Faculty of Informatics
Masaryk University

Analysis and Visualization of Biomolecules

Habilitation Thesis
(Collection of Articles)

Barbora Kozlíková

September 2016
Brno, Czech Republic

Abstract

All living organisms consist of biomolecules and understanding the complex processes and interactions between them helps also to understand the reasons of their deficiencies and malfunctions. Therefore, in many fields the researchers are striving for such understanding. By analyzing the molecular structure, observing its behavior over time and its reactivity with other molecules they are trying to design new molecules with desired properties and function. Such analysis was traditionally conducted by in-vitro experiments. However, in the last decades this costly and lengthy procedure started to be supported by newly emerging in-silico predictive methods. These methods help to explore the molecular structure and its properties prior to the laboratory experiments and thus significantly reduce the number of these experiments. In order to follow this trend, in our research we are focusing on the design and development of such methods, focusing namely on the detection of void paths leading from the molecular environment to its reactive site. These paths, called tunnels, play a crucial role in the reactivity of a protein with other small molecules penetrating to its reactive site. The importance of tunnels has been confirmed by numerous research studies namely in the fields of protein engineering and drug design. The algorithms used for tunnel detection are mostly based on the principles of computational geometry. An indispensable part of our research focuses also on the visualization and visual exploration of these tunnels.

In this thesis I aim to provide the readers with a comprehensive overview of our research results in the field of molecular analysis and visualization in the last decade. The thesis is structured as a collection of relevant papers accompanied with a commentary putting our contributions in the context of the state-of-the-art in the area and summarizing our achievements. The thesis consists of two main parts. In the first part I focus on the description of the algorithms we designed for the detection of tunnels and other types of void space inside protein structures and also our approaches to the computation of ligand transportation through a tunnel. The second part contains our achievements in the field of molecular visualization, namely focusing on visualization and visual analysis of tunnels and exploring the ligand trajectory. As these parts are tightly related, some of the corresponding papers are touching both fields. In such cases we explicitly mention this overlap in the text.

Keywords: Biomolecule, protein, analysis, visualization, visual analysis, tunnel.

Abstrakt

Všechny živé organismy jsou tvořeny molekulami a tudíž správné pochopení jejich vnitřních procesů a vzájemných interakcí vede rovněž k porozumění jejich dysfunkcí. Je tedy přirozené, že se výzkum v mnoha oblastech zaměřuje právě na snahu pochopit složité funkce biomolekul. Pomocí analýzy struktury molekul, jejich chování v čase a zkoumání reaktivity s jinými molekulami se vědci snaží navrhnout nové molekuly či změnit odpovídajícím způsobem strukturu těch stávajících tak, aby vykazovaly požadované vlastnosti a funkce. Tradičně se tato analýza provádí experimentálně přímo v laboratořích, což je však velmi finančně i časově náročný proces. Proto se v posledních desítkách let objevila řada algoritmů a metod, které se snaží celý proces analýzy zjednodušit zredukováním počtu nutných laboratorních experimentů. V našem dlouholetém výzkumu se zabýváme problémem výpočetní analýzy proteinových struktur a vyvíjíme vlastní metody pro detekci volných cest uvnitř molekul, které vedou z vnějšího prostředí molekuly do jejího reakčního místa. Tyto volné cesty, nazývané též tunely, mají zásadní vliv na reaktivitu dané molekuly s dalšími malými molekulami, které se prostřednictvím těchto tunelů dostávají právě do jejího reakčního místa. Významnost tunelů v proteinových strukturách byla potvrzena v řadě vědeckých studií, zejména v oblasti proteinového inženýrství a návrhu nových léčiv. Algoritmy zaměřující se na detekci tunelů jsou založeny na principech výpočetní geometrie a nezbytnou součástí celého procesu je následná vizualizace a vizuální prozkoumávání spočtených tunelů.

Tato práce si klade za cíl seznámit čtenáře s uceleným přehledem našich výsledků v oblasti molekulární analýzy a vizualizace, jichž jsme za poslední desetiletí našeho působení v této oblasti dosáhli. Práce je strukturována jako kolekce relevantních vědeckých publikací doplněných o úvodní komentář, ve kterém zasazuji naše výsledky do kontextu aktuálního stavu v dané oblasti a sumarizuji dosažené cíle. Práce je rozdělena na dvě hlavní části. V první části se zaměřuji na popis algoritmů, které jsme navrhli pro detekci tunelů a dalších typů volného prostoru uvnitř proteinů. Dále tato část obsahuje dva přístupy pro plánování průchodu ligandu tunelem. Druhá část práce obsahuje naše výsledky v oblasti molekulární vizualizace, zejména pak vizualizace a vizuální analýzy tunelů, včetně průzkumu komplexní trajektorie ligandu uvnitř proteinu. Obě části práce jsou vzájemně úzce propojeny, tudíž i některé uvedené publikace se dotýkají obou uvedených oblastí. V těchto případech tento překryv explicitně uvádím v textu.

Klíčová slova: Biomolekula, protein, analýza, vizualizace, vizuální analýza, tunel.

Acknowledgements

I would like to express my appreciation to all the mentors I have had along my journey—to Jiří Sochor for guiding me during my Ph.D. studies, to Ivan Viola and Eduard M. Gröller for helping me to dive into the amazing field of visualization and to grow as a scientist, and to all my colleagues, co-authors, and all amazing people from the Human-Computer Interaction Laboratory at the Masaryk University, from the Institute of Computer Graphics and Algorithms at TU Wien, from the Visualization Group at the University of Bergen, and from Loschmidt Laboratories at the Masaryk University. You are my inspiration.

Finally and foremost, I wish to thank my family and close friends for their support, understanding, and patience.

Barbora Kozlíková

Contents

Part I

Commentary

Chapter 1

Introduction

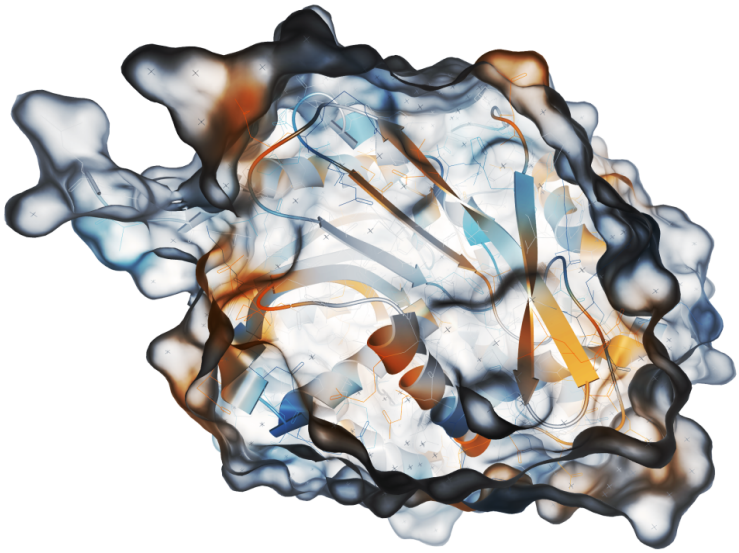
Proteins are biomolecules responsible for driving the machinery of life. Proper understanding of their structure, behavior, and function helps to reveal the fundamentals of biochemical processes taking place in living cells. The gained knowledge subsequently helps to design new protein structures and other chemical substances with desired properties and function.

One of the largest fields of biochemical research focusing on studying protein structures is protein engineering. Here the main task is to understand how the protein amino acids influence its behavior. By subsequent proposing of mutations of selected amino acids they are able to alter the protein function. The function is tightly related to protein's ability to react with other molecules. In protein engineering, these reactions undergo often deeply inside the protein structure where a small molecule has to be transported from the outer environment. In other words, the reaction between a protein and a small molecule, called substrate or ligand, can be performed only when there exists a void path connecting the outer environment with the protein reaction site, called the active site. The existence of such void path, called tunnel, is therefore crucial. There are numerous studies confirming the importance of tunnels. Koudeláková et al. [25] presented a research study revealing that the mutations of amino acids along the main protein tunnel lead to a substantial increase of desired protein properties, whereas mutations close to the protein surface have almost no effect.

In the last decade several research groups have been focusing on the problem of detection of tunnels and other types of void space inside proteins. Their efforts led to the development of several computational tools for protein analysis. A comprehensible overview of them was published by Damborský and Brezovský [10] or more recently also by García-Guevara et al. [14].

Recent studies confirm that the protein function is not influenced purely by its structure but also by its dynamic behavior [16]. Therefore, simulations of molecular dynamics are nowadays playing a crucial role in protein analysis. Protein movements highly influence the also the shape and openness of its tunnels. It naturally leads to the substantial increase

of demands for computational power because the size of simulations can currently reach hundreds of thousands of time steps. The results of such analysis of molecular dynamics then contain an enormous amount of data which should be explored and understood. Therefore, proper visualization and visual analysis of these results are crucial and open huge possibilities for research in these fields as well. Only a comprehensible and intuitive representation can help the biochemists to interpret the computational results correctly so many researchers in visualization and visual analysis currently tackle this challenge.



1.1 Focus of the thesis

This thesis summarizes my contributions to the progress within the field of protein analysis and visualization. Along with the cooperation partners focusing on protein engineering, computational geometry, visualization, and visual analysis, we reached within the last decade several achievements contributing to better understanding of protein structure and function.

In the text I first focus on the problem of protein analysis, namely with respect to detection of tunnels and other types of void space inside proteins. Second, I present the related field of molecular visualization where we contributed to both visualization of molecules as well as tunnels. Prior to our contributions I introduce the reader to the state-of-the-art techniques in the respective field. This thesis can be taken as an extension and follow up of my doctoral thesis [26] which presented our early results in visualization of proteins and their tunnels. Since that we reached a substantial progress in analysis, visualization, as well as visual analysis of protein tunnels which resulted in several journal publications and two software tools widely used by the international biochemical community.

1.2 Thesis structure

The thesis is structured with respect to my contributions to protein analysis and visualization. Therefore, it contains two main sections. Each section starts with the presentation of the state-of-the-art within the particular domain, followed by my contributions to its progress and the list of selected articles I have co-authored and that are attached to this text to exemplify my contributions. First section focuses on the problem of analysis of protein structures with respect to its reactivity with a ligand. It contains several contributions to the detection of tunnels and other types of void space inside proteins as well as our methods for analysis of the ligand transportation to the protein active site.- In the second section I present several approaches to visualization and visual analysis of biomolecules and their tunnels, along with our new method for exploration of ligand trajectory within a tunnel and our software tools for molecular analysis and visualization.

As both areas are tightly connected, some of the presented articles contain results both in analysis and visualization. In such cases I explicitly mention and explain this overlap.

Finally, the overall collection of articles is listed in Part II of this thesis.

Chapter 2

Analysis of Biomolecules

The aim of this chapter is to provide the readers with a brief introduction to protein analysis, to present the state-of-the-art in this area, and to summarize my contribution to it.

In the field of protein analysis there are several challenges regarding the exploration of protein void space. According to the properties of such a space, it can be classified into several categories. Figure 2.1 illustrates this classification.

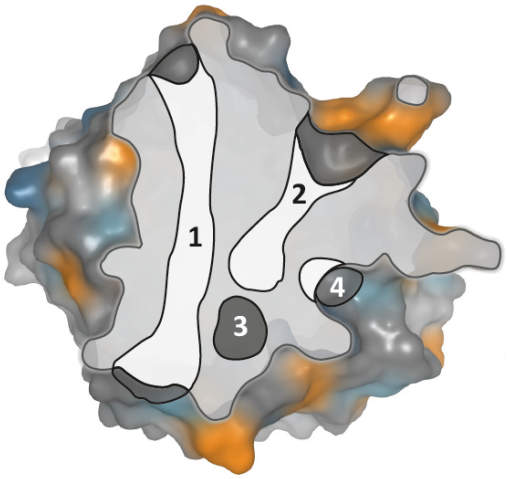


Figure 2.1: Illustration of different types of void space in protein structure: 1 – channel or pore, 2 – tunnel, 3 – inner cavity, 4 – pocket.

Although the terminology differs in literature, the biochemists basically distinguish between four types of void space. In this thesis I am using the following notation. Channels or pores are paths present often in transmembrane proteins which traverse through the whole protein structure and connect two sites on the protein surface. They can serve as transportation paths for ligands, water molecules, or other chemical compounds. Inner

cavities are buried deeply in the structure and can contain the protein active site. They are connected with the outer solvent via tunnels which play a crucial role in the reactivity of the protein with ligands traveling from the outer solvent to the active site. In some cases the chemical reactions can undergo also on the protein surface, in pockets. These shallow clefts are also able to adopt a small ligand molecule and to react with it under specific conditions.

In some resources the void space present in the molecule is denoted as cavity in general. Then the abovementioned inner cavities are marked as closed cavities, tunnels and pockets as single-entry open cavities, and channels and pores as multiple-entry open cavities. This classification is used also in one of our contributions in the area of analysis of biomolecules which is a state-of-the-art report [33] summarizing and explaining the basic principles of different algorithms designed to solve this problem.

According to our contributions to this area, this chapter is divided into three sections. First section briefly introduces different approaches to protein analysis and presents our algorithms for the detection of inner cavities and channels connecting more active sites in protein structures. In the second section I focus on the detection of protein tunnels which have been in the scope of our interdisciplinary research from its very beginning. Here I present also our significant contributions to this area. Third section presents the consequential problem of ligand transportation through a detected tunnel. This can finally confirm the compatibility between the ligand and tunnel and the possibility to transport the ligand into the protein active site.

Figure 2.2 illustrates the structure of this chapter along with references to our corresponding contributions.

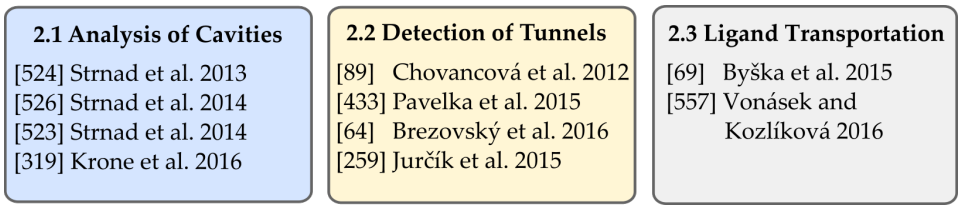


Figure 2.2: Structure of Chapter 2 with corresponding contribution papers.

2.1 Analysis of Cavities

As mentioned above, we can distinguish between different types of protein cavities based on their topological properties. Although their biological function is different, they do not differ significantly from the computational point of view. Almost all algorithms for the detection of cavities simplify the molecule by the hard sphere model. So the extraction of cavities can be described as a geometry processing problem. Therefore, the results of the analysis are highly dependent on the employed computational method. Historically the existing

approaches can be clustered into four main categories. These are approaches based on grids, Voronoi diagrams, molecular surfaces, and (usually spherical) probes. Some algorithms use also a combination of these to reach better results.

First algorithms for the detection of cavities were based on grids, due to their simplicity [36, 45]. However, at that time the algorithms were suitable only for small or medium size molecules (few thousands of atoms), namely because of the hardware limitations. One solution to this problem was to use Voronoi diagrams which proved to be the most suitable solution for the detection of paths in molecules [41, 44, 59, 7, 48]. Currently there are many existing solutions which are comprehensibly summarized and presented in our state-of-the-art report [33]. Therefore, I refer the interested readers to this report.

Contributions

Over the last years, I have contributed to the design and development of several algorithms for the detection of cavities. Here I will summarize all contributions except for those related to the computation of tunnels because this large area deserves its own section.

With my colleagues, in 2013 and 2014 we focused on the development of several algorithms for the detection of specific cavities inside proteins which were also tightly related to their visual representation and interactive exploration.

First, we published a paper presenting our approach to real-time visualization and exploration of protein empty space with varying parameters [53]. In this work we designed and implemented an algorithm for the detection of inner cavities, i.e., cavities without direct access to the molecular surface. The biggest contribution of this approach lies in the ability to adapt the computation parameters in real-time. It means that the user can anytime change the input parameters and the cavities are instantly recomputed and visualized. The algorithm also automatically marks the largest detected void as there is a high probability that it contains the active site.

The following two publications focused on geometrical detection of paths inside proteins leading among more user-defined binding sites. These algorithms take as an input a set of user-defined points inside the protein structure and calculate a transportation path of specific properties (e.g., minimal width) connecting these points in a given order. In the first publication [54] we introduce the principle of our algorithm which is based on localized Voronoi diagrams and Delaunay triangulation. Thanks to localization, our approach is applicable to proteins of all sizes because in each phase it uses only a subset of all atoms. The algorithm was primarily designed for protein structures but can be applicable to an arbitrary set of spheres in the 3D space. This concept was used also in our following paper [52] which introduced the algorithm for the detection of so called intramolecular tunnels in proteins. This time the proposed solution utilizes the CAVER 3.0 algorithm for computation of tunnels [6] which will be introduced in the following section. The idea of

this approach is the following. Using CAVER 3.0 we calculate tunnels connecting pairs of neighboring user-defined points of interest through which the intramolecular tunnel should lead. The first and last point are connected with the molecular surface. From these partial paths the resulting intramolecular tunnel is reconstructed. The benefit of this solution lies namely in the utilization of the thoroughly designed CAVER 3.0 algorithm which was tested on many real scenarios confirming its biochemical relevance.

To the category of contributions in the analysis of cavities belongs also our latest state-of-the-art report about visual analysis of biomolecular cavities [33]. This report is a product of cooperation between experts in the field of molecular analysis and visualization. It aims to summarize the existing approaches to the detection of cavities in biomolecules and to introduce a comprehensible taxonomy to distinguish between them. It should serve not only the biochemists to understand the basic principles of these approaches and to select the most suitable one for their purpose, but also experts in the visualization field. The first part introduces a classification of the existing approaches according to the algorithm they use. Then, for each category we provide the readers with the explanation of its basic principle and a list of available approaches. The second part focuses on the visualization and visual analysis of detected cavities. It presents techniques for spatial and non-spatial visualization of cavities as well as enhanced visualization and visual analysis of molecular shape. The report also contains a table summarizing the software tools which are currently available to the biochemical community or which significantly contributed to the field.

Articles in Collection

[53] O. Strnad, B. Kozlíková, V. Šustr, and J. Sochor. Real-time visualization and exploration of protein empty space with varying parameters. *International Journal on Advances in Life Sciences*, 5(3 & 4), 2013

I cooperated on the design of the algorithm and wrote the main part of the text. Contribution 35%.

[54] O. Strnad, V. Šustr, B. Kozlíková, and J. Sochor. Geometrical detection of pathways in protein structures leading among more binding sites. In H. H. Ali, editor, *BIOTECHNO 2014 : The Sixth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies*, pages 93–98. IARIA XPS Press, 2014

I cooperated on the design of the algorithm and was responsible for the text. Contribution 30%.

[52] O. Strnad, B. Kozlíková, and J. Sochor. Detection of intramolecular tunnels connecting sequence of sites in protein structures. In J. Saez-Rodriguez, editor, *Advances in Intelligent Systems and Computing*, pages 73–80. Springer International Publishing, 2014

I participated on the design of the algorithm and paper writing. Contribution 40%.

[33] M. Krone, B. Kozlíková, N. Lindow, M. Baaden, D. Baum, J. Parulek, H.-C. Hege, and I. Viola. Visual analysis of biomolecular cavities: State of the art. *Computer Graphics Forum*, 35(3):527–551, 2016

I coordinated the writing process, wrote several sections of the article, and was responsible for final corrections. Contribution 25%.

2.2 Detection of Tunnels

This section presents our largest contribution to the field of protein analysis because the problem of detection of tunnels has been in our scope from the beginning of the collaboration with the biochemists from the Loschmidt Laboratories at the Masaryk University in 2005. Within the following years, several approaches to tunnel detection have been proposed. The earliest solution published by Petřek et al. [45] was based on a simple grid-based approach. The main advantage of grid-based methods is that they usually require only simple data structures. However, their geometrical accuracy as well as the computational time and memory requirements are highly dependent on the grid resolution. Therefore, the researchers were searching for more robust solution and found it in Voronoi diagrams [41, 44, 59]. This solution overcomes the problems of grid-based algorithm and the edges of Voronoi diagrams automatically provide geometrically optimal paths (tunnels) based on given restrictions (e.g., tunnel minimal width). These restrictions can be taken as input parameters for the Dijkstra's graph search algorithm which detects the "cheapest" path leading from a given starting point to the protein outer environment. Basic Voronoi diagrams are designed for points and do not take into account different radii of atoms representing chemical elements. This was improved by introducing an approximation of larger spheres by a set of spheres with constant radii by Yaffe et al. [59]. This idea was also adopted by our CAVER 3.0 approach [6]. Alternative solution is to use the additively weighted Voronoi diagrams [39].

The first Voronoi-based solutions were designed for static molecular structures. However, with the possibility to simulate movements of proteins also the necessity to observe the movements of tunnels emerged. Therefore, some of the existing solutions are applicable to molecular dynamics as well [59, 6, 39].

As for the representation of the detected tunnels, the first solutions represented a tunnel by its centerline and a set of spheres whose centers were positioned on this centerline and whose radius was maximal with respect to the surrounding atoms (contained the maximal free space). This solution was very simple but mostly did not capture the shape of the void space around the centerline correctly (see Figure 2.3).

Therefore, we came with the definition of so called asymmetric tunnels [2, 20]. The idea is based on the detection of amino acids along the centerline which form the size and shape of the tunnel and to use them for the more precise tunnel representation.

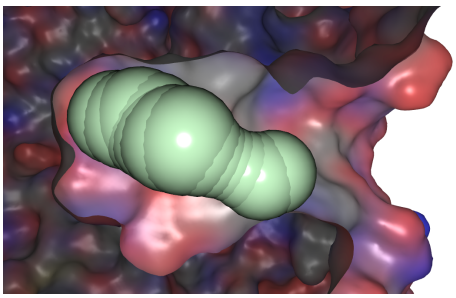


Figure 2.3: Protein tunnel represented as a sequence of intersecting spheres which does not cover the entire void space along the tunnel centerline (taken from [20]).

Contributions

Our largest contribution to the field of tunnel detection in proteins is surely the CAVER 3.0 algorithm presented in the PLoS Computational Biology journal in 2012 [6]. Its main benefit lies in the ability to detect tunnels in molecular dynamics simulations and track their behavior. Therefore, the biochemists have the information about the tunnel stability over time which was impossible to get from a static protein. Along with the publication we also released the CAVER 3.0 software tool (in a standalone form and as a PyMOL plugin). The relevance and usefulness of this tool was confirmed by numerous citations in top biochemical and biological journals presenting real applications of CAVER 3.0.

As the PLoS publication did not contain details about the algorithm we recently published its thorough description in the IEEE Transactions on Computational Biology and Bioinformatics journal [43]. Additionally, the paper presents the improved clustering solution for finding tunnels which is included in the CAVER 3.02 release of the software (available at *www.caver.cz*).

Most recently we also summarized the information about our released CAVER product as well as a detailed computational protocol employing the CAVER software [1] into a book chapter which is currently in print and will appear in the Springer series of Methods in Molecular Biology.

Another contribution is tightly related to the asymmetric tunnels capturing more realistic shape of tunnels. Our visibility-based approach [20] is an extension of the paper by Byška et al. [2]. Here we overcome the main limitation of the previous approach – its dependence on the user-defined parameter which substantially influences the resulting shape and volume of the detected asymmetric tunnel. As the setting of this parameter requires non-trivial knowledge from the domain experts we aimed to set it automatically. The second benefit of this paper is the introduction of a novel approach to the detection of amino acids lining the tunnel centerline and thus determining the tunnel boundary. The approach is based on the visibility of amino acids from the tunnel centerline when we project each atom onto each tunnel sphere and detect the overlaps.

Articles in Collection

[6] E. Chovancová, A. Pavelka, P. Beneš, O. Strnad, J. Brezovský, B. Kozlíková, A. W. Gora, V. Šustr, M. Klvaňa, P. Medek, L. Biedermannová, J. Sochor, and J. Damborský. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLOS Computational Biology*, 8(10), 2012

I participated on the design and implementation and proofreading of the manuscript. Contribution 10%.

[43] A. Pavelka, E. Šebestová, B. Kozlíková, J. Brezovský, J. Sochor, and J. Damborský. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE ACM T. Comput. Bi.*, 2015

I contributed namely to the paper writing. Contribution 30%.

[1] J. Brezovský, B. Kozlíková, and J. Damborský. Computational analysis of protein tunnels and channels. In *Methods in Molecular Biology*. Springer International Publishing, 2016

I participated on the paper writing and proofreading. Contribution 20%.

[20] A. Jurčík, J. Byška, J. Sochor, and B. Kozlíková. Visibility–based approach to surface detection of tunnels in proteins. In J. Jorge, L. P. Santos, and R. Ďurikovič, editors, *31th Proceedings of Spring Conference on Computer Graphics*, pages 85–92, Bratislava, Slovakia, 2015. Comenius University

I contributed to the algorithm design and namely was responsible for paper writing. Contribution 30%.

2.3 Ligand Transportation

The third section related to the analysis of cavities in proteins contains the algorithms based on the path-planning approach. Path-planning represents an alternative solution to the detection of tunnels as well as subsequent simulation of ligand transportation through these tunnels. In this case the task is to find a collision-free path for a robot between a start and an end position. In our case the start point is determined by the position of the protein active site and the end position is located anywhere on the protein surface. The robot is defined as ligand or its spherical approximation. When solving this problem in molecular dynamics, we are working with a high-dimensional configuration space (4D for the spherical and 7D for the non-spherical robot). As the exact solution is very time and memory consuming, there are randomized sampling solutions [35].

This field is still very challenging so there are only few methods which take the geometry and dynamics of a ligand into account. One of them is the method by Cortés et al. [8] which

is using the Rapidly-exploring Random Trees [34]. The authors further improved this first solution and presented it as a part of the BioMove3D package [9].

Contributions

Until now we contributed to the analysis of tunnels utilizing path-planning approach with two publications. In the first publication [3] we present a simple and fast path-planning algorithm for transportation of a tightly connected sphere objects (representing a ligand) through a narrow gap. This should simulate the passage of a non-spherical ligand through a bottleneck which is the narrowest part of the tunnel. In comparison with other methods our approach significantly reduces the number of generated samples and the cost of path planning. To reach this, introduce our novel approach to the so called Local Planning Method which was designed in tight cooperation with the Department of Computer Science and Engineering at the University of West Bohemia.

The second paper [58] presents the first results of our newly started cooperation with the Intelligent and Mobile Robotics Group at the Czech Technical University in Prague. It proposes a novel method for tunnel detection in dynamic proteins which is based on Rapidly-exploring Random Trees. It is based on the idea of building a single configuration tree describing the void space inside protein. When searching for tunnels in molecular dynamics, this configuration tree is pruned accordingly. To verify the correctness of the solution we present also the comparison of our results with tunnels detected by CAVER 3.0. Currently the solution enables to find the tunnels for a spherical approximation of a ligand. However, in the future we plan to extend the solution to non-spherical ligands. In the easier variant, the non-spherical ligand will be rigid but the final solution should take into account also the ligand movements.

Articles in Collection

[3] J. Byška, I. Kolingerová, B. Kozlíková, and J. Sochor. Path–planning algorithm for transportation of molecules through protein tunnel bottlenecks. In J. Jorge, L. P. Santos, and R. Ďurikovič, editors, *31th Proceedings of Spring Conference on Computer Graphics*, pages 101–108, Bratislava, Slovakia, 2015. Comenius University

I participated on the design of the approach, paper writing, and proofreading. Contribution 20%.

[58] V. Vonásek and B. Kozlíková. Application of sampling-based path planning for tunnel detection in dynamic protein structures. In *MMAR: 21st International Conference on Methods and Models in Automation and Robotics*, Miedzyzdroje, Poland, 2016

I contributed to the design, paper writing, and proofreading. Contribution 40%.

Chapter 3

Visualization and Visual Analysis of Biomolecules

Visualization and visual analysis enabling the exploration of computed tunnels is the irreplaceable part of our research. Using a proper visual representation the domain experts are able to explore the tunnels and their properties and evaluate their biochemical relevance. Additionally, when dealing with long simulations of molecular dynamics, currently consisting of hundreds of thousands of time steps, it becomes crucial to provide the biochemists with dedicated visualizations conveying the required information about tunnels in a fast and comprehensible way. This involves various abstracted and integrated representations and visual analysis tools which link different views onto the data. As already mentioned in Chapter 2.2, we recently published the state-of-the-art report [33] about methods for the detection of cavities and their visualization and visual analysis.

The necessity of proper visualizations and visual analysis tools holds also for biomolecules. Except for traditionally used visual representations of molecules utilizing different levels of abstraction also several specific visualizations appeared [42, 57]. The comprehensible overview of the existing molecular representations forms the content of our another state-of-the-art report [30] so the interested readers are referred to this publication.

This chapter summarizes our contributions to the fields of visualization and visual analysis of biomolecules and their tunnels. According to the nature of the contributions I decided to divide the chapter into four sections. In the first section I present five papers focusing on visualization of biomolecules. The second section presents our proposed approaches to visualization and visual analysis of tunnels. The third section contains first results of our latest research in the field of visualization of ligand path through a protein. Finally, in the fourth section I present our contributions to software tools for visualization of molecules with focus on tunnels.

Figure 3.1 presents an overview of the structure of this chapter along with corresponding papers presented as our contribution.

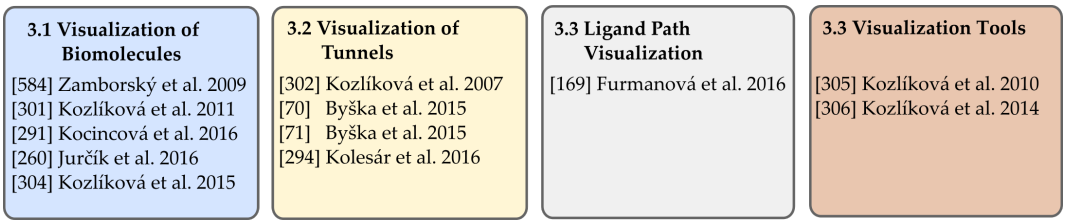


Figure 3.1: Structure of Chapter 3 with corresponding contribution papers.

3.1 Visualization of Biomolecules

Molecular visualization is one of the oldest branches of data visualization which developed the most during the past two decades. In that time several representations appeared and started to be frequently used by the domain experts. Within these belong molecular models with different levels of abstraction, spanning from models showing all atoms to highly abstracted representations of molecular chains (known as cartoon). Each of these models was designed for specific purposes – to get the overview of the molecular structure or to get deep insight into a specific part of the molecule. Therefore, these models are often combined to fulfill these requirements [57].

Among one of the most frequently used representations belongs the visualization of molecular surface which presents the interface between the molecule and outer solvent. The surface representation also enables to visually explore the gorges of potential tunnels or pockets. Therefore, several algorithms for the detection of different types of molecular surfaces have been already proposed. Figure 3.2 shows examples of different types of molecular surfaces.

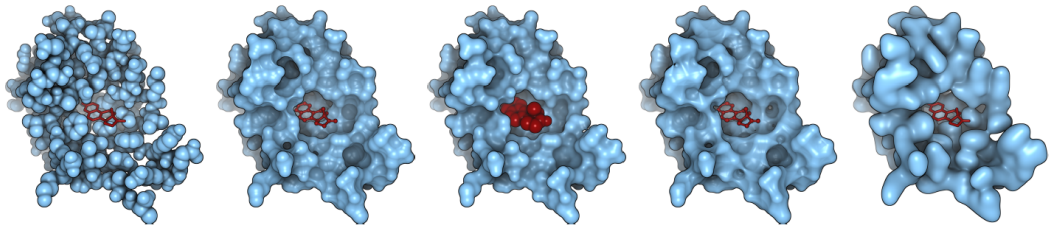


Figure 3.2: Different molecular surface representations (taken from [30]).

Another very popular representation is the visualization of secondary structures, known as cartoon or ribbon model (see Figure 3.3). This highly abstracted representation omits the visualization of individual atoms and uses helices and strands with arrows to convey the

information about the relationships (hydrogen bonds) between amino acids. This representation incorporates the algorithm for calculation of the type of secondary structure for each amino acid [56] and these structures are subsequently rendered. For large molecules this representation can contain enormous amount of triangles to be rendered, therefore several enhanced rendering techniques were proposed recently [17].

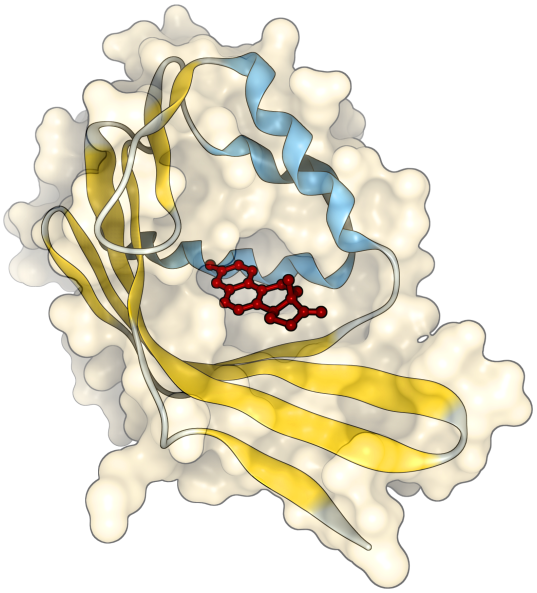


Figure 3.3: Cartoon representation depicting the protein secondary structures by helices (blue) and strands (yellow). Additionally, the image contains the transparent surface representation and ligand (red) visualized using the Balls & Sticks model (taken from [30]).

An integral part of molecular visualization is also the visual appearance and the quality of rendered images. Therefore, researchers are incorporating advanced real-time rendering and shading methods which enhance not only the image quality but also the perception of geometric shapes and depth complexity of the scene. One of the currently most preferred techniques adopted successfully to molecules is ambient occlusion [55, 15, 38]. These techniques can be further enhanced, such as in the case of Skånberg et al. [50] who proposed a combination of ambient occlusion and diffuse interreflections to highlight the entrance of the ligand to the active site located on the molecular surface.

Contributions

Already in the early stages of our interdisciplinary research in protein analysis we realized the gaps in the field of molecular visualization, namely in the visualization of secondary structures and surfaces. Therefore, we aimed to address them and the results were presented in two research papers published in 2009 and 2011. As these papers solved specific problems related also to the performance of computers at that time, the results are today rather

outdated and therefore should be taken as side contributions to the field. The first paper by Zamborský et al. [60] is focused on dynamic visualization of secondary structures. Here we solve the problem of real-time animation of movements of the cartoon representation. This is reached by animating the protein backbone onto which we position given objects representing parts of secondary structures (e.g., one turn of helix). These objects are then replicated as many times as necessary and are closely connected to form a solid secondary structure. High frame rates of the animation are reached by implementing the algorithm on GPU.

The second paper presents an algorithm for computation and visualization of molecular surface which is based on the state-of-the-art reduced surface approach by Sanner et al. [46]. Here we introduce several enhancements to the original approach which help to handle namely the singularities. These enhancements were designed with respect to the applicability of our modified solution to tunnels in proteins as well. The algorithm was also integrated to the CAVER Viewer application [31] which is presented in Section 3.4.

Most recently, we published also two contributions related to cartoon and surface representation of molecular structures. The first contribution will be presented at the IEEE BioVis conference in October 2016 and focuses on comparative visualization of protein secondary structures. We are facing the problem of comparison of more protein chains when the biochemists are aiming to explore differences and similarities in protein chains. This is reached by analyzing the sequences of amino acids and their structural or sequential alignment [18, 61, 49]. The resulting aligned structure can be traditionally visualized in 1D or 3D. But these representations suffer from several problems. The 1D representation lacks the information about the spatial arrangement of secondary structures of aligned proteins. The 3D spatial view provides this information but suffers from the occlusion problem when aligning more proteins. Therefore, we proposed a method to overcome these problems which is based on the 1D representation and equips it with the information about mutual spatial arrangement of corresponding secondary structures.

The second contribution is again related to the visualization of molecular surfaces. Here we propose an algorithm for accelerated visualization of transparent molecular surfaces as the transparency enables to get the insight into the structure. Our algorithm is implemented using GLSL in order to visualize the temporal changes of the surface in real-time. This is reached by our new accelerated rendering pipeline improving the performance of visualization of transparent surfaces proposed by Kauker et al. [22]. Our approach is applicable also to inner cavities.

My last contribution to the field of biomolecular visualization is the state-of-the-art report [30] presented at the EuroVis conference in 2015 which is currently under the review process in the Computer Graphics Forum journal. The report was compiled in cooperation with the top experts in visualization from seven universities. The report presents a comprehensive overview of techniques that have been developed in the field of molecular visualization. We proposed a taxonomy demonstrating the areas of molecular visualiza-

tion which have been already extensively investigated. Among others, our aim was also to provide the readers with an outlook on promising research topics to enable further success in advancing the knowledge about molecular structures.

Articles in Collection

[60] M. Zamborský, T. Szabó, and B. Kozlíková. Dynamic visualization of protein secondary structures. In *Proceedings of the 13th Central European Seminar on Computer Graphics (CESCG 2009)*, Vienna, 2009. Vienna University of Technology, Insitute of Computer Graphics and Algorithms

I contributed to the algorithm design and implementation and was responsible for the paper writing. Contribution 40%.

[27] B. Kozlíková, I. Aleksandrowicz, and J. Sochor. Computation and visualization of surface of proteins and their channels. In I. Press, editor, *The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2011*, pages 99–106. IADIS Press, 2011

I participated in the algorithm design and implementation and was responsible for the paper writing. Contribution 40%.

[23] L. Kocincová, M. Jarešová, J. Byška, J. Parulek, H. Hauser, and B. Kozlíková. Comparative visualization of protein secondary structures. In *BioVis*, 2016

I participated in the design of the algorithm and was responsible for paper writing. Contribution 40%.

[21] A. Jurčík, J. Parulek, J. Sochor, and B. Kozlíková. Accelerated Visualization of Transparent Molecular Surfaces in Molecular Dynamics. In *IEEE Pacific Visualization Symposium*, pages 112–119, 2016

I contributed to the design of the algorithm and significantly participated in the paper writing and proofreading. Contribution 30%.

[30] B. Kozlíková, M. Krone, N. Lindow, M. Falk, M. Baaden, D. Baum, I. Viola, J. Parulek, and H.-C. Hege. Visualization of Biomolecular Structures: State of the Art. In *Eurographics Conference on Visualization - STARs*, pages 61–81, 2015

I participated in the writing process, coordinated paper writing, wrote several sections of the article, and was responsible for final corrections. Contribution 25%.

3.2 Visualization of Tunnels

Except for the visualization of molecular structures, we focused extensively on the visualization and visual analysis of detected tunnels. Visualization of tunnels can be taken as a subfield of molecular visualization and to some extent also the methods are applicable to both areas. On the other hand, tunnels are specific structures and namely in molecular dynamics simulations they deserve their own methods as well.

The first visual representation of tunnels was rather simple. It depicts a tunnel as a set of overlapping spheres which are positioned onto the tunnel centerline. The radii of these spheres are limited by the atoms surrounding the tunnel. However, this representation can underestimate the void tunnel space significantly (see Figure 2.3). One of the straightforward extensions was to cover the spherical representation by a surface. But this did not solve the underestimation problem. Therefore, Byška et al. [2] proposed the algorithm for detection of so called asymmetric tunnels which better describe the tunnel void space. Further enhancement of this original approach [20] was presented as one of my contributions in Section 2.2 in Chapter 2.

With the emerge of techniques for the detection of tunnels in molecular dynamics also the necessity for proper visual representation appeared. The straightforward animation of the movements of tunnel surface is meaningful only for a limited amount of time steps. But for the biochemists it is impossible to observe hundreds of thousands of time steps and to spot significant events happening in the simulation. Therefore, several specialized methods for visualization and visual analysis of such data started to appear [37, 40].

Contributions

Our contribution to the field of tunnel visualization and visual analysis consists of one technique proposed for tunnel representation and three approaches to the exploration of tunnels in molecular dynamics simulations.

Our first contribution was published in 2007 [28] and presents two techniques for visualization of tunnels. Both were designed to overcome the problems of the existing representation of a tunnel as a set of spheres. The first method based on the representation of a tunnel as a set of subsequent tetrahedra which we obtained as an alternative tunnel representation when calculating the tunnel using Voronoi diagrams and Delaunay triangulation [41]. We subsequently applied the Loop subdivision to these tetrahedra and obtained the resulting representation. The main advantage of this approach was that we were able to capture the mutual circumfluence of tunnels which are located in a close vicinity. The second proposed technique aimed to visualize a tight tunnel boundary, i.e., to extend the tunnel surface to capture the maximal void space. This was reached by taking the tetrahedra and excluding the parts around the vertices which belong to some atom, similarly to the constructive solid geometry technique.

In the last two years our research focused namely on the visualization and visual analysis of tunnels in molecular dynamics. Here the information which should be conveyed to the biochemists is complex – they aim to explore the tunnels in order to observe namely the changes in their shape and physico-chemical properties of amino acids lining the tunnels. Therefore, we proposed three approaches enhancing these tasks. The first method, called MoleCollar [4], proposes visualization methods supporting the biochemical workflow for tunnel exploration. When analyzing tunnels and their behavior in molecular dynamics, the results contain the information about the whole system of tunnels and their temporal changes. The task is to explore the results and to detect the most biochemically relevant tunnel(s) with desired properties. One of the basic criteria is that such a tunnel has to be relatively stable over time, i.e., it is opened for a significant part of the simulation. To support this task we proposed two methods based on heat maps which help to recognize the most stable tunnel. In a subsequent exploration of a selected tunnel the aim is to observe the amino acids lining this tunnel and to observe the changes of their physico-chemical properties (e.g., hydrophobicity or partial charge of their atoms). The most critical part is the tunnel narrowest site (bottleneck) because it highly determines the size and properties of ligand penetrating through this tunnel. Therefore, we proposed a method for fast and intuitive exploration of the bottleneck shape and the properties of amino acids forming this bottleneck. This representation first calculates the contour of the bottleneck of the tunnel in each time step and then superposes these contours to get the information about bottleneck stability over time. The contour representation is then surrounded by the bars representing individual amino acids lining the bottleneck. The color and size of these bars corresponds to different physico-chemical properties of these amino acids and their temporal stability wrt. tunnel influence.

The second proposed method [5] extends the idea of exploration of the amino acids surrounding the bottleneck to the whole tunnel. For each time step, we first plot the tunnel profile in order to detect the most stable and unstable parts of the tunnel along its centerline as well as the position of the bottleneck(s). Then for each amino acid lining the tunnel we calculate its extent of influence of the tunnel and for each time step we visualize it as one line. Such amino acid in molecular dynamics is then represented by a strip of these lines which helps the biochemist to understand not only the importance of the amino acid but also how its importance changes over time. Additionally we support several coloring schemes according to different physico-chemical properties of the amino acids as well as sorting the strips with respect to given criteria.

Both presented methods, illustrated in Figure 3.4, were thoroughly tested by the biochemists from the Loschmidt Laboratories. For these tests they used real scenarios from their research in protein engineering when they are searching for the amino acids lining the tunnel and influencing it significantly. Such amino acids are then best candidates for mutations in order to change a desired property of the protein. In both cases the biochemists concluded that the proposed methods helped to reveal the candidate amino acids for mutations in a fast and intuitive way.

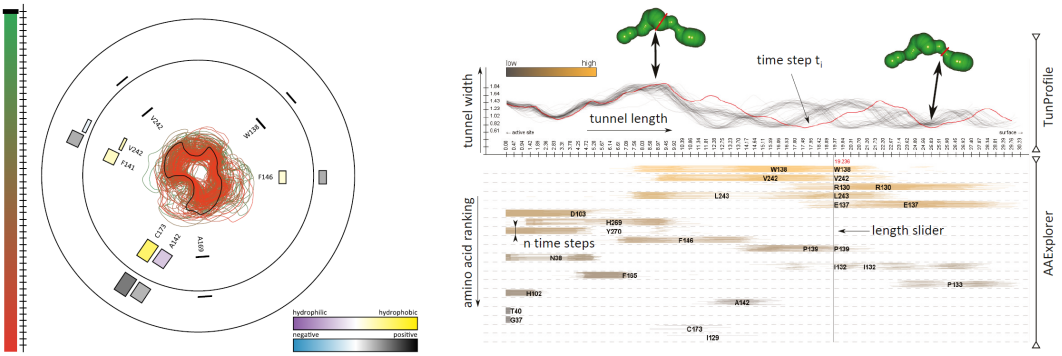


Figure 3.4: MoleCollar (left) and AnimoAminoMiner (right) visualization methods showing the evolution of tunnel over time (taken from [4, 5]).

Our latest contribution [24] presented at the Eurographics Workshop on Visual Computing for Biology and Medicine in 2016 is also related to the exploration of tunnel changes over time. This time we focused even more on the influence of the tunnel by individual amino acids and suggesting the time periods when the tunnel changed significantly and it is thus worthwhile to explore it in more detail. Our method first unfolds the surface representation of a selected tunnel and then describes the resulting 2D image by image moments which are commonly used in image processing. This way we are able to describe a set of images of tunnels and to detect similarities and outliers which we subsequently visualize as clusters in a scatterplot graph. To visualize also the changes in the length of the tunnel over time, we use bar chart representation. All supported views are linked which enables the user to explore the dataset more intuitively. The usability of our proposed approach was demonstrated on two case studies – exploration of a set of mutants of a haloalkane dehalogenase and exploration of tunnels in time steps of molecular dynamics.

Articles in Collection

[28] B. Kozlíková, F. Andres, and J. Sochor. Visualization of tunnels in protein molecules. In *Proceedings of the 5th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '07, pages 111–118, New York, NY, USA, 2007. ACM

I participated in the design and implementation of the algorithm and was responsible for paper writing. Contribution 60%.

[4] J. Byška, A. Jurčík, M. E. Gröller, I. Viola, and B. Kozlíková. MoleCollar and Tunnel Heat Map visualizations for conveying spatio–temporo–chemical properties across and along protein voids. *Computer Graphics Forum*, 34(3):1–10, 2015

I was responsible for coordinating the activities related to the design of the method and was responsible for paper writing. Contribution 35%.

[5] J. Byška, M. Le Muzic, M. E. Gröller, I. Viola, and B. Kozlíková. AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):747–756, 2015

I coordinated the activities related to the design of the method and was responsible for paper writing. Contribution 35%.

[24] I. Kolesár, J. Byška, J. Parulek, H. Hauser, and B. Kozlíková. Unfolding and interactive exploration of protein tunnels and their dynamics. In *EG VCBM 2016 Eurographics Workshop on Visual Computing for Biology and Medicine*, 2016

I participated in the design of the method, wrote several sections of the paper, and did the proofreading. Contribution 25%.

3.3 Ligand Path Visualization

This rather short section describes the area of visual exploration of the ligand transportation through a tunnel to the active site. This area is rather new because the simulations of molecular dynamics capturing also the ligand movements started to appear only recently. However, this area represents a next crucial step in protein analysis because it utilizes the detected tunnels for their purpose. So I consider it as one of the most significant directions of our future research.

The complexity of the visual exploration of ligand transportation lies namely in the complexity of the input data. The simulations of ligand movements are very long and the movement itself is very scattered – the ligand can follow different directions before reaching the active site. As it is hard to understand the significant movements of the ligand from the original data, advanced simplification and visualization methods have to be incorporated.

Except for the exploration of existing simulations there were already some attempts to calculate the ligand penetration to the active site and thus simulate the ligand docking. Among these belong the MoMA-LigPath tool [12] which calculates the ligand unbinding trajectory based on a simplified model considering mechanistic representation with partial flexibility. There were also attempts to apply path-planning approaches from robotics to this problem [9].

Contributions

In our research we currently focused on the problem of visual exploration of simulations of molecular dynamics containing the ligand motion as we obtained this data from our co-operating group of biochemists from Loschmidt Laboratories. This data contains dozens of thousands of time steps capturing protein and ligand movements. The main task was to simplify the ligand trajectory in order to be able to observe the trends in its movement and to

enable the biochemists to explore the trajectory using advanced visual analysis techniques. Therefore, we proposed a visual analysis system described in [13] which will be presented at the IEEE BioVis conference in 2016. The system consists of a multiscale simplification model for the trajectory and three linked views – 3D representation of the simiplified trajectory, scatterplot matrix, and bar charts helping to convey the information about different attributes of ligand and amino acids which are interacting with the ligand along its way to the active site. The simplification method combines interactive and automated solutions. First the input ligand trajectory is processed by our algorithm for automatic simplification and then the user is allowed to interactively select parts of the simplified curve and to apply the proposed interactive simplification. The linked views then provide the users with the information about the evolution of different ligand properties, such as its current distance to the active site, changes of amino acids lining the ligand, or its "stuckness", i.e., the situation when the ligand stays for a significant portion of time in some position.

Articles in Collection

[13] K. Furmanová, M. Jarešová, J. Byška, A. Jurčík, J. Parulek, H. Hauser, and B. Kozlíková. Interactive exploration of ligand transportation through protein tunnels. In *BioVis*, 2016

I participated in the design of the presented method and was responsible for writing of several parts of the paper and proofreading. Contribution 20%.

3.4 Visualization Tools

We presented several algorithms and methods for analysis and visualization of biomolecular structures and their cavities. However, without their integration to some software tool their use would be very limited. Similarly to the plethora of algorithms which were designed in the last decades also dozens of dedicated visualization tools appeared. Many of them were serving for a specific purpose and were not maintained for a longer period (e.g., QuteMol [55]). However, some of the tools became very robust and are still available to the biochemical community. Among these belong namely PyMOL [11, 47] and VMD [19, 51]. The list of other tools (also those which are not maintained but formed a milestone in molecular visualization) can be found in our state-of-the-art report [30].

Contributions

As already mentioned, there are several visualization tools available which mostly support also the integration of plugins in order to perform different analysis tasks. In the first stages of our research we used this opportunity and published a PyMOL plugin version

of our CAVER 3.0 algorithm (as well as its previous versions) for tunnel computation in order to visualize the results and get the computational tool to the wide community of PyMOL users. However, PyMOL supported only the basic visualizations of tunnels as a set of spheres and surface covering these spheres which proved to be insufficient. Therefore, with the design of our new visualization algorithms we started to develop also our own tool integrating both the computation of tunnels and their visualization. The first version of this tool, called CAVER Viewer, was first published in 2010 [31] and was released to the community. However, thanks to its basic functionality, this tool served at the end mainly as an environment for fast prototyping of our new algorithms.

Later we significantly changed the platform on which we were implementing the tool and started to develop a new tool, CAVER Analyst 1.0 [32]. This time the multiplatform Java-based tool was already successfully adopted by the community and currently we have dozens of users worldwide. The 1.0 version contains methods supporting the tunnel computation (it integrates the CAVER 3.0 algorithm), detection of inner cavities, different visual representations of molecules and tunnels, comprehensive overview of statistical information about tunnels in a form of interactive graphs, and other functions. Internally the tool still serves also for fast prototyping of our new methods and algorithms.

Most of the algorithms and methods presented in this thesis have been integrated to CAVER Analyst and will be a part of the planned next release of the software.

Articles in Collection

[31] B. Kozlíková, J. Sochor, T. Szabó, and M. Zamborský. Caver viewer-new tool enhancing computation and visualization of channels in proteins. In *The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2010*. IADIS Press, 2010

I was responsible for coordinating the design and implementation activities, implemented several algorithms, and wrote the paper. Contribution 50%.

[32] B. Kozlíková, E. Šebestová, V. Šustr, J. Brezovský, O. Strnad, L. Daniel, D. Bednář, A. Pavelka, M. Maňák, M. Bezděka, P. Beneš, M. Kotry, A. Gora, J. Damborský, and J. Sochor. CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics*, 30(18):btu364, 2014

I was responsible for coordinating the design and implementation activities as well as the team of developers and wrote the paper. Contribution 40%.

Chapter 4

Conclusion

In this text, I have presented my research contributions to the progress within the area of analysis and visualization of biomolecules, and put these in the context of the state-of-the-art in the area. In this commentary, the individual research contributions were presented as pieces of one puzzle, which spans across several tightly related areas. The individual research contributions were accompanied with selected representative articles I have co-authored, which are also attached to this text[1].

In the future I would like to focus namely on the design of visualization and visual analysis methods for exploration of ligand and water movements in molecular dynamics and to continue in the ongoing research in the alternative path-planning detection of tunnels taking into account the ligand shape and movements. Additionally, recently I started to explore the area of visualization of interactions between proteins forming protein complexes. In this case the task is to explore and visually convey the information about the interactions occurring on molecular surfaces of the interacting proteins. This challenging and still rather unexplored area gives us another high potential to the future.

[1] The fulltexts of the articles are excluded from the public version of this text to avoid copyright violation.

Bibliography

[1] J. Brezovský, B. Kozlíková, and J. Damborský. Computational analysis of protein tunnels and channels. In *Methods in Molecular Biology*. Springer International Publishing, 2016.

[2] J. Byska, A. Jurcik, and J. Sochor. Geometry-based algorithm for detection of asymmetric tunnels in protein molecules. In S. Czanner and W. Tang, editors, *TPCG*, pages 17–24. Eurographics Association, 2013.

[3] J. Byška, I. Kolingerová, B. Kozlíková, and J. Sochor. Path–planning algorithm for transportation of molecules through protein tunnel bottlenecks. In J. Jorge, L. P. Santos, and R. Ďurikovič, editors, *31th Proceedings of Spring Conference on Computer Graphics*, pages 101–108, Bratislava, Slovakia, 2015. Comenius University.

[4] J. Byška, A. Jurčík, M. E. Gröller, I. Viola, and B. Kozlíková. MoleCollar and Tunnel Heat Map visualizations for conveying spatio–temporo–chemical properties across and along protein voids. *Computer Graphics Forum*, 34(3):1–10, 2015.

[5] J. Byška, M. Le Muzic, M. E. Gröller, I. Viola, and B. Kozlíková. AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):747–756, 2015.

[6] E. Chovancová, A. Pavelka, P. Beneš, O. Strnad, J. Brezovský, B. Kozlíková, A. W. Gora, V. Šustr, M. Klvaňa, P. Medek, L. Biedermannová, J. Sochor, and J. Damborský. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLOS Computational Biology*, 8(10), 2012.

[7] E. Chovancová, A. Pavelka, P. Beneš, O. Strnad, J. Brezovský, B. Kozlíková, A. Gora, V. Šustr, M. Klvaňa, P. Medek, L. Biedermannová, J. Sochor, and J. Damborský. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.*, 8(10):e1002708, 2012.

[8] J. Cortés, T. Siméon, V. R. De Angulo, D. Guieysse, M. Remaud-Siméon, and V. Tran. A path planning approach for computing large–amplitude motions of flexible molecules. *Bioinformatics*, 21(suppl 1):i116–i125, 2005.

[9] J. Cortés, S. Barbe, M. Erard, and T. Siméon. Encoding Molecular Motions in Voxel Maps. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 8(2):557–563, 2011.

[10] J. Damborský and J. Brezovský. Computational tools for designing and engineering enzymes. *Current Opinion in Chemical Biology*, 19:8–16, 2014.

[11] W. L. DeLano. PyMOL: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter On Protein Crystallography*, 40, 2002. http://pymol.sourceforge.net/.

[12] D. Devaurs, L. Bouard, M. Vaisset, C. Zanon, I. Al-Bluwi, R. Iehl, T. Siméon, and J. Cortés. MoMA-LigPath: a web server to simulate protein-ligand unbinding. *Nucl. Acids Res.*, 41:297–302, 2013.

[13] K. Furmanová, M. Jarešová, J. Byška, A. Jurčík, J. Parulek, H. Hauser, and B. Kozlíková. Interactive exploration of ligand transportation through protein tunnels. In *BioVis*, 2016.

[14] F. García-Guevara, M. Avelar, M. Ayala, and L. Segovia. Computational tools applied to enzyme design − a review. *Biocatalysis*, 1(1):109–117, 2016.

[15] S. Grottel, G. Reina, C. Dachsbacher, and T. Ertl. Coherent culling and shading for large molecular dynamics visualization. *Comput. Graph. Forum*, 29:953–962, 2010.

[16] U. Hensen, T. Meyer, J. Haas, R. Rex, G. Vriend, and H. Grubmueller. Exploring protein dynamics space: the dynasome as the missing link between protein structure and function. *PLoS ONE*, 7(5):e33931, 2012.

[17] P. Hermosilla, V. Guallar, A. Vinacua, and P.-P. Vázquez. Instant Visualization of Secondary Structures of Molecular Models. In *VCBM*, 2015.

[18] L. Holm and P. Rosenström. Dali server: conservation mapping in 3D. *Nucleic acids research*, 38(Web Server issue):W545–W549, July 2010.

[19] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[20] A. Jurčík, J. Byška, J. Sochor, and B. Kozlíková. Visibility–based approach to surface detection of tunnels in proteins. In J. Jorge, L. P. Santos, and R. Ďurikovič, editors, *31th Proceedings of Spring Conference on Computer Graphics*, pages 85–92, Bratislava, Slovakia, 2015. Comenius University.

[21] A. Jurčík, J. Parulek, J. Sochor, and B. Kozlíková. Accelerated Visualization of Transparent Molecular Surfaces in Molecular Dynamics. In *IEEE Pacific Visualization Symposium*, pages 112–119, 2016.

[22] D. Kauker, M. Krone, A. Panagiotidis, G. Reina, and T. Ertl. Rendering molecular surfaces using order-independent transparency. In *Eurographics Symposium on Parallel Graphics and Visualization*, volume 13, pages 33–40, 2013.

[23] L. Kocincová, M. Jarešová, J. Byška, J. Parulek, H. Hauser, and B. Kozlíková. Comparative visualization of protein secondary structures. In *BioVis*, 2016.

[24] I. Kolesár, J. Byška, J. Parulek, H. Hauser, and B. Kozlíková. Unfolding and interactive exploration of protein tunnels and their dynamics. In *EG VCBM 2016 Eurographics Workshop on Visual Computing for Biology and Medicine*, 2016.

[25] T. Koudeláková, R. Chaloupková, J. Brezovský, Z. Prokop, E. Šebestová, M. Hesseler, M. Khabiri, M. Plevaka, D. Kulik, I. Kutá Smatanová, P. Řezáčová, R. Ettrich, U. T. Bornscheuer, and J. Damborský. Engineering enzyme stability and resistance to an organic cosolvent by modification of residues in the access tunnel. *Angewandte Chemie International Edition*, 52(7), 2013.

[26] B. Kozlíková. Visualization techniques for static and dynamic protein molecules and their channels. *PhD Thesis, Masaryk University, Czech Republic*, 2011.

[27] B. Kozlíková, I. Aleksandrowicz, and J. Sochor. Computation and visualization of surface of proteins and their channels. In I. Press, editor, *The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2011*, pages 99–106. IADIS Press, 2011.

[28] B. Kozlíková, F. Andres, and J. Sochor. Visualization of tunnels in protein molecules. In *Proceedings of the 5th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '07, pages 111–118, New York, NY, USA, 2007. ACM.

[29] B. Kozlíková, A. Jurčík, J. Byška, O. Strnad, and J. Sochor. Visualizing Movements of Protein Tunnels in Molecular Dynamics Simulations. In I. Viola, K. Buehler, and T. Ropinski, editors, *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2014.

[30] B. Kozlíková, M. Krone, N. Lindow, M. Falk, M. Baaden, D. Baum, I. Viola, J. Parulek, and H.-C. Hege. Visualization of Biomolecular Structures: State of the Art. In *Eurographics Conference on Visualization - STARs*, pages 61–81, 2015.

[31] B. Kozlíková, J. Sochor, T. Szabó, and M. Zamborský. Caver viewer-new tool enhancing computation and visualization of channels in proteins. In *The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2010*. IADIS Press, 2010.

[32] B. Kozlíková, E. Šebestová, V. Šustr, J. Brezovský, O. Strnad, L. Daniel, D. Bednář, A. Pavelka, M. Maňák, M. Bezděka, P. Beneš, M. Kotry, A. Gora, J. Damborský, and J. Sochor. CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics*, 30(18):btu364, 2014.

[33] M. Krone, B. Kozlíková, N. Lindow, M. Baaden, D. Baum, J. Parulek, H.-C. Hege, and I. Viola. Visual analysis of biomolecular cavities: State of the art. *Computer Graphics Forum*, 35(3):527–551, 2016.

[34] S. M. Lavalle. *Rapidly-exploring random trees: A new tool for path planning.* 1998.

[35] S. M. LaValle. *Planning Algorithms.* Cambridge University Press, Cambridge, U.K., 2006.

[36] D. G. Levitt and L. J. Banaszak. POCKET: A Computer Graphics Method for Identifying and Displaying Protein Cavities and Their Surrounding Amino Acids. *J. Mol. Graph.*, 10(4):229–234, 1992.

[37] N. Lindow, D. Baum, A.-N. Bondar, and H.-C. Hege. Exploring cavity dynamics in biomolecular systems. *BMC Bioinformatics*, 14(Suppl 19):S5, 2013.

[38] N. Lindow, D. Baum, and H.-C. Hege. Interactive rendering of materials and biological structures on atomic and nanoscopic scale. *Comput. Graph. Forum*, 31:1325–1334, 2012.

[39] M. Manak. Exploration of Empty Space among Spherical Obstacles via Additively Weighted Voronoi Diagram. *Computer Graphics Forum*, 2016.

[40] T. B. Masood, S. Sandhya, N. Chandra, and V. Natarajan. CHEXVIS: a tool for molecular channel extraction and visualization. *BMC Bioinformatics*, 16:119, 2015.

[41] P. Medek, P. Beneš, and J. Sochor. Computation of tunnels in protein molecules using Delaunay triangulation. In *Journal of WSCG*, pages 107–114, 2007.

[42] J. Parulek, D. Jönsson, T. Ropinski, S. Bruckner, A. Ynnerman, and I. Viola. Continuous levels-of-detail and visual abstraction for seamless molecular visualization. *Comput. Graph. Forum*, 33(6):276–287, 2014.

[43] A. Pavelka, E. Šebestová, B. Kozlíková, J. Brezovský, J. Sochor, and J. Damborský. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE ACM T. Comput. Bi.*, 2015.

[44] M. Petřek, P. Košinová, J. Koča, and M. Otyepka. MOLE: A Voronoi Diagram-Based Explorer of Molecular Channels, Pores, and Tunnels. *Structure*, 15(11):1357–1363, 2007.

[45] 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 Bioinform.*, 7(1):316, 2006.

[46] M. F. Sanner, A. J. Olson, and J.-C. Spehner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38(3):305–320, 1996.

[47] L. Schrödinger. The PyMOL Molecular Graphics System, Version 1.8. 2016.

[48] D. Sehnal, R. Svobodová Vařeková, K. Berka, L. Pravda, V. Navrátilová, P. Banáš, C.-M. Ionescu, M. Otyepka, and J. Koča. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminform.*, 5(1), 2013.

[49] I. Shindyalov and P. Bourne. Protein structure alignment by incremental combinatorial extension of the op-timum path. *Protein Engineering*, 11:739–747, 1998.

[50] R. Skanberg, P.-P. Vázquez, V. Guallar, and T. Ropinski. Real-time molecular visualization supporting diffuse interreflections and ambient occlusion. *IEEE Trans. Vis. Comput. Graph.*, 22(1):718–727, 2016.

[51] J. E. Stone, A. Kohlmeyer, K. L. Vandivort, and K. Schulten. Immersive molecular visualization and interactive modeling with commodity hardware. In *Advances in Visual Computing*, volume 6454 of *Lecture Notes in Computer Science*, pages 382–393. Springer, 2010.

[52] O. Strnad, B. Kozlíková, and J. Sochor. Detection of intramolecular tunnels connecting sequence of sites in protein structures. In J. Saez-Rodriguez, editor, *Advances in Intelligent Systems and Computing*, pages 73–80. Springer International Publishing, 2014.

[53] O. Strnad, B. Kozlíková, V. Šustr, and J. Sochor. Real-time visualization and exploration of protein empty space with varying parameters. *International Journal on Advances in Life Sciences*, 5(3 & 4), 2013.

[54] O. Strnad, V. Šustr, B. Kozlíková, and J. Sochor. Geometrical detection of pathways in protein structures leading among more binding sites. In H. H. Ali, editor, *BIOTECHNO 2014 : The Sixth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies*, pages 93–98. IARIA XPS Press, 2014.

[55] M. Tarini, P. Cignoni, and C. Montani. Ambient Occlusion and Edge Cueing for Enhancing Real Time Molecular Visualization. *IEEE Trans. Vis. Comput. Graphics*, 12(5):1237–1244, 2006.

[56] W. G. Touw, C. Baakman, J. Black, T. A. H. te Beek, E. Krieger, R. P. Joosten, and G. Vriend. A series of PDB-related databanks for everyday needs. *Nucleic Acids Research*, 43(Database issue):D364–D368, 2014.

[57] M. van der Zwan, W. Lueks, H. Bekker, and T. Isenberg. Illustrative molecular visualization with continuous abstraction. *Comput. Graph. Forum*, 30(3):683–690, 2011.

[58] V. Vonásek and B. Kozlíková. Application of sampling-based path planning for tunnel detection in dynamic protein structures. In *MMAR: 21st International Conference on Methods and Models in Automation and Robotics*, Miedzyzdroje, Poland, 2016.

[59] E. Yaffe, D. Fishelovitch, H. J. Wolfson, D. Halperin, and R. Nussinov. MolAxis: a server for identification of channels in macromolecules. *Nucleic Acids Res.*, 36(Web Server issue):W210–215, Jul 2008.

[60] M. Zamborský, T. Szabó, and B. Kozlíková. Dynamic visualization of protein secondary structures. In *Proceedings of the 13th Central European Seminar on Computer Graphics (CESCG 2009)*, Vienna, 2009. Vienna University of Technology, Insitute of Computer Graphics and Algorithms.

[61] Y. Zhang and J. Skolnick. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7):2302–2309, 2005.

Part II

Collection of Articles

Appendix A

Journal articles and chapters

This appendix together with Appendix B and C list contains the total of 23 research articles that were selected as the representatives of my contributions within the studied research field. The fulltexts of the articles are inserted into the corresponding appendixes of the printed version of this thesis[1] and referenced via the article numbers assigned in the list below (replacing page numbers). The same holds for Appendix B and C.

Article A: E. Chovancová, A. Pavelka, P. Beneš, O. Strnad, J. Brezovský, B. Kozlíková, A. W. Gora, V. Šustr, M. Klvaňa, P. Medek, L. Biedermannová, J. Sochor, and J. Damborský. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLOS Computational Biology*, 8(10), 2012

Article B: A. Pavelka, E. Šebestová, B. Kozlíková, J. Brezovský, J. Sochor, and J. Damborský. CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. *IEEE ACM T. Comput. Bi.*, 2015

Article C: B. Kozlíková, E. Šebestová, V. Šustr, J. Brezovský, O. Strnad, L. Daniel, D. Bednář, A. Pavelka, M. Maňák, M. Bezděka, P. Beneš, M. Kotry, A. Gora, J. Damborský, and J. Sochor. CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics*, 30(18):btu364, 2014

Article D: M. Krone, B. Kozlíková, N. Lindow, M. Baaden, D. Baum, J. Parulek, H.-C. Hege, and I. Viola. Visual analysis of biomolecular cavities: State of the art. *Computer Graphics Forum*, 35(3):527–551, 2016

Article E: J. Byška, A. Jurčík, M. E. Gröller, I. Viola, and B. Kozlíková. MoleCollar and Tunnel Heat Map visualizations for conveying spatio–temporo–chemical properties

[1] The fulltexts of the articles are excluded from the publicly available electronic version of this text to avoid copyright violation.

across and along protein voids. *Computer Graphics Forum*, 34(3):1–10, 2015

Article F: J. Byška, M. Le Muzic, M. E. Gröller, I. Viola, and B. Kozlíková. AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):747–756, 2015

Article G: O. Strnad, B. Kozlíková, V. Šustr, and J. Sochor. Real-time visualization and exploration of protein empty space with varying parameters. *International Journal on Advances in Life Sciences*, 5(3 & 4), 2013

Article H (Book Chapter): O. Strnad, B. Kozlíková, and J. Sochor. Detection of intramolecular tunnels connecting sequence of sites in protein structures. In J. Saez-Rodriguez, editor, *Advances in Intelligent Systems and Computing*, pages 73–80. Springer International Publishing, 2014

Article I (Book Chapter): J. Brezovský, B. Kozlíková, and J. Damborský. Computational analysis of protein tunnels and channels. In *Methods in Molecular Biology*. Springer International Publishing, 2016

Article J: B. Kozlíková, F. Andres, and J. Sochor. Visualization of tunnels in protein molecules. In *Proceedings of the 5th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '07, pages 111–118, New York, NY, USA, 2007. ACM

Article K: V. Vonásek and B. Kozlíková. Application of sampling-based path planning for tunnel detection in dynamic protein structures. In *MMAR: 21st International Conference on Methods and Models in Automation and Robotics*, Miedzyzdroje, Poland, 2016

Article L: M. Zamborský, T. Szabó, and B. Kozlíková. Dynamic visualization of protein secondary structures. In *Proceedings of the 13th Central European Seminar on Computer Graphics (CESCG 2009)*, Vienna, 2009. Vienna University of Technology, Insitute of Computer Graphics and Algorithms

Article M: B. Kozlíková, J. Sochor, T. Szabó, and M. Zamborský. Caver viewer-new tool enhancing computation and visualization of channels in proteins. In *The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2010*. IADIS Press, 2010

Article N: B. Kozlíková, I. Aleksandrowicz, and J. Sochor. Computation and visualization of surface of proteins and their channels. In I. Press, editor, *The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP) 2011*, pages 99–106. IADIS Press, 2011

Article O: B. Kozlíková, A. Jurčík, J. Byška, O. Strnad, and J. Sochor. Visualizing Movements of Protein Tunnels in Molecular Dynamics Simulations. In I. Viola, K. Buehler, and T. Ropinski, editors, *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2014

Article P: O. Strnad, V. Šustr, B. Kozlíková, and J. Sochor. Geometrical detection of pathways in protein structures leading among more binding sites. In H. H. Ali, editor, *BIOTECHNO 2014 : The Sixth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies*, pages 93–98. IARIA XPS Press, 2014

Article Q: A. Jurčík, J. Byška, J. Sochor, and B. Kozlíková. Visibility–based approach to surface detection of tunnels in proteins. In J. Jorge, L. P. Santos, and R. Ďurikovič, editors, *31th Proceedings of Spring Conference on Computer Graphics*, pages 85–92, Bratislava, Slovakia, 2015. Comenius University

Article R: J. Byška, I. Kolingerová, B. Kozlíková, and J. Sochor. Path–planning algorithm for transportation of molecules through protein tunnel bottlenecks. In J. Jorge, L. P. Santos, and R. Ďurikovič, editors, *31th Proceedings of Spring Conference on Computer Graphics*, pages 101–108, Bratislava, Slovakia, 2015. Comenius University

Article S: B. Kozlíková, M. Krone, N. Lindow, M. Falk, M. Baaden, D. Baum, I. Viola, J. Parulek, and H.-C. Hege. Visualization of Biomolecular Structures: State of the Art. In *Eurographics Conference on Visualization - STARs*, pages 61–81, 2015

Article T: A. Jurčík, J. Parulek, J. Sochor, and B. Kozlíková. Accelerated Visualization of Transparent Molecular Surfaces in Molecular Dynamics. In *IEEE Pacific Visualization Symposium*, pages 112–119, 2016

Article U: I. Kolesár, J. Byška, J. Parulek, H. Hauser, and B. Kozlíková. Unfolding and interactive exploration of protein tunnels and their dynamics. In *EG VCBM 2016 Eurographics Workshop on Visual Computing for Biology and Medicine*, 2016

Article V: K. Furmanová, M. Jarešová, J. Byška, A. Jurčík, J. Parulek, H. Hauser, and B. Kozlíková. Interactive exploration of ligand transportation through protein tunnels. In *BioVis*, 2016

Article W: L. Kocincová, M. Jarešová, J. Byška, J. Parulek, H. Hauser, and B. Kozlíková. Comparative visualization of protein secondary structures. In *BioVis*, 2016

Appendix B

Collection of Articles

Paper A

CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures

Eva Chovancová, Antonín Pavelka, Petr Beneš, Ondřej Strnad, Jan Brezovský, Barbora Kozlíková, Artur Gora, Vilém Šustr, Martin Klvaňa, Petr Medek, Lada Biedermannová, Jiří Sochor, Jiří Damborský

Masaryk University, Brno, Czech Republic & St. Anne's University Hospital Brno, Czech Republic

PLoS Computational Biology, volume 8, http://dx.doi.org/10.1371/journal.pcbi.1002708, 2012.

Abstract

Tunnels and channels facilitate the transport of small molecules, ions and water solvent in a large variety of proteins. Characteristics of individual transport pathways, including their geometry, physico-chemical properties and dynamics are instrumental for understanding of structure-function relationships of these proteins, for the design of new inhibitors and construction of improved biocatalysts. CAVER is a software tool widely used for the identification and characterization of transport pathways in static macromolecular structures. Herein we present a new version of CAVER enabling automatic analysis of tunnels and channels in large ensembles of protein conformations. CAVER 3.0 implements new algorithms for the calculation and clustering of pathways. A trajectory from a molecular dynamics simulation serves as the typical input, while detailed characteristics and summary statistics of the time evolution of individual pathways are provided in the outputs. To illustrate the capabilities

of CAVER 3.0, the tool was applied for the analysis of molecular dynamics simulation of the microbial enzyme haloalkane dehalogenase DhaA. CAVER 3.0 safely identified and reliably estimated the importance of all previously published DhaA tunnels, including the tunnels closed in DhaA crystal structures. Obtained results clearly demonstrate that analysis of molecular dynamics simulation is essential for the estimation of pathway characteristics and elucidation of the structural basis of the tunnel gating. CAVER 3.0 paves the way for the study of important biochemical phenomena in the area of molecular transport, molecular recognition and enzymatic catalysis. The software is freely available as a multiplatform command-line application at http://www.caver.cz.

Paper B

CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules

Antonín Pavelka, Eva Šebestová, Barbora Kozlíková, Jan Brezovský, Jiří Sochor, Jiří Damborský

Masaryk University, Brno, Czech Republic

IEEE ACM Transactions on Computational Biology and Bioinformatics, volume 13, number 3, pages 505–517, http://dx.doi.org/10.1109/TCBB.2015.2459680, 2016.

Abstract

The biological function of a macromolecule often requires that a small molecule or ion is transported through its structure. The transport pathway often leads through void spaces in the structure. The properties of transport pathways change significantly in time; therefore the analysis of a trajectory from molecular dynamics rather than of a single static structure is needed for understanding the function of pathways. The identification and analysis of transport pathways are challenging because of the high complexity and diversity of macromolecular shapes, the thermal motion of their atoms, and the large amount of conformations needed to properly describe conformational space of protein structure. In this paper, we describe the principles of the CAVER 3.0 algorithms for the identification and analysis of properties of transport pathways both in static and dynamic structures. Moreover, we introduce the improved clustering solution for finding tunnels in macromolecules, which is included in the latest CAVER 3.02 version. Voronoi diagrams are used to identify potential pathways in each snapshot of a molecular dynamics trajectory and clustering is then used to find the correspondence between tunnels from different snapshots. Furthermore, the geometrical properties of pathways and their evolution in time are computed and visualized.

Paper C

CAVER Analyst 1.0: Graphic Tool for Interactive Visualization and Analysis of Tunnels and Channels in Protein Structures

Barbora Kozlíková[1], Eva Šebestová[1], Vilém Šustr[1], Jan Brezovský[1], Ondřej Strnad[1], Lukáš Daniel[1], David Bednář[1], Antonín Pavelka[1], Martin Maňák[2], Martin Bezděka[1], Petr Beneš[1], Matúš Kotry[1], Artur Gora[1], Jiří Damborský[1], Jiří Sochor[1]

[1]Masaryk University, Brno, Czech Republic
[2]University of West Bohemia, Pilsen, Czech Republic

Bioinformatics, volume 30, number 18, 10.1093/bioinformatics/btu364, 2014.

Abstract

Summary: The transport of ligands, ions or solvent molecules into proteins with buried binding sites or through the membrane is ena-bled by protein tunnels and channels. CAVER Analyst is a software tool for calculation, analysis and real-time visualization of access tunnels and channels in static and dynamic protein structures. It provides an intuitive graphic user interface for setting up the calcula-tion and interactive exploration of identified tunnels/channels and their characteristics.

Availability and Implementation: CAVER Analyst is a multi-platform software written in JAVA. Binaries and documentation are freely available for non-commercial use at http://www.caver.cz.

Paper D

Visual Analysis of Biomolecular Cavities: State of the Art

Michael Krone[1], Barbora Kozlíková[2], Norbert Lindow[3], Marc Baaden[4], Daniel Baum[1], Julius Parulek[5], Hans-Christian Hege[3], Ivan Viola[6]

[1] University of Stuttgart, Germany
[2]Masaryk University, Brno, Czech Republic
[3]Zuse Institute Berlin, Germany
[4]Laboratoire de Biochimie Théorique, France
[5]University of Bergen, Norway
[6]TU Wien, Austria

Computer Graphics Forum, volume 35, number 3, 10.1111/cgf.12928, 2016.

Abstract

In this report we review and structure the branch of molecular visualization that is concerned with the visual analysis of cavities in macromolecular protein structures. First the necessary background, the domain terminology, and the goals of analytical reasoning are introduced. Based on a comprehensive collection of relevant research works, we present a novel classification for cavity detection approaches and structure them into four distinct classes: grid-based, Voronoi-based, surface-based, and probe-based methods. The subclasses are then formed by their combinations. We match these approaches with corresponding visualization technologies starting with direct 3D visualization, followed with non-spatial visualization techniques that for example abstract the interactions between structures into a relational graph, straighten the cavity of interest to see its profile in one view, or aggregate the time sequence into a single contour plot. We also discuss the current state of methods

for the visual analysis of cavities in dynamic data such as molecular dynamics simulations. Finally, we give an overview of the most common tools that are actively developed and used in the structural biology and biochemistry research. Our report is concluded by an outlook on future challenges in the field.

Paper E

MoleCollar and Tunnel Heat Map Visualizations for Conveying Spatio-Temporo-Chemical Properties Across and Along Protein Voids

Jan Byška[1], Adam Jurčík[1], Eduard M. Gröller[2,3], Ivan Viola[2,3], Barbora Kozlíková[1,2]

[1]Masaryk University, Brno, Czech Republic
[2]Vienna University of Technology, Austria
[3]University of Bergen, Norway

Computer Graphics Forum, volume 34, number 3, pages 1-10, 10.1111/cgf.12612, 2015.

Abstract

Studying the characteristics of proteins and their inner void space, including their geometry, physico-chemical properties and dynamics are instrumental for evaluating the reactivity of the protein with other small molecules. The analysis of long simulations of molecular dynamics produces a large number of voids which have to be further explored and evaluated. In this paper we propose three new methods: two of them convey important properties along the long axis of a selected void during molecular dynamics and one provides a comprehensive picture across the void. The first two proposed methods use a specific heat map to present two types of information: an overview of all detected tunnels in the dynamics and their bottleneck width and stability over time, and an overview of a specific tunnel in the dynamics showing the bottleneck position and changes of the tunnel length over time. These methods help to select a small subset of tunnels, which are explored individually and in detail. For this stage we propose the third method, which shows in one static image the temporal

evolvement of the shape of the most critical tunnel part, i.e., its bottleneck. This view is enriched with abstract depictions of different physico-chemical properties of the amino acids surrounding the bottleneck. The usefulness of our newly proposed methods is demonstrated on a case study and the feedback from the domain experts is included. The biochemists confirmed that our novel methods help to convey the information about the appearance and properties of tunnels in a very intuitive and comprehensible manner.

Paper F

AnimoAminoMiner: Exploration of Protein Tunnels and their Properties in Molecular Dynamics

Jan Byška[1], Mathieu Le Muzic[2], Eduard M. Gröller[2,3], Ivan Viola[2,3], Barbora Kozlíková[1]

[1]Masaryk University, Brno, Czech Republic [2]Vienna University of Technology, Austria [3]University of Bergen, Norway

IEEE Transactions on Visualization and Computer Graphics, volume 22, number 1, pages 747-756, 10.1109/TVCG.2015.2467434, 2015.

Abstract

In this paper we propose a novel method for the interactive exploration of protein tunnels. The basic principle of our approach is that we entirely abstract from the 3D/4D space the simulated phenomenon is embedded in. A complex 3D structure and its curvature information is represented only by a straightened tunnel centerline and its width profile. This representation focuses on a key aspect of the studied geometry and frees up graphical estate to key chemical and physical properties represented by surrounding amino acids. The method shows the detailed tunnel profile and its temporal aggregation. The profile is interactively linked with a visual overview of all amino acids which are lining the tunnel over time. In this overview, each amino acid is represented by a set of colored lines depicting the spatial and temporal impact of the amino acid on the corresponding tunnel. This representation clearly shows the importance of amino acids with respect to selected criteria. It helps the biochemists to select the candidate amino acids for mutation which changes the protein function in a desired way. The AnimoAminoMiner was designed in close cooperation with domain experts. Its usefulness is documented by their feedback and a case study, which are included.

Paper G

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

Ondřej Strnad, Barbora Kozlíková, Vilém Šustr, Jiří Sochor

Masaryk University, Brno, Czech Republic

International Journal on Advances in Life Sciences, volume 5, number 3 & 4, https://www.thinkmind.org/download.php?articleid=biotechno_2013_4_20_60063, 2013.

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.

Paper H

Detection of Intramolecular Tunnels Connecting Sequence of Sites in Protein Structures

Ondřej Strnad, Barbora Kozlíková, Jiří Sochor

Masaryk University, Brno, Czech Republic

Advances in Intelligent Systems and Computing, Springer International Publishing, pages 73-80, 10.1007/978-3-319-07581-5_9, 2014.

Abstract

Proteins are essential for functioning of all living organisms and studying their inner structure and functions has been of a high importance. Many studies concentrated on detection of various inner structures inside macromolecules (e.g., tunnels, channels, pores) which play an essential role in the functioning of a large number of proteins. Here we present a novel approach to a detection of intramolecular tunnels. These pathways may facilitate the transport of reaction intermediates among buried active sites. The results obtained by the proposed algorithm were compared to intramolecular tunnels whose presence in given structures is already known. The algorithm is able to also identify other inner structures, such as channels or pores.

Paper I

Computational Analysis of Protein Tunnels and Channels

Jan Brezovský, Barbora Kozlíková, Jiří Damborský

Masaryk University, Brno, Czech Republic

To appear as chapter in Methods in Molecular Biology, Springer International Publishing, ISSN 1064-3745, 2016.

Abstract

Protein tunnels connecting the functional buried cavities with bulk solvent and protein channels enabling the transport through biological membranes represent the structural features that govern the exchange rates of ligands, ions and water solvent. Tunnels and channels are present in a vast number of known proteins and provide control over their function. Modification of these structural features by protein engineering frequently provides proteins with improved properties. Here we present a detailed computational protocol employing the CAVER software that is applicable for: (i) the analysis of tunnels and channels in protein structures, and (ii) the selection of hot-spot residues in tunnels or channels that can be mutagenized for improved activity, specificity, enantioselectivity, or stability.

Paper J

Visualization of Tunnels in Protein Molecules

Barbora Kozlíková, Filip Andres, Jiří Sochor

Masaryk University, Brno, Czech Republic

ACM Afrigraph – Proceedings of the 5th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa, pages 111-118, 10.1145/1294685.1294704, 2007.

Abstract

This paper presents two novel techniques for visualization of tunnels in complex molecules of proteins. Long-term research in the field of protein analysis proved that the reactivity of the protein molecule depends on the presence of tunnels. These structures are very important mainly in the process of finding new pharmaceuticals. Visualization of a tunnel is the next very important step after the analysis because it enables the biochemists to determine the crucial regions of the tunnel which can have a substantial effect in the process of designing new medication.

Previous methods for the visualizaton of tunnels are based on the definition of a tunnel as a set of intersecting spheres. Our approach exploits tetrahedra obtained from the process of tunnel analysis that is based on the Voronoi diagrams and Delaunay tetrahedrization. Thanks to the results of the analysis we obtain more precise definition of the tunnel so we can use it in the process of visualization.

We proposed two algorithms for the tunnel visualization. Both of them are based on the Delaunay tetrahedrization and visualize a tunnel as a surface. The surface is derived from the tetrahedra which form the boundary constraint of the tunnel in the space of the molecule.

Paper K

Application of Sampling-based Path Planning for Tunnel Detection in Dynamic Protein Structures

Vojtěch Vonásek[1], Barbora Kozlíková[2]

[1]Czech Technical University in Prague, Czech Republic
[2]Masaryk University, Brno, Czech Republic

Proceedings of MMAR: 21st International Conference on Methods and Models in Automation and Robotics, Miedzyzdroje, Poland, 10.1109/MMAR.2016.7575276, 2016.

Abstract

Behavior and properties of proteins as well as other bio-macromolecules is influenced by internal void space such as tunnels or cavities. Tunnels are paths leading from an active site inside the protein to its surface. Knowledge about tunnels and their evolution in time provides an insight into protein properties (e.g., stability or resistance to a co-solvent). Tunnels can be found using Voronoi diagrams (VD). To consider protein dynamics, that is represented by a sequence of protein snapshots, correspondences between VD in these snapshots need to be found. The computation of these correspondences is however time and memory consuming. In this paper, we propose a novel method for tunnel detection in dynamic proteins based on Rapidly Exploring Random Tree (RRT). The method builds a single configuration tree describing free space of the protein. The nodes of the tree are pruned according to protein dynamics. The proposed approach is compared to CAVER 3.0, one of the widely used freely available tools for protein analysis.

Paper L

Dynamic Visualization of Protein Secondary Structures

Matúš Zamborský, Tibor Szabó, Barbora Kozlíková

Masaryk University, Brno, Czech Republic

Proceedings of the 13th Central European Seminar on Computer Graphics, pages 147-152, http://www.cescg.org/CESCG-2009/papers/BrnoMU-Zamborsky-Matus.pdf, 2009.

Abstract

Visualization of molecular structures and their characteristics represents a very popular and extensive area of computer graphics, in which the researchers are intensively interested for the last decades. During this time there have been developed many methods for visualization of molecules, which are trying to satisfy the needs of biochemists. These methods are mainly designed for the visualization of the particular molecule in a static position. For the more complex visualization methods special techniques have to be implemented in order to obtain a plausible method for visualization of secondary structures in time space.

This paper presents the possible solution of this problem by introducing the animation of the main backbone of the protein molecule onto which the particular objects representing the secondary structures are bound. These objects are replicated as many times as necessary and are closely connected to form a solid structure representing the whole molecule. In order to achieve high frame rates we are using the advanced GPU features, such as fragment and vertex shaders.

Paper M

CAVER Viewer – New Tool Enhancing Computation and Visualization of Channels in Proteins

Barbora Kozlíková, Jiří Sochor, Tibor Szabó, Matúš Zamborský

Masaryk University, Brno, Czech Republic

Proceedings of the The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP), ISBN: 978-972-8939-22-9, 2010.

Abstract

Protein analysis and visualization is one of the most expanding fields in the bioinformatics research. Many applications were developed till now. These programs usually concern on some specific features of proteins and provide rich functionality showing different aspects of protein structure and behavior. However, the complex tool for protein analysis, which would provide users with advanced and integrated analytic and visualization methods, is still missing. In this article, we describe the CAVER Viewer as the next tile in the mosaic of protein visualization programs. CAVER Viewer provides enhanced interface for channel analysis and a set of traditional and new techniques for visualization of the results. With the CAVER Viewer, users can explore specific paths in protein structure, which are called channels. Channels may be explored in the three-dimensional space of the molecule using the two-dimensional information and statistics, visualized as 2D graphs and connected with the 3D displaying window. In comparison to the existing applications, our program also enables working with large data sets obtained from dynamic simulations of protein behavior.

Paper N

Computation and Visualization of Surface of Proteins and their Channels

Barbora Kozlíková, Igor Aleksandrowicz, Jiří Sochor

Masaryk University, Brno, Czech Republic

Proceedings of the The IADIS Computer Graphics, Visualization, Computer Vision and Image Processing (CGVCVIP), pages 99-106, ISBN: 978-972-8939-48-9, 2011.

Abstract

In this paper we are introducing our approach to the computation of molecular surface. This challenge occupies the researchers for many years because of its complexity and many solutions have been proposed. Our algorithm comes from the generally known reduced surface algorithm but in our case we innovate this approach using various simplifications and improvements, especially when dealing with various problems. This is done with respect to the demands to utilize our algorithm not only for the computation of molecular surfaces but also for the surface detection of so called channels inside proteins. Our approach is namely used as a part of the complex CAVER Viewer application designed specially for the localization, visualization and further exploration of channels in protein molecules.

Paper O

Visualizing Movements of Protein Tunnels in Molecular Dynamics Simulations

Barbora Kozlíková, Adam Jurčík, Jan Byška, Ondřej Strnad, Jiří Sochor

Masaryk University, Brno, Czech Republic

Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM), Vienna, Austria, 10.2312/vcbm.20141188, 2014.

Abstract

Analysis and visualization of molecules and their structural features help biochemists and biologists to better understand protein behavior. Studying these structures in molecular dynamics simulations enhances this understanding. In this paper we introduce three approaches for animating specific inner pathways composed of an empty space between atoms, called tunnels. These tunnels facilitate the transport of small molecules, water solvent and ions in many proteins. They help researchers understand the structure-function relationships of proteins and the knowledge of tunnel properties improves the design of new inhibitors. Our methods are derived from selected tunnel representations when each stresses some of the important tunnel properties – width, shape, mapping of physico–chemical properties, etc. Our methods provide smooth animation of the movement of tunnels as they change their length and shape throughout the simulation.

Paper P

Geometrical Detection of Pathways in Protein Structures Leading Among More Binding Sites

Ondřej Strnad, Vilém Šustr, Barbora Kozlíková, Jiří Sochor

Masaryk University, Brno, Czech Republic

Proceedings of BIOTECHNO 2014: The Sixth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, Chamonix, France, pages 93-98, https://www.thinkmind.org/download.php?articleid=biotechno_2014_5_20_60047, 2014.

Abstract

In this paper, we present a novel algorithm for the detection of pathways connecting two or more specific user defined binding sites, which are deeply buried in a protein macromolecule. These pathways can play an important role in the protein reactivity and overall behavior. However, our new algorithm can be generalized and used for computation of pathways inside an arbitrary set of spheres in three-dimensional space, leading through an ordered set of user-defined sites. Our approach is based on the localized Voronoi diagram approach and the Delaunay triangulation. The greatest benefit of our approach is its independence on the size of the input data set. This is achieved by using only a subset of all atoms in the macromolecule in each phase. This substantially reduces the size of the processed space. The method can also be utilized for determination whether pathways wide and straight enough exist among determined binding sites. This information then serves as the guideline for assessing the migration of products of chemical reaction between these binding sites.

Paper Q

Visibility-Based Approach to Surface Detection of Tunnels in Proteins

Adam Jurčík, Jan Byška, Jiří Sochor, Barbora Kozlíková

Masaryk University, Brno, Czech Republic

Proceedings of Spring Conference on Computer Graphics (SCCG), Smolenice, Slovakia, pages 85-92, 10.1145/2788539.2788548, 2015.

Abstract

Structural properties of proteins substantially influence their reactivity. Thus, the presence of pathways serving for transportation of a small molecule to the protein active site is crucial. These pathways, called tunnels, are defined by their surroundings – tunnel lining amino acids (or residues). In consequence, studying these amino acids and their properties is tightly connected with protein reactivity.

The set of tunnel lining amino acids detected for a given tunnel can differ with respect to selected algorithm for their computation. The criteria for evaluating their biochemical relevance are different as well, as they can depend on various physico-chemical properties. In this paper we firstly present a novel approach to the detection of tunnel lining amino acids. This approach is more robust in comparison with the existing methods. It is based on the visibility of atoms of the amino acids from the tunnel surface, often derived from the Voronoi diagram. Moreover, the detected set of amino acids is further utilized for an automatic detection of the asymmetric tunnel surface. The results are compared with already existing approaches and the benefits are discussed.

Paper R

Path-planning Algorithm for Transportation of Molecules through Protein Tunnel Bottlenecks

Jan Byška[1], Ivana Kolingerová[2], Barbora Kozlíková[1], Jiří Sochor[1]

[1]Masaryk University, Brno, Czech Republic
[2]University of West Bohemia, Plzeň, Czech Republic

Proceedings of Spring Conference on Computer Graphics (SCCG), Smolenice, Slovakia, pages 101-108, 0.1145/2788539.2788550, 2015.

Abstract

We present a simple and fast path planning algorithm for transportation of a set of tightly connected sphere objects (a small molecule) through a narrow gap. In our approach we are using common sampling-based path planning, however, instead of sampling the entire configuration space, we estimate which subsets of this space must be crossed on the desired path. In comparison with other methods using minimal bounding volumes, we improve the algorithm accuracy for arbitrary shaped molecules and significantly reduce the number of generated samples as well as time cost of path planning. We have accomplished a number of tests on scenes formed by proteins and ligand molecules. The results suggest that the proposed method works well in practice and the number of generated samples is substantially lower then the proved upper bound.

Paper S

Visualization of Biomolecular Structures: State of the Art

Barbora Kozlíková[1], Michael Krone[2], Norbert Lindow[3], Martin Falk[4], Marc Baaden[5], Daniel Baum[1], Ivan Viola[6,7], Julius Parulek[6], Hans-Christian Hege[3],

[1]Masaryk University, Brno, Czech Republic
[2] University of Stuttgart, Germany
[3]Zuse Institute Berlin, Germany
[4] Linköping University, Sweden [5]Laboratoire de Biochimie Théorique, France
[6]University of Bergen, Norway
[7]TU Wien, Austria

EuroVis – Eurographics Conference on Visualization – STARs, pages 61-81, 10.2312/eurovisstar.20151112, 2015.

Abstract

Structural properties of molecules are of primary concern in many fields. This report provides a comprehensive overview on techniques that have been developed in the fields of molecular graphics and visualization with a focus on applications in structural biology. The field heavily relies on computerized geometric and visual representations of three-dimensional, complex, large, and time-varying molecular structures. The report presents a taxonomy that demonstrates which areas of molecular visualization have already been extensively investigated and where the field is currently heading. It discusses visualizations for molecular structures, strategies for efficient display regarding image quality and frame rate, covers different aspects of level of detail, and reviews visualizations illustrating the dynamic aspects of molecular simulation data. The survey concludes with an outlook on promising and important research topics to foster further success in the development of tools that help to reveal molecular secrets.

Paper T

Accelerated Visualization of Transparent Molecular Surfaces in Molecular Dynamics

Adam Jurčík[1], Julius Parulek[2], Jiří Sochor[1], Barbora Kozlíková[1]

[1]Masaryk University, Brno, Czech Republic
[2]University of Bergen, Norway

Proceedings of the IEEE Pacific Visualization Symposium, 10.1109/PACIFICVIS.2016.7465258, 2016.

Abstract

The reactivity of the biomolecular structures is highly influenced by their structural features. Thus, studying these features along with the exploration of their dynamic behavior helps to understand the processes ongoing in living cells. This can be reached by the visual representation of these processes as visualization is one of the most natural ways to convey such information. However, none of the currently available techniques provides the biochemists with an intuitive real-time representation of the dynamic movements of molecules and precise geometrical based extraction of their structural features performed instantly. In this paper we introduce such a technique enabling the user to compute and also to visualize the molecular surface along with inner voids. To obtain a better insight into the molecule, our technique enables to visualize the molecular surface transparently. The opacity can be adjusted by changing user-defined parameters in order to enhance the perception of the surfaces of inner voids. All integrated algorithms run in real-time which gives the user a big variety of exploration possibilities. The importance of our approach is even amplified with respect to the fact that currently the size of molecular dynamics simulations is increasing dramatically and offline rendering thus becomes impracticable. The usability of our technique was evaluated by the domain experts.

Paper U

Unfolding and Interactive Exploration of Protein Tunnels and their Dynamics

Ivan Kolesár[1], Jan Byška[1,2], Julius Parulek[1], Helwig Hauser[1], Barbora Kozlíková[2]

[1]University of Bergen, Norway
[2]Masaryk University, Brno, Czech Republic

Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine (VCBM), Bergen, Norway, 2016.

Abstract

The presence of tunnels in protein structures substantially influences their reactivity with other molecules. Therefore, studying their properties and changes over time has been in the scope of biochemists for decades. In this paper we introduce a novel approach for the comparative visualization and exploration of ensembles of tunnels. Our goal is to overcome occlusion problems with traditional tunnel representations while providing users a quick way to navigate through the input dataset and to identify potentially interesting tunnels. First, we unfold the input tunnels to a 2D representation enabling to observe the mutual position of amino acids forming the tunnel surface and the amount of surface they influence. These 2D images are subsequently described by image moments commonly used in image processing. This way we are able to detect similarities and outliers in the dataset, which are visualized as clusters in a scatterplot graph. The same coloring scheme is used in the linked bar chart enabling to detect the position of the cluster members over time. These views provide a way to select a subset of potentially interesting tunnels that can be further explored in detail using the 2D unfolded view and also traditional 3D representation. The usability of our approach is demonstrated by case studies conducted by domain experts.

Paper V

Interactive Exploration of Ligand Transportation through Protein Tunnels

Katarína Furmanová[1], Miroslava Jarešová[1], Jan Byška[1,2], Adam Jurčík[1], Julius Parulek[2], Helwig Hauser[2], Barbora Kozlíková[1]

[1]Masaryk University, Brno, Czech Republic
[2]University of Bergen, Norway

To appear in the Proceedings of the IEEE Symposium on Biological Data Visualization (BioVis), Baltimore, USA, 2016.

Abstract

Background Protein structures and their interaction with ligands have been in the focus of biochemistry and structural biology research for decades. The transportation of ligand into the protein active site is often complex process, driven by geometric and physico-chemical properties, which renders the ligand path full of jitter and impasses. This prevents understanding of the ligand transportation and reasoning behind its behavior along the path.

Results To address the needs of the domain experts we design an explorative visualization solution based on a multi-scale simplification model. It helps to navigate the user to the most interesting parts of the ligand trajectory by exploring different attributes of the ligand and its movement, such as its distance to the active site, changes of amino acids lining the ligand, or ligand "stuckness". The process is supported by three linked views – 3D representation of the simplified trajectory, scatterplot matrix, and bar charts with line representation of ligand-lining amino acids.

Conclusions The usage of our tool is demonstrated on molecular dynamics simulations provided by the domain experts. The tool was tested by the domain experts from protein engineering and the results confirm that it helps to navigate the user to the most interesting parts of the ligand trajectory and to understand the ligand behavior.

Paper W

Comparative Visualization of Protein Secondary Structures

Lucia Kocincová[1], Miroslava Jarešová[1], Jan Byška[1,2], Julius Parulek[2], Helwig Hauser[2], Barbora Kozlíková[1]

[1]Masaryk University, Brno, Czech Republic
[2]University of Bergen, Norway

To appear in the Proceedings of the IEEE Symposium on Biological Data Visualization (BioVis), Baltimore, USA, 2016.

Abstract

Background Protein function is determined by many factors, namely by its constitution, spatial arrangement, and dynamic behavior. Studying these factors helps the biochemists and biologists to better understand the protein behavior and to design proteins with modified properties. One of the most common approaches to these studies is to compare the protein structure with other molecules and to reveal similarities and differences in their polypeptide chains.

Results We support the comparison process by proposing a new visualization technique that bridges the gap between traditionally used 1D and 3D representations. By introducing the information about mutual positions of protein chains into the 1D sequential representation the users are able to observe the spatial differences between the proteins without any occlusion commonly present in 3D view. Our representation is designed to serve namely for comparison of multiple proteins or a set of time steps of molecular dynamics simulation.

Conclusions The novel representation is demonstrated on two case studies. The first study aims to compare a set of proteins from the family of cytochromes P450 where the position of the secondary structures has a significant impact on the substrate channeling. The second study focuses on the protein flexibility when by comparing a set of time steps our representation helps to reveal the most dynamically changing parts of the protein chain.