Identification of protein pockets and cavities by Euclidean Distance Transform

Sebastian Daberdaku

Department of Information Engineering - University of Padova

sebastian.daberdaku@unipd.it

Abstract

Protein pockets and cavities usually coincide with the active sites of biological processes, and their identification is significant since it constitutes an important step for structure-based drug design and protein-ligand docking applications. This research presents PoCavEDT, an automated purely geometric technique for the identification of binding pockets and occluded cavities in proteins based on the 3D Euclidean Distance Transform. Candidate protein pocket regions are identified between two Solvent-Excluded surfaces generated with the Euclidean Distance Transform using different probe spheres, which depend on the size of the binding ligand. The application of simple, yet effective geometrical heuristics ensures that the proposed method obtains very good ligand binding site prediction results. The method was applied to a representative set of protein-ligand complexes and their corresponding unbound protein structures to evaluate its ligand binding site prediction capabilities. Its performance was compared to the results achieved with several purely geometric pocket and cavity prediction methods, namely SURFNET, PASS, CAST, LIGSITE, LIGSITE^{CS}, PocketPicker and POCASA. Success rates PoCavEDT were comparable to those of POCASA and outperformed the other software.

Introduction

Proteins play a wide range of biological functions such as signal transmission, immune defence, structural support, transport, storage, biochemical reaction catalysis and motility processes, usually by interacting with other molecules (Berggård et al. 2007). Interaction with small molecules such as substrates, coenzymes and drugs, generally occurs in concave regions of the protein surface known as pockets or in voids enveloped by the protein surface called cavities. Therefore, the identification of protein pockets and cavities is of great interest for furthering our understanding of protein interfaces and interactions, and especially for applications such as protein-ligand docking (Krivák et al. 2015) or structure-based drug design (Y. Lee et al. 2018).

Given its importance, several *in silico* methods have been developed over the years for the protein pocket and cavity detection task, which can be roughly divided into three broad categories: (1) evolutionary-based algorithms (Pupko et al. 2002) that rely on multiple sequence alignments to find pockets and cavities on the protein surface, (2) energy-based algorithms (Rodney et al. 2007; Soga et al. 2007) that detect pockets and cavities by computing the interaction potential between the protein and the ligand molecule, and (3) geometric algorithms (Brady et al. 2000; Weisel et al. 2007) that detect pockets and cavities starting from the protein's three-dimensional structure information as retrieved from a Protein Data Bank (Berman et al. 2000) (PDB) file. A few consensus methods that combine the results of two or more techniques to improve the success rate in predicting the location of binding pockets or cavities have also been introduced (Huang 2009; Zhang et al. 2011).

The geometric algorithms are the most commonly used search methods since they do not require any other non-geometric knowledge but the 3D structure of the protein. In the current literature there is no consensus about the definition or classification of protein pockets and/or cavities (Simões et al. 2017), and the two terms are often used interchangeably. For the sake of clarity, in this work, a pocket is defined as a concave protein surface feature accessible to the outer solvent; a cavity is defined as an inner void inside the protein surface that is not accessible to the outer solvent.

Kuntz et al. pioneered the field by introducing a sphere-based method for the geometric docking of macromolecules with small ligands (Kuntz et al. 1982). A similar approach was later implemented in SURFNET (Laskowski 1995), where pockets/cavities and the ligand are filled with probe spheres of varying sizes, and the candidate sites are identified by evaluating the degree of overlap among spheres. Putative Active Sites with Spheres (PASS) (Brady et al. 2000) is another sphere-based method where the pockets and cavities are filled with layers of probe spheres of decreasing size. For each layer, only spheres with a certain degree of "buriedness" (low solvent exposure) are retained, and the filling procedure stops when no new spheres can be added. Potential binding sites are then identified as the centres of probe sphere clusters called Active Site Points.

The CAST method (Jie et al. 1998) implements a pocket identification approach based on the theory of 3D α-shapes (Edelsbrunner et al. 1994, 1998). The method builds two α-shape envelopes for a given protein: the convex hull of the atomic centres as the outer envelope and the dual subcomplex of the union of the atom spheres (i.e. the α-shape that is entirely inside such union of spheres); protein pockets are identified in the space between these two envelopes. The LIGSITE method (Hendlich et al. 1997) is a grid-based algorithm (the protein is mapped onto an axis-aligned regular cubic grid) that searches for cavities along the *x*, *y*, and *z* axes and along the Cartesian cubic diagonals for a total of 7 directions (14 oriented directions). LIGSITE^{CS} (Huang et al. 2006) further improves accuracy over LIGSITE by capturing surface-solvent-surface events on Connolly's protein surface (Connolly 1983), instead of capturing atom-solvent-atom events. Weisel et al. later introduced PocketPicker (Weisel et al. 2007) a method similar to LIGSITE that carries out scans along 30 directions equally distributed on a sphere. A "buriedness" index is computed for each grid point by counting the number of directions (out of 30) that come into contact with a protein atom: a pocket consists of connected grid points with a buriedness index greater than a certain threshold.

More recently, Yu et al. introduced the Pocket-Cavity Search Application (POCASA), which implements a sphere-based grid algorithm called Roll (J. Yu et al. 2010). The protein is initially placed inside a regular 3D grid, and a large probe sphere is rolled on the outer atoms of the protein. The surface generated by the probe works as a second envelope of the protein: pockets are identified between the protein's surface and the probe surface. Dias et al. introduced a similar method which uses two Gaussian surfaces instead (Dias et al. 2017).

This paper presents a novel purely geometric algorithm for the detection of ligand binding protein pockets and cavities based on the Euclidean Distance Transform (EDT). The EDT can be used to compute the Solvent-Excluded surface for any given probe sphere radius value at high resolutions and in a timely manner (Daberdaku et al. 2016, 2018). The algorithm is adaptive to the specific candidate ligand: it computes two voxelised protein surfaces using two different probe sphere radii whose sizes depend on the shape of the candidate ligand. The pocket regions consist of the voxels located between the two surfaces, which exhibit a certain minimum depth value from the outer surface. The distance map values computed by the EDT algorithm during the second surface

computation can be used to elegantly determine the depth of each candidate pocket and to rank them accordingly. Cavities on the other hand, are identified by scanning the inside of the protein for voids. The algorithm determines and outputs the best *k* candidate pockets and cavities, i.e. the ones that are more likely to bind to the given ligand. The proposed approach is experimentally compared to other purely geometric pocket and cavity search algorithms and is shown to outperform most of the previously developed methods.

Methods

Protein surface definitions

Several definition of protein surface have been introduced, each one having different characteristics and various degrees of detail. The most commonly encountered surface definitions are: the van der Waals surface (vdW) (Bash et al. 1983), the Lee-Richards surface, also known as the Solvent-Accessible surface (SAS), (B. Lee et al. 1971) and the Connolly surface or Solvent-Excluded surface (SES) (Connolly 1983) (see Fig. 1).

We can imagine a protein as being represented by a set of possibly overlapping spheres, with each sphere representing a specific atom in the protein. The vdW of that protein is then defined as the topological boundary of this set of spheres. The SAS is traced by the centre of a probe sphere with radius equal to the radius of the solvent molecule as it rolls over the van der Waals surface. Thus, the outer space consists of the points at which the probe sphere can be placed without overlapping with the atoms of the molecule. The SES is defined as the union of the contact surface and the re-entrant surface: the contact surface is the part of the vdW touched by the probe sphere while it rolls over it; the re-entrant surface is composed of the inward-facing surface portions of the probe when it touches two or more atoms.

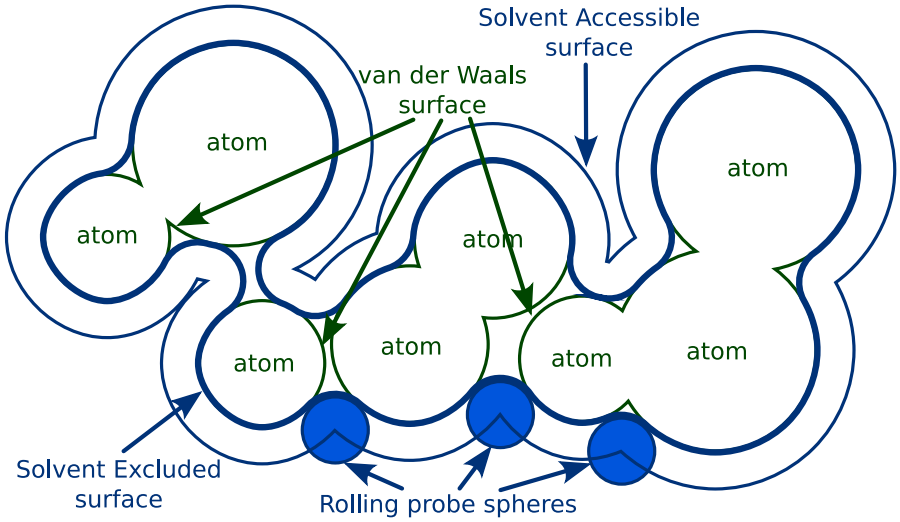


***Figure 1**: Surface definitions: van der Waals, Solvent-Accessible and Solvent-Excluded surfaces.*

The SES represents a continuous functional surface of the protein, i.e. the surface that is available to interact with. The SAS is displaced outward from the SES by a distance equal to the probe sphere radius. The SAS and SES nomenclatures will be maintained throughout the paper although probe spheres of various radii will be used (not necessarily of the size of the solvent molecule). Both surfaces used for the cavity detection procedure are SE surfaces obtained using two different probe sphere radii.

Surface calculation algorithm

The SES computation starts with the acquisition of the 3D representation of a molecule from the related PDB file. The algorithm then calculates the tightest axis-aligned bounding-box enclosing the whole protein by determining the minimal and maximal coordinates of all atoms. The bounding-box is iscretised into a voxel grid, given the user-defined resolution parameter, i.e. the number of voxels per $Å^3$. All the atomic coordinates are then translated, scaled and quantised into the new coordinate system defined by the voxel grid. The algorithm then computes the voxelised space-filling model of the molecule. Each atom in the molecule is represented by a ball having a radius equal to the atom's radius increased by the probe sphere's radius. The algorithm computes the voxelised representation of each atom and marks the corresponding voxels in the voxel grid as occupied using an adaptation of the Midpoint Circle Algorithm (Bresenham 1965) to efficiently determine the voxels needed to represent a ball in a discrete 3D grid.

The SAS is extracted directly from the voxelised space-filling model. The algorithm extracts all boundary voxels from the space-filling model by implementing a fast and efficient 3D seed-filling algorithm, first introduced by (W.-W. Yu et al. 2010). The calculation of the SES is more complex as it includes the re-entrant surface portions. There is a strict geometrical relationship between the Solvent Accessible and Solvent Excluded surfaces that can be exploited for the computation of the latter: for each point in the SES, its nearest SAS point is at exactly one probe-sphere-radius distance. This means that the SES can be calculated starting from the SAS by employing a surface smoothing algorithm such as the Euclidean Distance Transform (EDT).

The Euclidean Distance Transform

A distance transform (also known as distance map or distance field), is a derived representation of a digital binary image. Distance maps are images where the value of each free voxel is the distance to the nearest occupied voxel. Let $B \in \{0,1\}^{l \times w \times h}$ be a binary voxel grid of length l, width w and height h. There are exactly $l \times w \times h$ voxels in B, each one identified by the ordered triple $v = (i,j,k) \in V = \{1,\dots,l\} \times \{1,\dots,w\} \times \{1,\dots,h\}$. Also, let $I_B: V \to \{0,1\}$ be the image function of B defined as $I_B(i,j,k) = b_{i,j,k} \in \{0,1\}$, where $b_{i,j,k}$ is the value of voxel (i,j,k) in B. Let V_O be the set of occupied voxels of B, i.e.

$$V_O = \{v = (i,j,k) \in V \vee I_B(i,j,k) = 1\}.$$

Also, let $NBV_B: V \to V_O$, such that $\forall v \in V$, $NBV_B(v)$ is a nearest occupied voxel of B to v, that is

$$NBV_B(v) \in arg\min_{w \in V_O} d(w,v) = \{w \in V_O \vee \forall y \in V_O: d(w,v) \leq d(y,v)\},$$

according to some distance metric d. $NBV_B(v)$ is called the nearest boundary voxel (NBV) of v in B. Clearly, if $v \in V_O$ then $NBV_B(v) = v$.

Finally, the distance transform of B is defined as a real-valued voxel grid $DT_B \in R^{l \times w \times h}$ such that

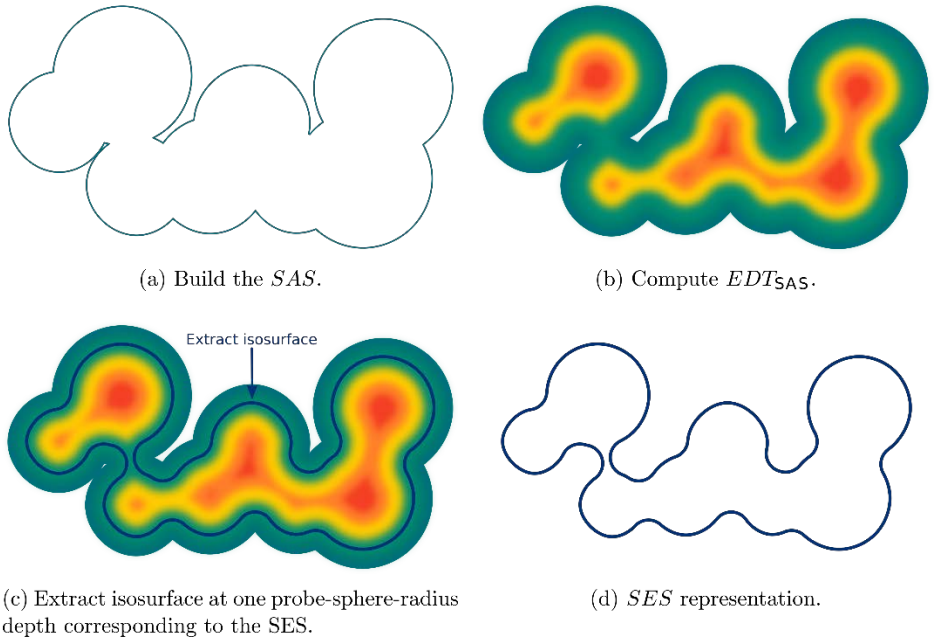
$$I_{DT_B} = d\big(v, NBV_B(v)\big), \forall v \in V,$$

where $I_{DT_B}: V \to R$ is the image function of DT_B.

When the chosen distance metric is the Euclidean Distance, we talk about Euclidean Distance Transform, however the squared Euclidean distance is usually used in order to avoid time-consuming square root calculations.

SES computation with EDT

Let SAS be the voxelised representation of the Solvent Accessible surface of a given protein, and let EDT_{SAS} be the Euclidean Distance transform of SAS. Because the SAS is displaced outward from the SES by a distance equal to the probe-sphere radius, the voxelised representation of the latter can be obtained from EDT_{SAS} by extracting all voxels with a distance value equal to one probe-sphere-radius (see Fig. 2). The SES computation algorithm is described in detail in (Daberdaku et al. 2016, 2018).



(a) Build the SAS. (b) Compute EDT_{SAS}.

(c) Extract isosurface at one probe-sphere-radius depth corresponding to the SES. (d) SES representation.

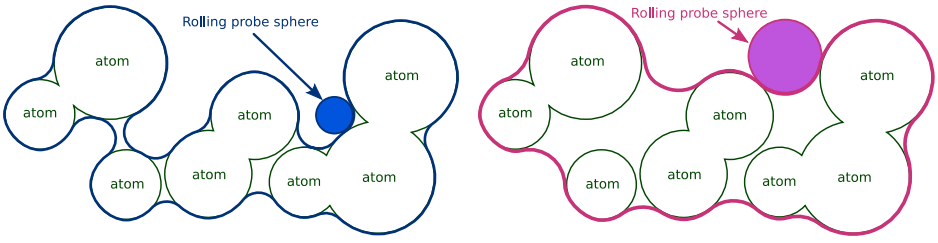
***Figure 2**: 2D representation of the SES calculation by EDT starting from the SAS. In Figs. 2b and 2c the transition of the gradient from blue to red corresponds to increasing distance values.*

Pocket and cavity detection

The pocket detection procedure for a given protein is carried out by computing two SE surfaces using two different probe-sphere-radii. To compute the inner surface, the smaller radius is either set to the radius of the water molecule, i.e. 1.4Å, or to the radius of the smallest atom in the ligand molecule if it is larger than 1.4Å. This way, the inner surface determines the protein's surface that is available to interact with the ligand while remaining impenetrable to the latter. The outer SE surface is computed using a larger probe-sphere-radius which also depends on the candidate ligand. Given a candidate ligand, Principal Component Analysis is performed on its atom centres in order to determine its length, width and height (where the length is the largest principal component, and the height is the smallest), and the probe-sphere-radius of the outer SE surface is heuristically set to either 4.0Å, or to half the height of the ligand if this quantity is greater than 4.0Å. Pocket regions are then identified between the outer and inner SE surfaces using the fast 3D seed-filling algorithm mentioned earlier: Fig. 3 schematically describes the procedure and Fig. 4 describes the ligand-binding pocket identification for PDB entry 1APU.

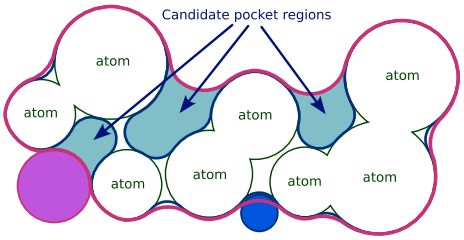
On the other hand, cavity detection is very straightforward and can be carried out immediately after the inner SES calculation using the same efficient 3D seed-filling algorithm. Please note that only internal cavities that can fit the probe sphere are detected during this phase: inter-atomic voids that cannot fit the probe sphere are eliminated during the SAS computation step.

To avoid considering pocket regions that are too shallow to effectively be able to host the ligand molecule, a minimum depth threshold was set in order for a voxel to belong to a candidate pocket region. An empty voxel between the inner and outer SE surfaces should be at a distance of at least one third of the ligand's height from the outer SES in order for it to be considered as belonging to a pocket region. This threshold was also chosen heuristically in order to guarantee a minimum degree of penetration of the ligand inside the outer SES.



(a) Inner SES computation.

(b) Outer SES computation.



(c) Identif cation of candidate pocket regions between the two SE surfaces.

Figure 3*: 2D representation of the pocket identification procedure. Figs. 3a and 3b represent, respectively, the inner and outer SES computation: pockets are identified between these two SE surfaces (Fig. 3c).*

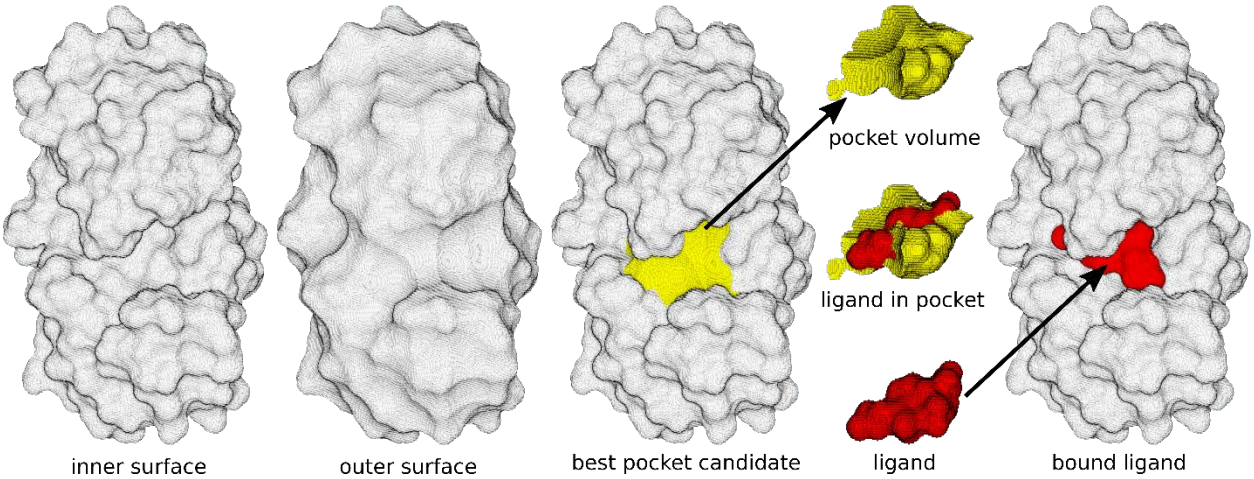


Figure 4*: Ligand-binding pocket identification for PDB entry 1APU.*

Ranking candidate pockets and cavities

In general, several candidate pockets (and cavities) are detected for a given target protein. Therefore, one must identify the candidates with the highest likelihood of being the actual ligand binding site. Candidate pockets can be ranked by their volume, following the intuition that pockets with larger volume are more likely to bind the ligand. This ranking scheme however would assign

the same score to pockets with the same volume but with varying depths. Large but shallow candidate pockets are less likely to host the ligand compared to the ones that go deep inside the protein, although they might have similar volume. For this reason, for each candidate pocket the proposed method computes the *weighted volume score*, i.e., the sum of the corresponding distance map values (squared distance values from the nearest voxel of the outer SAS which are available from the EDT computation step) of all the voxels belonging to the current pocket. The EDT algorithm used to compute the two surfaces also provides a very elegant means to sort candidate pockets. Candidate cavities, on the other hand, are sorted solely by their volume (sum of their voxels).

Results and discussion

The presented methodology was implemented in a program named PoCavEDT, and its ligand binding site prediction capabilities were compared with other purely geometric pocket and cavity identification software. The employed evaluation method is the same that was used in several previous studies (Brady et al. 2000; Huang et al. 2006; Weisel et al. 2007; J. Yu et al. 2010). Each predicted pocket and/or cavity is represented with a single point in the 3D space: a prediction is regarded as successful if any atom of the ligand is located within a 4.0Å distance from the point representing the candidate pocket or cavity; otherwise, the prediction is regarded as a failure. Each candidate cavity X (which can be seen as the set of voxels it is made of, scaled and translated back from the voxel grid coordinate system to the molecule's original one) is represented by its centre of mass c_X, formally given by: $c_X = \frac{1}{|X|}\sum_{v \in X} v$. Each candidate pocket P, on the other hand, can be represented by its *weighted centre of mass* w_P, computed by weighting each voxel in P with its corresponding distance map value from the outer SES computation: $w_P = \sum_{v \in P} d_{map}(v) \times v / \sum_{v \in P} d_{map}(v)$.

Candidate pockets and cavities are ranked separately at first since different scores are used for their comparison. Then, the top-3 pockets and top-3 cavities are selected from the corresponding ranked lists and are inserted in a new list which is then re-ranked using the inverse distance from their representative points to the protein's centroid as a score, so that the closest to the centre of the protein is considered the most likely to bind the ligand. A resolution of 125 voxels per $Å^3$ was used throughout all experiments.

The ligand binding site prediction capability of PoCavEDT was evaluated on a test set of 48 protein-ligand complexes and the respective unbound protein structures (Brady et al. 2000). This is the most widely used dataset for comparing pocket detection methods. The comparison of pocket and cavity predictions in bound and unbound structures is of special interest for geometric search algorithms in order to evaluate the degree of conformational changes the proteins undergo upon binding with their ligands. The prediction results were divided into two categories for quality assessment: "Top1-hits" indicating correct predictions and "Top3-hits" indicating predictions where the respective ligand is found within the top-3 scoring candidate pockets and cavities. The success rates of pocket predictions are summarized in Table 1. Table 2 shows the best results obtained by PoCavEDT on each protein in the testing set, indicating the rank of the predicted binding site and the distance between the representative pocket/cavity point and the nearest ligand atom.

***Table 1**: Binding site prediction success rate for 48 bound and unbound protein-ligand complexes.*

Method	Top1-hits		Top3-hits	
	Unbound (%)	Bound (%)	Unbound (%)	Bound (%)
PoCavEDT	75	79	92	96
(fixed parameters)*	69	71	88	94
POCASA	75	77	88	94
PocketPicker	69	72	85	85
LIGSITE^{CS}	60	69	77	87
LIGSITE	58	69	75	87
CAST	58	67	75	83
PASS	60	63	71	81
SURFNET	52	54	75	78

** Results obtained with PoCavEDT by using the following fixed parameters for all predictions: probe- sphere-radius for inner SES = 1.4Å, probe-sphere-radius for outer SES = 4.0Å, minimum pocket depth threshold = 2.67Å*

***Table 2**: Binding site prediction results for 48 bound and unbound protein-ligand complexes. Only the best hit is shown. Dashes indicate that the actual binding site was not found within the top-3 candidate pockets and/or cavities.*

Bound			Unbound		
PDB ID	Rank	Distance (in Å)	PDB ID	Rank	Distance (in Å)
1A6W	2	1.13	1A6U	3	0.85
1ACJ	1	1.36	1QIF	1	2.16
1APU	1	1.78	3APP	1	0.89
1BID	1	1.95	3TMS	1	2.41
1BLH	2	1.04	1DJB	1	0.87
1BYB	1	1.07	1BYA	1	2.54
1CDO	1	1.16	8ADH	1	1.63
1DWD	1	0.85	1HXF	1	0.91
1FBP	3	0.97	2FBP	2	0.73
1GCA	1	1.00	1GCG	1	1.25
1HEW	1	0.83	1HEL	1	0.75
1HFC	1	0.81	1CGE	1	0.72
1HYT	1	0.99	1NPC	1	0.50
1IDA	1	1.59	1HSI	1	1.96
1IGJ	3	0.79	1A4J	3	1.08
1IMB	1	1.54	1IME	1	1.43
1INC	1	1.09	1ESA	1	0.62
1IVD	2	1.23	1NNA	3	1.39
1MRG	1	0.26	1AHC	2	0.45
1MTW	1	0.57	2TGA	-	-
1OKM	1	0.52	4CA2	1	1.01
1PDZ	1	1.82	1PDY	1	1.41

1PHD	1	1.28	1PHC	1	1.28
1PSO	1	0.96	1PSN	1	1.18
1QPE	2	1.24	3LCK	2	1.19
1RBP	1	0.43	1BRQ	1	0.98
1RNE	1	1.47	1BBS	1	1.28
1ROB	1	1.05	8RAT	1	0.40
1SNC	1	0.33	1STN	1	2.74
1SRF	1	1.05	1PTS	1	0.82
1STP	1	1.27	1SWB	1	1.21
1ULB	2	1.14	1ULA	-	-
2CTC	1	0.54	2CTB	1	0.73
2H4N	1	1.11	2CBA	1	0.42
2IFB	1	2.27	1IFB	1	2.08
2PK4	1	1.53	1KRN	1	1.26
2SIM	2	0.52	2SIL	2	0.39
2TMN	1	1.44	1L3F	1	0.70
2YPI	1	1.02	1YPI	1	2.65
3GCH	1	0.85	1CHG	-	-
3MTH	-	-	6INS	2	2.99
3PTB	1	1.16	3PTN	1	0.99
4DFR	1	2.95	5DFR	1	2.39
4PHV	1	1.26	3PHV	-	-
5CNA	-	-	2CTV	1	0.79
5P2P	1	2.06	3P2P	1	1.56
6RSA	1	0.84	7RAT	1	0.40
7CPA	1	0.97	5CPA	1	0.50

The proposed pocket and cavity search algorithm depends on four parameters: the probe-sphere-radii of the inner and outer SE surfaces, the minimum pocket depth parameter and the resolution. A relatively high resolution parameter was chosen for all experiments (125 voxels per $Å^3$) in order to ensure that the pocket and cavity search procedures did not depend on the orientation of the protein in the voxel grid. The implemented EDT algorithm allows the efficient computation of voxelised protein surfaces at very high resolutions. None of the pocket and cavity detection steps in the proposed algorithm depend on the particular orientation of the protein inside the voxel grid, excluding the effects of the discretisation error which are minimized by using a high resolution parameter. Several combinations of the remaining three parameters were tested, and the best results were obtained by using a probe-sphere-radius of 1.4Å for the inner SES, a probe-sphere-radius of 4.0Å for outer SES and minimum pocket depth threshold of 2.67Å, as shown in Table 1. It is worth noticing that, in general, the adaptive parameter strategy based on the ligand size yields better prediction results.

This strategy however fails to correctly predict the ligand binding site in a few instances, as for 3MTH (Methylparaben Insulin) whose ligand binding pocket is rather wide and shallow. The algorithm automatically determines the following set of parameters: inner SES probe-sphere-radius = 1.7Å, outer SES probe-sphere-radius = 4.0Å, minimum pocket depth = 1.80Å. By doubling the probe-sphere-radius for the outer SES (8.0Å) and using a minimum pocket depth of 3.0Å, the algorithm identifies the correct ligand binding pocket of 3MTH at rank 1. This indicates that with a more complex parameter selection strategy, it might be possible to further enhance the automatic ligand binding site prediction capabilities of the proposed method.

Conclusion

The purely geometric pocket detection method PoCavEDT was developed and successfully applied to the ligand binding site prediction task in proteins. The search routine of PoCavEDT is capable of identifying the active site within a protein structure with a high success rate on a representative test set. The adaptive parameter selection strategy based on the ligand size yields better prediction results than any tested set of fixed parameters. A more complex parameter selection scheme could further improve the prediction capabilities of the proposed method.

The proposed method outputs the voxelised representation of protein pockets and cavities that are potential ligand binding sites. Voxelised representations are well-suited to represent multiple physicochemical and geometrical properties of molecular surfaces, as each voxel can describe multiple properties of a portion of the 3D space. Descriptor-based approaches could also benefit from the voxelised representation of pockets and cavities enriched with local physicochemical properties in order to implement fast and accurate pocket and cavity ranking, comparison and classification. These topics will be explored as future research.

Acknowledgements

This research has been partially supported by the University of Padova project CPDR150813/15 “Models and Algorithms for Protein-Protein Docking”.

The datasets, python scripts, binaries and supplementary materials generated and/or analysed during the current study are available in the figshare repository doi:10.6084/m9.figshare.6794231.

References

Armon A, Graur D, Ben-Tal N. (2001). ConSurf: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface Mapping of Phylogenetic Information. *Journal of Molecular Biology* 307 (1): 447–63.

Bash PA, Pattabiraman N, Huang C, Ferrin TE, Langridge R. (1983). Van Der Waals Surfaces in Molecular Modeling: Implementation with Real-Time Computer Graphics. *Science* 222 (4630): 1325–1327.

Berggård T, Linse S, James P. (2007). Methods for the Detection and Analysis of Protein–Protein Interactions. *Proteomics* 7 (16): 2833–42.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000). The Protein Data Bank. *Nucleic Acids Research* 28 (1): 235–42.

Brady GP, Stouten PFW. (2000). Fast Prediction and Visualization of Protein Binding Pockets with PASS. *Journal of Computer-Aided Molecular Design* 14 (4): 383–401.

Bresenham JE. (1965). Algorithm for Computer Control of a Digital Plotter. *IBM Systems Journal* 4 (1): 25–30.

Connolly ML. (1983). Analytical Molecular Surface Calculation. *Journal of Applied Crystallography* 16 (5): 548–558.

Daberdaku S, Ferrari C. (2016). Computing Discrete Fine-Grained Representations of Protein Surfaces. In *Computational Intelligence Methods for Bioinformatics and Biostatistics - 12th International Meeting, CIBB 2015, Naples, Italy, September 10-12, 2015, Revised Selected*

Papers, edited by Claudia Angelini, Paola MV Rancoita, and Stefano Rovetta, 9874:180–195. Lecture Notes in Bioinformatics. Cham: Springer International Publishing.

———. (2018). Computing Voxelised Representations of Macromolecular Surfaces: A Parallel Approach. *The International Journal of High Performance Computing Applications* 32 (3): 407–32.

Dias SED, Martins AM, Nguyen QT, Gomes AJP. (2017). GPU-Based Detection of Protein Cavities Using Gaussian Surfaces. *BMC Bioinformatics* 18 (1): 493.

Edelsbrunner H, Facello M, Liang J. (1998). On the Definition and the Construction of Pockets in Macromolecules. *Discrete Applied Mathematics* 88 (1): 83–102.

Edelsbrunner H, Mücke EP. (1994). Three-Dimensional Alpha Shapes. *ACM Transactions on Graphics* 13 (1): 43–72.

Hendlich M, Rippmann F, Barnickel G. (1997). LIGSITE: Automatic and Efficient Detection of Potential Small Molecule-Binding Sites in Proteins. *Journal of Molecular Graphics and Modelling* 15 (6): 359–363.

Huang B. (2009). MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. *OMICS A Journal of Integrative Biology* 13 (4): 325–330.

Huang B, Schroeder M. (2006). LIGSITEcsc: Predicting Ligand Binding Sites Using the Connolly Surface and Degree of Conservation. *BMC Structural Biology* 6 (1): 19.

Jie L, Clare W, Herbert E. (1998). Anatomy of Protein Pockets and Cavities: Measurement of Binding Site Geometry and Implications for Ligand Design. *Protein Science* 7 (9): 1884–1897.

Krivák R, Hoksza D. (2015). Improving Protein-Ligand Binding Site Prediction Accuracy by Classification of Inner Pocket Points Using Local Features. *Journal of Cheminformatics* 7 (1): 12.

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. (1982). A Geometric Approach to Macromolecule-Ligand Interactions. *Journal of Molecular Biology* 161 (2): 269–288.

Laskowski RA. (1995). SURFNET: A Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *Journal of Molecular Graphics* 13 (5): 323–330.

Lee B, Richards FM. (1971). The Interpretation of Protein Structures: Estimation of Static Accessibility. *Journal of Molecular Biology* 55 (3): 379–IN4.

Lee Y, Basith S, Choi S. (2018). Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *Journal of Medicinal Chemistry* 61 (1): 1–46.

Levitt DG, Banaszak LJ. (1992). POCKET: A Computer Graphies Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino Acids. *Journal of Molecular Graphics* 10 (4): 229–234.

Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. (2002). Rate4Site: An Algorithmic Tool for the Identification of Functional Regions in Proteins by Surface Mapping of Evolutionary Determinants within Their Homologues. *Bioinformatics* 18 (suppl_1): S71–S77.

Rodney H, J OA, S GD. (2007). Automated Prediction of Ligand-Binding Sites in Proteins. *Proteins: Structure, Function, and Bioinformatics* 70 (4): 1506–17.

Simões T, Lopes D, Dias S, Fernandes F, Pereira J, Jorge J, Bajaj C, Gomes A. (2017). Geometric Detection Algorithms for Cavities on Protein Surfaces in Molecular Graphics: A Survey. *Computer Graphics Forum* 36 (8): 643–83.

Soga S, Shirai H, Kobori M, Hirayama N. (2007). Use of Amino Acid Composition to Predict Ligand-Binding Sites. *Journal of Chemical Information and Modeling* 47 (2): 400–406.

Weisel M, Proschak E, Schneider G. (2007). PocketPicker: Analysis of Ligand Binding-Sites with Shape Descriptors. *Chemistry Central Journal* 1 (1): 7.

Yu J, Zhou Y, Tanaka I, Yao M. (2010). Roll: A New Algorithm for the Detection of Protein Pockets and Cavities with a Rolling Probe Sphere. *Bioinformatics* 26 (1): 46–52.

Yu W-W, He F, Xi P. (2010). A Rapid 3D Seed-Filling Algorithm Based on Scan Slice. *Computers and Graphics* 34 (4): 449–459.

Zhang Z, Li Y, Lin B, Schroeder M, Huang B. (2011). Identification of Cavities on Protein Surface Using Multiple Computational Approaches for Drug Binding Site Prediction. *Bioinformatics* 27 (15): 2083–2088.