

Real-time Visualization and Exploration of Protein Empty Space with Varying Parameters

Ondřej Strnad, Barbora Kozlíková, Vilém Šustr, and Jiří Sochor

Faculty of Informatics

Masaryk university, Brno, Czech Republic

Email: xstrnad2@fi.muni.cz, xkozlik@fi.muni.cz, xsustr@fi.muni.cz, sochor@fi.muni.cz

Abstract—Long-term research in the area of protein analysis proved the importance of an empty space situated inside these macromolecular structures. This empty space influences the protein function, characteristics or reactivity. Many algorithms enabling computation of these empty spaces (or voids) have been published and their results were evaluated by protein engineers to confirm their chemical relevance. However, not all detected voids inside protein are of the same importance. Thus, the examination and assessment of all voids must follow to reveal the important ones. In this phase the visual representation of voids is very valuable and substantially decreases the time spent in this evaluation phase. In this paper we present an extension of the algorithm for the visualization and further evaluation of protein voids in real-time. The user-driven approach enables to compute and display empty space that satisfies the input parameters instantly. The values of these parameters can be changed by the user anytime and the changes are immediately displayed and prepared for further exploration. Our improvements involve an exclusion of selected atom or group of atoms (ligands, ions) from the computation, which can change the size and shape of the detected void. Another improvement is related to the detection of the binding site which is usually located in one of the largest voids. So the algorithm suggests and visually separates (by different coloring) the largest void of given area. Several improvements were also made in the field of real-time exploration – currently the interaction on large structures is fluent. In consequence, the current version of the algorithm provides the biochemists with very adjustable and precise algorithm for detection of inner voids in a user-defined region of protein structures.

Keywords-protein, empty space, void, visualization, real-time, cavity, volume, Voronoi diagram, Delaunay triangulation

I. INTRODUCTION

This paper is an extended version of the conference paper entitled "Real-Time Visualization of Protein Empty Space with Varying Parameters" presented at the BIOTECHNO 2013: The Fifth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, March 24 – 29, 2013, Lisbon, Portugal [1]. In this extended version, we provide a substantially revised paper describing the algorithm for real-time visualization of protein empty space supplied with functions for its further exploration. These additional features include the visual evaluation of detected voids, automatic detection and highlighting of the largest void (it requires the computation of the volume of voids) or

several improvements of the implementation, which facilitate the fluent user interaction even with large macromolecules.

The biochemical research concentrates on the identification and characterization of structures and processes, which are present in all living cells. Investigation of these structures, their function, causes and effects, are performed on a daily basis in labs. This process is usually very time and resources demanding so the current trend in this field is based on searching for various in-silico predictive software tools, which enable to explore and evaluate the structure prior to in-vitro testing in the laboratory.

Proteins belong to the category of structures that are present in all living organisms so their importance is indisputable. Analysis of their inner arrangement and effects of such arrangement to their function form a basis of many biochemical disciplines, such as protein engineering. In this case, the protein is studied with respect to its function. The order of its amino acids in the polypeptidic chain and its folding determines the characteristics of given protein (see Figure 1).

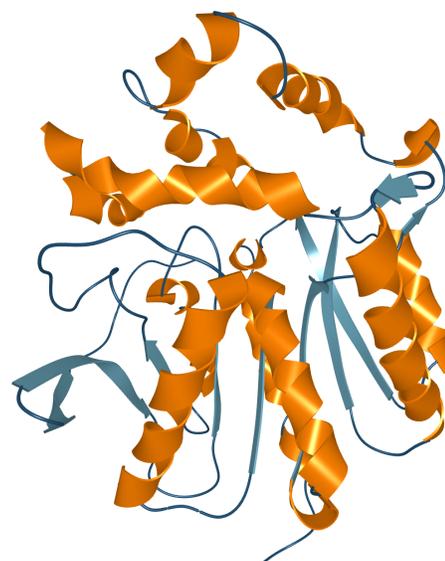


Figure 1. Polypeptidic chain of DhaA haloalkane dehalogenase visualized using the Cartoon method.

Directed mutagenesis of some amino acids causes changes in the protein structure as well as in its behavior. Using this approach, various important characteristics of proteins can be influenced – such as their activity, stability or enantioselectivity [2][3].

The other approach, which is utilized in the drug design, is based on the transportation of small molecule of substrate (ligand, ion or solvent molecule) to the protein binding site. The binding site (also called an active site) is formed by a cavity, which is deeply buried inside the protein structure and it is a place where chemical reaction between protein and substrate undergo. The product of such reaction then forms a basis of some chemical compound, such as new medication.

We advocated that studying of protein inner structure is important in many areas of biochemical research. Moreover, these studies can be performed using software tools, which are also able to visualize the protein structure along with results of the analysis. When analyzing the protein inner structure, biochemists concentrate on an empty space, which is present in the protein. The amount of the empty space determines also the size of substrates entering the active site or possible mutations. Moreover, the binding site is also defined by an empty space marked as a void (or more specific, a cavity).

Many algorithms that compute the empty spaces in proteins (often also marked as voids), have been published and the chemical relevance of computed results was proven by biochemists. However, not each void can play a role of a binding site. So it is necessary to enable the user also visual exploration of detected voids to better assess their importance. To fulfill these requirements, we developed a novel approach to the computation, visualization and further evaluation of voids in protein. The low complexity of the algorithm enables the user to perform all these steps in real-time, which substantially decreases time demands on the evaluation phase. Our user-driven approach enables to compute and display empty space that satisfies the input parameters, which influence the minimal width of voids. The values of these parameters can be changed anytime during the exploration and changes are immediately projected.

A. Empty space classification

The empty space present in protein structures can be further qualified according to various criteria and marked as a cavity, pocket, tunnel, channel, pore or other specific structure (see Figure 2). As mentioned above, inner cavities can play a role of binding sites and thus serve as the destination for a small ligand molecule that can follow a pathway from the outside environment of the protein. In such cavity the chemical reaction between protein and ligand can take place. Pockets are defined as a (concave, cleft, hole)-shaped region on the protein surface [4]. They are also of high importance because they can also serve as potential

ligand binding sites. Thanks to their large binding surface area they are more easily accessible than the deeply buried binding sites.

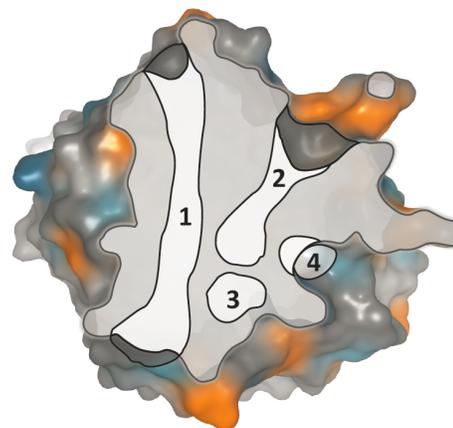


Figure 2. Illustration of channel (1), tunnel (2), cavity (3) and pocket (4).

Channels and pores are specific pathways crossing the whole protein. Channels can be used for the transport of substrates, products, water molecules and other compounds through the protein. Pores are present in transmembrane channel proteins and are essential in mediating the transport of ions and molecules through biological membranes [5].

Our long-term research in this field was concerned mainly with the detection of tunnels. These structures represent a path leading from a specific protein cavity (the binding or active site) to the molecular surface. Prior to tunnel detection it is crucial to determine the starting point for their computation – the active site. Thus it is necessary to analyze and evaluate protein cavities and assess that one containing the active site. Small substrate molecules entering the active site determine the limitation of properties of computed tunnels (e.g., their width or curvature).

The derivation of an empty space from the 3D structure of a protein's amino-acid sequence introduces a very complex task. Proteins with known three-dimensional structure are stored in the PDB database [6]. This archive contains information about experimentally-determined structures of proteins, nucleic acids, and complex assemblies. Some of the structures involved into the PDB database were analyzed and their binding sites were detected. These sites are stored in the CSA (Catalytic Site Atlas) database [7] or in the UniProt database [8]. However, active sites of most of the structures still have not been revealed or published. This situation creates the necessity of using other semi-automated tools or even manual detection of the active site.

B. Cavities

Our novel approach to the detection of cavities is able to identify all cavities in a given region (its size is defined by the *distance* parameter, which determines the radius of

a sphere of interest with user-defined center) or in the whole protein structure. First of all, all cavities inside the structure or region of interest are detected. However, when operating with large protein complexes or even ribosomes, computation of all cavities in such molecule is very time and memory consuming. Thus we introduce a novel method for detection and visualization of inner cavities focusing on minimizing the memory and time requirements. This technique is designed to operate in real-time, enabling users to interactively change the inner and outer size of a spherical probe utilized for detection of cavities.

Detection of cavities should be followed by their proper visualization. Otherwise, it is very complicated for biochemists to evaluate the resulting cavities. To be able to determine the cavity containing the active site properly, biochemists should be equipped with a powerful visualization tool enabling not only displaying of detected voids but also allowing real-time alternations of parameters of cavities. When combining this method with other chemical properties of given protein (such as the knowledge of partial charges of atoms), biochemists are able to recognize the ligand binding site easily. To be more specific, when the atoms surrounding the cavity have neutral or small partial charges, this cavity probably will not be marked as an active site.

A brief outline of the paper follows. Section II presents existing algorithms for detection and classification of empty spaces in protein structures, and common approaches to computation and visualization of protein surfaces. Section III gives a detailed description of the proposed real-time algorithm for visualization of protein voids. The next section, IV, discusses several possibilities of visual evaluation of detected voids. Section V concludes the paper with computational results and their analysis. Last section, VI, introduces some ideas for the future improvements of the proposed algorithm.

II. RELATED WORK

Detection and classification of the empty space inside proteins has been in the scope of biochemists for the last decades. Many algorithms have been proposed and published in this field. Although the aim of this article is to, above all, present a novel approach to detection and visualization of inner cavities, in this section we will introduce a related research focused on the detection of empty space in general. Techniques mentioned in this section can be mostly adapted to the computation of more specific structures, which were mentioned (tunnels, channels, etc.).

Algorithms detecting an empty space inside proteins are based on the similar principle – they are all based on computational geometry using the three-dimensional protein structure (positions and radii of atoms) as the input. These algorithms can be divided into two groups according to their approach to space representation:

- algorithms based on a grid approach

- algorithms utilizing Voronoi diagram and Delaunay triangulation

The main difference between these two approaches lies in their precision, speed and memory consumption. More detailed description of these approaches along with their representatives follow.

A. Detection of empty space

1) *Grid method*: In this approach, the entire protein is enclosed in an axis aligned bounding box, which is subsequently regularly sampled to a voxel grid. Each vertex of the voxel grid is classified according to its collision with an atom. Non-colliding voxels form the empty space used for construction of cavities, tunnels and other structures. The quality of results is strongly influenced by the sampling density. Too sparse sampling can lead to a situation where all vertices of voxels are colliding with an atom and no empty space is detected. On the other hand, too dense a sampling causes an enormous increase in time and memory demands. The main advantage of this approach is its simplicity; the disadvantage, as already mentioned, comes from its computational complexity $\mathcal{O}(n^3)$, with n depending on the sampling density.

The grid approach was adopted for tunnel computation in CAVER 1.0 [9]. Another tool using the grid approach for computation of specific cavities (pores) inside proteins is called CHUNNEL [10]. Each voxel is marked according to its distance to the nearest atom. On this structure, the Dijkstra algorithm is launched and the tunnel with highest voxel values (the widest tunnel) is detected.

Kleywegt et al. [11] presented their grid approach applied to the detection of cavities. Their implementation is presented in the VOIDOO application. The first step of the algorithm maps the protein onto a 3D grid with a spacing between 0.5 and 1.0 Ångströms. Each point of the grid is noted by the zero value. Then, each grid point is processed and when the distance to the nearest atom is less than the sum of the atom radius and the probe radius, its value is set to one. This method is also known as the flood-fill algorithm. Finally, points inside cavities still have a zero value so they can be easily detected and their volume can be measured.

2) *Voronoi diagram and Delaunay triangulation*: Another approach to protein 3D space inspection is based on the Voronoi diagram (VD) and its dual structure – the Delaunay triangulation (DT). The main benefit of this approach is based on the division of space, which is independent on any user-defined settings. That is the main reason why this method overcomes the main disadvantage of the previous grid approach. The detailed description of VD construction can be found, e.g., in [12]. The dual structure to VD, the Delaunay triangulation (tetrahedrization in the three-dimensional space), can be constructed by connecting neighboring points sharing the Voronoi edge (see Figure 6). Tetrahedra of the Delaunay tetrahedrization fulfills the

condition that no point is presented inside the circumsphere of any tetrahedron.

Voronoi diagrams and Delaunay tetrahedra were utilized by various software tools for tunnel and channel computation [13] (see Figure 3), such as CAVER 2.0 [14], MolAxis [15] or MOLE [16].

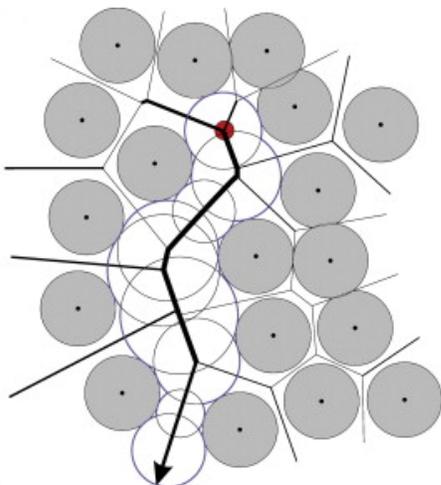


Figure 3. Tunnel detected using the Voronoi diagram (taken from [13]).

Another approach to cavity detection using the Delaunay triangulation and the alpha complex was implemented in the CAST application [17] (CASTp is its online version). This tool is able to measure the area and volume of cavities analytically as well.

In [18], Voronoi diagrams were extended to the Additively weighted Voronoi diagrams (AVD). AVDs were originally designed for environments containing non-uniform objects. This is also the case of protein structures because atoms of various chemical elements vary also in the size of spheres representing them. Compared to traditional VDs, AVDs gain the more adequate space subdivision through the specification of weight w attached to each site point. According to their weight values the respective points attract ($w > 1$) or repel ($w < 1$) the corresponding Voronoi edges. Resulting Voronoi edges have curvilinear shapes (see Figure 4). AVD construction is more complex in comparison to traditional VD and thus the time and space complexity increases substantially. AVD were used in the protein visualization tool called Voroprot [19].

The second part of the related work is connected with the protein outer environment and its differentiation from the inner part of the protein structure. The outer environment can be considered a void as well. To distinguish the inner environment from the outer one, the protein has to be encapsulated in some bounding object – a protein surface. Detection of voids inside protein structures is highly influenced by the protein surface, which gives an overview of the protein's compactness. It makes sense to detect empty

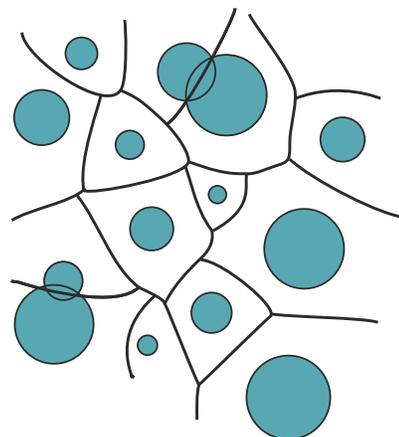


Figure 4. Additively weighted Voronoi diagram on a set of atoms.

space only within the volume that is defined by the protein surface. Therefore, in the next section, we concentrate on existing approaches to detection of protein surfaces.

B. Protein surface and its detection

Computation and visualization of surfaces play an important role not only in the case of detection of voids. The resulting shape of the surface also gives the biochemists a fair overview of protein constitution, presence of pockets and clefts etc (see Figure 5). Generating a smooth molecular surface is also important for a number of applications, including molecular recognition, drug design, electrostatics, molecular graphics, etc. Thus, protein surface detection has been in researchers' scope for decades and many approaches have been proposed. Two main groups of existing algorithms employ either analytical or numerical approach.

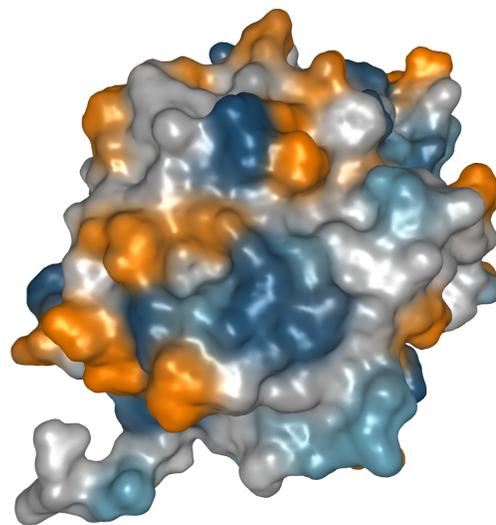


Figure 5. Visualization of molecular surface, which enables to observe the overall shape of the structure.

1) *Analytical surface construction:* The input set contains objects (atoms) that should be encapsulated by the surface. The analytical approach describes the surface using a set of mathematical equations. For protein exploration there are two basic analytical approaches to generation of surfaces. The Reduced surface [20] is constructed by rolling a probe of specific radius over the protein starting in the outer environment. Inwards facing parts of the probe surface combined with parts of atoms' surfaces on the boundary create the resulting solvent-accessible surface. The second approach is based on the alpha-shapes theory [21]. Totrov and Abagyan [22] introduced the contour-buildup algorithm for analytic calculation of molecular surface. The core part of this algorithm is based on the sequential build up of multi-arc contours on the van der Waals spheres representing individual atoms.

The main disadvantage of the analytical representation of the surface comes from its complexity. Thus its utilization on large datasets (e.g., macromolecular structures) cannot be performed in real-time or can even fail.

2) *Numerical surface construction:* The accuracy of numerically based algorithms is strongly dependent on initial user settings. The basic principle is the division of the scrutinized space into a uniform voxel grid. Each voxel is classified according to its intersection with objects in space. Subsequently, the marching cubes algorithm [23] can be utilized for visualization of the surface. The marching cubes method was designed primarily for a simple and fast construction of isosurfaces in volume data sets. This approach is widely used, e.g., in MRI or other medical applications. One of the representatives of this approach is the LSMS algorithm [24]. This algorithm is able to compute not only the protein surface – it can detect the interior cavities as well.

III. REAL-TIME VISUALIZATION OF PROTEIN VOIDS

In comparison with existing algorithms for visualization of empty voids, our novel approach does not require any additional time for their recomputation when the input parameters change. The empty space corresponding to these changes is visualized instantly. The original version of our approach was able to perform the real-time update of voids on small and middle size molecular systems. But when dealing with large macromolecules (e.g., ribosomes) we had to introduce several improvements. Current version of our algorithm handles also cases operating on macromolecules.

In the rest of this section, the basic principle of our algorithm along with current improvements will be described. The main aim of our approach is the real-time visualization of inner voids. Naturally, these voids must be computed in the first phase. For this task we utilize the standard Voronoi diagram, which omits the differences between radii of atoms (contrary to AVD approach) since our priority is the speed of the algorithm. Of course, VD does not provide users with

as precise results as AVD does, but from our experience the difference is acceptable for purpose of cavities detection. The main difference between the results obtained by VD and AVD is in the exact representation of the surface of voids. However, the set of detected cavities is equal.

In the visualization phase, the algorithm is divided into five basic steps. Steps 1 and 2 represent preprocessing and they are performed only once during the initialization phase. Steps 3 to 5 (represented by subsections C to E) are iteratively repeated for any change of input parameters and are considered as one of the main contributions of this paper.

Algorithm Real-Time Visualization of Protein Voids

Input: set of atoms
 1. compute Delaunay triangulation
 2. convert it to a graph
 3. **while** user is changing parameters **do**
 4. determine center point of the bounding box
 5. select empty space inside the protein
 6. visualize selected empty space
 7. **end while**

A. Construction of Delaunay triangulation

input: set of atoms A

output: Delaunay triangulation T

The input set A consists of all atoms of the protein. Since we do not take into account the difference between atom radii, the atomic centers were selected as representatives of atoms. These atomic centers then form the input set of points marked as P , which is subsequently processed. For the set P , the Delaunay triangulation T is constructed using the QuickHull 4D algorithm described in [25].

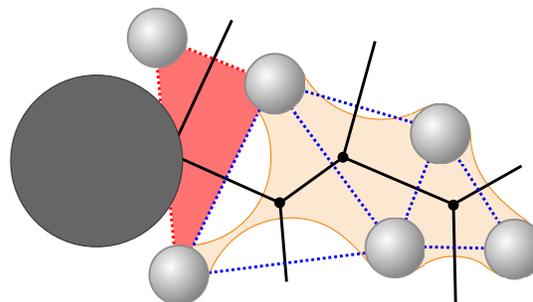


Figure 6. Tetrahedron (red) accessible by a probe (dark gray) is removed from the triangulation (blue dotted). The molecular surface defined by the probe is highlighted (orange).

The triangulation T is afterwards refined so that all tetrahedra intersecting the molecular surface of the protein are removed. In other words, all surface tetrahedra that are accessible from the outside by a probe with radius 2.8\AA (double the van der Waals radius of oxygen) are removed from T (see Figure 6). This ensures that the tunnel throat will not contain any excessive boundary spheres.

B. Construction of the graph G

input: Delaunay triangulation T

output: evaluated graph G

For each tetrahedron $t_i \in T$ a node N_i is inserted into a newly constructed graph G . An edge e_{jk} connecting nodes N_j and N_k is added into G if their referenced tetrahedra t_j and t_k share a face f_{jk} . For every edge $e_{jk} \in G$, we define its center point $c(e_{jk})$ and width $w(e_{jk})$ as follows. The center point $c(e_{jk})$ is defined as a point in f_{jk} where sphere with maximal possible radius not intersecting any atom from t_j or t_k can be placed. The width $w(e_{jk})$ is then defined by the radius of such a sphere. The evaluation process is illustrated in Figure 7.

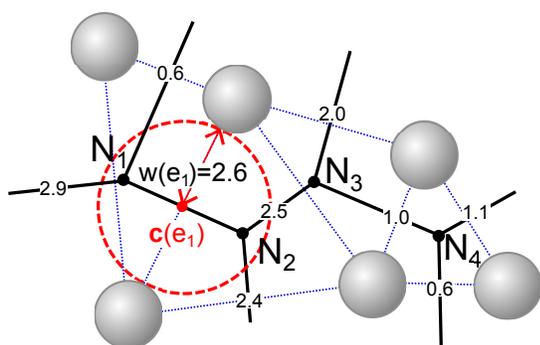


Figure 7. Illustration of a part of a graph G . Thick lines represent Voronoi edges. Every edge is evaluated by the value representing its distance to the nearest atom.

C. Selection of center point

input: Delaunay triangulation T

output: center point C

The algorithm was designed to operate with large macromolecules. In this case, computation and visualization of all inner voids usually leads to complex and ambiguous results, which the biochemist cannot properly explore, thanks to the huge amount of visualized data. In order to avoid this situation, we allow computing inner voids from a starting point C that represents the center of the bounding sphere. The empty space is then visualized only inside this bounding sphere, which represents the area of interest. The point C set by the user can be determined in two ways. The user can enter the space coordinates of the point directly or, in most cases, the binding site loaded from the CSA (Catalytic Site Atlas) database [7] can define the point C . Once the center point C is set, it can be stored for further iterations of the algorithm.

D. Selection of relevant edges

input: graph G , point C , distance d , parameter w_{min}

output: set of filtered edges E

In this phase, the iteration process is started. The goal is to select a set E of edges from G , which satisfy the condition of thickness (driven by the parameter w_{min} representing the minimal width of the edge) and proximity (parameter d defining the bounding sphere radius). For remark, every edge e_{jk} connecting two nodes N_j and N_k is evaluated by a width $w(e_{jk})$.

The set of filtered edges E consists of all edges having the $w(e_{jk})$ greater or equal to w_{min} and with the distance to C lower than d . More formally, let G_E is the set of all edges from G . The set of filtered edges is then $E = \{e_{jk} \in G_E | dist(C, c(e_{jk})) < d \wedge w_{min} \leq w(e_{jk})\}$.

E. Visualization

input: set of edges E , selected visualization method(s)

Firstly, the set E has to be transformed into geometrical objects, which are possible to render. Every edge e_{jk} is transformed into a sphere s_{jk} with center in $c(e_{jk})$ and with radius equal to $w(e_{jk})$. The set S of all such spheres is then prepared as an input for selected visualization method(s).

For our case of protein visualization, we utilized two basic methods effectively describing the empty space inside macromolecules.

1) *Rendering of spheres:* This option represents the most intuitive visualization method as well as the fastest one (see Figure 8). It displays all spheres of the set S . From the construction introduced above, all spheres fill the empty space inside the molecule and do not intersect with any atom. Using this method, the empty space is highlighted, but it looks distracting and for user it can sometimes be difficult to distinguish between an atom of the molecule and a sphere highlighting the empty space. Thus also another approach was introduced.

2) *Grid sampling:* This technique enables users to visualize a continuous surface of discrete voids, which gives more intuitive and user friendly results. To construct such surface, a grid based approach is performed. All spheres from the set S are enclosed into an axis aligned bounding-box. Afterwards, this bounding-box is regularly sampled with a user defined *density*. It is obvious that the higher density leads to more precise surface. On the other hand, the number of samples directly influences the memory and time complexity of the computation. We found out that for exploring of local neighborhood the empirically obtained *density* = 200 (i.e., grid 200x200x200) is optimal. Subsequently, each vertex of each cell in the grid is evaluated according to its intersection with any sphere from S . When all vertices are processed, the fully evaluated grid serves as the input for the marching cubes algorithm. For a notice, this algorithm operates with a predefined set of configurations, thus it is very straightforward and fast when constructing the resulting surface (see Figure 8).

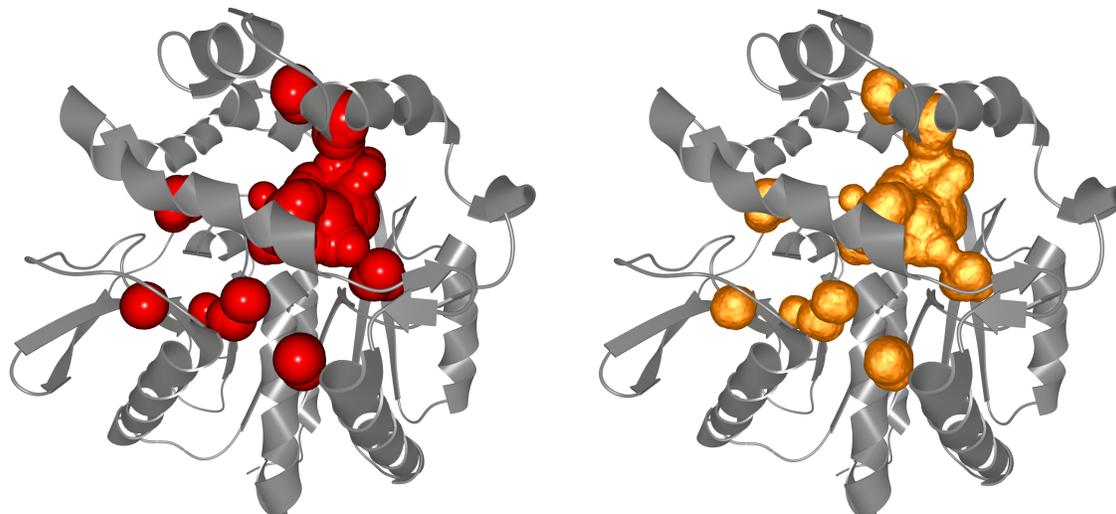


Figure 8. DhaA haloalkane dehalogenase with computed cavities (parameter settings: distance = 13.4 Å, minimal width = 1.5 Å). Empty space visualized as a set of spheres (left) or a surface (right).

IV. VISUAL EVALUATION OF DETECTED VOIDS

The main aim of the visualization methods for displaying computed voids is the subsequent evaluation of these voids. Without proper visualization it is very complicated to observe the size and shape of individual voids along with their surrounding amino acids. For better exploration of the cavity inner environment, we designed a clip plane with an arbitrary orientation – see Figure 9. It displays the largest cavity of the DhaA haloalkane dehalogenase, which corresponds to the binding site of this structure (it is specified by the cross with axes).

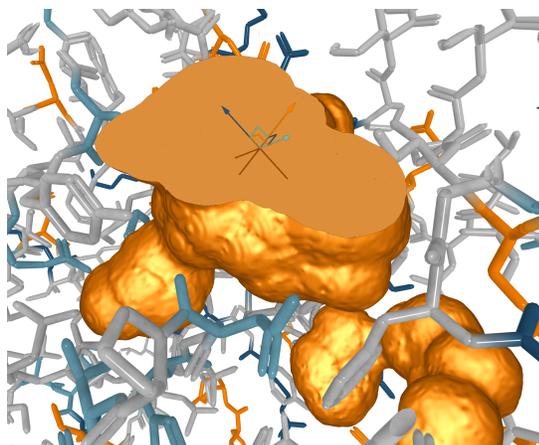


Figure 9. Largest cavity of DhaA haloalkane dehalogenase containing the active site is cut by a clip plane to explore the inner environment. The clip plane can be shifted in both directions in order to go through the whole cavity.

The fast recomputation and real-time observation of the impact of parameter changes also helps to explore voids

with respect to changes of the input set. This means that users are able to include or exclude any atom from the computation and the results are instantly displayed to the user. This situation is illustrated in Figure 10. Molecules of water and ions of iodine, which are originally present in the PDB file of given structure are excluded from the computation of voids.

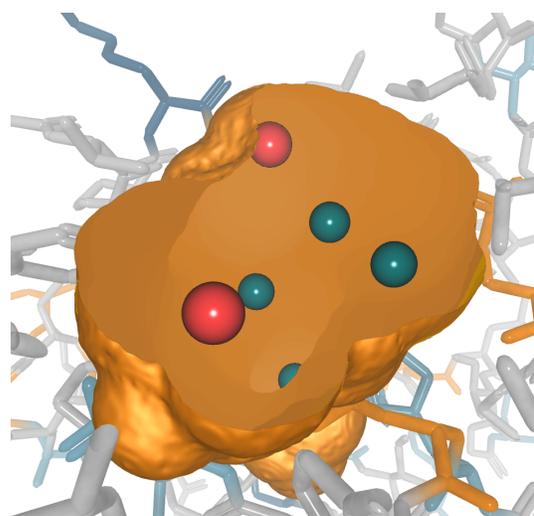


Figure 10. Cavity containing molecules of water (blue spheres) and red spheres representing ions of iodine.

Figure 11 illustrates the high importance of atoms included into the computation. The left part of the figure shows the largest cavity of the DhaA haloalkane dehalogenase. In this case only the amino acids forming the polypeptidic chain were taken into account. Blue spheres in the neighborhood of the cavity represent water molecules. The middle

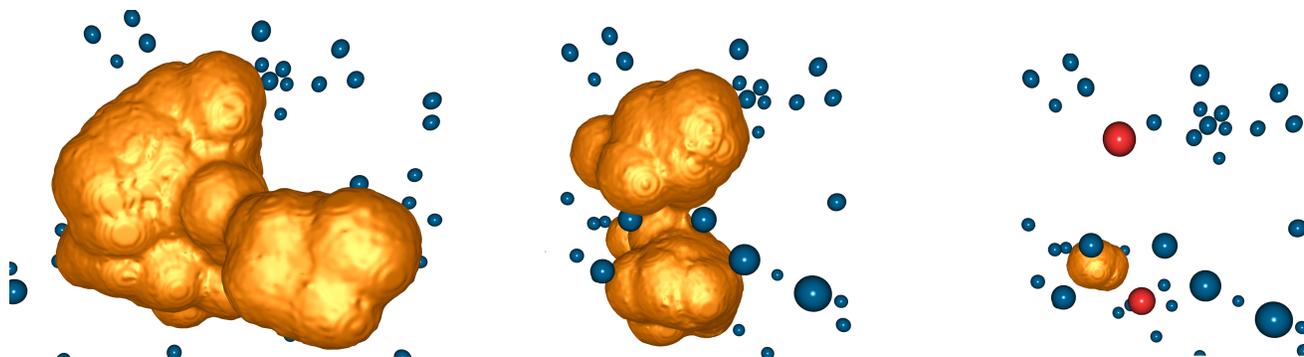


Figure 11. Largest cavity of DhaA haloalkane dehalogenase with different computation settings. Left - only protein amino acids are taken into account during computation. Blue spheres represent water molecules present in the PDB file of the structure. Middle - amino acids and water molecules (blue) present in the structure are considered. This causes reduction of the cavity. Right - amino acids, waters and ions of iodine (red) present in the structure are involved into computation. The cavity is even more contracted.

part displays the same cavity when the surrounding waters are considered during computation. Finally, in the right part the cavity almost disappeared because water molecules and also ions of iodine were taken into account.

The improved version of our algorithm is also able to compute the volume of detected cavities and according to this metric to determine and mark the largest cavity. Figure 12 shows the result of such evaluation when the largest cavity is highlighted using red color.

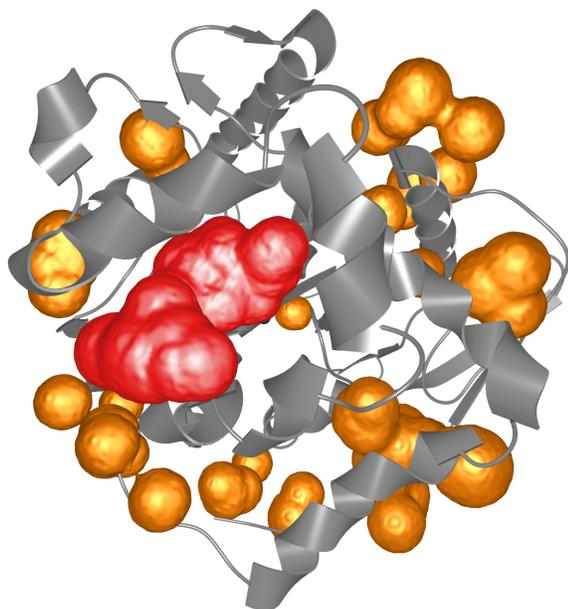


Figure 12. All cavities of the DhaA haloalkane dehalogenase with minimal width 1.9 Å. The largest cavity is detected and marked with red color.

V. RESULTS AND DISCUSSION

In this paper, we presented a novel method for real-time visualization of empty space inside macromolecules, which

concentrates on user-driven evaluation of computed voids. The method is not limited by the size of the molecule (the number of atoms) as the encapsulation of displayed voids into a bounding sphere allows to restrict the amount of processed data respectively. The current version is able to process also the whole macromolecule in real-time. Thanks to introducing adaptive refinement of the surface we are able to react on changing of parameters instantly because of visualization of the rough surface (with the low density of the sampling grid). When parameters are set and not changing, the algorithm immediately starts to calculate more precise surface. This calculation is performed in three levels of detail. So when the parameters remain the same for a few seconds, the smooth surface is generated.

The implementation has not any special hardware or software requirements, the algorithm was implemented in 32-bit Java environment. The performance was tested on a common single-threaded 2.66GHz computer. Both rendering strategies, as well as various types of macromolecules (ranging from proteins to ribosomes) underwent this test. Examples of tested protein structures and combinations of parameters are summarized in table in Figure 13. Results correspond to the first level of detail surface – the rough one.

To illustrate the robustness of our algorithm we selected four representatives of molecules containing common to large number of atoms.

To illustrate the importance of input parameters, Figure 14 shows how these parameters influence the resulting voids. Left part of the figure shows the protein structure where cavities in the area of interest defined by a sphere with radius 13.4 Å, where the minimal width (defined by the size of the inscribed sphere) is set to 1.14 Å. Right part of the figure shows a cavity detected in the same area of interest but the minimal width changed to 1.6 Å.

The relevance of computed cavities was tested on many protein structures with well known inner arrangement. To

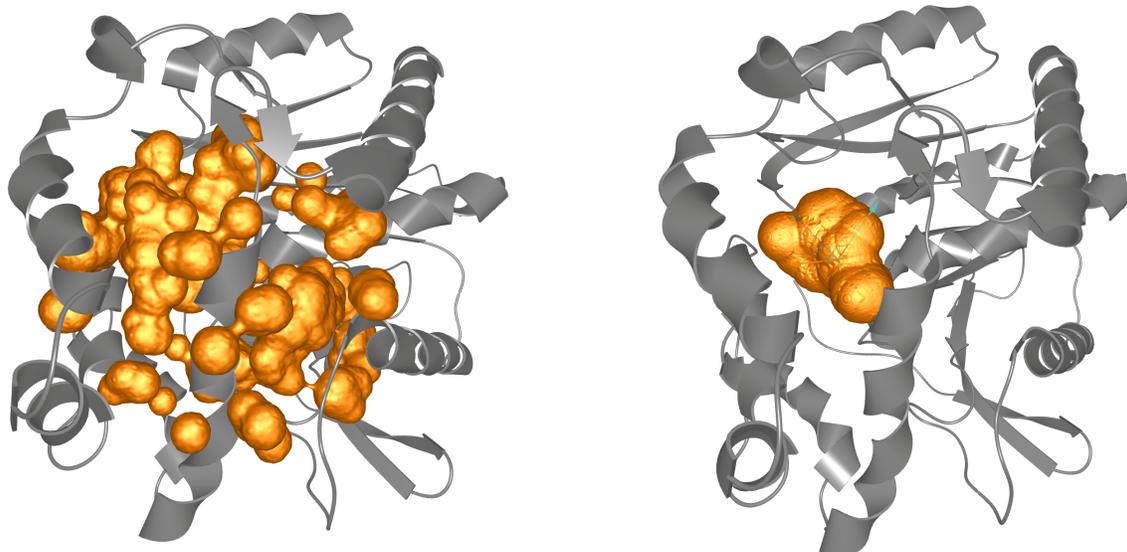


Figure 14. DhaA haloalkane dehalogenase with computed cavities. Left: All cavities detected in the region defined by a sphere with radius 13.4 Å, minimal width (size of the inscribed sphere) is set to 1.14 Å. Right: The largest cavity of minimal width 1.6 Å. This cavity contains the active site - marked with the cross with three axes.

Structure	d (Å)	$w(e)_{min}$ (Å)	spheres (FPS)	surface (FPS)
1cq w 2754 atoms	15	1.4	> 100	78
	<i>max</i>	1.4	> 100	62
1cq z 8218 atoms	15	1.4	> 100	77
	<i>max</i>	1.4	> 100	58
2oau 13580 atoms	15	1.4	> 100	36
	<i>max</i>	1.4	88	19
1aon 58874 atoms	15	1.4	73	16
	<i>max</i>	1.4	28	2

Figure 13. Performance of the algorithm measured in frames per second. Parameter d represents radius of the bounding sphere, $w(e)_{min}$ shows the minimal width of an edge.

evidence the relevance of empty space detected and visualized using our approach, we performed a comparison with results obtained by the well acknowledged CAVER algorithm. CAVER was designed for the detection of tunnels inside proteins and results were thoroughly tested by the community of protein engineers [26]. Thus, to manifest the relevance of voids detected by our new approach, the computed voids must contain all detected structures such as tunnels, cavities etc. We verified that tunnels detected by the CAVER algorithm lead through the empty space highlighted by our method (see Figure 15).

Figure 16 shows more results of our algorithm – cavities computed on structures with PDB IS's 2OAU and 1M4X.

VI. FUTURE

Further extension of our implementation should lead to the parallelization of the marching cubes algorithm on the

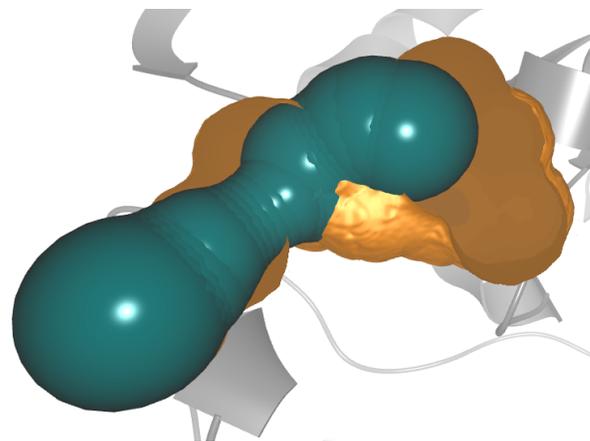


Figure 15. Cavity containing the binding site along with the computed tunnel.

modern graphic cards [27]. Such implementation would substantially increase the performance of the rendering phase. Moreover, to obtain even more precise results, the additively weighted Voronoi diagrams should be utilized.

Another step towards the better exploration of empty voids in proteins can involve implementing more variants of shapes of area of interest. Users will not be limited only by a bounding sphere but could create e.g., slices of user-defined width. More importantly, they will be able to explore the local neighborhood of various inner structures, such as tunnels, channels or pores.

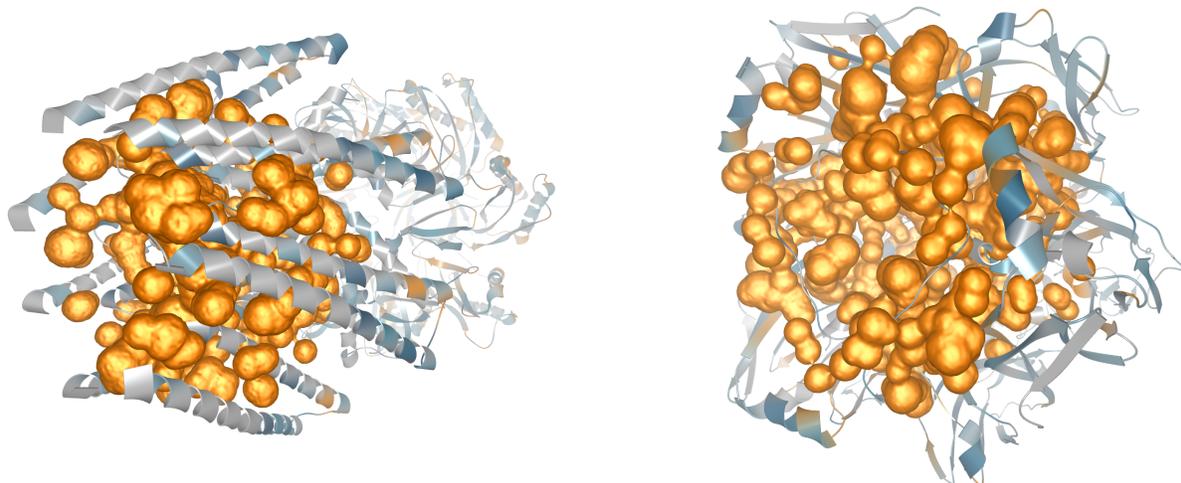


Figure 16. Resulting cavities computed on different structures. Left: Protein structure with PDB ID 20AU. Right: Structure marked with PDB ID 1M4X.

ACKNOWLEDGMENT

This work was supported by The Grant Agency of the Czech Republic, Contract No. P202/10/1435.

REFERENCES

- [1] O.Strnad, V. Šustr, B. Kozlíková, and J. Sochor, "Real-time visualization of protein empty space with varying parameters," *Proceedings of BIOTECHNO, IARIA XPS Press*, pp. 65–70, March 2013.
- [2] M. Pavlova, M. Klvana, Z. Prokop, R. Chaloupkova, P. Banas, M. Otyepka *et al.*, "Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate," *Nature Chemical Biology*, vol. 5, pp. 727–733, 2009.
- [3] T. Koudelakova, R. Chaloupkova, J. Brezovsky, Z. Prokop, E. Sebestova, M. Hesseler *et al.*, "Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel," *Angewandte Chemie International Edition*, vol. 52, pp. 1959–1963, 2013.
- [4] T. Kawabata and N. Go, "Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites," *PROTEINS: Structure, Function, and Bioinformatics*, vol. 68, pp. 516–529, 2007.
- [5] M. Pellegrini-Calace, T. Maiwald, and J. M. Thornton, "Pore-walker: A novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure," *PLoS Computational Biology*, vol. 5, no. 7, 2009.
- [6] F. C. Bernstein, T. F. Koetzle, G. J. Williams, E. E. M. Jr., M. D. Brice, J. R. Rodgers *et al.*, "The protein data bank: A computer-based archival file for macromolecular structures," *Journal of Molecular Biology*, vol. 112, no. 535, 1977.
- [7] C. T. Porter, G. J. Bartlett, and J. M. Thornton, "The catalytic site atlas: a resource of catalytic sites and residues identified in enzymes using structural data," *Nucleic Acids Research*, vol. 32, no. Database-Issue, pp. 129–133, 2004.
- [8] The UniProt Consortium, "Reorganizing the protein space at the universal protein resource (uniprot)," *Nucleic Acids Research*, vol. doi: 10.1093/nar/gkr981, 2011.
- [9] M. Petřek, M. Otyepka, P. Banáš, P. Košinová, J. Koča, and J. Damborský, "Caver: A new tool to explore routes from protein clefts, pockets and cavities," *BMC Bioinformatics*, vol. 7, p. 316, 2006.
- [10] R. G. Coleman and K. A. Sharp, "Finding and characterizing tunnels in macromolecules with application to ion channels and pores," *Biophysical Journal*, vol. 96, no. 2, pp. 632 – 645, 2009.
- [11] G. J. Kleywegt and T. A. Jones, "Detection, delineation, measurement and display of cavities in macromolecular structures," *Acta Crystallographica Section D*, vol. 50, no. 2, pp. 178–185, Mar 1994. [Online]. Available: <http://dx.doi.org/10.1107/S0907444993011333>
- [12] F. Aurenhammer, "Voronoi diagrams a survey of a fundamental geometric data structure," *ACM Comput. Surv.*, vol. 23, no. 3, pp. 345–405, sep 1991. [Online]. Available: <http://doi.acm.org/10.1145/116873.116880>
- [13] J. Brezovsky, E. Chovancova, A. Gora, A. Pavelka, L. Biedermannova, and J. Damborsky, "Software tools for identification, visualization and analysis of protein tunnels and channels," *Biotechnology Advances*, vol. 31, pp. 38–49, 2013.
- [14] P. Medek, P. Beneš, and J. Sochor, "Computation of tunnels in protein molecules using delaunay triangulation," *Journal of WSCG*, vol. 15(1-3), pp. 107–114, 2007.
- [15] E. Yaffe, D. Fishelovitch, H. J. Wolfson, D. Halperin, and R. Nussinov, "MolAxis: efficient and accurate identification of channels in macromolecules." *Proteins*, vol. 73, no. 1, pp. 72–86, oct 2008. [Online]. Available: <http://dx.doi.org/10.1002/prot.22052>
- [16] M. Petřek, P. Košinová, J. Koča, and M. Otyepka, "Mole: a voronoi diagram-based explorer of molecular channels, pores, and tunnels," *Structure*, vol. 15, no. 11, pp. 1357 – 1363, 2007.

- [17] J. Liang, H. Edelsbrunner, and C. Woodward, "Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design," *Protein science : a publication of the Protein Society*, vol. 7, pp. 1884–1897, sep 1998.
- [18] H. Edelsbrunner, *Algorithms in combinatorial geometry*. New York, NY, USA: Springer-Verlag New York, Inc., 1987.
- [19] K. Olechnovič, M. Margelevičius, and C. Venclovas, "Voroprot," *Bioinformatics*, vol. 27, no. 5, pp. 723–724, mar 2011. [Online]. Available: <http://dx.doi.org/10.1093/bioinformatics/btq720>
- [20] M. F. Sanner, A. J. Olson, and J. C. Spehner, "Reduced surface: an efficient way to compute molecular surfaces," *Biopolymers*, vol. 38, pp. 305–320, Mar 1996.
- [21] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on Information Theory*, vol. 29, pp. 551–559, jul 1983.
- [22] M. Totrov and R. Abagyan, "The contour-buildup algorithm to calculate the analytical molecular surface," *J Struct Biol*, no. 1, pp. 138–43, 1995.
- [23] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *SIGGRAPH Comput. Graph.*, vol. 21, no. 4, pp. 163–169, Aug. 1987. [Online]. Available: <http://doi.acm.org/10.1145/37402.37422>
- [24] T. Can, C. I. Chen, and Y. F. Wang, "Efficient molecular surface generation using level-set methods," *Journal of Molecular Graphics & Modelling*, vol. 25, pp. 442–454, 2006.
- [25] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, dec 1996. [Online]. Available: <http://doi.acm.org/10.1145/235815.235821>
- [26] E. Chovancová, A. Pavelka, P. Beneš, O. Strnad, J. Brezovský, B. Kozlíková *et al.*, "Caver 3.0: A tool for the analysis of transport pathways in dynamic protein structures," *PLoS Comput Biol*, vol. 8, no. 10, p. e1002708, 2012. [Online]. Available: <http://dx.doi.org/10.1371/journal.pcbi.1002708>
- [27] C. Dyken, G. Ziegler, C. Theobalt, and H. P. Seidel, "High-speed marching cubes using histopyramids," *Computer Graphics Forum*, vol. 27, no. 8, pp. 2028–2039, 2008. [Online]. Available: <http://dx.doi.org/10.1111/j.1467-8659.2008.01182.x>