

FWF-GAČR Project Proposal

between

TU Wien, Institute of Computer Graphics and Algorithms

Austria

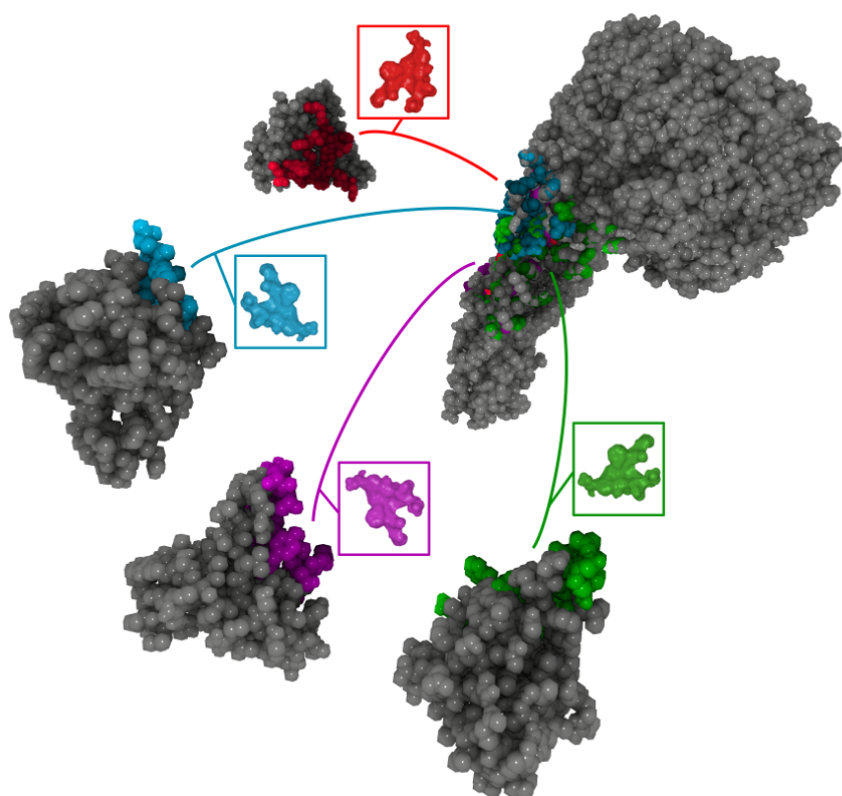
and

Masaryk University Brno, Department of Computer Graphics and Design

and Department of Functional Genomics and Proteomics

Czech Republic

Visual Analysis of Protein-Protein Interactions



submitted by

Assist. Prof. RNDr. Barbora Kozlíková, Ph.D., Assist. Prof. Dipl.-Ing. Dr.techn. Ivan Viola,

Assoc. Prof. Jan Paleček, Dr. rer. nat.

Vienna, Brno, March 2016

Contents

1	Introduction	1
2	Objectives	4
3	State of the Art	4
3.1	Computational Analysis of Proteins	4
3.2	Visualization of a Protein Surface	5
3.3	Interactive Visual Analysis and Visual Guidance	6
3.4	Our Own Previous Work and Competences	7
4	Technical Description	9
4.1	Stage 1: Comparative Visualization of Docking Ensembles	10
4.1.1	Visualizing Multidocking Search-Space	11
4.1.2	Comparison of Conformations from Different Computational Tools	12
4.1.3	Visual Analysis of Interactions	13
4.2	Stage 2: Formalization of Visual Analysis Systems	13
4.3	Integration of Visualization Technologies	15
4.4	Usability Testing in Proteomics Research	15
5	Project Management	16
6	Timetable	16
6.1	Dissemination Plan	17
6.2	Risk Analysis	17
7	Research Institution	18
7.1	Location	18
7.2	Personnel	18
7.3	Available Equipment	19
8	Requested Funding	19
	References	21

MISSION STATEMENT

To develop novel visualization technologies that will revolutionarize protein-protein interaction analysis workflows.

1 Introduction

Understanding the biological functions of proteins is a basal knowledge that contributes to advances in medicine, pharmaceuticals, but also agriculture and even energy sectors. The ultimate goal of proteomic research is to describe functions of all protein complexes in cellular space and life time. Protein interaction patterns are, however, very complex, and revealing these is a challenging and time-consuming scientific process. Most of the proteins critical for cellular life act in a cooperative manner forming multiprotein complexes. It is estimated that about 800 complexes exist in just one yeast cell. All complexes are composed of subunits which constitute the complex via protein-protein interactions. The main goal in studying such protein-protein interactions is to identify an appropriate spatial conformation of interacting proteins (see Figure 1). These studies span from investigating interactions between two protein structures to exploring large multi-molecular complexes. Every subunit of such complexes makes its contacts with neighboring subunits under a specific spatial orientation that is critical for the function of the whole complex.

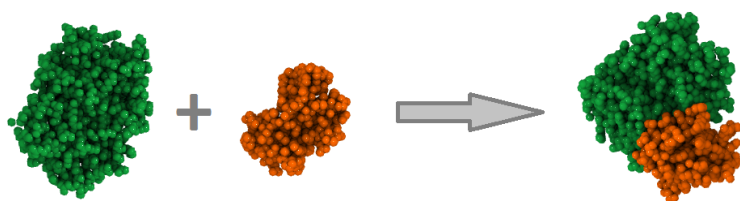


Figure 1: *Two protein chains and their mutual interaction.*

Protein structures studied with respect to their mutual interactions are primarily defined by the sequences of amino acids. This 1D sequential information does not give any explicit information about the spatial conformation of the protein chain. However, for studying interactions between two or more proteins, the knowledge of the 3D protein structure is crucial because it reveals the potential contact

surfaces between proteins. The process of creating the protein spatial conformation is called folding.

Subsequently, the set of these 3D conformations of proteins is tested for their mutual interaction. This process is called *molecular docking*. Existing approaches calculate a mutual conformation of two interacting proteins and suggest several most probable contact zones [14]. The subsequent task is to evaluate the biochemical relevance of these results. Such an evaluation is nowadays performed using manual comparisons or utilizing traditional 3D visualization methods which have been used in molecular visualization for decades. The main

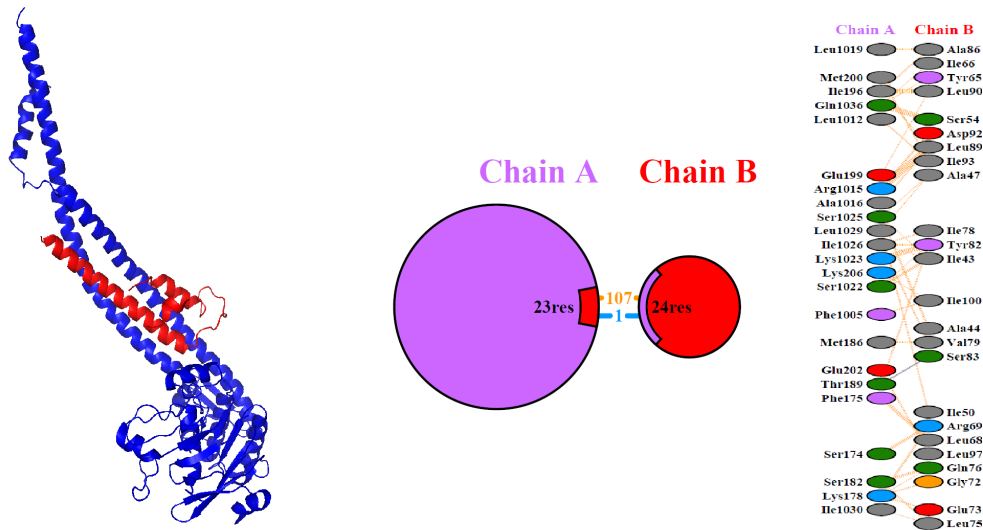


Figure 3: Existing visual representations of the crystal structure of complex SMC3-Scc1 with ID 4UX3 stored in the PDBsum database. Left: 3D structure, middle: abstracted representation of the contact zones between the chains, right: detailed interactions between individual amino acids as represented in the PDBsum database.

limitations of the existing methods are the occlusion caused by the 3D environment, hidden contact zones, and high overlaps when comparing more conformations (see Figure 2).

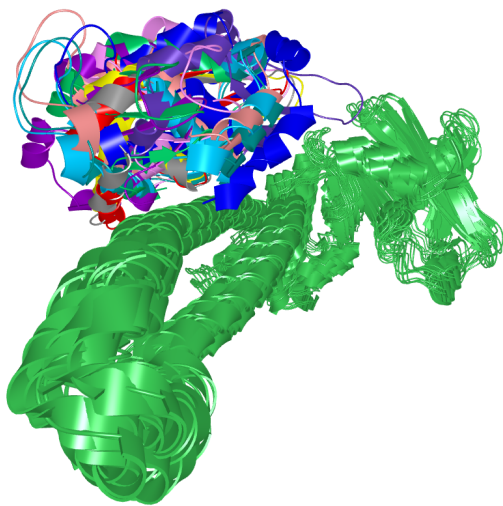


Figure 2: Superposition of several possible conformations between two proteins visualized using a traditional cartoon method. The set of green protein instances corresponds to one of the proteins in the interaction, the colored components represent the second protein in different conformations. Due to the high visual overlap, this visualization does not allow a clear comparison.

While there are many existing algorithms searching for protein-protein docking, researchers suffer from a lack of appropriate solutions for visual representations, explorations, and comparisons of the results. A similar situation holds for the already determined contacts between proteins which are available, e.g., in the PDBsum database [27]. The database currently contains thousands of entries for protein-protein interfaces which capture the 3D mutual orientation of the docked proteins using Jmol [1] and provides the users with an abstracted static representation of the contact zones (see Figure 3).

However, the existing visual representations do not allow the user to interactively explore the contact zones nor to compare different protein-protein interfaces. The information from the existing databases can serve as a reference or guidance and the visual comparison is a necessity. This functionality is not available in the existing visualization tools. Current visual workflows employ the widely used tools for molecular visualization such as PyMol [8] or VMD [16] which are targeting a broad user base and do not focus on specialized visualizations for a given task. In

consequence, the usage of general techniques is often impractical as it does not provide the necessary information to support a particular workflow. An example can be the superposition of several possible conformations of two interacting proteins illustrated in Figure 2. In this case, the proteins are represented by the cartoon visualization method and the colored proteins represent different possible positions with respect to the green (aligned) ones. It is obvious that such a cluttered representation cannot be used for an intuitive comparison of individual solutions.

With respect to the docking problem, the current solutions focus mainly on searching for possible interactions between two proteins. However, proteins can also form more complicated *complexes of proteins* where several protein structures interact with each other. For this process, called *multidocking*, the set of possible contact zones increases exponentially. In this case existing visualization tools do not even offer a minimally sufficient solution, and research for effective visual explorations, comparisons, and ranking is urgently needed.

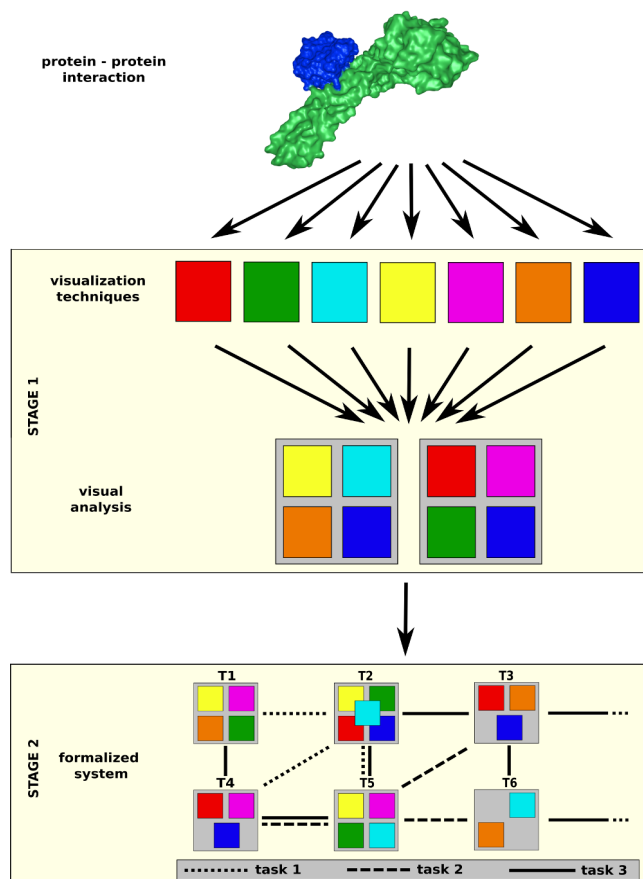


Figure 4: Overview of the project outline. In Stage 1 we will focus on designing specific visualizations for given tasks and building visual analysis systems consisting of these visualizations. Stage 2 proposes a formalized model of the visual analysis systems which will focus on specific tasks T_i .

As a part of the research project, the novel visualization technology will be tested with the ongoing biological research at the Department of Functional Genomics and Proteomics (DFGP), located at the Faculty of Science at the Masaryk University in Brno. The focus of interest of researchers at DFGP is on SMC (structure maintenance of chromosome), telomere-associated, and other chromatin complexes. The SMC complexes are key players in chromatin organization and dynamics (e.g., the condensin complex ensures a proper condensation of chromosomes – see Figure 5 taken from Paleček and Gruber [34]) and are the main focus of the Paleček’s group. Paleček’s group (group of Structural Proteins of Eukaryotic Chromosomes; SPEC) group analyzes the architecture and function of such complexes using a variety of experimental approaches. Their goal is to uncover the way how the subunits of these complexes interact with each other and execute unique function(s) of these complexes. Therefore, a visual representation of such information is highly beneficial because it helps to reveal the spatial relationships between the subunits in an intuitive way.

2 Objectives

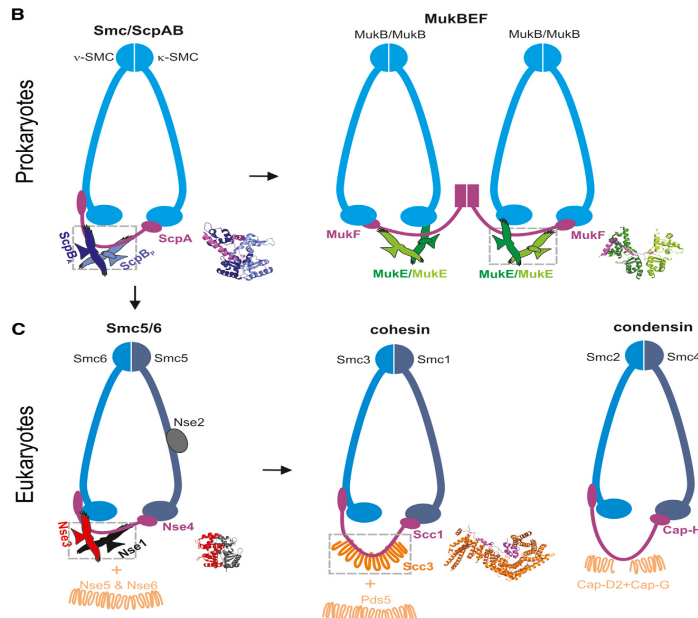


Figure 5: Architecture and composition of SMC complexes (taken from Paleček and Gruber [34]). These complexes will serve in the evaluation of the usability of our proposed visualizations and visual analysis tools.

3 State of the Art

The previous work related to the scope of our project consists of several parts. First, we briefly describe the basic principles of existing solutions for the computational analysis of protein structures, including protein-protein interactions. Then we review the existing approaches for the visualization of protein surfaces as the contact zones between proteins can be visualized as surfaces in 3D. This is followed by surveying methods for an interactive visual analysis and visual guidance through large molecular systems, which combine several representations in order to get a comprehensive view on the whole system.

3.1 Computational Analysis of Proteins

The spatial conformation of a protein greatly influences its reactivity and properties, thus many publications are related to the geometrical analysis of protein structure. Proteins can react with other structures, such as small molecules or other proteins, via chemical reactions or interactions occurring deeply inside the protein or on the protein surface. In the first case, the protein inner void space gives the opportunity for small molecules to enter the protein and perform a chemical reaction. The inner voids (called cavities) along with their entrance

The first objective O1 is to design novel visualization metaphors supporting particular tasks. The second objective O2 is to build system for visual analysis combining the proposed visualization methods into an intuitive and interactive environment where information from several visualizations are related to each other. The third objective O3 is to formalize the theory for visual analytics frameworks design given a specific set of data and analytical goals. The O1 advances visualization research with specific new methods, O2 provides an enabling technology for the domain of structural biology and O3 lays down the theory for visual analytics design. Figure 4 outlines the whole project and shows that this forms only one part of the planned activities.

paths (tunnels, channels, or pores), which are traversed by small molecules, have been intensively studied in the last decade. Most of the existing algorithms for the detection of entrance paths are based on simple Voronoi diagrams [5, 40, 43] or Voronoi diagrams of spheres [19, 32]. Benkaidali et al. [2] presented an alternative method using modified alpha shapes. In 2013, Krone et al. [25] published another approach for the detection of cavities, channels, and pockets, which was based on the concept of ambient occlusion. Krone et al. [23] further introduced a method for the detection and interactive exploration of protein cavities. Lindow et al. [30] presented a Voronoi based algorithm for the computation of cavities and channels from molecular dynamics trajectories. Their dynamic channels are computed by analyzing the temporal evolution of components of the cavity structure.

If a protein interacts with another protein structure, the computational analysis solves the problem of detecting interaction contact zones on the surfaces of these proteins. The currently existing search strategies for protein-protein docking can be divided into four categories. The first category consists of methods using an exhaustive global search which requires enormous computational resources. The other categories use local shape feature matching, randomized search, or post-docking approaches employing advanced scoring functions or consider protein flexibility. Details about these approaches as well as a list of existing solutions can be found in the survey of Huang [14].

Existing visual representations of computed structures (pathways or contact zones) mostly deal with their 3D spatial arrangement. However, such a 3D visualization is rather limited, mainly with respect to depicting many properties (physico-chemical, geometric) of the amino acids forming the pathways or contact zones. In such cases an abstracted representation is more appropriate. An example of this was presented by Heinrich et al. [13] where the authors use parallel coordinates to convey many attributes of amino acids (see Figure 6).

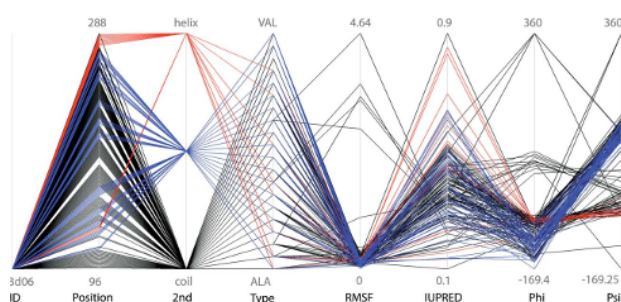


Figure 6: Parallel coordinates showing different amino acid attributes. Each amino acid is represented as a polyline. Image taken from Krone et al. [24].

Furthermore, when dealing with several proteins at once, the visual comparison of results in 3D is impossible to reach in a clear and comprehensible way.

3.2 Visualization of a Protein Surface

The protein surface plays a crucial role in interactions with the outer environment, including other molecules. In case of protein-protein interactions, the most important parts of the proteins are the mutual contact zones. These zones can be represented by one of the traditional models for visualization of their amino acids (e.g., space-filling model). One of the most common methods is to represent the contact zone by its surface.

There are many publications focusing on the visualization of molecular surfaces. Here we will mention only the latest results in this field. In 2012, Parulek and Viola [38] presented the first ray-casting of the so called solvent excluded surface that does not need a pre-computation of the analytical description of the surface. They use a modified sphere tracing and directly compute the implicit description of the surface based on the local neighborhood of a ray. In contrast to these direct ray-casting approaches, Decherchi and Rocchia [7] presented a new triangulation method based on ray-casting. It performs ray-casting through a 3D grid from which the surface is extracted using the Marching Cubes algorithm.

In our case the surface of a protein can be divided into two parts – contact zone (focus) and the remaining part (context). The contact zone has been studied in a method published by Lee and Varshney [29]. Their approach uses the hypothesis that the intermolecular volume and the area between docked molecules is useful as a measure of the shape complementarity. Therefore they propose an algorithm for the interactive computation of a so called intermolecular negative volume. In combination with dedicated visual representations (see Figure 7), it can serve the domain experts as an interactive tool for studying possible docking conformations.

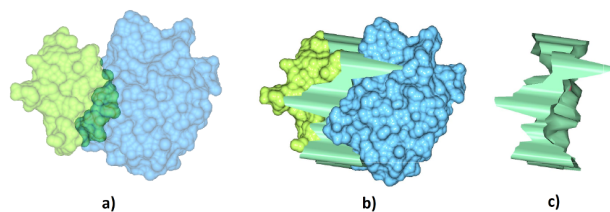


Figure 7: Visual representations of the contact zones proposed by Lee and Varshney. a) Translucent surfaces, b) Intermolecular negative volume between molecules, c) Intermolecular volume only. Image taken from Lee and Varshney [29]

The context part can be represented using some level of abstraction. Cipriano and Gleicher [6] presented a visualization technique which provides an abstracted view of the shape of molecules. Their approach suppresses small details to facilitate rapid comprehension but marks the location of significant but removed features so they remain perceivable.

Along with the algorithms and techniques for both analysis and visualization of macromolecules, several tools integrating individual techniques emerged. Some of them are quite widespread and currently serve the chemists in the visualization of results from various

analyses. The PyMol software tool [8] enables the users to integrate and render the results of their own analyses by employing Python scripting. The VMD tool [16] is a molecular visualization program which enables displaying, animating, and analyzing large biomolecular systems. UCSF Chimera [39] is an application for the interactive visualization and analysis of molecular structures and related data.

3.3 Interactive Visual Analysis and Visual Guidance

Here we mention several existing methods for an effective visual guidance through various tasks related to interactions between proteins (e.g., comparison or multidocking). Lindow et al. [31] introduced a tool for the interactive exploration of the dynamics of cavities and their volumes. Krone et al. [24] presented a visual

analysis application for protein simulations and inner cavities. Parulek et al. [36] utilize implicit surfaces for an interactive graph-based cavity-analysis of molecular simulations. They introduce linked views together with a 3D visualization of extracted cavities coupled with graph statistics in the context of the molecular surface (see Figure 8). In 2013, Parulek et al. [37] published an approach for the extraction of cavities and their visual exploration in molecular dynamics simulations.

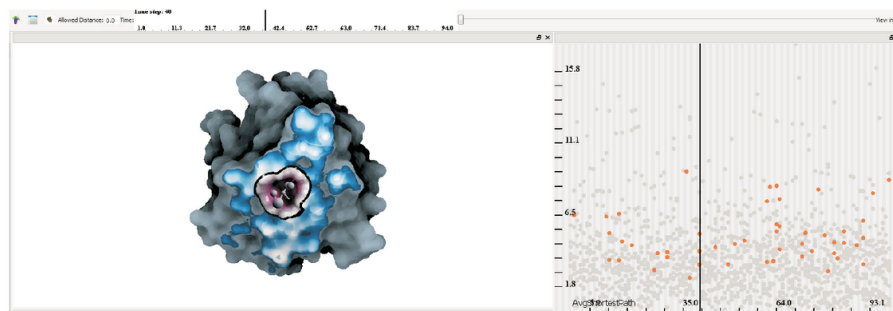


Figure 8: Combination of a 3D view onto the structure containing a cavity with scatterplots [36].

3.4 Our Own Previous Work and Competences

To reach the goals of this project, which focuses on the expressive visualizations of interactions between proteins and their comparison, their integration into the visual analysis tool, and the formalizing of visual analysis systems, we need to involve researchers with complementary expertise. The project requires specialists in computational analysis and visualization, visual analysis, and functional proteomics. Therefore our team has the following structure.

The project will be coordinated by Barbora Kozlíková from the Department of Computer Graphics and Design (DCGD) at the Masaryk University. Her research group is already actively cooperating with other project partners. The DCGD group has several years of experiences with research focusing on in-silico protein analysis and visualization. In cooperation with biochemists from the Loschmidt Laboratories at the Masaryk University, they have developed the CAVER 3.0 [5] tool for the computation of tunnels, channels, and pores inside protein structures. This tool was positively and widely accepted by the target community which is confirmed by many citations of the corresponding publication [5] (139 in WoS, to date 25-02-2016). Along with the computational tool, the DCGD team also developed the CAVER Analyst visualization tool [22] that enables the analytical visualization of detected pathways and their further evaluation through various statistical and abstracted representations. Here Kozlíková played a key role in the design of the tool and currently is the leader of the research group of CAVER Analyst. The tool integrates algorithms for fast rendering of molecular structures as well as their tunnels, both for static snapshots as well as for molecular dynamics simulations. The tool is multiplatform and is freely available to the community. The download package and the installation instructions are available from the software webpage (<http://www.caver.cz>).

The second partner, i.e., researchers from the SPEC group at the Masaryk University led by Jan Paleček, has a long term experience in studying protein complexes (see Paleček's CV). Their efforts are oriented towards SMC complexes [11, 15, 34]. The DCGD and SPEC groups already established a fruitful mutual cooperation in the field of protein-protein interactions which is realized via an internal grant of the Masaryk University, started in 2016. The cooperation involves also PhD students performing a thorough study of existing approaches for protein-protein docking. Both research groups are affiliated with the same university so personal meetings are already happening and are easily organized on a regular basis.

On the Austrian side, the project will be led by Ivan Viola, from the Institute of Computer Graphics and Algorithms (ICGA) at the TU Wien. Viola worked on the acceleration of rendering of very large molecular systems [26] based on geometry shaders. Later on, the group has focused on novel geometric representations using implicit surfaces [38], which is an enabling technology for the interactive visual analysis of surface cavities [36, 37]. The group also focused on an implicit hierarchical representation of molecules for the purpose of a seamless level of detail transition [35]. The latest publications focus on the illustrative visualization of large molecular systems and the visual guidance concerning ongoing reactions in these systems [28]. The group also recently published the cellVIEW tool [33] focusing on illustrative and multi-scale rendering of large biomolecular datasets. Large molecular systems were also studied with respect to attention guidance in dynamic narrative visualizations [42]. In cooperation with the University of Bergen, they presented a method for interactively illustrating polymerization using a three-level model fusion [20] which depicts physiological polymerization processes at different time scales.

Between October 2014 and March 2015, Barbora Kozlíková was a member of the team of Prof. Viola at ICGA during a research stay, establishing a close cooperation in molecular visualization. The success of this effort is already evidenced by four joint papers. The first paper is related to the abstracted representation of temporal changes in a protein tunnel bottleneck (narrowest site) and its surroundings [3]. It was presented at the EuroVis 2015 conference and also published in the Computer Graphics Forum journal ¹. The second paper again focused on protein tunnels and their temporal evolution [4]. This time the study concentrated on the whole tunnel along its centerline, changes of the tunnel shape, and exploration of the surrounding amino acids with respect to their properties and extent of influence of the tunnel. This paper was presented at the IEEE VIS 2015 conference and was published in the IEEE TVCG (Transactions on Visualization and Computer Graphics) journal ². The third paper was a state-of-the-art report on molecular visualization where nine authors from several institutions joined their efforts [21]. This report has been coordinated by Dr. Kozlíková and was accepted and presented at the EuroVis 2015 conference. The report was also invited for publication in the Computer Graphics Forum journal. The report was followed by a survey submitted to the EuroVis 2016 conference which is currently under review. It focuses on the analysis and visualization of molecular cavities.

¹<https://www.cg-forum.org/index.php/publications/computer-graphics-forum>

²<http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=2945>

The DCGD and ICGA groups were also successful in obtaining a mutual mobility project (the OeAD ICM through project CZ 17/2015) for 2015 and 2016.

4 Technical Description

The project is driven by the needs of the functional proteomics research, which to date does not take advantage of effective analytical workflows based on visualization. It is unclear how to extend existing molecular visualization approaches reviewed in the previous section that would lead to a satisfactory solution of the yet untackled protein-protein interaction problem. In our proposed solutions, we will start out from relevant real-world problems the researchers are facing when exploring and comparing different conformations of interacting proteins. In tight cooperation between the project partners, these problems will be systematized into a set of tasks. The proposed visual representations will then be designed to fulfill these tasks. Such tasks can be, e.g., T1: explore all possible conformations and easily filter out those conformations which are probably biochemically irrelevant, T2: briefly explore the individual conformations and their pairs of amino acids to support the decision process of task T1, T3: explore several conformations in a 3D view, along with the position of their contact zones, T4: explore the shape of the contact zones, T5: explore individual amino acids in the contact zones, their interaction and physico-chemical properties, T6: explore all possible conformations for multidocking, etc.

In the following sections we present the three main goals, divided into two main stages of the project, in detail. These stages are:

- Defining and systematizing the set of tasks and proposing novel visualization techniques enabling to fulfill these tasks. It will be followed by the design of an interactive visual analysis environment which integrates the selected proposed visualizations into one compact tool serving mainly for fast and intuitive exploration of the contact zones between proteins. The complexity of the proposed environment will be reduced by tracking the interaction history with individual views of the visual analysis system.
- Proposing a formalism of visual analysis systems dealing with heterogeneous data and complex tasks. Such a formalism will categorize the individual visualizations according to the tasks they are able to support. Then it will enable to automatically suggest a set of the most suitable visualizations according to a given set of tasks and these visualizations will be combined into a visual analysis tool.

The algorithms suggested on the basis of the below delineated research ideas will be validated by a prototypical implementation in the CAVER Analyst visualization framework. Their potential utility factor will be evaluated in the context of the proteomics domain problem.

4.1 Stage 1: Comparative Visualization of Docking Ensembles

One of the biochemist’s task in this stage is to select the most appropriate conformation for a particular pair of proteins. Typically a set of automated computational procedures is launched (see Section 3), resulting in an ensemble of potential contact zones. The other possible scenario for the use of the newly proposed visualization methods is to employ the existing information about known interactions between proteins (taken, e.g., from the PDBsum database). This information can be used as a reference when searching for the unknown contact between similar proteins (e.g., from the same family of proteins). A promising strategy is to combine these two approaches and, if possible, use the database entries to find the most biochemically relevant solutions in the ensemble calculated with a computational tool, such as HADDOCK [10] or PyDOCK [18].

For this purpose, we plan to design several visual representations which will convey different types of information about the interacting proteins. On the highest level, the users should have an overview of all conformations which should be explored. For this purpose, we suggest an abstracted interactive representation which enables the user to filter out conformations not fulfilling a given set of criteria (task T1).

As an example, one such crite-

tion could be the required pairs of interacting amino acids. Such a representation will combine a **matrix-based visualization** with additional views highlighting the interacting amino acids in a selected conformation (task T2). Figure 9 illustrates these ideas.

As the exploration of molecules in 3D space is very natural for the biochemists, we plan to focus on it as well. However, as already shown in Figure 2, simple superposition is not an acceptable solution as this leads to severe clutter and does not scale well with the number of conformations. Therefore we propose to adapt the technique of **exploded views** (see Figure 9) which will be equipped with a preview window (task T3). In these previews, positioned close to the moved (exploded) protein, we will be able to present different information about the contact zones (tasks T4 and T5). Contact zones are very important but easily affected by occlusion problems.

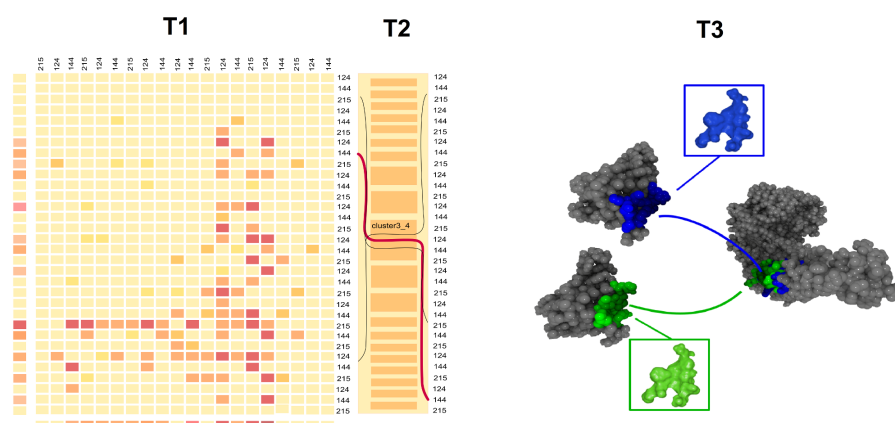


Figure 9: T1: matrix view showing an overview of all conformations, T2: lens view highlighting individual pairs of amino acids for the conformation in focus, T3: 3D exploded view enabling to show several solutions at once and to highlight the contact zones.

Therefore we propose to present a trimmed **semi-flattened view** of the contact zone (positioned perpendicularly to the user; Fig. 10a), the **neighborhood map** of individual amino acids in the contact zone (Fig. 10b), and their mutual distances using a **height map** (Fig. 10c). Here we suggest to embed a simplified contact zone representation into a comprehensive 3D context with the addition of supporting guiding lines. For individual conformations, these maps can be also visualized as illustrated in Figure 10d.

Among the approaches used for solving occlusion problems we plan to use aggregated visual representations which will find the best representative of a set of aligned reference proteins which are very similar with respect to their spatial orientation (see the set of aligned reference proteins in green in Figure 2). The representative can be selected, e.g., as the median, average, or the most central from all the representations of the reference protein. For distributions of a single random variable, box plots are already widely used (and already extended to curves). However, we have to extend these concepts to complex geometric structures like molecular conformations, ensembles of molecular surfaces, and ensembles of secondary structures.

Another possible solution will be the usage of reformations when we translate the problem to lower-dimensional representations which are simpler than the original 3D situation and which allow the user a standardized comparison. In the context of protein-protein interactions, this means, e.g., to display the contact zones directly on the molecular surface of the interacting proteins, flatten these surfaces and show many contact zones as overlapping contours in this simplified representation. As the data has to be viewed with different levels of detail, we will introduce domain-semantic based "levels-of-complexity". This will enable the user to explore, e.g., a contact zone, on different levels of abstraction or detail. To give an example, the contact zone can be viewed just as an entry in the list of all possible contact zones, it can be visualized as a surface in 3D, a 3D surface with indicated interacting amino acids, 3D surface with indicated properties of these amino acids, a 2D-flattened polygon, etc.

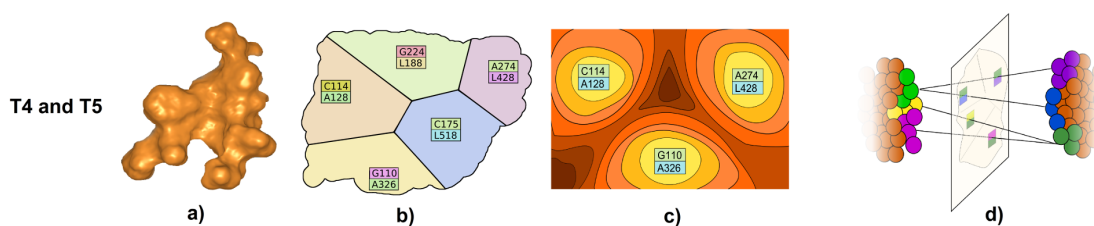


Figure 10: a) View on the contact zone (should be semi-flattened). b) Neighborhood map of individual amino acids in the contact zone. c) Height map showing the mutual distances between amino acids in the contact zone. d) For individual conformations, the preview can be positioned between the interacting proteins.

4.1.1 Visualizing Multidocking Search-Space

Initially, we will use available docking computations to generate ensembles of pair-wise dockings among all involved proteins forming the complex. Some contact zone candidates will naturally conflict with each other by competing for the same surface area. We will investigate visualization metaphors such as a pairwise mul-

tidocking matrix (see Figure 11), where the analyst can choose a particular contact among two proteins (task T6). Attention will be guided to contact zones with other proteins from the same complex, which are affected by *occupying* the surface area by selected interaction. As the interacting proteins form complexes consisting of more than two proteins (results of multidocking), the corresponding visual representations will also face this increased clutter and occlusion challenge. It is an open question how to tackle it and whether or not exploded views are still a meaningful metaphor for conveying structural propoerties of the interaction.

Multidocking establishes a relationship between several molecules that form the molecular complex. In this sense, the relationship in the contact zone set is no longer binary, but $n > 2$ -nary, and an edge becomes a hyperedge in a hypergraph. We envision a contact-zone hypergraph to depict all possible relationships in the molecular complex. A hyperlink will essentially define one possible multidocking complex. From such a visual representation, we can see all possible contact scenarios, which contact zones are central, and we can rank these according to the overall or minimal stability.

We will experiment with taking the hypergraph as a basis for a heterogeneous visualization and will augment it with structural details. Even if these will be severely distorted, the visualization will still communicate important geometric, physical, and chemical properties. This way, we will experiment with heterogeneous visualizations where the underlying relationships serve as the basis, and the structure augments further aspects on top of the basis.

During the analysis, the history of contact-zone evaluation will be provided so that the analyst avoids reevaluating already assessed surfaces. Visual guidance has the potential to improve this analytical stage substantially, as minimal-conflict high-probability contact complexes can be suggested to the analyst for a prioritized assessment over scenarios that are far less likely to form a molecular complex.

After a particular selection of a preferred multidocking scenario, MD simulations will further assess the stability of the given setup. Should the tested complex scenario turn out to be unstable, this information is then propagated back into the contact-zone hypergraph representation, so that other constellations will be prioritized in ranking and consequent visual analysis.

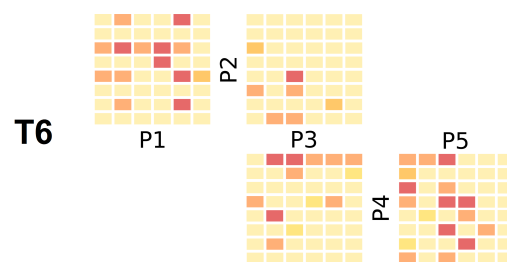


Figure 11: Interactive multidocking matrix enabling to show the overview of all conformations between proteins P1 – P5.

4.1.2 Comparison of Conformations from Different Computational Tools

If the source of the input data is made-up by the results of various computational tools, it is not advisable to simply combine them into one common input. The main reason is that these tools produce slightly different

output parameters (e.g., each tool computes its own scores). Therefore, the proposed visualizations should need to be extended. For example, each cell in the matrix visualization illustrated in Figure 9 will be divided according to the number of used computational tools and each part will correspond to one tool. Visual reduction operations based on semantic domain knowledge will reduce visual complexity in the context region. Such operations could be: take the worst/best solution, just highlight high-variance regions, indicate biased areas, etc.

4.1.3 Visual Analysis of Interactions

Subsequently, we will propose several combinations of the designed visualization techniques into visual analysis tools aiming to provide the users with an intuitive and interactive environment supporting their tasks. However, the input data as well as the tasks can be highly complex which will in consequence lead to too complex visual analysis tools containing too many linked views. As a solution to this problem we will propose a semi-automatic simplification of such tools through adaptation. It will be based on the tracking of user interaction with a given tool when performing a specific task. Based on the interaction history with individual views (i.e., how often the view was used by the user) we will decide which views should be removed or replaced by other ones. This user-centered approach to the design of the visual analysis tool tailored to a specific set of tasks will enable to utilize such a tool very effectively.

All visualizations used in the tool will be interlinked and by using advanced brushing methods the tool will serve exploration, comparison, simplification, and other required functionality.

4.2 Stage 2: Formalization of Visual Analysis Systems

If we look back in the visual analytics (VA) research in the past decade, there have been many VA frameworks proposed, which aid the completion of a particular visualization task. Many of them appear to have a similar high-level pattern, the mostly utilized pattern is the juxtaposition of different views which reveal one or few aspects of a multi-dimensional phenomenon. A VA framework typically provides views for each aspect that is necessary to understand to accomplish a given analytical task. These aspects are linked together so that their mutual relationships can be understood as well. These are the famous multiple coordinated views.

Our task analysis, of which preliminary results are listed in the previous section, offers several scenarios for which a tailor-made VA framework can be designed. Until now, this has been done by visualization experts themselves. However, in case we have a clear structure about the aspects which describe a given phenomenon and at the same time we have clearly defined tasks, we hypothesize that it should be possible to design the VA framework algorithmically. In the second stage of the proposed research we will focus on developing a formalism for a VA design and contribute to the theoretical foundations to this emerging subfield of visualization.

To propose a formal systematization for such structured data, the overall goal is to enable a systematic

exploration of the underlying design space of possible visual representations and interactions. Until now, when designing visualizations for a specific target domain, the typical procedure is to perform an analysis of requirements and according to them design the most suitable visualizations for the whole visual analysis system. However, this obviously leads to a constantly repeating process (even for slight task changes) which is based on the same principles and mostly ad-hoc approaches have been used to propose a specific visual representation for a specific task situation at hand. Therefore, a drill down into relevant subparts of the highly structured data must be done with typical database queries but with a strong domain semantic from our application area. Operations used for the extraction of these subparts could be, e.g., slicing, dicing, projections, aggregations, etc. For each of these operations the appropriate visual templates need to be developed.

The selection of relevant subparts will be driven by the user-defined tasks. These tasks are likewise highly complex and are usually formed by the combination of much simpler basic tasks, such as T1, ..., T6 mentioned in the Stage 1 of the proposal. As the set of basic tasks can be large and combining them creates a vast space of possible queries, we need a systematic formalization here as well. In order to propose the formal system based on different user tasks, the definition of these tasks has to be structured as well. We will base our considerations on the recent task taxonomies in the visualization research literature.

Instead of defining the VA design explicitly, the specification will be of a declarative nature so that even a domain specialist would be able to implicitly author the VA framework. For specification we will create a domain-specific language for the definition of these tasks which will serve as a meta-language for the automatic selection of relevant visualizations and their integration into the overall visual analysis system. The syntax of the language will combine traditional commands of programming languages with naming conventions of the given domain. The functions derived from a set of basic as well as complex tasks will form the basis of the semantic meaning of the language. Here we will be inspired by a feature definition language proposed by Doleisch et al. [9] who presented a framework for flexible and interactive specification of high-dimensional and complex features in simulation data. It allows the definition of one or several features by brushing multiple dimensions and highlighting the selected features in all linked views. In our case, such features of the input data are, e.g., two interacting proteins, a set of interacting proteins (larger than two), contact zones, interacting amino acids, temporal changes of the proteins, etc.

A naive approach to build a VA workflow would be to include all possible visualizations from the predefined set to show all possible kinds of relationships. If the display estate is huge, such approach is possible to implement. However, this would result in too much of visual information to be effectively consumable during the analysis workflow. Humans have only a limited field of view, they could be easily overwhelmed, so the display space needs to be utilized that only most crucial visualization views are included in the VA framework design. The VA design would then be constrained from both sides, the necessity to show particular visual information and the availability of visual estate. We will investigate whether force-based approaches would be

suitable for resolving the constraints. At the first glance procedural modeling of architecture might sound like not very relevant for automated design of VA frameworks. Especially the recent works on interior design or scaffold structure deal with a similar problem. On one hand there is the desired functionality, on the other hand there are technical norms and space limitations that constraint the design choice. While it is at this moment rather open what is a feasible solution for automated VA design, we will seek for inspiration in related visual computing domains to tackle such high-risk high-gain research objective.

4.3 Integration of Visualization Technologies

The visualization research described in the previous sections is motivated by the needs of analyzing data in proteomics and structural biology. The ambition is to utilize these technologies in new analytical workflows. We will therefore build a meaningful selection of the proposed approaches into one software framework, integrated into the CAVER Analyst visualization software. As a part of this system integration, we plan to perform studies on the usability of the entire system, and we will employ iterative design strategies for making the system intuitive to use by the experts. This will be primarily done during the technological development. In relation to the validation part of the proposed research described below, where the system will be employed in an actual scientific workflow, we will assess the system's accuracy and stability with respect to the experimental data provided by the domain experts.

4.4 Usability Testing in Proteomics Research

The goal of this research focus is to apply and test the above visualization technologies on real SMC complexes. The SMC complexes have a unique architecture suited for their role as chromatin organizers [34]. The SMC subunits form rings (see Figure 5) which can be temporarily opened to load or release chromatin fibre [12]. Specific bindings of non-SMC (NSE) subunits at one end of the SMC ring regulates its closure. Mapping of these interactions provides the information about the positioning of the NSE subunits and may suggest loading/exit paths for DNA fibres.

The new software framework developed within this project will enable us to compose and visualize a complete hypothetical SMC complex with its dynamics. Such an insight will help us to advance the understanding of the SMC architecture and its dynamics.

In this validation stage of the project, we will analyze all the available (>10 PDB) crystal structures of prokaryotic and eukaryotic SMC complexes [34]. We will model protein structures of their homologs (e.g., yeast proteins will be modeled based on bacterial and human crystal structures), combine partial information from these different structures and make a general model of a complete SMC complex. Similarly, other complexes studied at the group of DFGP (e.g., telomere-associated shelterin complex [17, 41]) will be analyzed and presented within the semestral course of "Protein complexes" (taught by Dr. Paleček).

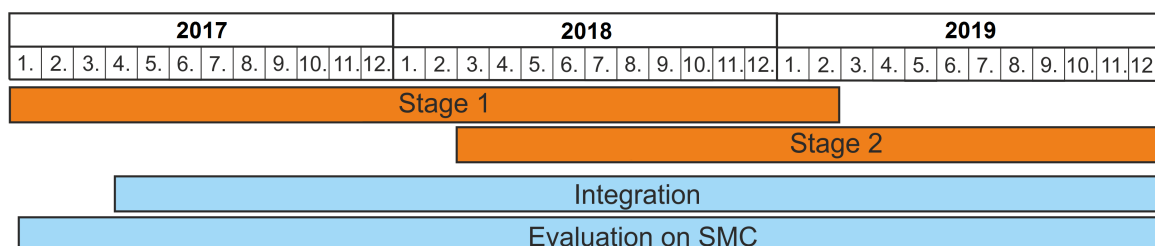


Figure 12: The proposed timetable.

5 Project Management

The project management can be divided into research activities and project administration. The research activities will be coordinated by Barbora Kozlíková and Ivan Viola. They will be responsible for the scientific results and progress of the project. The administration of the project will be managed by Barbora Kozlíková and Ivan Viola, the representatives of the project on the Czech and Austrian side, respectively. Dr. Kozlíková and Prof. Viola will be responsible for the scientific and financial reports, which will be prepared jointly by both partners and their financial administrators.

The coordination of both partners will be carried out by frequent meetings. Because of the geographical distribution and interdisciplinary nature of the project, special coordination and communication strategies are necessary. This includes a database for storing joint literature and publications and joint repositories for easy software deployment. The shared database and repositories will be maintained using the graphical workstation required for this project. Moreover, we will organize regular project meetings (four meetings of the whole project team per year) and yearly workshops, where our latest results will be presented to colleagues and students at the corresponding departments. We will also establish a collaboration on an institutional level, where we will arrange several reciprocal internships of young researchers (via mobility projects or the ERASMUS+ study program).

The initial project meeting will be organized in January 2017 in Brno, the project end meeting will be held in Vienna. Regular meetings will be held in Vienna, in Brno or half way between these two cities, in Mikulov.

6 Timetable

The project timetable, illustrated in Figure 12, details the time spread of the individual stages along with the integration and evaluation steps.

Stage 1 will be coordinated by Dr. Kozlíková. In this stage, a very tight cooperation with experts from SPEC will be crucial, so the personal meetings between members of DCGD and SPEC will be held on a regular (weekly) basis. The design of the novel methods will gain from the experience of the ICGA group, and the preliminary results will be immediately tested and evaluated by the SPEC group. Stage 2 will be coordinated by Prof. Viola and along with Dr. Kozlíková and their teams they will cooperate on the design of the proposed

formal system. The design of the system will highly depend on the user requirements which will come from Dr. Paleček who will be also responsible for coordinating the evaluation and testing phase.

All proposed methods will be integrated into one software framework, CAVER Analyst. This activity will be coordinated and mostly implemented by Dr. Kozlíková as her research group is the author of this software tool. The methods will be intensively tested on the SMC complexes and this testing will take place throughout the entire project.

6.1 Dissemination Plan

This basic research project on visualization will be published first in primary scientific venues for visualization research. These include IEEE TVCG (IEEE VIS Conference) and EG CGF (EG EuroVis Conference). Focused events on visualization of biological data, such as IEEE BioVis (in particular in connection with BMC Bioinformatics) or EG VCBM will be also considered for publication of the project's results. We expect approximately 8-10 visualization papers published in primary venues on visualization.

The integration into a system for visualization and analysis of protein-protein interactions will be published in a bioinformatics journal. Currently we have BMC Bioinformatics in mind. In case the new workflow enables a major scientific breakthrough, we will prepare a Nature Methods paper.

New tools will be presented to regular master students subscribed to the Proteomics course and the course on Structure and function of protein complexes taught by Dr. Palecek. For presentation and educational purposes, we also aim to print the sub-complexes in a selected spatial conformation on a 3D printer.

6.2 Risk Analysis

Risks with respect to the quality and usability of project outputs are related to the performance of the created framework, individual proposed methods, and the accuracy and scalability of the developed methods. Other risks are: the graphical user interface (GUI) of the framework will not be accepted by domain experts, some of the proposed visualization methods will suffer from visual clutter, the proposed interaction techniques will be too complex, there will be still unknown limitations in the target domain preventing to create a sufficient formal system. To avoid these problems, we will profit from our previous experience in creating optimized accurate software tools [5], as well as in designing domain specific visualization, GUI and interaction [22], and many years of expertise in the visualization and visual analysis domain (ICGA).

Numerous structures of protein complexes are available in the PDBsum database, therefore, if the SMC complex focus fails we can refocus on to the other DFGP studied complexes or any PDBsum (chromatin-associated) complex familiar to Dr. Paleček.

7 Research Institution

7.1 Location

On the Austrian side, the research will take place at the Computer Graphics Group of ICGA, Vienna University of Technology, which performs extensive fundamental and applied research in computer graphics and visualization. Its areas of expertise are modeling and rendering, scientific visualization, virtual environments, and computer animation. In the field of scientific visualization the research is focused on advancing medical visualization, in particular medical ultrasound, illustrative visualization, knowledge-assisted visualization, visual analysis of parameter spaces, and point-based visualization. The Visualization Group at ICGA has also developed and implemented an open-source visualization prototyping environment, i.e., VolumeShop. In the area of scientific visualization there are long standing cooperations with the Austrian Academy of Sciences, the General Hospital in Vienna, VRVis, Austria; University of California, Stanford University, USA; University of Tübingen, University of Stuttgart, University of Rostock, University of Magdeburg, Germany; INRIA Research Institute, University Joseph Fourier, France; University of Girona, Spain; TU Budapest, Hungary; CTU Prague, Masaryk University in Brno, Czech Republic; Comenius University in Bratislava, Slovakia; and University of Bergen, Norway.

The Czech part of the team consists of two departments at the Masaryk University in Brno. The first research group comes from the Department of Computer Graphics and Design (DCGD) located at the Faculty of Informatics of the Masaryk University. The computer graphics group consists of two associate professors, two assistant professors and nine PhD students. The main research areas of the group are human-computer interfaces, virtual reality, haptic interfaces, computational geometry and visualization.

The second group of researchers from the Masaryk University involved in this project is the SPEC group of Dr. Paleček. The main research areas of Paleček's group are architecture, structure, function, and evolution of protein complexes. Its members focus on studying protein-protein and DNA-protein interactions using a wide range of biochemical approaches (e.g., yeast two-hybrid, pull-down, immunoprecipitation, peptide libraries, mutagenesis, etc.). In addition, the Paleček lab collaborates with NMR and cryoEM (CEITEC) and crystallographic groups (at Sussex University, United Kingdom).

7.2 Personnel

Dr. Barbora Kozlíková, the project coordinator, is an assistant professor at DCGD, Faculty of Informatics, Masaryk University in Brno. In 2011 she has received her PhD from the same university. She focuses mainly on the interdisciplinary CAVER project which integrates computational geometry and visualization. She is the co-leader of the Human-Computer Interaction Laboratory (www.decibel.fi.muni.cz).

Prof. Dr. Ivan Viola is an assistant professor at ICGA, Vienna University of Technology (VUT) and adjunct professor at the University of Bergen, Norway. In 2005 he has received his PhD from VUT. In 2011,

he received an WWTF Vienna Research Groups award which provides basic financing of his group for eight years. His research interests include computer graphics, visualization, with particular focus on illustrative visualization. He is co-heading the visualization group at ICGA. The group performs basic and applied research projects in the area of scientific visualization.

Assoc. Prof. Jan Paleček is currently heading his own Structural Proteins of Eukaryotic Chromosomes group at DFGP (Faculty of Science) at the Masaryk University. In addition, he is a senior researcher in Genomics and Proteomics of the Plant Systems program at CEITEC MU. He received his Dr.rer.nat. (equivalent to a PhD in biochemistry) from the Vienna University (Austria) in 1998 and habilitated (in molecular biology) at the Masaryk University (Czech Republic) in 2012. He is interested in protein-protein interactions and protein complexes. Since his postdoctoral stay at Sussex University (United Kingdom, from 2002 till 2006), he focuses on the architecture and evolution of the SMC and MAGE protein complexes. Dr. Paleček was awarded with a Czech Science Foundation (CSF) grant to study a "Role of non-SMC proteins in stabilization and processing of stalled replication forks" (GA13-00774S, 2013–2016).

7.3 Available Equipment

All participating institutes will provide the basic research infrastructure, including workplaces, basic hardware, licenses for operating systems and development tools, internet connection, and office supplies for their team members. Computation time on a grid architecture will be provided free of charge by the NGI Metacentre <http://www.metacentrum.cz/> if needed.

8 Requested Funding

Personnel Apart from the applicants and the staff at the institutions who will participate in this research project, we request two full-time PhD positions, who will conduct the proposed research full time. One position will be covered by three part-time contracts for Katarína Furmanová, Adam Jurčík and one of the new PhD student. They will be supervised by Dr. Barbora Kozlíková. Additionally to this, Dr. Paleček will be 20% engaged in the project. For the junior researcher at ICGA we plan to recruit a suitable candidate from new graduates of ICGA or graduates from other universities, who will meet the requirements. This researcher will be supervised by Prof. Dr. Ivan Viola.

Total personnel costs AT = 109 980,00 EUR

Total personnel costs CZ = 87 651,17 EUR

Total personnel costs (including insurance): 197 631,17 EUR

Equipment To efficiently conduct our research project we request one high-performance graphics workstation, software + updates (mainly graphics cards).

Workstation = 11 826,77 EUR

Materials On the Czech side, the material costs will be used for high-performance graphics cards. As we plan to produce high-quality color posters to present our project at conferences and important local events, we involve these costs into the budget as well. Here we include also costs for organizing yearly project workshops with the participation of the involved researchers and participation of external experts.

Posters and 3D prints and workshops AT = 1 500,00 EUR

Graphics cards, posters, 3D prints, workshops CZ = 5 991,35 EUR

Total material costs: 7 491,34 EUR

Travel The results of the project will be presented at international conferences. We expect to publish about 5 conference papers per year, resulting in a total of 15 active participations of project members, 9 of which we expect to be in the US.

Travel costs AT = 9 000,00 EUR

Travel costs CZ = 13 200,00 EUR

Total travel costs: 22 200,00 EUR

Other Costs Other costs consist of conference fees (in CZ they fall into this category) and open-access publication fees. We include also overheads for the project administration, allowed by GAČR (20% of the CZ budget) and FWF (5% of the AT budget).

Publications AT = 4 500,00 EUR

Conference fees CZ = 4 438,03 EUR

Overheads AT = 6 840,34 EUR

Overheads CZ = 27 737,71 EUR

Total other costs: 43 516,08 EUR

Summary of Requested Funding

	1 st year	2 nd year	3 rd year	Total	
Personnel costs	65 877,06	65 877,06	65 877,06	197 631,17	EUR
Equipment costs	11 826,77	0,0	0,0	11 826,77	EUR
Material costs	2 497,11	2 497,11	2 497,11	7 491,34	EUR
Travel costs	7 400,00	7 400,00	7 400,00	22 200,00	EUR
Other costs	14 505,36	14 505,36	14 505,36	43 516,08	EUR
Yearly costs	102 106,3	90 279,53	90 279,53	282 665,36	EUR
Total requested costs:				282 665,36	EUR

References

- [1] Jmol: an open-source java viewer for chemical structures in 3D. URL: <http://jmol.sourceforge.net/>.
- [2] Lydia Benkaidali, François André, Boubekeur Maouche, Pridi Siregar, Mohamed Benyettou, François Maurel, and Michel Petitjean. Computing cavities, channels, pores and pockets in proteins from non-spherical ligands models. *Bioinformatics*, 30(6):792–800, 2014.
- [3] Jan Byska, Adam Jurcik, M. Eduard Groeller, Ivan Viola, and Barbora Kozlikova. MoleCollar and Tunnel Heat Map Visualizations for Conveying Spatio-Temporo-Chemical Properties Across and Along Protein Voids. *Comput. Graph. Forum*, 34(3):1–10, 2015.
- [4] Jan Byska, Mathieu Le Muzic, M. Eduard Groeller, Ivan Viola, and Barbora Kozlikova. AnimoAminoMiner: Exploration of Protein Tunnels and their Properties in Molecular Dynamics. *IEEE Trans. Vis. Comput. Graph.*, 22(1):747–756, 2016.
- [5] Eva Chovancová, Antonín Pavelka, Petr Beneš, Ondřej Strnad, Jan Brezovský, Barbora Kozlíková, Artur Gora, Vilém Šustr, Martin Klvaňa, Petr Medek, Lada Biedermannová, Jiří Sochor, and Jiří Damborský. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. *PLoS Computational Biology*, 8(10), 2012. URL: <http://dblp.uni-trier.de/db/journals/ploscb/ploscb8.html#ChovancovaPBSBKSGKMBSD12>.
- [6] G. Cipriano and M. Gleicher. Molecular surface abstraction. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1608–1615, Nov 2007.
- [7] Sergio Decherchi and Walter Rocchia. A general and robust ray-casting-based algorithm for triangulating surfaces at the nanoscale. *PLoS ONE*, 8(4):e59744, 2013. URL: <http://dx.doi.org/10.1371/journal.pone.0059744>.
- [8] W. DeLano. PyMOL: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography*, 40, 2002. <http://pymol.sourceforge.net/>.
- [9] Helmut Doleisch, Martin Gasser, and Helwig Hauser. Interactive feature specification for focus+context visualization of complex simulation data. In *Proceedings of the Symposium on Data Visualisation 2003, VISSYM '03*, pages 239–248, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association. URL: <http://dl.acm.org/citation.cfm?id=769922.769949>.
- [10] Cyril Dominguez, Rolf Boelens, and Alexandre M. J. J. Bonvin. Haddock: a protein–protein docking approach based on biochemical or biophysical information. *Journal of the American Chemical Society*, 125(7):1731–1737, 2003. PMID: 12580598.
- [11] Marc Guerineau, Zdenek Kriz, Lucie Kozakova, Katerina Bednarova, Pavel Janos, and Jan Palecek. Analysis of the Nse3/MAGE-binding domain of the Nse4/EID family proteins. *PLoS ONE*, 7(4):e35813, 04 2012. URL: <http://dx.doi.org/10.1371/journal.pone.0035813>.
- [12] C. H. Haering and S. Gruber. SnapShot: SMC Protein Complexes Part II. *Cell*, 164(4):818–818, Feb 2016.
- [13] Julian Heinrich, Michael Krone, Daniel Weiskopf, and Sean O’Donoghue. Visualising intrinsic disorder and conformational variation in protein ensembles. *Faraday Discussions*, 169:179–193, 2014.
- [14] S. Y. Huang. Search strategies and evaluation in protein-protein docking: principles, advances and challenges. *Drug Discov. Today*, 19(8):1081–1096, Aug 2014.
- [15] Jessica J. R. Hudson, Kateřina Bednářová, Lucie Kozáková, Chunyan Liao, Marc Guerineau, Rita Colnaghi, Susanne Vidot, Jaromír Marek, Sreenivas R. Bathula, Alan R. Lehmann, and Jan Paleček. Interactions between the Nse3 and Nse4 components of the SMC5-6 complex identify evolutionarily conserved interactions between MAGE and EID families. *PLoS ONE*, 6(2):e17270, 02 2011. URL: <http://dx.doi.org/10.1371/journal.pone.0017270>.
- [16] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.
- [17] E. Janoušková, I. Nečasová, J. Pavloušková, M. Zimmermann, M. Hluchý, V. Marini, M. Nováková, and C. Hofr. Human Rap1 modulates TRF2 attraction to telomeric DNA. *Nucleic Acids Res.*, 43(5):2691–2700, Mar 2015.

- [18] Brian Jimenez-Garcia, Carles Pons, and Juan Fernandez-Recio. pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics*, 29(13):1698–1699, 2013.
- [19] Deok-Soo Kim, Youngsong Cho, Jae-Kwan Kim, and Kokichi Sugihara. Tunnels and voids in molecules via Voronoi diagrams and Beta-complexes. In Marina L. Gavrilova, C.J. Kenneth Tan, and Bahman Kalantari, editors, *Transactions on Computational Science XX*, volume 8110 of *Lecture Notes in Computer Science*, pages 92–111. Springer Berlin Heidelberg, 2013.
- [20] Ivan Kolesár, Julius Parulek, Ivan Viola, Stefan Bruckner, Anne-Kristin Stavrum, and Helwig Hauser. Interactively illustrating polymerization using three-level model fusion. *BMC Bioinformatics* 2014, 15(345):1–16, October 2014. URL: http://www.cg.tuwien.ac.at/research/publications/2014/Viola_Ivan_IIP/.
- [21] Barbora Kozlikova, Michael Krone, Norbert Lindow, Martin Falk, Marc Baaden, Daniel Baum, Ivan Viola, Julius Parulek, and Hans-Christian Hege. Visualization of Biomolecular Structures: State of the Art. In *Eurographics Conference on Visualization - STARs*, pages 61–81, 2015.
- [22] Barbora Kozlíková, Eva Šebestová, Vilém Šustr, Jan Brezovský, Ondřej Strnad, Lukáš Daniel, David Bednář, Antonín Pavelka, Martin Maňák, Martin Bezděka, Petr Beneš, Matúš Kotry, Artur Gora, Jiří Damborský, and Jiří Sochor. CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics*, 2014.
- [23] M. Krone, M. Falk, S. Rehm, J. Pleiss, and T. Ertl. Interactive exploration of protein cavities. *Comput Graph Forum*, 30(3):673–682, 2011.
- [24] M. Krone, D. Kauker, G. Reina, and T. Ertl. Visual analysis of dynamic protein cavities and binding sites. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, pages 301–305, March 2014.
- [25] M. Krone, G. Reina, C. Schulz, T. Kulschewski, J. Pleiss, and T. Ertl. Interactive extraction and tracking of biomolecular surface features. *Computer Graphics Forum*, 32(3pt3), 2013.
- [26] Ove Daae Lampe, Ivan Viola, Nathalie Reuter, and Helwig Hauser. Two-level approach to efficient visualization of protein dynamics. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1616–1623, 2007.
- [27] Roman A. Laskowski, Gail E. Hutchinson, Alex D. Michie, Andrew C. Wallace, Martin L. Jones, and Janet M. Thornton. PDBsum: a web-based database of summaries and analyses of all PDB structures. *Trends Biochem.Sci.*, 22(12):488–490, 1997.
- [28] M. Le Muzic, J. Parulek, A.K. Stavrum, and I. Viola. Illustrative visualization of molecular reactions using omniscient intelligence and passive agents. *Comput. Graph. Forum*, 33(3):141–150, 2014. URL: <http://dx.doi.org/10.1111/cgf.12370>.
- [29] Chang Ha Lee and Amitabh Varshney. Computing and displaying intermolecular negative volume for docking. *Scientific Visualization: The Visual Extraction of Knowledge from Data*, pages 49–64, 2006.
- [30] Norbert Lindow, Daniel Baum, Ana-Nicoleta Bondar, and Hans-Christian Hege. Dynamic channels in biomolecular systems: Path analysis and visualization. In *BioVis*, pages 99–106. IEEE, 2012.
- [31] Norbert Lindow, Daniel Baum, Ana-Nicoleta Bondar, and Hans-Christian Hege. Exploring cavity dynamics in biomolecular systems. *BMC Bioinformatics*, 14(S-19):S5, 2013.
- [32] Norbert Lindow, Daniel Baum, and Hans-Christian Hege. Voronoi-based extraction and visualization of molecular paths. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2025–2034, 2011.
- [33] Mathieu Le Muzic, Ludovic Autin, Julius Parulek, and Ivan Viola. cellVIEW: a Tool for Illustrative and Multi-Scale Rendering of Large Biomolecular Datasets. In Katja Bühler, Lars Linsen, and Nigel W. John, editors, *Eurographics Workshop on Visual Computing for Biology and Medicine*. The Eurographics Association, 2015.
- [34] J. J. Paleček and S. Gruber. Kite Proteins: a Superfamily of SMC/Kleisin Partners Conserved Across Bacteria, Archaea, and Eukaryotes. *Structure*, 23(12):2183–2190, Dec 2015.
- [35] Julius Parulek, Timo Ropinski, and Ivan Viola. Seamless abstraction of molecular surfaces. In *Proceedings of the 29th Spring Conference on Computer Graphics, SCCG '13*, pages 120–127, 2013.
- [36] Julius Parulek, Cagatay Turkay, Nathalie Reuter, and Ivan Viola. Implicit surfaces for interactive graph based cavity analysis of molecular simulations. In *Biological Data Visualization (BioVis), 2012 IEEE Symposium on*, pages 115–122. IEEE, 2012.

- [37] Julius Parulek, Cagatay Turkay, Nathalie Reuter, and Ivan Viola. Visual cavity analysis in molecular simulations. *BMC Bioinformatics*, 14(Suppl 19):S4, 2013. URL: <http://www.biomedcentral.com/1471-2105/14/S19/S4>.
- [38] Julius Parulek and Ivan Viola. Implicit representation of molecular surfaces. In *IEEE Pacific Visualization Symposium (PacificVis)*, pages 217–224, 2012.
- [39] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004. URL: <http://dblp.uni-trier.de/db/journals/jcc/jcc25.html#PettersenGHCGMF04>.
- [40] Martin Petřek, Pavlína Košinová, Jaroslav Koča, and Michal Otyepka. MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, 15(11):1357–1363, 2007.
- [41] P. P. Schrumpfova, I. Vychodilova, M. Dvorakova, J. Majerska, L. Dokladal, S. Schorova, and J. Fajkus. Telomere repeat binding proteins are functional components of Arabidopsis telomeres and interact with telomerase. *Plant J.*, 77(5):770–781, Mar 2014.
- [42] Manuela Waldner, Mathieu Le Muzic, Matthias Bernhard, Werner Purgathofer, and Ivan Viola. Attractive flicker: Guiding attention in dynamic narrative visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2456–2465, December 2014. URL: <http://www.cg.tuwien.ac.at/research/publications/2014/waldner-2014-af/>.
- [43] E. Yaffe, D. Fishelovitch, H.J. Wolfson, D. Halperin, and R. Nussinov. MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins*, 73(1):72–86, 2008.