.; Stock, A. M. Structure of a cholesterol-binding protein deficient in Niemann-Pick type C2 disease. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2512–2517.
- (80) Wright, C. S.; Zhao, Q.; Rastinejad, F. Structural analysis of lipid complexes of GM2-activator protein. *J. Mol. Biol.* **2003**, *331*, 951–964.
- (81) Mueller, G. A.; Benjamin, D. C.; Rule, G. S. Tertiary structure of the major house dust mite allergen Der p 2: sequential and structural homologies. *Biochemistry* **1998**, *37*, 12707–12714.
- (82) Nicholls, A.; Sharp, K. A.; Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **1991**, *11*, 281–296.
- (83) Herre, J.; Gronlund, H.; Brooks, H.; Hopkins, L.; Waggoner, L.; Murton, B.; Gangloff, M.; Opaleye, O.; Chilvers, E. R.; Fitzgerald, K.; Gay, N.; Monie, T.; Bryant, C. Allergens as Immunomodulatory Proteins: The Cat Dander Protein Fel d 1 Enhances TLR Activation by Lipid Ligands. *J. Immunol.* **2013**, *191*, 1529–1535.
- (84) Hii, S. L.; Tan, J. S.; Ling, T. C.; Ariff, A. B. Pullulanase: role in starch hydrolysis and potential industrial applications. *Enzyme Res.* **2012**, *2012*, 921362.
- (85) Doman-Pytka, M.; Bardowski, J. Pullulan degrading enzymes of bacterial origin. *Crit. Rev. Microbiol.* **2004**, *30*, 107–121.
- (86) Lee, S. P.; Morikawa, M.; Takagi, M.; Imankaka, T. Cloning of the aapT gene and characterization of its product, alpha-amylase-pullulanase (AapT), from thermophilic and alkaliphilic *Bacillus* sp. strain XAL601. *Appl. Environ. Microbiol.* **1994**, *60*, 3764–3773.
- (87) Liebl, W.; Stemplinger, I.; Ruile, P. Properties and gene structure of the *Thermotoga maritima* alpha-amylase AmyA, a putative lipoprotein of a hyperthermophilic bacterium. *J. Bacteriol.* **1997**, *179*, 941–948.
- (88) Mikami, B.; Iwamoto, H.; Malle, D.; Yoon, H. J.; Demirkiran-Sarikaya, E.; Mezaki, Y.; Katsuya, Y. Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site. *J. Mol. Biol.* **2006**, *359*, 690–707.
- (89) Yamashita, M.; Matsumoto, D.; Murooka, Y. Amino acid residues specific for the catalytic action towards α -1,6-glucosidic linkages in *Klebsiella* pullulanase. *J. Ferment. Bioeng.* **1997**, *84*, 283–290.
- (90) Hille, B. *Ion channels of excitable membranes*, 3rd ed.; Sinauer Associates, Inc.: Sunderland, MA, 2001.
- (91) Catterall, W. A. Structure and function of voltage-gated ion channels. *Annu. Rev. Biochem.* **1995**, *64*, 493–531.
- (92) Sands, Z.; Grottesi, A.; Sansom, M. S. Voltage-gated ion channels. *Curr. Biol.* **2005**, *15*, R44–47.
- (93) Bezanilla, F. Voltage-gated ion channels. *IEEE Trans. Nanobioscience* **2005**, *4*, 34–48.
- (94) Ulmschneider, M. B.; Bagneris, C.; McCusker, E. C.; Decaen, P. G.; Delling, M.; Clapham, D. E.; Ulmschneider, J. P.; Wallace, B. A. Molecular dynamics of ion transport through the open conformation of a bacterial voltage-gated sodium channel. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 6364–6369.
- (95) Payandeh, J.; Gamal El-Din, T. M.; Scheuer, T.; Zheng, N.; Catterall, W. A. Crystal structure of a voltage-gated sodium channel in two potentially inactivated states. *Nature* **2012**, *486*, 135–139.
- (96) Le Guilloux, V.; Schmidtko, P.; Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **2009**, *10*, 168.
- (97) Barber, C. B.; Dobkin, D. P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Software* **1996**, *22*, 469–483.