

Molecular Sombreros: Abstract Visualization of Binding Sites within Proteins

K. Schatz¹ and M. Krone² and T. L. Bauer³ and V. Ferrario³ and J. Pleiss³ and T. Ertl¹

¹Visualization Research Center (VISUS), University of Stuttgart, Stuttgart, Germany

²Big Data Visual Analytics in Life Sciences Group, University of Tübingen, Tübingen, Germany

³Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

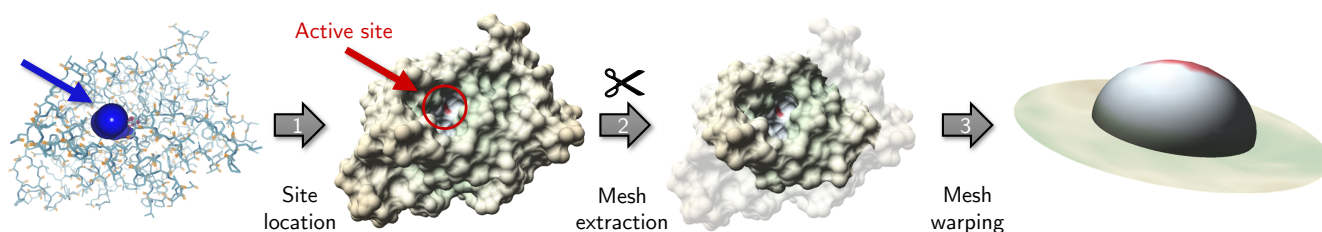


Figure 1: Overview of the three major steps of our abstract visualization for molecular binding sites. First, the location of the active site (marked in red) and a path through a tunnel towards it (blue spheres) are extracted. Secondly, the tunnel containing the active site is cut out of the molecular surface mesh. As third and final step the resulting geometry is then deformed into a hat-like shape. The molecular surface coloring was inspired by topographic maps, where blue denotes low elevation and brown high (i.e., distance to the protein center). Leftmost image made with CAVER Analyst [Kv^v*14].

Abstract

We present a novel abstract visualization for the binding sites of proteins. Binding sites play an essential role in enzymatic reactions and are, thus, often investigated in structural biology. They are typically located within cavities. The shape and properties of the cavity influence whether and how easily a substrate can reach the active site where the reaction is triggered. Molecular surface visualizations can help to analyze the accessibility of binding sites, but are typically prone to visual clutter. Our novel abstract visualization shows the cavity containing the binding site as well as the surface region directly surrounding the cavity entrance in a simplified manner. The resulting visualization resembles a hat, where the brim depicts the surrounding surface region and the crown the cavity. Hence, we dubbed our abstraction Molecular Sombrero, using the Spanish term for 'hat'. Our abstraction is less cluttered than traditional molecular surface visualizations. It highlights important parameters, like cavity diameter, by mapping them to the shape of the sombrero. The visual abstraction also facilitates an easy side-by-side comparison of different data sets. We show the applicability of our Molecular Sombreros to different real-world use cases.

CCS Concepts

• **Human-centered computing** → Scientific visualization; • **Applied computing** → Molecular structural biology;

1. Introduction

Molecular surfaces are a widely used representation for the visual analysis of biomolecular structures, especially to gain knowledge about the geometric properties of proteins and to investigate the interface of a protein with respect to other molecules. Especially relevant are the so-called *binding sites* of a protein, as these are one of the key contributors to the function of a protein while it interacts with its surroundings. Enzymes, for example, contain binding

sites that are able to trigger and accelerate chemical reactions of other, smaller molecules in the vicinity. These small molecules are called ligands, or, more specifically in the case of an enzymatic binding site, substrate molecules. Despite their importance, visualizations that concentrate only on binding sites are still not very common. By comparing the shape of binding sites, the biochemical properties of enzymes can be predicted [PFS98]. However, due to the complex structure of proteins, a direct visual comparison of the shape of different binding sites can be challenging and time-

consuming. A tool for a systematic analysis of binding sites that simplifies such comparisons would, therefore, have potential application in the selection of optimal enzymes for specific tasks and the design of improved enzyme variants. In order to develop a comparable representation for enzymatic binding sites, it is important to note that most of the binding sites are located inside tunnels of the protein [PBV*14]. Due to occlusion and other hindering effects it would, therefore, not be sufficient to only show the outer surface of a protein to enable the analysis of binding sites.

We developed a novel approach to systematically extract and visualize binding sites within proteins. More specifically, our method targets enzymes, which are not only involved in almost all processes in living organisms, but are nowadays also commonly used in biotechnical applications. The basic goal of our method was to create a visual abstraction of the binding site that shows a highly simplified version of the binding site in order to reduce visual clutter, while still showing the relevant details needed for a visual analysis and comparison of different data sets. To do so, we start with the so-called Solvent Excluded Surface (SES), a molecular surface that depicts the interface of the protein with respect to a certain substrate, and a path to the active site where the enzymatic reaction will happen. We then extract the part of the surface that forms the binding site, namely the tunnel containing it. These extracted surface parts may have vastly different shapes, which might make the visual comparison process more difficult for the viewer. Therefore, we deform the original shape into an abstract representative. This representative form includes the tunnel in which the binding site is located, as well as the surroundings of the tunnel exit, which may also influence the behavior of an incoming substrate molecule.

We designed our visual abstraction such that the shape of the final representative resembles that of a hat. We call this abstraction the *Molecular Sombrero*, using the Spanish term for ‘hat’. While the crown of the sombrero represents the tunnel containing the binding site, the brim depicts the surrounding surface. The shape of the sombrero encodes the length of the tunnel, the width of the surrounding surface, an user-selected radius measurement. Additional variables can be depicted by the use of color. An example outlining the construction of our abstract binding site visualization is shown in Figure 1. The exact dimensions of our sombrero primitive correspond to the original sizes of the depicted tunnel, allowing for a direct comparison via juxtaposition of two or more sombreros that are shown in coordinated views. Multiple sombreros can either originate from different proteins that exhibit the same binding site, or from different time steps of a simulation. In the latter case, the sombreros allow the user to investigate the dynamics of the binding site, that is, how it changes over time. Furthermore, we show how our newly developed representative can be applied to real-world application cases that were developed in collaboration with domain experts from the field of biochemistry.

2. Structural Biology Background

Our work targets enzymes, thus, we briefly introduce the required biological background. Enzymes are catalytically active proteins that facilitate a chemical reaction. They are essential for the metabolism in all organisms and are widely used in industrial appli-

cations, such as biosynthesis, food processing, or biomedical applications. Therefore, engineering of enzymes to increase their stability, catalytic rate, and substrate profile is an active field of research.

Like all proteins, enzymes are biological macromolecules formed by one or more chains of amino acids. All amino acids have an identical backbone part and a specific side chain, which determines the physico-chemical properties of the amino acid and, therefore, also of the respective region in the protein. In an enzyme, a small number of amino acids forms the so-called *active site*. If substrate molecules bind to the active site, the chemical reaction is accelerated or inhibited. Therefore, the position and orientation of the substrate in the active site is crucial.

For the function of an enzyme, its 3D structure is important. The amino acid chain folds into an energetically favorable, stable 3D configuration, which determines the physical and chemical properties. For example, the amino acids forming the active site have to be accessible to the substrate, otherwise, the enzyme would be inactive. The active site is often located within a so-called cavity. Following the classification provided by Krone et al. [KKL*16], there are three types of protein cavities: (i) enclosed *holes* within the protein, which are not accessible from the outside; (ii) *pores* that form a channel through the protein; (iii) *tunnels* or *pockets* that lead into the protein. In our case, the last category is of interest, since active sites are typically located within such pockets or tunnels. The catalytic function of an enzyme depends not only on the presence of the amino acids forming the active site, but also on the shape of the substrate binding site, which binds the preferred substrate like a lock recognizing the appropriate key (*keyhole-lock-key* model [PGB*12]). In lipases, for example, the active site is located at the bottom of a deep substrate binding tunnel, and the shape and physico-chemical properties of this substrate binding site determines the substrate specificity. Generally speaking, the whole tunnel containing the active site is known as *binding site*, since this region of the molecule is responsible for the substrate binding.

3. Related Work

Our proposed visual abstraction is related to previous methods from molecular visualization, comparative visualization, and geometry processing. In this section, we give an overview of the most relevant related literature.

Protein Visualization: The visualization of molecular structures is a very diverse field. For an overview, we refer to the recent reports of Kozlíková et al. [KKF*16], Alharbi et al. [AAM*17], and Schatz et al. [SKPE19]. For the visual analysis of tunnels and the accessibility of binding sites, molecular surfaces are typically used. Specifically, the Solvent Excluded Surface (SES) [Ric77] is well-suited as it shows the interface of a molecule with respect to a specific type of substrate molecule approximated as a spherical probe. A multitude of efficient algorithms to compute the SES have been proposed [KKF*16]. One of the few publicly available programs that compute the SES as a triangle mesh is *MSMS* [SOS96], which we used for our implementation. Newer approaches usually render the SES using shader-based ray casting [KBE09, LBPH10, JBSK15], which would be unfeasible for our intended use as basis for an abstraction.

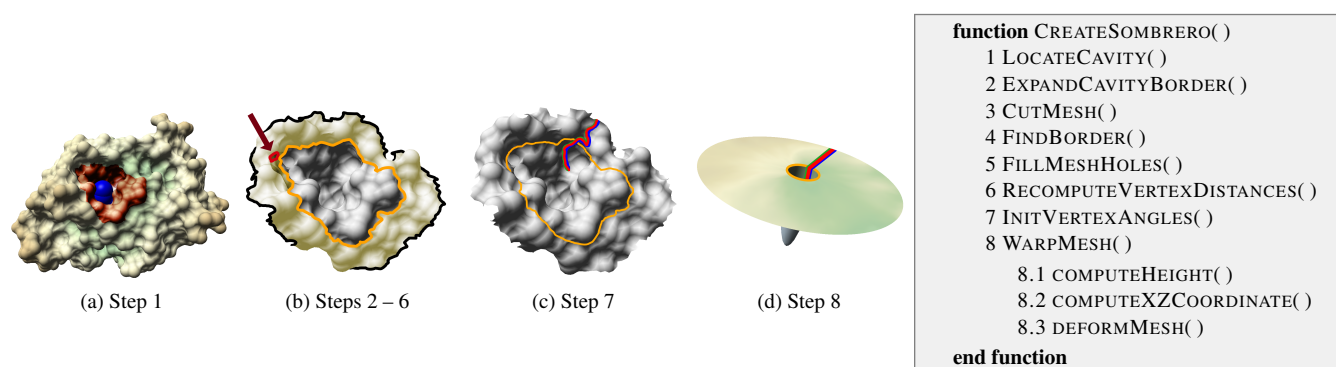


Figure 2: Overview of our Molecular Sombroso method. It shows the different steps of the cutout and morphing process for molecule 1TCA (left) and the algorithmic overview (right). (a) The relevant tunnel (brown) is extracted by combining the results of CAVER [PvK*15] (blue) and a modified version of the dual-SES algorithm [VG10]. (b) Region growing on the original mesh is performed (2), expanding the outer border (yellow). The resulting mesh is cut out (3) and the new outer border (black) is detected (4). Holes in the mesh (red arrow) are either closed during the region growing step or detected and closed with an extra step (5). (c) The meridian (red) and two directions (blue, green) are detected. Additionally, the sweatband (yellow) is computed. Based on this information, the vertex angles are initialized. (d) The deformation of the mesh is performed by computing the new coordinates for each vertex, resulting in a straightened meridian and a circular sweatband. In this example, the bottleneck radius of the tunnel is used to compute the sombrero radius.

One of the drawbacks of the SES is that the resulting molecular surfaces have a rather complicated shape with many small ridges and bulges due to its definition (cf. Figure 2(a)). Therefore, an abstract visualization that shows fewer details to reduce visual clutter while still highlighting important geometric features can be desirable. Gaussian molecular surfaces [Bli82] can be parametrized so that they generate smoother surfaces [KSES12]. In this case, however, the surface will not faithfully represent the cavities anymore.

Cavity Extraction: As mentioned above, we focus on tunnels, since these often constitute the binding site where an enzymatic reaction takes place. Different methods for the extraction and visualization of cavities have been presented. In their recent survey, Krone et al. [KKL*16] classified these methods into four main categories, based on the algorithmic background: Voronoi-diagram-based, probe-based, grid-based, and surface-based algorithms. Besides these four main categories, hybrid algorithms that combine two of the aforementioned approaches exist. In general, Voronoi-diagram-based methods are widely accepted as being stable and reliable. They are for example used by Lindow et al. [LBH11] and in the CAVER 3.0 tool [PvK*15]. Another approach is to extract cavities based on two SES. The first, inner SES is extracted using a probe radius corresponding to the solvent or ligand of interest. The second, outer ‘shell’ SES is extracted using a much larger probe. The space between these two surfaces denotes the cavities. A grid-based version of this algorithm was presented by Voss and Gerstein [VG10]. While the Voronoi-based approaches have the advantage that they provide paths to the binding site, the dual-SES approach provides the whole space of a cavity that is accessible to a ligand of a specific size.

Cavity Visualization: Visualizations of the previously extracted cavities support users during analysis. Simple visualizations typically render a line for the ligand path, fill the extracted cavity with spheres to illustrate its volume, or color the molecular surface to show parts that belong to a cavity [KKL*16]. Advanced

visualizations show additional information to enhance the analysis of binding sites within cavities. Especially the visualization of the dynamic behavior of cavities has recently been addressed in a number of works. Parulek et al. [PTRV13] employed scatter plots for the visual analysis of cavity attributes. They used a graph-based cavity search and mainly plot the attributes of the cavity graphs like length or degree. Lindow et al. [LBBH13] visualized the accessibility of a binding site by tracking piecewise accessibility over time, plotted in a relational graph and shown as an aggregated cavity. Masood et al. [MSCN15] presented a method to extract tunnels and visualize them as a 2D profile by straightening the main axis of the tunnel, showing the width of the pore as well as biochemical properties. While this visualization provides useful information for cavity analysis, the spatial correlation between tunnel-lining amino acids is lost. Byška et al. [BJG*15] proposed a visualization of tunnels that uses cross-sections to show bottlenecks of a tunnel and enriched it with information about tunnel-lining amino acids. In a subsequent work, Byška et al. [BLMG*16] plotted time-dependent tunnel parameters like length and diameter together with tunnel-lining amino acids to show the dynamics of a tunnel. Kolesár et al. [KBP*16] presented a visualization that unfolds protein tunnels to create a 2D map. The length and width of the tunnel as well as tunnel-lining amino acids are thus visualized. The similarity of the tunnels is shown in a scatterplot, while the tunnels can be visually compared using juxtaposition. Malzahn et al. [MKR17] presented a similar approach to create a 2D map from a cavity. They add three-dimensional ball-and-stick representations of the tunnel-lining amino acids as context. The aforementioned methods have in common that they either represent the original, complex 3D shape of the cavity, which makes it hard to compare different data sets, or they only show a 2D representation of the cavity, thus completely removing spacial information. Our method is situated in between, as it only partially abstracts the 3D shape.

Geometric Projection and Mesh Deformation: Our approach

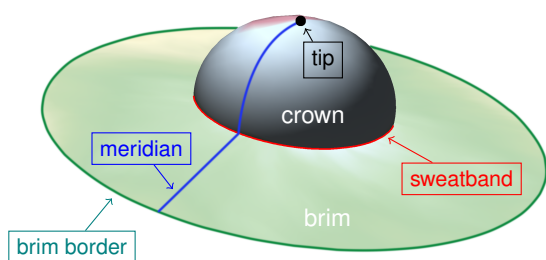


Figure 3: The sombrero terminology. A sombrero primitive is divided into the round crown and the flat brim. The tip (black) describes the highest point of the crown, while the meridian (blue) is the formerly shortest path from the tip to the brim border (green). The sweatband (red) forms the border between brim and crown.

to map the SES mesh of a tunnel to a hat-like structure requires deforming one shape into another. Mesh morphing has for example been used for molecular visualization by Scharnowski et al. [SKR*14] to compare two molecular surfaces by mapping one surface to the other. Their method works well if the two surface meshes are similar—that is, comparable in shape. Other methods have been proposed to map the molecular surface to a sphere. Hass and Koehl [HK14] used a conformal mapping that provides a measure of the ‘roundness’ of a protein, which can also be used to compare two surfaces. In our case, however, we are not directly interested in the geometric differences between two proteins. Krone et al. [KFS*17] proposed a technique to create a 2D map of the molecular surface. They first reparametrize the whole molecular surface to a sphere and then create the 2D map using projection methods from cartography. Their method, however, is not tailored to cavities. In fact, their algorithmic pipeline optionally removes cavities with multiple openings prior to the reparametrization to lower the distortion of the surface.

In medical visualization, 2D representations of tubular 3D structures are also used for flattening vessels or colons (e.g., [KFW*02, GSZ*13, MK16]). While these methods create good results in their intended application scenario, these methods are not directly applicable to protein tunnels, since these are often not tubular and their radius and shape exhibit a larger variation.

4. Data Preparation and Algorithm Overview

At first, the choice of a hat-like shape seems somewhat arbitrary for the representation of a cavity containing an active site. The final shape of our abstraction was in fact directly motivated by our project partners from technical biochemistry. As the possible radius variation of a cavity was mostly irrelevant to them, it allowed us to use a more general geometric shape. We discussed the use of cylinders as well as the currently used half-spheroids with them and they decided that the half-spheroids offer an easier way to locate the active site. Additionally, cylinders would tend to over-emphasize the area surrounding the active site as their tip is an area and not a singular point. To show the surrounding of the cavity entrance was an important additional requirement of our project partners. As we did not want to over-emphasize any direction on the surface, a circular shape with a cut-out for the cavity entrance itself was obvious. The

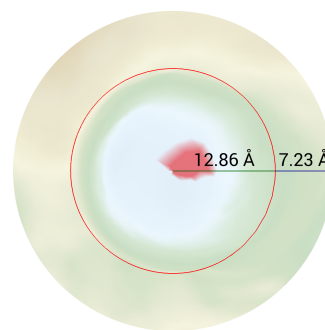


Figure 4: Flat version of the Molecular Sombrero shown in Figure 1. Coloring by distance to protein center and active site location. The sweatband that denotes the border between crown and brim is shown as red circle. Text labels show the tunnel radius and the brim width.

result of this thought process was a shape that resembled the shape of a hat, so the final naming was also obvious.

Before describing our Molecular Sombrero algorithm, we want to reiterate the terms to describe the different parts of a hat. Since our method creates a hat-like abstraction of the binding site, it seems natural to adopt the terminology commonly used for this piece of clothing. The part that covers the wearer’s head is called *crown*. The horizontal part, typically protecting the wearer from the sun or other atmospheric conditions, is the *brim*. At the inside of a hat, directly at the border between brim and crown, a ribbon is typically sewn in, which is called the *sweatband*. Therefore, we will describe the border between brim and crown as sweatband, ignoring the breadth of real-world sweatbands. The highest point on top of the crown of our sombrero is called *tip* (see Figure 3).

Our approach consists of the three major steps shown in Figure 1. A more detailed description is given in Figure 2. First, the user has to choose the tunnel of interest by providing information about the amino acids that form the active site, as mentioned in Section 2. We apply the established and widely-used tool *CAVER 3.0* [PvK*15], which extracts the path to the active site using a Voronoi diagram. Since there can be multiple paths leading to the active site, we use the visual analysis tool *CAVER Analyst* [Kvv*14], which integrates *CAVER*, to choose the desired one. The result, that is, the chosen path and the protein atoms along it, are used as input for our Molecular Sombrero generation.

Secondly, our approach requires a mesh of the molecular surface. We use the tool *MSMS* by Sanner et al. [SOS96], which generates a high-quality mesh of the SES. In contrast to *CAVER*, we incorporated *MSMS* directly into our visualization application. The output of *CAVER* is used to determine the vertices that are close to the chosen path to the active site. That is, we obtain the part of the surface mesh that forms the tunnel leading to the active site. This part will later form the crown of the Molecular Sombrero. Additionally, a certain region around the tunnel exit is extracted, which will later form the brim of the hat. The size of this region is user-definable; in most cases, the size is directly derived from the diameter of the assumed ligand for the visualized binding site.

Table 1: Modified default values of the CAVER algorithm, including the range of the modification (range), the most typical chosen value (typical), and a short explanation of the meaning of the value. All numerical values are given in Ångström (Å).

Parameter	Range	Typical	Explanation
probe_radius	[0.9, 2]	1	minimum cavity radius
shell_depth	[3, 6]	4	branching dead zone
shell_radius	[8, 12]	10	outer SAS probe radius
max_distance	[2, 6]	3	dist (calculated & chosen start)
desired_radius	[2, 6]	5	start vertex radius

Once all relevant parts of the surface mesh are extracted, the third and final step of our pipeline, namely the morphing into the sombrero shape, can be performed. Finally, if desired by the user, a flattened version of the sombrero can be generated. This flat version provides an occlusion-free 2D representation, but conveys less information than the original hat-like 3D shape (see Figure 4).

5. Algorithmic Details and Implementation

In the first step of our pipeline, the path to the active site is extracted via CAVER. The cavity detection performed by CAVER depends on dozens of parameters. We list all parameters of which we modified the default values in Table 1. To ensure that the final results are comparable, a restriction on the input parameters of CAVER has to be taken into account: all parameters that vary the exit point of the detected tunnels should be the same for all visualized data sets (e.g., probe_radius or shell_radius). However, CAVER is not always able to find a Voronoi vertex close enough to the location of the user-defined active site, which is necessary to start the tunnel detection. In this case, the parameters that control the location of the starting point (e.g. desired_radius and max_distance) have to be modified by the least possible amount (normally not more than 1 Å) that leads to the successful detection of a starting Voronoi vertex. Note that this issue is not related to our Molecular Sombrero visualization, as this is always necessary when using CAVER.

5.1. Binding Site Mesh Extraction

To be able to deform the mesh corresponding to the binding site onto the sombrero, it has to be segmented from the molecular surface mesh first. In order to facilitate the mapping onto the sombrero, we have to ensure that the resulting vertices and triangles form a planar object without any holes. The results of CAVER contain the center line of the selected tunnel, the width of the tunnel on several selected positions along this center line, and the tunnel-lining amino acids and atoms. Using this information, the vertices belonging to the final mesh are selected. In the following, we consider binding site mesh vertices as *marked*, all others as *unmarked*.

Given the tunnel-lining atoms from CAVER and the per-vertex atom identifiers provided by the MSMS mesh generation, it is possible to mark all vertices that are in contact with a tunnel-lining atom. These vertices—or rather the triangles they belong to—are forming the mesh of the binding site. However, a mesh consisting only of those vertices will not be ideal, since it is possible that

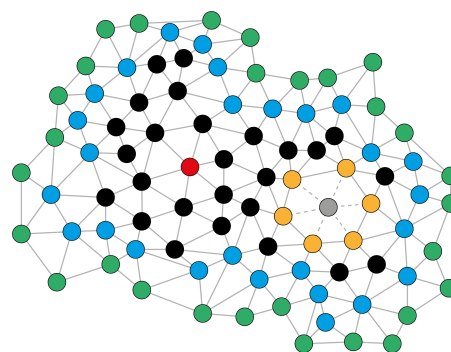


Figure 5: Possible mesh after mesh cutting. Red vertex: Closest vertex to active site. Orange vertices: Were forming a hole in the mesh. This hole is closed using the grey vertex (newly created triangles shown as dashed edges). Blue vertices: Outer border before region growing. These will be the sweatband vertices. Green vertices: Added by region growing with step count 1. These will be the brim border vertices.

CAVER lists only one side of the tunnel in the results. This happens for example for cleft-like cavities that are narrower in one direction. To overcome this issue, we decided to incorporate a second cavity extraction method based on the difference between two SES with different probe sizes. Our approach is based on the 3V algorithm by Voss and Gerstein [VG10] (see Section 3). Contrary to their original work, we developed a mesh-based implementation. That is, we use two SES meshes as input, not surface-representing volumes. The first of the two meshes is the already calculated SES, typically generated using a probe radius between 1.2 and 3 Å depending on the size of the solvent molecule considered. Additionally, the second mesh—called the shell mesh—has to be calculated using a larger probe radius. Voss and Gerstein empirically determined a range of 6 to 10 Å as suitable values. In our experiments, we found that a range of 8 to 12 Å delivers better results for most of our use cases, depending on the radius of the inner SES. Using their radii we noticed that most of the found cavities were smaller than expected when comparing them with the results of their original method. Therefore we increased the used radius. The lower values of Voss and Gerstein tend to leave out a part of the cavity area close to the opening when using them for our mesh-based approach. Our mesh-based variant of the algorithm then extracts the cavities by detecting vertices of the inner mesh that are farther away from all vertices of the shell mesh than a distance threshold d_t . All triangles with vertices of the shell mesh closer than d_t belong to the cavity mesh. Submeshes, which are forming individual cavities, are found by detecting connected components. Too small values for d_t could lead to the detection of cavities where there are in fact only small bumps. A default value of $d_t = 4$ Å gives good results for the enzymes shown in this work.

As pre-calculated vertex labels make it possible to compare the vertices of CAVER with the current results, we can now find the cavity submesh that shares the most vertices with the previously detected one. That is, the new binding site mesh is the union of the submeshes found by the two cavity extraction methods. Con-

sequently, our method combines the benefits of both algorithms: CAVER provides the exact active site access paths, while our 3V-like algorithm provides the complete geometry of the tunnel.

The combined binding site mesh will later form the crown of the desired hat-like abstraction. Therefore, we will also refer to it as *crown mesh*, consisting of crown vertices V_c . However, two issues still have to be addressed: First, no vertices for the brim have been defined yet. Secondly, it is possible that the crown mesh has a genus higher than zero, i.e. the mesh has holes in it. These holes occur if a secondary tunnel leading to the binding site exists. This tunnel is not part of the binding site, but will lead to a hole in the mesh. In order to close these holes and to generate the brim, a region growing is performed. That is, additional vertices are added to the binding site mesh iteratively. In the first iteration, all vertices of the SES mesh that do not belong to V_c but have an edge that connects them to a vertex in V_c will be marked (green vertices in Figure 5). In each subsequent iteration, all unmarked SES mesh vertices connected to a marked vertex will also be marked. Since this process defines the width of the hat's brim, the number of steps is user-defined.

After the region growing was performed, it is still possible that holes in the mesh were not yet correctly closed. In order to close these holes, the border vertices and edges of the mesh have to be detected. Border edges are all edges that only belong to a single triangle, while border vertices are defined as the vertices connected by border edges. If more than one connected set of border edges is found, all sets except one (the set forming the outer border) belong to a hole in the mesh. We identify the outer border set by averaging the vertex positions of the borders and comparing these values with the end position of the tunnel description produced by CAVER. The set with the closest average point is the one representing the outer border. Holes are then closed by inserting a triangle fan with its center vertex at the average position of the corresponding border vertices (gray vertex in Figure 5).

The final result of the mesh cutting process is a genus zero mesh. Figure 5 shows an overview of this mesh, highlighting the different stages using color. All vertices in the figure (except the grey one) are also present in the original surface of the whole protein produced by MSMS.

5.2. Geometry Morphing

Once the mesh of the protein is cut to only represent the relevant tunnel chosen by the user, the remaining geometry can be deformed to the actual sombrero shape. Since the active site should always correspond to the tip of the hat, the closest vertex to the active site v_a has to be identified. Therefore, the exact location of the active site is needed. A typical definition for the active site position would be the average position of all side chain atoms of the responsible amino acids. As this definition can be arbitrarily interchanged, our project partners used a far more specific description based on the properties of the specific active site type.

After v_a is found, the outer border of the sombrero brim V_{bb} is detected. This follows the same process as in Section 5.1 so the results from there can be adopted. Note that all vertices of the mesh are now divided into two classes. The first class are the crown vertices V_c the second class V_b is composed of the brim vertices,

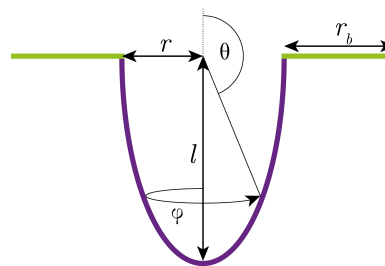


Figure 6: Governing values for the parametrization of the hat-like shape. The crown (purple) is one half of a spheroid with radius r in x - and z -direction and radius l in y -direction.

of which V_{bb} is a direct subset. Additionally to these classes, the sweatband vertices V_s have to be determined. They are defined as follows, where E is the set of all edges of the triangle mesh:

$$V_s = \{v \in V_c | \exists e = (v, a) \in E : a \in V_b\} \quad (1)$$

In other words, all vertices of the crown that have an outgoing edge towards a brim vertex are considered as sweatband vertices. Given these arrangements it is now possible to determine all relevant parameters for the final sombrero shape, which will be, for the sake of simplicity, represented by a half-spheroid in the case of the crown and a circular disk containing a cut-out for the brim. A spheroid can be parametrized with two radii r and l and two angles φ and θ . For the final sombrero shape, additionally the diameter of the brim r_b is required. All of these parameters are depicted in Figure 6.

The length l of the sombrero is the length of the tunnel detected by CAVER. While the brim width r_b is calculated from the properties of the undeformed SES mesh, the choice of the calculation method for r is left open for the user. Depending on the desired property, there currently exist three options for the calculation of r_b . First, the sombrero radius r can be derived from the length l_s of the sweatband, which describes circumference of the opening of the crown. Therefore, it is calculated as $r = l_s/2\pi$. Second, r can be derived from the bottleneck radius of the inspected tunnel, which is given by CAVER. Opposed to the first option, this leads to smaller radius values that may not preserve the original proportions of the tunnel. Despite this deformation, this radius can be useful for certain use cases. The third and last option is to directly adopt the radius of the tunnel exit Voronoi vertex of CAVER. This approach serves as a compromise between the former two as the resulting value typically lies between the others.

The brim width r_b is calculated by averaging the minimum distances from each brim vertex to its closest sweatband vertex, following the edges of the mesh graph. We compute the shortest path for each vertex of V_{bb} to the closest vertex in V_s using an iterative algorithm similar to the region growing. That is, the computational complexity is the same as for the brim finding.

After calculating the parameters of the final sombrero, the target position of each vertex has to be derived. It is not possible to calculate the angular values of the vertices directly, as their distribution is not uniform in all directions. The position of v_a , which lies typically not in the center (Figure 5), makes this even more difficult.

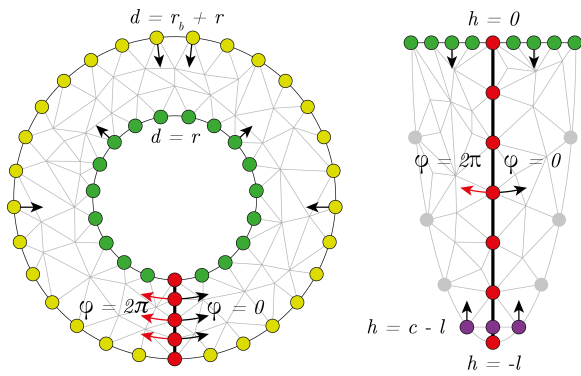


Figure 7: Value diffusion for the brim (left, seen from top) and for the crown (right, seen from the side). The fixed values are the outer brim (yellow), the sweatband (green), the meridian (red), and the tip neighbours (purple). Additionally to the shown fixed values, the φ -values of the yellow and purple vertices are already set to their intended angle. While the value diffusion for d and h runs on the separated crown and brim, the φ -diffusion is performed on the whole mesh.

For our sombrero calculation, we adapt the method developed by Rahi and Sharp [RS07] that solves the problem for spherical target geometry. It assigns spherical coordinates to all vertices followed by a mapping to a sphere. Since both shapes—spheroid and circular disk—can be easily parametrized with polar coordinates, only few modifications are necessary to tailor the method to our needs. After assigning fixed coordinates for certain vertices, the method of Rahi and Sharp diffuses these values over the whole surface, resulting in a smooth value distribution. This is done via a Gauss-Seidel relaxation, which calculates the value for all calculated vertices by iteratively averaging the values of all of their neighbors. Using a sphere as target geometry, fixed values are assigned to the north and south pole, their direct neighbors as well as to the meridian vertices that connect the two poles. These fixed vertices serve as seeds for the diffusion process, providing a constant value flux. In the original work, the diffusion is performed in two directions, one for the φ angle and one for the height z of the vertex. The final θ angle is derived from z using the inverse Mercator projection.

In our case, the φ value is determined as described by Rahi and Sharp, with the small modification that we have to insert an artificial north pole vertex v_{north} that is connected to all vertices of V_{bb} using a triangle fan. While the active site vertex v_a serves as south pole, the meridian is determined by calculating the shortest path between v_a and v_{north} , using a constant edge weight of 1. The diffusion of the φ values is performed on the complete mesh. For this, the values of the meridian vertices, as well as the ones of the neighbors of v_{north} (V_{bb}) are fixed. These, and all other fixed values are shown in Figure 7. While the artificial north pole is needed for the φ value diffusion, the diffusion of the height values is possible without it. The original approach for the z value diffusion is utilized to calculate two different values. One is the height of the vertex on the crown h , the other one is the distance from the center of the brim, called d . For the value calculation, brim and crown are

considered as separate meshes. The sweatband vertices have to be present in both meshes, since they form the border between the two vertex sets and provide fixed values for the diffusion. For the diffusion of the crown's height values h is set to 0 for the sweatband vertices and to $c - l$ for the neighbors of v_a . c is a constant offset from the length l of the sombrero. In our tests a value of $c = l/2m$, where m is the count of meridian vertices, appeared to produce the most appealing results. After the height h of each crown vertex is known, their θ angles can be determined via the inverse Mercator projection, analogously to the approach of Rahi and Sharp. Note that the range of the θ values is $[\pi/2, \pi]$, since we are mapping only the lower half of an ellipsoid. The third and last value diffusion is the one for the distance to the brim center. Here, the values of the sweatband vertices are set to r while the values of the brim border vertices are set to $r_b + r$. Since the Gauss-Seidel relaxation needed for the value diffusion is the most time consuming part of the whole pipeline it was parallelized using CUDA.

After all necessary values are known, the movement of all vertices into their final position can be performed. With the y-axis being the rotational axis of the sombrero, the position of the crown vertices is determined using the following formula:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r \cdot \cos(\theta) \cdot \cos(\varphi) \\ l \cdot \sin(\theta) \\ r \cdot \cos(\theta) \cdot \sin(\varphi) \end{pmatrix} \quad \begin{array}{l} \theta \in [\pi/2, \pi] \\ \varphi \in [0, 2\pi] \end{array} \quad (2)$$

Contrary to the crown, the brim vertices do not require a specific calculation for their position on the y-axis, since they are all residing in the $y = 0$ plane. Given the newly calculated distance from the center d , their position will be the following:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d \cdot \cos(\varphi) \\ 0 \\ d \cdot \sin(\varphi) \end{pmatrix} \quad \varphi \in [0, 2\pi) \quad (3)$$

Due to the deformation, the surface normals of the spheroid have to be recomputed to achieve a correct lighting. For a correct lighting, crown and brim have to be rendered separately, allowing the sweatband vertices to have different normals in each draw call.

The optional flat variant of the sombrero, depicted in Figure 4, is calculated directly from the final sombrero shape. Setting all available y-coordinates to zero would generate highly crowded crown geometry near the sweatband, especially for narrow and deep tunnels. Therefore, we decided to recalculate the distance from the center for each crown vertex by interpolating it between 0 and r using the absolute value of their already known height h divided by the sombrero length as weighting factor. This ensures an even vertex distribution for the crown vertices. Additionally, a line connecting the sweatband vertices is rendered to provide orientation for the user. Judging from the look of the visual primitive, the flat version of the sombrero could be prone to so-called Delboeuf illusions (cf. [HHG07]) that distort the perception of circle radii. This is only the case if the user decides to compare several flat representations using different values for the brim width. As a variation of the brim width between comparison partners destroys the comparability, this case will not happen often. Nonetheless, we decided to add text labels for the radii to the flattened representation to counter this issue.

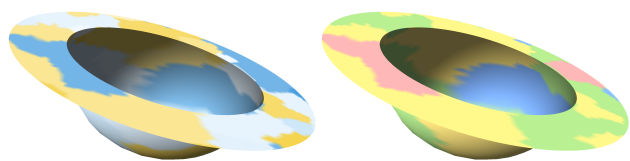


Figure 8: Molecular Sombrero of the active site of an esterase (PDB-ID 1JU4). Left: A color gradient shows the degree of hydrophilicity (blue) or hydrophobicity (yellow). Right: Colored by physico-chemical properties. The colors show if an amino acid is acidic (red), polar (green), nonpolar (yellow), or basic (blue).

5.3. Mapping of Physico-Chemical Properties

Projection of a potentially complicated surface onto a simple and smooth abstract shape obviously leads to some kind of information loss. This increases the importance of the remaining information carriers. Apart from the encoded tunnel features, only color remains as primary information carrier. Since our approach directly deforms the mesh representing the surface of a protein, all coloring techniques that are used to color protein surface meshes are also applicable for the Molecular Sombreros.

As for most protein visualizations, the mapping of physico-chemical properties is of importance. Values like the temperature factors or hydrophobicity values (cf. Figure 8) allow conclusions on the behavior of the visualized tunnel and its surroundings. In addition to the more general surface properties, especially the location of the binding site residing in the shown tunnel is of interest. We show it by either coloring all responsible amino acids or only the responsible atoms. From the construction of the sombrero shape, it should be clear that a part of the site should cover the tip of the primitive.

In addition to the more traditional physico-chemical coloring modes, we decided to include the heightmap coloring for molecules presented by Krone et al. [KFS*17] for their Molecular Surface Maps. This coloring mode is inspired by traditional topographic maps of the earth that try to incorporate the typical color of vegetation into the map. Areas at lower altitudes are shown in blue (water), medium ones (woods) are green, and the highest areas, typically mountains, become brown (cf. Figure 1). In our case, all vertices are colored by their distance to the geometric center of the protein mesh. The geometric center is the average position of all the represented vertices. That is, this scheme maintains some of the structural information that was lost in the projection process. This possibly allows for the identification of ligand access paths on the brim. A higher part of the surface may block ligand movement while a lower one may not.

Apart from the static values described above, some quantities like the Root Mean Square Fluctuation (RMSF) [KZ10] describe the movement of the shown protein over time. As shown in Figure 9, we use a color gradient from white (low RMSF) to brown (high RMSF). These visualization modes are of specific interest since these movements may lead to an opening or closing of the visualized tunnel. Tunnels with consistently low RMSF values will be

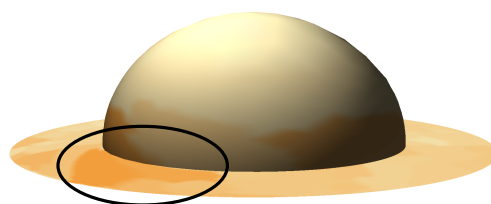


Figure 9: Sombrero of 1TCA using the Root Mean Square Fluctuation coloring option. The values were accumulated over 10,000 time steps of a simulation. White corresponds to a low value, brown to a high one. Note the marked, more flexible region on the brim, which can potentially influence the catalytic rate.

more stable, while tunnels with higher values may have a tendency to close eventually. Please note that these movements sometimes may be crucial to transport the substrate molecule to the active site.

Given the aforementioned coloring modes there is still no direct visual link between the original SES and the sombrero representation. This issue is solved by the introduction of a brushing and linking approach. It is possible for the user to display the original SES or the cutout (cf. Figure 2(c)) alongside with the sombrero. Now the user is able to paint on each of the displayed representations to mark certain areas. This painting is propagated to all other representations to create a direct visual link between the surfaces.

6. Results and Discussion

As mentioned in the introduction, our goal was to create an abstract representation of the binding site of a protein—that is, of the tunnel leading to the active site and its immediate surroundings—without losing the spatial context entirely, since the geometric properties are important for the function. Our Molecular Sombreros satisfy these requirements, as they provide a visual abstraction that only shows the aforementioned, most relevant aspects of the binding of an enzyme: the length of the tunnel, its width, and the biochemical properties that can influence the accessibility of the active site within the tunnel. That is, the overall shape is represented by the geometry of our abstract representation (brim width, crown height and diameter), whereas the local shape of the SES can be preserved by mapping them to color (see, for example, Figure 1).

Molecular Sombreros are an intermediate representation that fills the gap between the full atomistic detail provided by the SES and more complete abstractions like Molecular Surface Maps [KFS*17] or the tunnel flattening proposed by Kolesar et al. [KBP*16], which both map parts of the surface to a 2D domain. Following Shneiderman’s visual information seeking mantra [Shn96], our abstract visualization provides a first *overview* that allows domain experts to assess the key properties of a tunnel and to quickly identify regions of interest. Subsequently, the user can then switch other representations like the SES that provides *details on demand* for an in-depth analysis.

Due to the abstraction, Molecular Sombreros also allow for an efficient and straightforward comparison of key aspects of the binding sites for multiple proteins. Here, this visual complexity of the SES is not needed and might even hinder the initial comparison.

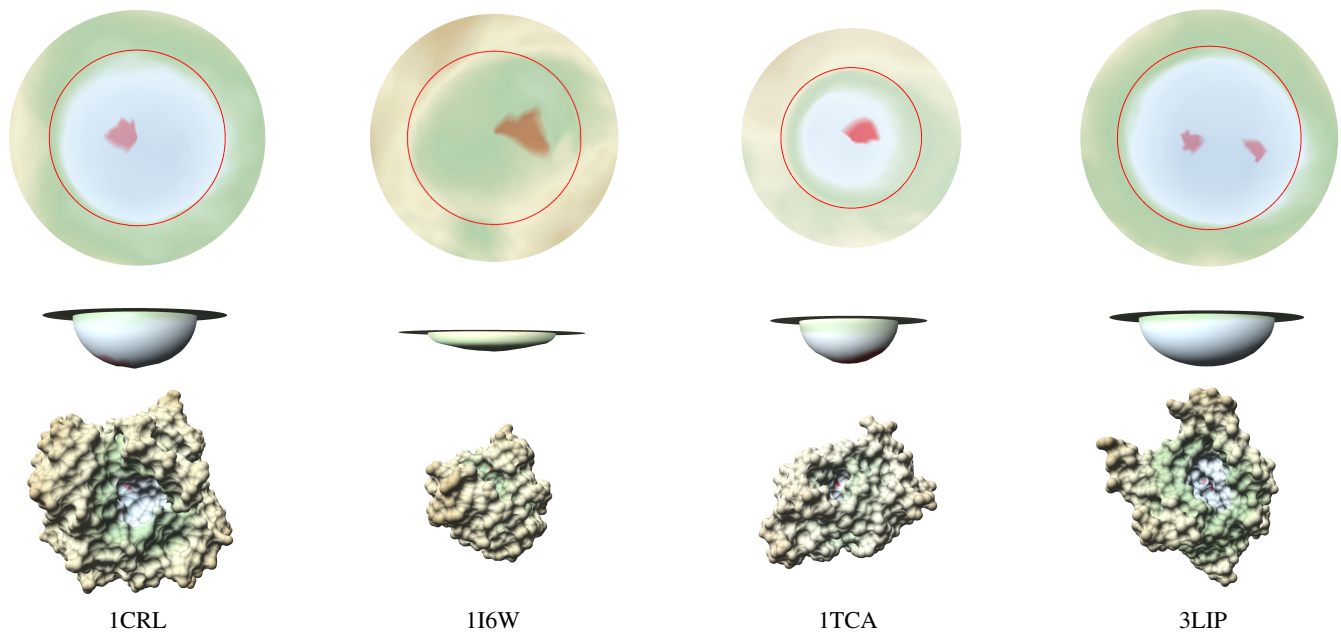




Figure 10: Comparison of the results for four different enzymes (name tags: PDB IDs). Top row: flattened sombrero view with the ‘sweatband’ shown as red line. Middle row: side view of the sombrero. Bottom row: original Solvent Excluded Surface of the respective protein. The representations in each row are shown in the same coordinate system, that is, they are shown in the correct size with respect to each other. Coloring by distance to the protein center (blue-green-brown: ) and active site location (red: ) . Note that the geometric properties of our abstract visualization allow the user to compare tunnel properties with respect to the other proteins, while the surface colors show the local geometric properties relative to the size of each individual protein.

In subsection 6.1, we discuss three typical use cases for Molecular Sombreros that were identified together with domain experts from the field of structural biology and biochemistry. We also show examples for all use cases using real-world data sets. The computation times of our prototypical implementation described in section 5 are given in subsection 6.2. Finally, we discuss the limitations of our method in subsection 6.3.

6.1. Use Cases and Discussion

In this section, we discuss the intended use cases for our Molecular Sombrero visualization. All of the presented use cases are motivated by our project partners from structural biology and use real world data. Most of the presented data sets are publicly available via the PDB [BWF*00].

Visual Analysis of Binding Site Accessibility: As explained in Section 2, the influence of the binding site on the orientation of an approaching substrate is of interest for bioengineering applications. This orientation is directly affected by the physico-chemical properties of the protein surface like charge or hydrophobicity. Hydrophobic parts, for example depicted in Figure 8 (left) as blue areas in the front and in the back, may repel some parts of a substrate, while the hydrophilic parts shown in yellow may act as attractors. As mentioned above, the traditional visualization of the binding site using the SES would potentially be cluttered, which would increase the burden on the viewer unnecessarily. The simplified visualiza-

tion offered by our Molecular Sombreros only shows the most important geometric properties of the binding site (radius and length of the tunnel) and maps a user-defined physico-chemical property to it using color, which allows for a fast and convenient visual analysis. The user can directly infer whether a substrate molecule can enter the binding site and reach the active site in the appropriate orientation, or whether it would be repelled.

Comparison of Homologous Proteins: The second use case is the direct comparison of the binding sites of different proteins with similar enzymatic function. This can be especially useful for cases where two proteins are able to bind to the same ligand type and where vastly different cavity shapes lead to the same ligand conformation. We implemented a prototypical application that allows to show the Molecular Sombreros of an arbitrary number of data sets in coordinated views that are juxtaposed. Juxtaposition was identified as a meaningful arrangement for comparison tasks by Gleicher et al. [GAW*11]. Typically, coordinated views of different proteins require an alignment. Molecules can for example be aligned by minimizing the Root Mean Square Deviation [Kab76]. Alignment can be problematic if the structures differ too much. In our case, however, the alignment of the sombreros is no issue since they are generated in the same reference coordinate system. The virtual camera of each view can be adjusted either independently or in a coordinated manner. Consequently, following the taxonomy of Gleicher et al. [GAW*11], our comparison employs the concept of *juxtaposition using similar layouts to aid comparison*.

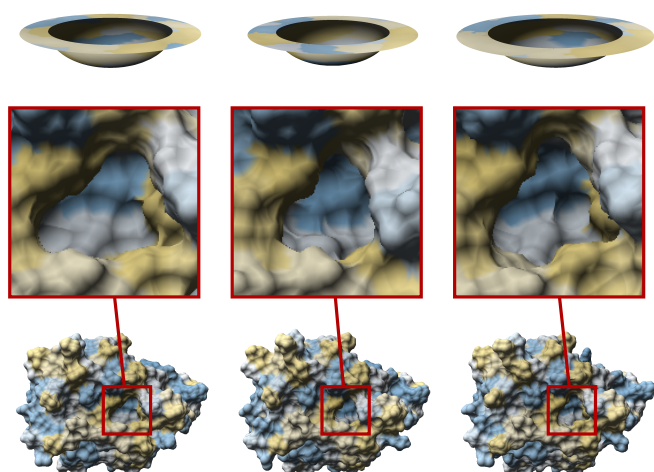


Figure 11: Comparison between different time steps of a simulation of 1TCA in water (not visible). Colored by hydrophobicity (blue: hydrophobic; yellow: hydrophilic). Additionally, the brightness is modulated according to the distance to protein center.

An example for this comparison is depicted in Figure 10. The shape of the sombreros as well as their coloring represent the properties of the visualized tunnels, making them easily accessible for visual comparison. It is clearly visible that the tunnel leading to the active site of 1TCA is considerably narrower than the ones of 1CRL or 3LIP, while having a similar depth. Our sombrero visualization of 1I6W shows that the tunnel leading to the active site is much more shallow but similar in width as in 1CRL and 3LIP. The height map coloring shows the original structure and can highlight possible substrate access paths (green ‘valleys’ on the brim). In the case of 1TCA, the coloring allows the user to identify a part of the brim that is lower than the surrounding geometry. Substrates coming from this direction may proceed faster towards the active site since they are not hindered by other atoms. That is, our proposed visual abstraction of binding sites can also help domain experts to reason about catalytic rates based on the visualized parameters.

Investigation of Protein Dynamics: The third use case for our Molecular Sombreros is the analysis of protein dynamics, which is especially interesting in the context of protein engineering. Over time, the atoms forming the path to the active site may move and change the shape of the tunnel. This can be depicted using the RMSF coloring mode, which shows the squared mean displacement of each atom relative to its average position. As shown in Figure 9, the RMSF helps to identify flexible parts of the protein. While the left part of the crown is relatively stable, the part shown on the right is more flexible. Depending on the task, this could either be a desired property of the protein or—if it is undesirable—the visualization offers a hint where to introduce a point mutation of an amino acid to stabilize the protein tunnel. With classical surface visualizations, this can be less obvious, since the shape of a tunnel is visually quite complex. Not only the stability of the tunnel is of interest, also the surrounding amino acids in the brim region influence the substrate binding probability. As depicted in Figure 9,

the visualized lipase has a flexible surface region close to the tunnel entrance. When investigating this more closely, one sees that this part actually corresponds to a loop in the amino acid chain that could block the access to the active site completely (e.g., a lid that makes the binding site inaccessible). Removing or at least stabilizing this part may increase the reaction rates of the enzyme. Note that this closer investigation of the flexible loop region requires switching to another representation, for example, the Cartoon representation [Ric81] visualizing the conformation of the amino acid chain, or a ball-and-stick representation showing all atoms. As with all other visualizations for molecular data, there is not one single ‘best’ representation that serves all purposes, but each representation shows a specific feature. The Molecular Sombreros provide a first overview to quickly identify the region of interest in this use case, which can then be further investigated using other representations providing details on demand, if necessary.

A second possibility to depict protein dynamics is juxtaposing several time steps, similar to the comparison use case discussed above. This helps to analyze the change of the binding site over time by using a juxtaposed comparison of selected configurations. Figure 11 shows an example where three different time steps of a simulation are depicted, colored by hydrophobicity. As observable, the tunnel entrance is not only considerably smaller in the second point in time, but the tunnel is also more shallow. The pattern of hydrophobic patches on the sombrero brim is the same in all time steps, that is, the influence on the orientation of an approaching substrate molecule will be the same. However, since the tunnel entrance is smaller, it will be potentially harder for the substrate to reach the active site. This rather obvious property of the tunnel could probably also be detected using the SES, even though it has higher visual clutter. However, the difference between the first time step and the third one is much more obvious when using our sombrero visualization. While both tunnels have similar shape, the first one is deeper but not as wide as the third one, which is hard to see using the SES.

Expert Feedback: We presented our results to our project partners from biochemistry and asked them to use the method to investigate different binding sites. Typically, the shape of a binding site in proteins allow for a prediction of specificity and biochemical activity of the protein. They described the Molecular Sombrero method as a powerful method for the visual classification of binding sites, because it supports them in obtaining an overview on the large number of structures in the protein family under investigation. It also allows the experts to analyze the flexibility of binding sites obtained from molecular dynamics simulations and to identify functionally relevant conformational states. In the future they would like to apply the Molecular Sombrero method as a useful first level abstraction for routine visual analysis. In the past, this visual analysis has been done using CAVER, so our method works as an extension to their already accustomed workflow and makes it easy for them to generate the needed input files. This visual analysis is then typically followed by molecular docking experiments or Molecular Dynamics simulations to further examine the assumptions made during the visual analysis. They expressed that the Molecular Sombreros will help them to select relevant conformations and ensemble members for additional investigations.

Table 2: Performance measurements for different proteins. All timings are given in milliseconds (averaged over five different runs). Timings are denoted as follows: t_{ce} : SES-based cavity extraction time (excluding MSMS computation time); t_{me} : Binding site mesh extraction; t_{ϕ} : ϕ -value diffusion; t_{θ} : Combined height value diffusion and θ -value computation for brim and crown; t_{total} : Total time. The vertex count is measured after mesh extraction.

	#atoms	#vert	t_{ce}	t_{me}	t_{ϕ}	t_{θ}	t_{total}
116W	1351	2709	17	25	1322	1991	3353
1TCA	2324	3406	33	37	1286	2025	3379
3LIP	2338	4470	38	46	1455	2017	3557
1JU4	4351	3735	47	53	1425	2068	3593

6.2. Performance Evaluation

As explained in section 5, we implemented all time-critical parts of our algorithmic pipeline to run on the GPU. We evaluated the performance of our implementation using four differently sized enzymes downloaded from the PDB [BWF*00]. All of the following evaluations were conducted on a desktop PC equipped with an Intel i7 6700 (3.4 GHz), 32 GB RAM, and a Nvidia Geforce GTX 980 graphics card. The timings are listed in Table 2. All measurements were performed with a brim width of ten, which is the width we also used for most of the figures throughout this paper. As can be observed, computation times are mostly independent of the atom count, since the most crucial part—the value diffusion—only operates on the mesh of the tunnel. Therefore, the limiting factor is the vertex count of the final mesh and, furthermore, the connectivity of these vertices. That is, the more triangles are connected to the vertices on average, the higher the computational cost. This is the case for 1JU4. With a lower vertex count than 3LIP but with a higher overall vertex connectivity, the computation times are slightly higher. The height value diffusion times t_{θ} divide almost equally between crown and brim. It is notable that the height value diffusion in total takes considerably longer than the angle diffusion, although they work on the same mesh. This is caused by the lower number of fixed value vertices in the height value diffusion case. With an overall computation time of several seconds, our current implementation cannot be considered as interactive. However, as it produces results after a reasonably short waiting time, it can be integrated into a typical visual analytics workflow, where the user expects almost instantaneous feedback after changing parameters.

6.3. Limitations

As detailed above, our Molecular Sombrero visualization is beneficial for different use cases due to the design of the visual abstraction. However, as most abstract visualizations, our method is not intended to completely replace the original visualization (in our case the SES), but rather to complement it where fitting (as already mentioned in the *Protein Dynamics* use case). Due to the reduced detail offered by the abstraction and the deformation of a complex geometry into a simple one, it has inherent drawbacks and limitations. For example, the projection onto the sombrero is not area-preserving, which is observable especially at the difference between brim and

crown. In areas where the vertex connectivity is high, the approach tends to decrease the size of triangles, leading to areas with varying vertex density. Another problem arises if a cavity has side tunnels. Since our algorithm maps the whole result from the adapted 3V method, side tunnels may be projected to the final surface, increasing the visual clutter by adding unwanted information. This could, in theory, be bypassed by cutting off these side tunnels at a certain position (e.g., based on the distance from the path found by CAVER). However, the detection of the cutting positions is no trivial task due to the possibly intertwined structure of some tunnels. This effect could be avoided by increasing the probe radius of the initially used SES and, therefore, closing the side tunnel, but this might also close the main access towards the active site in some cases. Finally, the mapping of the whole tunnel leads to another problem. If the active site does not lie at the bottom of the tunnel but in the middle, parts of the tunnel that should be farther away from the entrance than the active site will be mapped to a position above it. While this is no problem from an algorithmic point of view, it may cause confusion during the interpretation of the final images. In this case, the user should also look at the original SES for reference.

7. Summary and Future Work

We presented Molecular Sombreros, a novel abstract visualization for protein data that provides a visually simplified approach for rendering of molecular binding sites residing inside protein tunnels. This is achieved by morphing the geometry of the respective tunnel and its surroundings into a hat-like shape, which encodes relevant parameters of the cavity for visual analysis. To do so, the results of CAVER and a modified version of the 3V algorithm are combined to determine the mesh-based geometry corresponding to the tunnel. The morphing itself is then achieved by distributing polar coordinate values on the geometry, which then determine the final position of the molecular surface mesh vertices so that it resembles a hat.

Our method was applied to several real-world enzyme data sets. We demonstrated that all coloring modes typically used for molecular surface visualizations are also applicable to our visual abstraction method. This allows for the analysis of important physicochemical attributes using our representation, if it is possible to encode them into color. Our method is especially useful for the comparison of the binding sites of different proteins, as it encodes several characteristic values of the access channel to the proteins' active site. Using other methods, these values are not always easy to compare visually due to occlusion or complex tunnel structures. By neglecting the length of the tunnel, it is also possible to generate a real two-dimensional representation which can be useful in cases where the viewer is not able to interact with the visualization.

In the future, we aim to include the ligand that can bind to the visualized binding site explicitly. This should enable the analysis of the path of the ligand relative to the tunnel and especially the general approach direction that can be projected onto the brim. Since our method can distort the space inside the sombrero heavily under some circumstances (see subsection 6.3), a projection method to the new system will be required. Such a projection would also allow for other visualization possibilities, such as ligand density

maps accumulated over time. This would support the analysis of the reaction rates of the visualized protein. Another improvement would be to adapt the base geometry better to the ligand compatible to the binding site, since the SES only offers an approximation to the true shape of the ligands. The Ligand Excluded Surface (LES) developed by Lindow et al. [LBH14] uses the actual ligand geometry instead of a probe sphere to determine the surface that is actually accessible to the respective ligand. It would be interesting to investigate the difference between a sombrero based on the SES and one based on the LES. For the 3V-like cavity detection however, the SES would still be required to provide the outer geometry, since the LES does not really support operations like increasing the radius of the probe sphere. Finally, a more direct representation of the surrounding geometry would be helpful. A slightly tilted brim, based on the general orientation of the original surface might help to convey some of the information lost during the morphing process. Another possibility would be to use half-ellipsoids instead of half spheroids to depict tunnels with uneven measurements.

Acknowledgments

The authors want to thank M. Sanner for providing his *MSMS* software and B. Kozlíková and colleagues for providing the *CAVER Analyst* software package. This work was partially funded by Deutsche Forschungsgemeinschaft (DFG) as part of SFB 716 and PROLINT.

References

- [AAM*17] ALHARBI N., ALHARBI M., MARTINEZ X., KRONE M., ROSE A., BAA DEN M., LARAMEE R. S., CHAVENT M.: Molecular visualization of computational biology data: A survey of surveys. In *EuroVis Short Papers* (2017), vol. 1, pp. 133–137. 2
- [BJG*15] BYŠKA J., JURČÍK A., GRÖLLER M. E., VIOLA I., KOZLÍKOVÁ B.: MoleCollar and tunnel heat map visualizations for conveying spatio-temporo-chemical properties across and along protein voids. *Comput. Graph. Forum* 34, 3 (2015), 1–10. 3
- [Bli82] BLINN J. F.: A generalization of algebraic surface drawing. *ACM Trans. Graph.* 1, 3 (1982), 235–256. 3
- [BLMG*16] BYŠKA J., LE MUZIC M., GRÖLLER M. E., VIOLA I., KOZLÍKOVÁ B.: AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE Trans. Vis. Comput. Graphics* 22, 1 (2016), 747–756. 3
- [BWF*00] BERMAN H. M., WESTBROOK J., FENG Z., GILLILAND G., BHAT T. N., WEISSIG H., SHINDYALOV I. N., BOURNE P. E.: The Protein Data Bank. *Nucleic Acids Research* 28, 1 (2000), 235–242. 9, 11
- [GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309. 9
- [GSZ*13] GURIJALA K. C., SHI R., ZENG W., GU X., KAUFMAN A.: Colon flattening using heat diffusion riemannian metric. *IEEE Trans. Vis. Comput. Graphics* 19, 12 (Dec 2013), 2848–2857. 4
- [HHG07] HAMBURGER K., HANSEN T., GEGENFURTNER K. R.: Geometric-optical illusions at isoluminance. *Vision Research* 47, 26 (2007), 3276–3285. 7
- [HK14] HASS J., KOEHL P.: How round is a protein? exploring protein structures for globularity using conformal mapping. *Mathematics of Biomolecules* 1 (2014), 26. 4
- [JBSK15] JURČÍK A., BYŠKA J., SOCHOR J., KOZLÍKOVÁ B.: Visibility-based approach to surface detection of tunnels in proteins. In *Proc. 31st Spring Conference on Computer Graphics* (2015), pp. 65–72. 2
- [Kab76] KABSCH W.: A solution for the best rotation to relate two sets of vectors. *Acta Cryst. Section A* 32, 5 (1976), 922–923. 9
- [KBE09] KRONE M., BIDMON K., ERTL T.: Interactive visualization of molecular surface dynamics. *IEEE Trans. Vis. Comput. Graphics* 15, 6 (2009), 1391–1398. 2
- [KBP*16] KOLESÁR I., BYŠKA J., PARULEK J., HAUSER H., KOZLÍKOVÁ B.: Unfolding and interactive exploration of protein tunnels and their dynamics. In *Proc. EG Workshop on Visual Computing for Biology and Medicine* (2016), pp. 1–10. 3, 8
- [KFS*17] KRONE M., FRIESS F., SCHARNOWSKI K., REINA G., FADEMRECHT S., KULSCHEWSKI T., PLEISS J., ERTL T.: Molecular surface maps. *IEEE Trans. Vis. Comput. Graphics* 23, 1 (2017), 701–710. 4, 8
- [KFW*02] KANITSAR A., FLEISCHMANN D., WEGENKITTL R., FELKEL P., GRÖLLER M. E.: CPR: Curved Planar Reformation. In *Proceedings of the Conference on Visualization* (2002), IEEE, pp. 37–44. 4
- [KKF*16] KOZLÍKOVÁ B., KRONE M., FALK M., LINDOW N., BAA DEN M., BAUM D., VIOLA I., PARULEK J., HEGE H.-C.: Visualization of biomolecular structures: State of the art revisited. *Comput. Graph. Forum* 36, 8 (2016), 178–204. 2
- [KKL*16] KRONE M., KOZLÍKOVÁ B., LINDOW N., BAA DEN M., BAUM D., PARULEK J., HEGE H.-C., VIOLA I.: Visual analysis of biomolecular cavities: State of the art. *Comput. Graph. Forum* 35, 3 (2016), 527–551. 2, 3
- [KSES12] KRONE M., STONE J. E., ERTL T., SCHULTEN K.: Fast visualization of gaussian density surfaces for molecular dynamics and particle system trajectories. In *EuroVis Short Papers* (2012), pp. 67–71. 3
- [KvV*14] KOZLÍKOVÁ B., ŠEBESTOVÁ E., ŠUSTR V., BREZOVSKÝ J., STRNAD O., DANIEL L., BEDNÁŘ D., PAVELKA A., MAŇÁK M., BEZDĚKA M., BENEŠ P., KOTRY M., GORA A. W., DAMBORSKÝ J., SOCHOR J.: CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* (2014). 1, 4
- [KZ10] KUZMANIC A., ZAGROVIC B.: Determination of ensemble-average pairwise root-mean square deviation from experimental B-factors. *Biophys. J.* 98, 5 (2010), 861–871. 8
- [LBBH13] LINDOW N., BAUM D., BONDAR A.-N., HEGE H.-C.: Exploring cavity dynamics in biomolecular systems. *BMC Bioinformatics* 14, Suppl 19 (2013), S5. 3
- [LBH11] LINDOW N., BAUM D., HEGE H.-C.: Voronoi-based extraction and visualization of molecular paths. *IEEE Trans. Vis. Comput. Graphics* 17, 12 (2011), 2025–2034. 3
- [LBH14] LINDOW N., BAUM D., HEGE H.-C.: Ligand excluded surface: A new type of molecular surface. *IEEE Trans. Vis. Comput. Graphics* 20, 12 (2014), 2486–2495. 12
- [LBPH10] LINDOW N., BAUM D., PROHASKA S., HEGE H.-C.: Accelerated visualization of dynamic molecular surfaces. *Computer Graphics Forum* 29, 3 (2010), 943–952. 2
- [MK16] MARINO J., KAUFMAN A.: Planar visualization of treelike structures. *IEEE Trans. Vis. Comput. Graphics* 22, 1 (2016), 906–915. 4
- [MKR17] MALZAHN J., KOZLÍKOVÁ B., ROPINSKI T.: Protein tunnel reprojection for physico-chemical property analysis. In *Proc. EG Workshop on Visual Computing for Biology and Medicine* (2017). 3
- [MSCN15] MASOOD T. B., SANDHYA S., CHANDRA N., NATARAJAN V.: CHEXVIS: a tool for molecular channel extraction and visualization. *BMC Bioinformatics* 16, 1 (2015), 119. 3
- [PBV*14] PRAVDA L., BERKA K., VAŘEKOVÁ R. S., SEHNAL D., BANÁŠ P., LASKOWSKI R. A., KOČA J., OTYEPKA M.: Anatomy of enzyme channels. *BMC Bioinformatics* 15, 1 (2014), 379. 2

- [PFS98] PLEISS J., FISCHER M., SCHMID R. D.: Anatomy of lipase binding sites: the scissile fatty acid binding site. *Chem. Phys. Lipids* 93, 1 (1998), 67–80. 1
- [PGB*12] PROKOP Z., GORA A., BREZOVSKÝ J., CHALOUPKOVÁ R., ŠTĚPÁNKOVÁ V., DAMBORSKÝ J.: Engineering of protein tunnels: Keyhole-lock-key model for catalysis by the enzymes with buried active sites. In *Protein Engineering Handbook*, Lutz S., Bornscheuer U., (Eds.). Wiley-VCH, 2012, pp. 421–464. 2
- [PTRV13] PARULEK J., TURKAY C., REUTER N., VIOLA I.: Visual cavity analysis in molecular simulations. *BMC Bioinformatics* 14, Suppl 19 (2013), S4. 3
- [PvK*15] PAVELKA A., ŠEBESTOVÁ E., KOZLÍKOVÁ B., BREZOVSKÝ J., SOCHOR J., DAMBORSKÝ J.: CAVER: Algorithms for analyzing dynamics of tunnels in macromolecules. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 13, 3 (2015), 505–517. 3, 4
- [Ric77] RICHARDS F. M.: Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 6, 1 (1977), 151–176. 2
- [Ric81] RICHARDSON J. S.: The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* 34 (1981), 167–339. 10
- [RS07] RAHI S. J., SHARP K.: Mapping complicated surfaces onto a sphere. *Int. J. Comput. Geom. Appl.* 17, 04 (2007), 305–329. 7
- [Shn96] SHNEIDERMAN B.: The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), pp. 336–343. 8
- [SKPE19] SCHATZ K., KRONE M., PLEISS J., ERTL T.: Interactive visualization of biomolecules' dynamic and complex properties. *The European Physical Journal Special Topics* 227, 14 (2019), 1725–1739. 2
- [SKR*14] SCHARNOWSKI K., KRONE M., REINA G., KULSCHEWSKI T., PLEISS J., ERTL T.: Comparative visualization of molecular surfaces using deformable models. *Comput. Graph. Forum* 33, 3 (2014), 191–200. 4
- [SOS96] SANNER M. F., OLSON A. J., SPEHNER J.-C.: Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 38, 3 (1996), 305–320. 2, 4
- [VG10] VOSS N. R., GERSTEIN M.: 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Research* 38, suppl 2 (2010), W555–W562. 3, 5