Eur. Phys. J. Special Topics **227**, 1725–1739 (2019)
© EDP Sciences, Springer-Verlag GmbH Germany, part of Springer Nature, 2019
https://doi.org/10.1140/epjst/e2019-800162-y

THE EUROPEAN PHYSICAL JOURNAL
SPECIAL TOPICS

Review

Interactive visualization of biomolecules' dynamic and complex properties

State of the art and project review

Karsten Schatz[1,a], Michael Krone[1,2,b], Jürgen Pleiss[3], and Thomas Ertl[1]

[1] Visualization Research Center, University of Stuttgart, Stuttgart, Germany
[2] Big Data Visual Analytics in Life Sciences, University of Tübingen, Tübingen, Germany
[3] Institute of Technical Biochemistry, University of Stuttgart, Stuttgart, Germany

Received 1 October 2018 / Received in final form 19 November 2018
Published online 8 March 2019

Abstract. In this paper, we review the advances in molecular visualization over the last 12 years and put the development of the community in context with our own efforts in the DFG Collaborative Research Center (CRC) 716. This includes advances in the field of molecular surface computation and rendering, interactive extraction of protein cavities, and comparative visualization for biomolecules. Our main focus was on the development of methods that assist the interactive and explorative visual analysis of large, dynamic molecular data sets on single desktop computers. To meet this goal, we often developed GPU-accelerated algorithms, which is in line with the general research direction of the field. Over the last years, we made seminal contributions to the field of molecular visualization, which partially still constitute the state of the art development or provided the basis for follow-up works.

1 Introduction

Interactive molecular visualization has been an important area of research in the field of data visualization for more than five decades [1,2]. It has its foundations in the pre-computer era, where scientists were building physical models of molecules like the ball-and-stick representation [3]. This paper reviews interactive visualization of biomolecular structures, which has made tremendous advances during the last two decades. This review paper builds up on two recent surveys that discuss the visual analysis of biomolecular structures [4,5]. We summarize the state of the art development and specifically highlight the contributions that were made by project D.4 of the Collaborative Research Center CRC 716 over the last 12 years.

The goal of project D.4 was to advance the field of biomolecular visualizations by developing tailored visualizations for large, complex molecular dynamics simulations – mainly all-atom simulations of one or more proteins in a mixed solvent. This not only included the adaption and improvement of existing molecular visualization

[a] e-mail: karsten.schatz@visus.uni-stuttgart.de
[b] e-mail: michael.krone@uni-tuebingen.de

techniques to large, dynamic data sets, but also the development of novel visualization methods that allow users to gain a deeper understanding of the simulation results. A secondary goal was to make this interactive visual analyis possible with minimal or no preprocessing on modern consumer desktop PCs equipped with a high-end graphics card. Since the project focused on research in the fields of scientific visualization and visual analytics, the primary concern was not to create a new molecular visualization tool to compete with existing software like VMD [6], PyMOL [7,8], or UCSF Chimera [9], but rather to create working prototypes that could be used and tested by the project partners within CRC 716. Consequently, almost all research prototypes developed within D.4 are included in the visualization framework MegaMol [10], which was developed by another project of CRC 716 (D.3) as a common rapid prototyping platform for visualization research. Due to its flexibility and modular architecture, MegaMol greatly sped up the development of novel visualization techniques [11,12].

The rest of the paper is structured as follows: In Section 2, we briefly outline the basics, assumptions, and prerequisites for molecular visualization. Section 3 reviews the current state of the art technology in the field of biomolecular visualization. The contributions of CRC 716 D.4 to this field over the last 12 years with respect to the state of the art technology are summarized in Section 4. Finally, Section 5 concludes the paper and briefly discusses open questions and directions for future research.

2 Biochemical basics

Molecules are composed of nuclei and the core electrons belonging to the inner electron shells; the nuclei are held together by an outer electronic shell (valence shell), composed of molecular orbitals. Biomolecules are all molecules that play an essential role in living systems. This comprises large molecules (macromolecules) like proteins, lipids, DNA, and RNA, as well as small molecules such as metabolites. Weak bonds between these molecules are important, since they are not only essential for the formation of stable macromolecules, but also critical in maintaining their 3D structures.

Molecular visualization provides graphical depictions of the rich, complex material world on an atomistic level by making molecular structures, their properties, and their interactions visible and, therefore, understandable. The subfield of *biomolecular visualization* deals with the depiction of biomolecular systems, which are the molecular machines that drive the processes in all living cells. Thus, it provides means for integrated visual analysis of biomolecular structure data, an example being the visual exploration of biomolecular simulations.

The research in D.4 mainly focused on proteins. Proteins are linear macromolecules consisting of one or more chains of amino acids. Each amino acid consists of an identical part – the *backbone*, which consists of a central α-carbon, the amino group and the carboxyl group – and an individual side chain that defines the properties of the amino acid. In a protein, the amino group of one amino acid is covalently bound to the carboxyl group of the next amino acid, thereby forming the aforementioned chain. Biomolecules are often visualized using general molecular representations like ball-and-stick, space-filling, or molecular surfaces. However, specialized representations have been devised that show higher-level structural properties such as the well-known abstract depiction of the DNA double helix as a twisted ladder-like structure, or the Cartoon representation of the secondary structure of proteins [13].

The dynamic behavior of biomolecular systems is often investigated using Molecular Dynamics (MD) simulations, which provide a good approximation, even though the method has several restrictions (e.g., no covalent bonds are newly formed or

broken). In MD simulations, atoms are treated as classical objects that move under the influence of multi-body forces. These so-called force fields mimic quantum effects. Atoms are considered approximately as "hard" spheres characterized by their mass, radius, and the forces they exert on other atoms. This *hard sphere* model is the basis for many of the molecular representations used today in molecular visualization.

3 State of the art development in the visual analysis of biomolecular structures

The widespread availability of powerful parallel computing hardware in consumer desktop PCs – such as multi-core CPUs and programmable GPUs – that started about 12 years ago, same as the CRC 716. This ongoing development opens the door for many advances in the field of interactive molecular visualization and has been identified as a driving factor early on by Chavent et al. [14] in their 2011 position paper, which was co-authored by CRC 716 D.4 researchers. In this section, we briefly describe the current state of the art facilities concerning visual analysis methods for biomolecular simulation data. As mentioned in the introduction, this part is based on two recent surveys that were co-authored by D.4 researchers. The first one is a general review of the state of the art facilities in molecular graphics and biomolecular visualization [15], which was later extended and updated [5]. The second one is a survey of methods for the extraction and visualization of cavities in proteins [4]. Below, we summarize the state of the art development with respect to the goals of project D.4 and specifically highlight newer works that were published after the two survey papers. The two surveys are of course not the only works found in literature that summarize the state of the art development in molecular graphics and visualization. Recently, Alharbi et al. [16] presented a survey of surveys, which compares 11 surveys about molecular visualization published within the last 12 years that either appeared in a visualization-specific journal or a computational biology one. This work, which was also co-authored by a CRC 716 D.4 researcher, provides a perspective on the different focus points of the two communities as well as links to further reading material.

In the recent years some key challenges emerged in the field of biomolecular visualization. Biochemical data often bridge several length scales, from the smallest ligand molecule, consisting only of few atoms, to the largest cell comprising millions or even billions of elementary particles. Often, different representations are required to properly represent the relevant aspects of each scale, although the underlying structure, the atoms, does not change. In the context of proteins, especially their interaction with their environment is of importance. The surface of a protein forms the interface to the environment, so accurate and fast representations of these complicated shapes are required. Especially in the case of enzymes, cavities formed by the proteins' surface are of importance. As there exists no ground truth for the detection of the cavities, new methods improving either accuracy or detection speed have to be devised. While the detected cavities form the pathways for surrounding ligand or solvent molecules, domain scientiest are largely interested in the actual paths these smaller molecules take. The sheer amount of solvent or ligand molecules typically produces visual clutter in direct visualizations, so there is a demand for techniques that allow for an easy analysis of many different molecule pathways at the same time. A further challenge is the comparison of different biomolecules. Due to their complex and intertwined structure it is difficult to find representations that solve these comparison tasks well. Additionally, the complexity of the atom arrangements makes it difficult to correctly perceive them using only simple lighting methods. All of these representations typically have some limitations when the visualized data

set gets larger. As mentioned beforehand, the size of the data may be a challenge of its own, posing the need to bridge different scales, may it be via level-of-detail approaches or or abstracted methods to create a consistent picture even when using different visualization methods. Finally, emerging new display types and interaction methods offer new ways to view and interact with the generated visualizations. So incorporating them in a user-friendly way may benefit the whole analysis process. In the following we will summarize recent work not covered by the mentioned surveys due to later release date which are addressing these challenges.

3.1 Molecular surface visualization

Molecular surface representations are a popular way to display proteins. The three most used methods are the van der Waals (vdW) surface [17], the Solvent Accessible Surface (SAS) [18], and the Solvent Excluded Surface (SES) [17]. While the van der Waals method uses simple spheres to represent the surface of each atom, it serves as a basis for the other two representations. For these a so called *probe sphere* representing a previously specified ligand or solvent molecule is rolled around the vdW surface. The SAS is represented by the center of the probe, depicting the surface directly accessible for the represented solvent molecule. Solvent Excluded Surfaces, on the other hand, represent the part of the molecule that is inaccessible to the probe sphere. In the recent years, many methods for the computation and visualization of molecular surfaces, mostly for the mentioned three types, have been developed. Kozlíková et al. [5] provided and in-depth survey of molecular surface algorithms and visualization techniques. Additionally, Lindow et al. recognized that van der Waals radii obtained from literature are not always ideal to describe the true surface of a molecule [19]. Based on the bonding status of each atom, even atoms that have the same element may have different radii. Lindow et al. introduce the *atomic accessibility radius*, that takes this problem into account. Using these newly calculated atom radii, more accurate surface representations can be calculated.

3.2 Interactive visual analysis of protein cavities

Molecular surface representations are typically not round, uniform objects. The attracting and repelling forces between the atoms forming the molecule result in intertwined surface structures. This may lead to paths inside or even through the protein, which are called tunnels, or cavities, respectively. Cavities are crucial for many functional aspects of a protein, for example the accessibility of binding sites or transmembrane channels. Therefore, the detection of tunnels or cavities is an important part in the analysis of biomolecules. As already shown by Krone et al. [4], there exists a vast number of different methods to extract and visualize void zones in proteins. Depending on the application case they may fulfill different requirements like speed, robustness, or accuracy. Another survey by Simões et al. [20] further underlines the importance of protein cavities. However, their definition and classification of void zones inside proteins is vastly different to the ones published before.

Newer methods, like one by Vonásek et al. [21], propose the use of so called Radpidly Exploring Random Trees to efficiently calculate and track cavity structures over time. The same authors extend this method to improve the bevaviour in the case of temporally non-static tunnels [22]. To do so they exploit the vertices of voronoi diagrams of the subsequent data input frames. However, not only the identification and tracking of cavities is of importance but also the visualization. Malzahn et al. [23] propose an abstracted visualization of protein tunnels by projecting their surface to

a plane. This representation can be used to display and analyze physico-chemical properties of the tunnel. Since the detection and visualization of tunnels is of such large importance, several tools for this application case have been proposed in the past. The quasi-standard tool *CAVER Analyst* has been recently updated to its second major iteration [24], introducing new features like tunnel profile visualization and the handling of mutated proteins.

3.3 Abstract and multi-scale visualizations

The high complexity of biomolecules leads to a need of more abstract visualizations, hiding currently irrelevant information from the user. Typically such visualizations utilize the structure definitions used by the domain scientists to further simplify or enrich the visualizations. Vázquez et al. utilize the so-called *secondary structure* of a protein, that describes the order of amino acids forming the protein chain, in a circular manner to easily show interaction of the shown protein with an external ligand molecule [25]. The previously mentioned method of Malzahn et al. [23] can also be viewed as abstract visualization, since it transforms the layout of protein cavity into one that is understandable more easily.

Larger molecules typically need some kind of simplification to produce aliasing-free, understandable visualizations. This can be achieved by multi-scale approaches. Based on the previous work of Le Muzic et al. [26], who proposed a multi-scale cell visualization system, Klein et al. presented a system that is able to instantly construct cell renderings based on a given recipe, stating the distribution of certain molecules inside the cell [27]. Additionally, Kouřil et al. proposed a method that extends this idea from the visualization to the textual labeling of the visualization itself [28]. Based on the currently visualized level of detail, text labels describing the currently visualized features are shown or hidden in a smooth manner. The main targets for their implementation were visualizations of cells, but they should be easily extendable to smaller structures like proteins.

3.4 Comparative visualization and visual analytics

The evolution of all lifeforms is based on the mutation of DNA, and therefore, the gradual change of resulting proteins. In the recent years, scientists tried to mimic this phenomenon by developing modified proteins with slightly different properties. Before producing a modified protein in the real world, typically simulations of its behavior are performed. Therefore, comparative visualizations, able to constrast a biomolecule with mutated versions of itself, are needed. Most available methods juxtapose visualizations to achieve a comparison, which may become difficult when viewing more than two molecules. Kocincová et al. use juxtaposition and superimposition of protein cartoon renderings to analyze the secondary structure of multiple proteins [29]. They enrich classical 1D visualizations of the secondary structure with a second dimension to emphasize the difference between the visualized protein chains. This allows for an easy identification of different and common parts.

3.5 Advanced rendering and shading for molecular graphics

In scientific visualization shading is an important part of the rendering process. Shadows and light effects allow for an easier perception of the displayed scene and therefore accelerate the understanding of the visualization. Additionally, newly developed rendering methods are also of use. They either accelerate existing ideas to make them

available for weaker hardware or produce visually more appealing results. In the recent years new rendering and shading methods for biomolecules have become less frequent, but some of them appeared nonetheless.

Hermosilla et al., for example, proposed a special illumination model for biomolecules [30]. It incorporates global illumination effects into a model that is consistent among different visual representations. To do so they assume that most molecular visualizations are composed out of spheres and cylinders only. Lindow et al., on the other hand, presented a novel rendering model for the secondary structure of DNA and RNA [31]. Contrary to other existing methods, the rendered geometry is not composed from triangles or quads. All of the shown objects are directly raycasted and therefore produce a pixel-perfect representation of the molecule.

3.6 Visualizing solvent and ligand interactions

Most of a protein's function is expressed in its interactions with its surroundings. The hosting solvent as well as potential ligand molecules can influence the behavior of the protein. Investigating these interactions provide insight into the function of a protein and is therefore a frequently studied topic. The method of Furmanová et al. [32] is specifically tailored towards tracking ligand molecules. This enables the user to investigate the parts of the ligand's trajectory where it is able to interact with the protein. Vad et al. proposed a similar method for solvents, mostly water [33]. However, the visualization approaches of the two works do not perform very well with large number of molecules or very long trajectories. This issue was targeted by Duran et al. who solved this problem by the usage of an accumulating graph and selective processing of ligands [34]. Contrary to these approaches, the works of Vázquez et al. [25] and Hermosilla et al. [35] do not visualize the trajectory of ligands at all. Instead, they highlight the interactions with a ligand either directly in the visualization of the protein [35], or as abstracted, two-dimensional view [25].

Besides the interaction with other, smaller molecules, proteins can also contact with other proteins. Furmanová et al. therefore highlight the contact zones with other proteins on the surface of the visualized one [36]. All of the approaches presented here have one thing in common: they mainly investigate the bevavior of the amino acids on the protein's surface, since they are the ones with the possibility to get in contact with the environment. Lichtenberg et al. therefore provide a visual approach to show and track surface-facing amino acids [37]. This allows for an easier identification of highly contacting zones.

3.7 Immersive analytics

With newly emerging display technologies, new interaction possibilities become available. *Immersive Analytics* try to utilize these new opportunities to support decision making and reasoning. Technologies like VR-/AR-headsets, powerwalls or CAVEs, and even simple smartphones are used to create a more immersive experience. An overview over the topic is provided by Chandler et al. [38]. It is not immediately obvious how these technologies can be adapted for biomolecular data. Trellet et al. therefore recently proposed a pipeline specifically designed for molecular visualizations [39]. It allows for the usage of several display technologies in a molecular data context. A specific implementation of a visualization system for biomolecular data is given by Wiebrands et al. [40]. They utitilize the capabilities of the Unity game engine to visualize molecular trajectories and provide interactions on traditional screens as well as on larger installations like curved display walls. These are some of very few

publications dealing with the field of Immersive Analytics. The recency of this field opens up many unexplored directions for reasearch.

4 Contributions of the Collaborative Research Center 716

As discussed above, interactive molecular visualization has greatly advanced in the last 12 years, mainly due to the availability of new, parallel computing hardware that facilitated algorithmic improvements. In this section, we will show how the research conducted in project D.4 of CRC 716 contributed to these developments and how it helped to shape the current state of the art in interactive visual analysis of biomolecular simulations.

We used the open-source visualization framework MegaMol [10] as a development platform for the majority of visualizations we implemented. The development of MegaMol was started by project D.3 of CRC 716. It was designed as a rapid prototyping framework for visualization research. Over the years, MegaMol has evolved into a stable and very valuable tool for our research [11,12], which helped to speed up development times and to create sustainable research prototypes. It has also been used as a showcase in a scientific exhibition for the general public, which presented the results of CRC 716 to the general public, including our interactive visualizations of biomolecular simulations [41].

4.1 Molecular surface visualization

Early on, we focused on the interactive computation and visualization of Solvent Excluded Surfaces (cf., left Subfigure of Fig. 1), since these are widely used and very beneficial for the visual analysis of the accessibility of a molecule. In 2009, we published the first paper that used GPU-based ray casting for rendering the individual patches of the SES [42]. For an in-depth description of GPU-based ray casting, we refer to our recent textbook on this topic [43]. The SES was computed using a variant of the Reduced Surface (RS) algorighm [44], which allowed the interactive visualization of the SES for dynamic, medium-sized proteins where the RS has to be at least partially recomputed on the fly for each frame [45]. However, a bottleneck was that the original RS algorithm is inherently sequential; therefore, our attempts to parallelize it on the GPU only lead to a small speedup [46]. As a result of this, we turned to the Contour-Buildup (CB) algorithm [47], which can be parallelized quite straightforward on the CPU, as shown by Lindow et al. [48]. We adapted the CB algorithm to take advantage of the massively parallel architecture of modern, programmable GPUs, which resulted in a significant speedup [49]. Our CUDA implementation of the CB algorithm is to date still one of the fastest published algorithms to compute the SES analytically. Therefore it is now possible to compute the analytically correct SES for small to medium sized molecules at interactive frame rates, allowing for an analysis process without unnecessary waiting times. Recently, we published another algorithm that computes the SES on the GPU by sampling the atoms on a volumetric grid, which can be progressively refined. The SES is then extracted as an isosurface from this volume using direct volume rendering (i.e., GPU-based ray marching). Depending on the size of the data set and the resolution of the volume, this approach can be faster than the analytical CB computation. For the visual analysis of proteins, semi-transparent renderings of the SES are often beneficial since they also show the inner structure of the protein (e.g., cavities). While the volume rendering approach natively supports transparency, it is not straightforward for the analytically described SES. We, therefore, developed a method that enables transparency also for the GPU-based ray casting [50]. Reaching interactive frame rates of

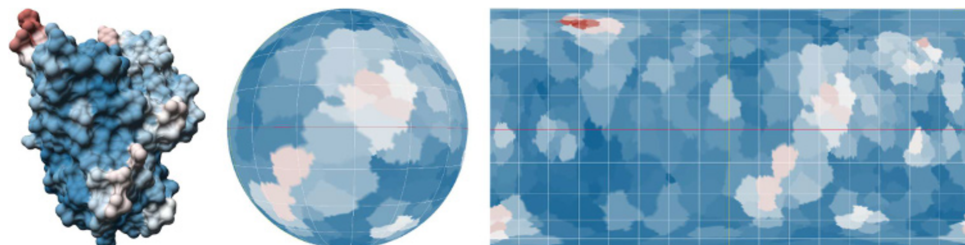

Fig. 1. Results of the Molecular Surface Maps algorithm [55]. Left: method is starting with the Solvent Excluded Surface depicting the B-Factor of the atoms. Middle: result is then mapped to a sphere. Right: finally, a map is generated from spherical representation using standard map projection techniques.

around 10 fps for medium sized proteins the approach outperformed other available implementations of that time, like Dual Depth Peeling [51], for example. This leads to a further refinement of the method by Jurcik et al. [52].

An alternative to the SES are Gaussian surfaces, which model the electron density of a molecule and can approximate the SES. We developed the QuickSurf algorithm [53], which computes Gaussian surfaces entirely on the GPU using a regular grid and isosurface extraction. Since this method is computationally less involved than the SES, it is ideally suited for the interactive visualization of surfaces for very large molecular complexes consisting of multiple millions of atoms (cf., left Subfigure of Fig. 2). Due to its scalability, the QuickSurf algorithm found its way into large, widely used packages for biomolecular visualization, like VMD [6]. We also showed that this surface extraction algorithm can be used to extract fluid surfaces for large Smoothed Particle Hydrodynamics (SPH) simulations [54]. In this case, a Wendland kernel function was used instead of Gaussians, which models the fluid more faithfully. The extension to SPH simulations was able to reach interactive frame rates even for large data sets (∼67 fps for 125 000 particles, ∼11 fps for 1 000 000 particles).

4.2 Interactive visual analysis of protein cavities

Molecular surfaces are ideally suited to analyze protein cavities. In an early work, we used a grid-based representation of the protein (similar to our QuickSurf method [53]). Users can select a cavity of interest in a 2D cutting plane, which is then extracted using a flood fill algorithm that stops when the isosurface is reached. We later refined this approach using only the molecular surface mesh [56]. Following the idea of Borland [57], we classify each mesh vertex as belonging to a cavity or not based on its Ambient Occlusion factor. The cavity extraction algorithm was implemented on the GPU and embedded into a multi-view visual analytics application for cavity analysis [58], which allows users to interactively refine the cavity extraction results and to analyze cavity properties over time for simulation data. Thanks to the GPU implementation, the presented method was the first approach to reach interactive frame rates for the real-time cavity extraction and visualization for typically-sized enzymes (more than 10 fps for a protein consisting of 6600 atoms or ∼330 amino acids).

4.3 Abstract and multi-scale visualizations

Abstract representations of the atomistic data are commonly used in molecular visualization. Since we were primarily interested in visualizing dynamic proteins, we

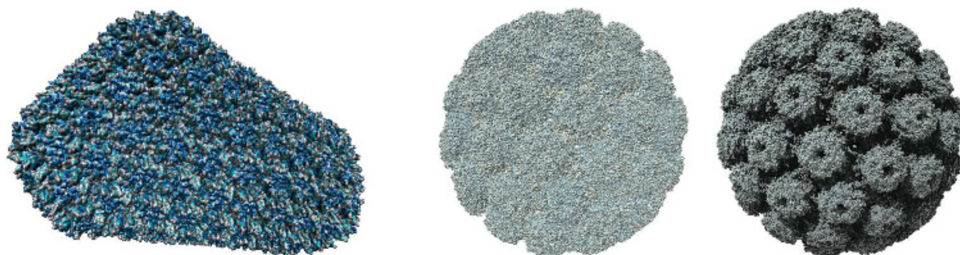

Fig. 2. Left: interactive Gaussian density surface visualization of a HIV capsid, containing over 2 million atoms, using the QuickSurf method [53]. Middle: depiction of a virus capsid with only local lighting. Right: highlighted structure of the virus capsid, using additional ambient occlusion [76] and shadows [77].

developed a fast, GPU-based rendering technique for the Cartoon representation that shows the secondary structure of proteins [59]. Our implementation was the first one to construct the Cartoon representation entirely on the GPU using the Geometry Shader (cf., left part of Fig. 2). Following up on our work, Hermosilla et al. [60] later showed that even higher performance can be obtained when using the new Tessellation Shaders, which were not available when we started our implementation. Recently, we extended the classical Cartoon representation to also show uncertainty in the data by adding a ripple distortion [61]. The frequency and amplitude of the ripples shows the amount of uncertainty in this region. This approach also addresses another major issue in the field, namely the possibility to encode further values into a representation without relying on color.

While the Cartoon rendering is a commonly used abstraction of proteins, we also investigated new ways of abstracting molecular data. One inherent problem of three-dimensional visualization is that parts of the molecule that are turned away from the user are not visible. Transparency is often not a good solution, since it makes the whole image very cluttered and confusing for complex representations. Therefore, we presented Molecular Surface Maps [55], which project the whole molecular surface on a 2D map. The challenge was to create a mapping that is easy to understand and introduces as little distortion as possible. Thus, we first morph the SES to a sphere and then use well-known map projection methods from cartography. In contrast to previous works [62–64], our approach consistently solves the problem of mapping a molecular surface of genus n to a sphere (genus 0) by cutting channels. An example of the Molecular Surface Maps pipeline is given in Figure 1.

Level-of-Detail representations are typically used to increase rendering performance by using simpler models in areas where full detail is not needed (e.g., for distant objects that have a image space footprint of only a few pixels). For very large, complex models, this can have the additional benefit that the simpler representation maintains the overall shape while removing fine details that would only lead to noise. Based on the work of Lindow et al. [65], we presented a Level-of-Detail rendering method for very large particle data sets, which enables rendering simulations of whole cells at atomistic detail using either spacefilling representations or molecular surfaces [66,67]. This method served as a benchmark for the previously mentioned and more recent systems of Le Muzic et al. and subsequently Klein et al. [26,27]. The atoms of each protein in the scene are sorted into an individual, uniform grid that subdivides the protein's bounding box. During rendering, either the atoms inside a cell are rendered or, if the image-space size of the grid cell is smaller than one pixel, only an occupancy test is made. The same is done for the bounding box. This improves rendering speed and proteins will not disappear in the final image if they are

smaller than one pixel. Due to the improvements compared to the method of Lindow et al. [65] data sets with 25 billion atoms can be rendered with around 3.6 fps. Without any optimizations over their method, only 1.8 fps can be reached. Additionally, the method is able to handle data sets consisting triangle meshes instead of atoms. This typically happens at the cost of a lower frame rate, due to the more expensive intersection tests. Subsequently, we developed a similar hierarchical rendering method for astronomical data sets of multiple trillions of atoms [68]. This method, which combines volume rendering and particle rendering, could also be used for whole cell data sets.

4.4 Comparative visualization and visual analytics

As mentioned in the last section, the comparison of different molecular data sets is an important topic. Previous methods often focused on computing the similarity of proteins (e.g., [69,70]). Another popular approach is using rigid alignment to superimpose proteins as good as possible for visual comparison, for example using RMSD alignment [71], which we also used [72]. We subsequently developed a method that compares two molecular surfaces by mapping one of them onto the other [73]. This allows to compare surface properties at corresponding points of the two molecular surfaces. The results are visualized using color coding and transparency.

Our uncertainty visualization mentioned above can also be used for visual comparison: we extracted the flexibility of two variants of a protein from a simulation and mapped the values as uncertainty. By juxtaposing the two resulting visualizations, users can assess differences in the flexiblity. As mentioned in Section. 3, Kocincová et al. [29] recently presented a comparative visualization of protein secondary structures, which shows structural differences. In contrast to our method, it compares individual time steps of a simulation, not aggregated properties like flexibility. Molecular Surface Maps [55] are also useful for comparing different data sets: users can juxtapose or superimpose the maps of different data sets to compare the depicted properties.

We also presented a visual analytics tool for the comparative analysis of a large ensemble of proteins [74]. The application consists of a 3D view, where the superimposed proteins are shown, and a 2D view showing a Parallel Coordinate Plot (PCP). Multiple physico-chemical properties are collected for each amino acid of each protein. In the PCP, an axis is drawn for each property; that is, each line in the PCP represents one amino acid. Using linking and brushing, the user can select and compare proteins based on their amino acids' properties. We also applied the concept of using a 3D visualization in concert with a PCP to analyze SPH simulations [75].

4.5 Advanced rendering and shading for molecular graphics

The shape of complex molecular models and the spacial arrangement of the molecules in simulations are often hard to perceive when using only local lighting, which is the standard for interactive GPU-accelerated rendering based on rasterization. This is due to the absence of effects like shadows or reflections. Effects that mimic global illumination, like Ambient Occlusion, greatly increases depth and shape perception; however, this is often computationally too expensive for interactive visualization. We developed an Ambient Occlusion method tailored to particle data sets, which can be computed interactively even for multiple millions of particles [76]. Recently, we also presented a method to render high-quality drop shadows in real time for ray-cast spheres – the Implicit Sphere Shadow Maps [77]. Results of this method

are depicted in Figure 2. Opposed to other existing shadow mapping methods, ours were specifically tailored towards particle-based data sets like proteins. Therefore, interactive frame rates above 30 fps are reachable, even for data sets comprising more than a million atoms.

The shape perception of can also be improved using non-photorealistic rendering (NPR) methods. We presented a novel method that uses line integral convolution on molecular surfaces to mimic the effect of hatching [78]. This can illustrate areas of high curvature.

4.6 Visualizing solvent and ligand interactions

When analyzing molecular dynamics simulations, the behavior of individual solvent molecules is not of interest, but rather the average behavior of the solvent during the whole simulation. In order to visually analyze the dynamics of water near a protein cavity, we aggregated the paths of all solvent molecules over time and clustered similar paths [79]. Only clusters above a certain size are then rendered in order to show the main water paths, which additionally reduces the produced visual clutter. This work served as early predecessor for the approach of Vad et al. [33], mentioned in the last section. Their concept mainly adds several aggregated views and improves the overall interaction with the visualization.

Instead of aggregating the motion of individual molecules (which results in paths), the motion vectors of the molecules can be aggregated on a 3D grid over time (which results in a 3D vector field). We first used this approach for dipole moments in a crystal simulation from material science [80], and later adapted it to visualize the ion flow around DNA [81]. The ion flow was shown by extracting streamlines from the vector field. The same approach was concurrently employed by Chavent et al. [82] to illustrate lipid motions in a membrane.

For cases where users are only interested in the presence of solvent molecules but not in their motion, we developed another visual analysis tool that extracts the probability of presence by temporally aggregating solvent molecules on a grid [83]. The presence of solvent molecules over time at the surface of a protein can also be visualized using the aforementioned Molecular Surface Maps. We aggregate the number of time steps in the simulation where a solvent molecule is in the vicinity of the molecular surface and color-code this onto the map top provide an overview [55].

4.7 Immersive analytics

Interactive visualization is commonly used as a tool for the analysis of molecular structures, especially for molecular simulations. Especially stereoscopic rendering has been adopted for molecular visualization early on, as it makes it easier to understand the spatial relations between these complex structures. We have used a Powerwall (a stereoscopic high-resolution tiled display [84]) for the visual analysis of an ensemble of molecular simulation data sets [85,86]. While the visualization framework MegaMol [10], in which we implemented our visualization, supports stereoscopic rendering on large, tiled displays, the user interaction is mainly tailored to classical desktops using mouse and keyboard. We, therefore, added remote control capabilites using an Android tablet, which allows users to move freely in front of the powerwall while interacting with the visualization [87].

In order to increase the degree of immersion beyond stereoscopic rendering, we also used sonification and spatial audio [88]. Important events in a molecular simulation are detected (e.g., a ligand reaching the active site of a protein) and corresponding sounds are played. Our sonification can also be used for attention guidance, that

is, to make users aware of events that occurred outside their field of view Due to the availability of modern head-mounted displays like HoloLens that enable see-through Augmented Reality, this technique has recently gained attention for molecular visualization. We implemented a prototype that visualizes proteins as spacefilling representation on HoloLens and evaluated the performance of the device for different rendering techniques [89], since high frame rates are important for AR.

5 Conclusion and outlook

Molecular visualization is one of the oldest branches of scientific visualization, yet it has advanced tremendously during the last two decades, mostly due to the constant improvement of computing hardware – especially the advent of freely programmable GPUs. In the first part of the paper, we discussed how the field of interactive biomolecular visualization has recently evolved. Specifically, this paper provides an update to two recent state of the art reports on molecular visualization [5] and visual analysis of protein cavities [4], as it explicitly focuses on new visualization methods that were published after these two literature surveys. In the second part, we discussed the contributions we made in the last 12 years as part of a larger research project, the Collaborative Research Center CRC 716, which was funded by the German Research Foundation (DFG). Some of the methods we developed are still considered as the state of the art development for molecular visualization or provided the basis for notable follow-up works. This includes especially the developed methods for molecular surface visualization.

For the future, we project that the trend towards bigger and more complex biomolecular simulation data will continue. Therefore, novel visualization methods that can efficiently support the visual analysis of these data are necessary. Nowadays, especially limited computing resources provide a challenge for experts of biology as well as biomolecular visualization. In contrast to particle physics, for example, the memory requirements are rather low but the computation of other properties apart from the physical ones requires additional resources. The sheer amount of different properties to visualize demands various different approaches as it is still unclear how to incorporate all this complexity in a single easily comprehensible visualization. To adapt to these challenges, this not only includes improving existing algorithms for the visualization of traditional molecular models, but also new representations that can show higher-order structures up to the level of whole cells or even organisms simulated at atomistic detail. This will require the development of new multi-scale visualizations that not only bridge spatial scales but also temporal ones and provide insight into multi-dimensional properties of the simulated biomolecular systems.

We want to thank Katrin Bidmon and Katrin Scharnowski, who worked with us on the project and made valuable contributions. We also want to thank all our collaborators, which include not only our colleagues from the other CRC projects, but also our external project partners. Finally, we want to thank the German Research Foundation (DFG) for funding our research as part of the Collaborative Research Center CRC 716.

Author contribution statement

All authors contributed equally to this work.

References

1. C. Levinthal, Sci. Am. **214**, 42 (1966)
2. E. Francoeur, Endeavour **26**, 127 (2002)
3. C. Meinel, Blick in die Wissenschaft **18**, 10 (2006)
4. M. Krone, B. Kozlíkova, N. Lindow, M. Baaden, D. Baum, J. Parulek, H.-C. Hege, I. Viola, Comput. Graph. Forum **35**, 527 (2016)
5. B. Kozlíková, M. Krone, M. Falk, N. Lindow, M. Baaden, D. Baum, I. Viola, J. Parulek, H.-C. Hege, Comput. Graph. Forum **36**, 178 (2017)
6. W. Humphrey, A. Dalke, K. Schulten, J. Mol. Graph. **14**, 33 (1996)
7. W.L. DeLano, CCP4 Newsletter on protein crystallography No 40 (2002), https://pymol.org/
8. L. Schrödinger, The PyMOL Molecular Graphics System, Version 1.8 (2016)
9. E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, J. Comput. Chem. **25**, 1605 (2004)
10. S. Grottel, M. Krone, C. Mouller, G. Reina, T. Ertl, IEEE Trans. Vis. Comput. Graphics **21**, 201 (2015)
11. S. Grottel, G. Reina, M. Krone, C. Müller, T. Ertl, *Workshop on Visualization in Practice* (IEEE, 2016)
12. M. Krone, S. Grottel, G. Reina, C. Müller, T. Ertl, IEEE Comput. Graphics App. **38**, 109 (2018)
13. J.S. Richardson, Adv. Protein Chem. **34**, 167 (1981)
14. M. Chavent, B. Lévy, M. Krone, K. Bidmon, J.-P. Nominé, T. Ertl, M. Baaden, Brief. Bioinform. **12**, 689 (2011)
15. B. Kozlíková, M. Krone, N. Lindow, M. Falk, M. Baaden, D. Baum, I. Viola, J. Parulek, H.-C. Hege, in *Eurographics Conference on Visualization (EuroVis) – STARs* (The Eurographics Association, 2015), pp. 61–81
16. N. Alharbi, M. Alharbi, X. Martinez, M. Krone, A.S. Rose, M. Baaden, R.S. Laramee, M. Chavent, in *EuroVis 2017 – Short Papers* (The Eurographics Association, 2017)
17. F.M. Richards, Annu. Rev. Biophys. Bio. **6**, 151 (1977)
18. B. Lee, F.M. Richards, J. Mol. Bio. **55**, 379 (1971)
19. N. Lindow, D. Baum, H.-C. Hege, in *Workshop on Molecular Graphics and Visual Analysis of Molecular Data* (The Eurographics Association, 2018)
20. T. Simões, D. Lopes, S. Dias, F. Fernandes, J. Pereira, J. Jorge, C. Bajaj, A. Gomes, Comput. Graph. Forum **36**, 643 (2017)
21. V. Vonásek, B. Kozlíková, in *11th International Workshop on Robot Motion and Control, 3–5 July 2017* (IEEE, Wasowo, Poland, 2017), pp. 185–192
22. V. Vonásek, A. Jurčík, K. Furmanová, B. Kozlíková, J. Intell. Rob. Syst., https://doi.org/10.1007/s10846-018-0877-6
23. J. Malzahn, B. Kozlíková, T. Ropinski, in *Eurographics Workshop on Visual Computing for Biology and Medicine* (The Eurographics Association, 2017)
24. A. Jurcik, D. Bednar, J. Byska, S.M. Marques, K. Furmanova, L. Daniel, P. Kokkonen, J. Brezovsky, O. Strnad, J. Stourac, A. Pavelka, M. Manak, J. Damborský, B. Kozlikova, Bioinformatics **34**, 3586 (2018)
25. P. Vázquez, P. Hermosilla, V. Guallar, J. Estrada, A. Vinacua, Comput. Graph. Forum **37**, 391 (2018)
26. M. Le Muzic, L. Autin, J. Parulek, I. Viola, in *EG Workshop on Visual Computing for Biology and Medicine* (The Eurographics Association, 2015), pp. 61–70
27. T. Klein, L. Autin, B. Kozlíková, D.S. Goodsell, A. Olson, M.E. Gröller, I. Viola, IEEE Trans. Vis. Comput. Graphics **24**, 862 (2018)
28. D. Kouřil, L. Čmolík, B. Kozlíková, H. Wu, G. Johnson, D.S. Goodsell, A. Olson, M.E. Gröller, I. Viola, IEEE Trans. Vis. Comput. Graphics **25**, 977 (2019)
29. L. Kocincová, M. Jarešová, J. Byška, J. Parulek, H. Hauser, B. Kozlíková, BMC Bioinformatics **18**, 23 (2017)
30. P. Hermosilla, P. Vázquez, A. Vinacua, T. Ropinski, Comput. Graph. Forum **37**, 367 (2018)

31. N. Lindow, D. Baum, M. Leborgne, H.-C. Hege, IEEE Trans. Vis. Comput. Graphics **25**, 967 (2019)
32. K. Furmanová, M. Jarešová, J. Byška, A. Jurčík, J. Parulek, H. Hauser, B. Kozlíková, BMC Bioinformatics **18**, 22 (2017)
33. V. Vad, J. Byška, A. Jurcík, I. Viola, E. Gröller, H. Hauser, S.M. Marques, J. Damborský, B. Kozlíková, in *Proceedings of Eurographics Workshop on Visual Computing for Biology and Medicine* (2017), pp. 33–42
34. D. Duran, P. Hermosilla, T. Ropinski, B. Kozlíková, À. Vinacua, P. Vázquez, IEEE Trans. Vis. Comput. Graphics **25**, 987 (2018)
35. P. Hermosilla, J. Estrada, V. Guallar, T. Ropinski, À. Vinacua, P.-P. Vázquez, IEEE Trans. Vis. Comput. Graphics **23**, 731 (2017)
36. K. Furmanová, J. Byška, E.M. Gröller, I. Viola, J.J. Paleček, B. Kozlíková, BMC Bioinformatics **19**, 125 (2018)
37. N. Lichtenberg, R. Menges, V. Ageev, A.P. George, P. Heimer, D. Imhof, K. Lawonn, Comput. Graph. Forum **37**, 379 (2018)
38. T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, E. Wilson, in *2015 Big Data Visual Analytics (BDVA)* (IEEE, Hobart, TAS, Australia, 2015), pp. 1–8
39. M. Trellet, N. Ferey, J. Flotynski, M. Baaden, P. Bourdot, JIB **15**, 20180004 (2018)
40. M. Wiebrands, C.J. Malajczuk, A.J. Woods, A.L. Rohl, R.L. Mancera, JIB **15**, 20180010 (2018)
41. M. Krone, K. Schatz, N. Hieronymus, C. Müller, M. Becher, T. Barthelmes, A. Cooper, S. Currle, P. Gralka, M. Hlawatsch, L. Pietrzyk, T. Rau, G. Reina, R. Trefft, T. Ertl, in *Proceedings of SIGRAD 2017* (2017), Vol. 143, pp. 17–24
42. M. Krone, K. Bidmon, T. Ertl, IEEE Trans. Vis. Comput. Graphics **15**, 1391 (2009)
43. M. Falk, S. Grottel, M. Krone, G. Reina, Interactive gpu-based visualization of large dynamic particle data, in *Synthesis lectures on visualization*, edited by N. Elmqvist, D. Ebert (Morgan & Claypool Publishers, 2016)
44. M.F. Sanner, A.J. Olson, J.-C. Spehner, in *Proceedings of Symposium on Computational Geometry (SCG '95)* (1995), pp. 406–407
45. M.F. Sanner, A.J. Olson, in *Pacific Symposium on Biocomputing* (1997), pp. 385–396
46. M. Krone, C. Dachsbacher, T. Ertl, in *Proc. of ACM International Conference on Bioinformatics and Computational Biology* (ACM, New York, NY, 2010), pp. 402–405
47. M. Totrov, R. Abagyan, J. Struct. Biol. **116**, 138 (1995)
48. N. Lindow, D. Baum, S. Prohaska, H.-C. Hege, Comput. Graph. Forum **29**, 943 (2010)
49. M. Krone, S. Grottel, T. Ertl, in *IEEE Symposium on Biological Data Visualization* (IEEE, Providence, RI, USA, 2011), pp. 17–22
50. D. Kauker, M. Krone, A. Panagiotidis, G. Reina, T. Ertl, in *Eurographics Symposium on Parallel Graphics and Visualization* (The Eurographics Association, 2013), Vol. 13, pp. 33–40
51. L. Bavoil, K. Myers, *Order independent transparency with dual depth peeling*, Tech. Rep. (NVIDIA Corp., 2008)
52. A. Jurcik, J. Parulek, J. Sochor, B. Kozlikova, in *IEEE Pacific Visualization Symposium* (IEEE, Taipei, Taiwan, 2016), pp. 112–119
53. M. Krone, J. E. Stone, T. Ertl, K. Schulten, in *EuroVis – Short Papers* (The Eurographics Association, 2012), pp. 67–71
54. M. Krone, M. Huber, K. Scharnowski, M. Hirschler, D. Kauker, G. Reina, U. Nieken, D. Weiskopf, T. Ertl, in *EuroVis – Short Papers* (The Eurographics Association, 2014), Vol. 3, pp. 109–113
55. M. Krone, F. Frieß, K. Scharnowski, G. Reina, S. Fademrecht, T. Kulschewski, J. Pleiss, T. Ertl, IEEE Trans. Vis. Comput. Graphics **23**, 701 (2017)
56. M. Krone, G. Reina, C. Schulz, T. Kulschewski, J. Pleiss, T. Ertl, Comput. Graph. Forum **32**, 331 (2013)
57. D. Borland, J. WSCG **19**, 17 (2011)
58. M. Krone, D. Kauker, G. Reina, T. Ertl, in *IEEE PacificVis – Visualization Notes* (IEEE, Yokohama, Japan, 2014), Vol. 1, pp. 301–305

59. M. Krone, K. Bidmon, T. Ertl, *Theory and practice of computer graphics* (The Eurographics Association, 2008), pp. 115–122
60. P. Hermosilla, V. Guallar, A. Vinacua, P.-P. Vázquez, in *Eurographics Workshop on Visual Computing for Biology and Medicine* (The Eurographics Association, 2015)
61. C. Schulz, K. Schatz, M. Krone, M. Braun, T. Ertl, D. Weiskopf, in *2018 IEEE Pacific Visualization Symposium* (IEEE, Kobe, Japan, 2018), pp. 96–105
62. S.J. Rahi, K. Sharp, Int. J. Comput. Geom. Appl. **17**, 305 (2007)
63. N. Postarnakevich, R. Singh, in *ACM Symposium on Applied Computing, SAC '09* (2009), pp. 782–787
64. J. Hass, P. Koehl, Front. Mol. Biosci. **1**, 26 (2014)
65. N. Lindow, D. Baum, H.-C. Hege, Comput. Graph. Forum **31**, 1325 (2012)
66. M. Falk, M. Krone, T. Ertl, in *EG Workshop on Visual Computing for Biology and Medicine* (The Eurographics Association, 2012), pp. 123–130
67. M. Falk, M. Krone, T. Ertl, Comput. Graph. Forum **32**, 195 (2013)
68. K. Schatz, C. Müller, M. Krone, J. Schneider, G. Reina, T. Ertl, in *6th IEEE Symposium on Large Data Analysis and Visualization (LDAV)* (IEEE, Baltimore, MD, USA, 2016), pp. 56–64
69. D.W. Ritchie, G.J.L. Kemp, J. Comput. Chem. **20**, 383 (1999)
70. L. Sael, D. La, B. Li, R. Rustamov, D. Kihara, Proteins: Struct. Funct. Bioinf. **73**, 1 (2008)
71. W. Kabsch, Acta Crystallogr. A **32**, 922 (1976)
72. D. Kauker, M. Krone, A. Panagiotidis, G. Reina, T. Ertl, Comput. Visual. Sci. **15**, 111 (2013)
73. K. Scharnowski, M. Krone, G. Reina, T. Kulschewski, J. Pleiss, T. Ertl, Comput. Graph. Forum **33**, 191 (2014)
74. J. Heinrich, M. Krone, S. I. O'Donoghue, D. Weiskopf, Faraday Discuss. **169**, 179 (2014)
75. S. Reinhardt, M. Huber, O. Dumitrescu, M. Krone, B. Eberhardt, D. Weiskopf, in *2017 21st International Conference Information Visualisation (IV)* (IEEE, London, UK, 2017), pp. 117–126
76. S. Grottel, M. Krone, K. Scharnowski, T. Ertl, in *IEEE Pacific Visualization Symposium* (2012), pp. 209–216
77. M. Krone, G. Reina, S. Zahn, T. Tremel, C. Bahnmüller, T. Ertl, in *2017 IEEE Pacific Visualization Symposium* (IEEE, Songdo, South Korea, 2017), pp. 275–279
78. K. Lawonn, M. Krone, T. Ertl, B. Preim, Comput. Graph. Forum **33**, 181 (2014)
79. K. Bidmon, S. Grottel, F. Bös, J. Pleiss, T. Ertl, Comput. Graph. Forum **27**, 935 (2008)
80. K. Scharnowski, M. Krone, F. Sadlo, P. Beck, J. Roth, H.-R. Trebin, T. Ertl, IEEE Comput. Graph. App. **33**, 9 (2013)
81. T. Ertl, M. Krone, S. Kesselheim, K. Scharnowski, G. Reina, C. Holm, Faraday Discuss. **169**, 167 (2014)
82. M. Chavent, T. Reddy, J. Goose, A.C.E. Dahl, J.E. Stone, B. Jobard, M.S.P. Sansom, Faraday Discuss. **169**, 455 (2014)
83. B. Thomaß, J. Walter, M. Krone, H. Hasse, T. Ertl, in *Vision, Modeling, and Visualization* (The Eurographics Association, 2011), pp. 301–308
84. C. Müller, G. Reina, T. Ertl, in *CHI POWERWALL 2013 Workshop* (2013)
85. C. Müller, M. Krone, K. Scharnowski, G. Reina, T. Ertl, in *International Symposium on Visual Information Communication and Interaction (VINCI)* (ACM, New York, NY, 2015), Vol. 8, pp. 131–136
86. C. Müller, M. Krone, K. Scharnowski, G. Reina, T. Ertl, IJSI **9**, 355 (2016)
87. M. Krone, C. Müller, T. Ertl, D. Kauker, A. C. Silva, D. Salsa, M. Gräber, M. Kallert, in *International Symposium on Visual Information Communication and Interaction (VINCI)* (ACM, New York, NY, 2015), Vol. 8, pp. 21–26
88. B. Rau, F. Friess, M. Krone, C. Müller, T. Ertl, in *IEEE International Workshop on Virtual and Augmented Reality for Molecular Science* (IEEE, Arles, France 2015), pp. 25–30
89. C. Müller, M. Krone, M. Huber, V. Biener, D. Herr, S. Koch, G. Reina, D. Weiskopf, T. Ertl, JIB **15**, 20180005 (2018)