Multiscale Molecular Visualization

Haichao Miao, Tobias Klein, David Kouřil, Peter Mindek, Karsten Schatz, M. Eduard Gröller, Barbora Kozlíková, Tobias Isenberg, Ivan Viola

▶ To cite this version:

HAL Id: hal-01873724

https://hal.inria.fr/hal-01873724

Multiscale Molecular Visualization

Haichao Miao, Tobias Klein, David Kouřil, Peter Mindek, Karsten Schatz, M. Eduard Gröller, Barbora Kozlíková, Tobias Isenberg, Ivan Viola*

TU Wien, Austria; Masaryk University Brno, Czechia; Inria, France; University of Stuttgart, Germany; VRVis Research Center, Austria

Abstract

We provide a high-level survey of multiscale molecular visualization techniques, with a focus on application-domain questions, challenges, and tasks. We provide a general introduction to molecular visualization basics and describe a number of domain-specific tasks that drive this work. These tasks, in turn, serve as the general structure of the following survey. First we discuss methods that support the visual analysis of molecular dynamics simulations. We discuss, in particular, visual abstraction and temporal aggregation. In the second part we survey multiscale approaches that support the design, analysis, and manipulation of DNA nanostructures and related concepts for abstraction, scale transition, scale-dependent modeling, and navigation of the resulting abstraction spaces. In the third part of the survey we showcase approaches that support interactive exploration within large structural biology assemblies up to the size of bacterial cells. We describe fundamental rendering techniques as well as approaches for element instantiation, visibility management, visual guidance, camera control, and support of depth perception. We close the survey with a brief listing of important tools that implement many of the discussed approaches and a conclusion that provides some research challenges in the field.

Keywords: molecular visualization, molecular dynamics, modelitics, DNA nanotechnology, visual abstraction

1. Introduction

Data visualization is a computer-supported methodology for conveying complex digital information to the user by visual means. Visualization is often realized as a pipeline [1] where the digital data undergo several transformation stages, before their visual representations are displayed on the computer screen. In the first stage, digital data is filtered according to relevance to reduce the overall amount that needs to be handled. Furthermore, data can be aggregated, derived, and subsampled to achieve a representation with a high expressive value. In the next stage, the data is mapped to geometric primitives

---

*Corresponding author
    Email address: viola@cg.tuwien.ac.at (Ivan Viola)

that have particular optical or visual properties. Finally, an image is synthesized in the rendering stage, which displays the visual representation on the screen. The image is viewed, perceived, and cognitively consumed by a human observer. To learn more about the specifics of the visualization pipeline and visualization in general, we refer to visualization textbooks, such as by Munzner [2], Ware [3], or Telea [4].

Domains studying molecular structures or molecular phenomena were among the first to introduce computer graphics into their workflows. Molecular graphics, or molecular visualization, is a field with a rich history. Data and models for molecular visualization are originating from X-ray crystallography, nuclear magnetic resonance, and electron-microscopic acquisitions. In the past, this field was focused on developing visual languages beyond the visual vocabulary of pre-computer chemistry. Some representations gave visual prominence to the atoms, some others were focusing on the bonds between them. Various abstractions have been introduced to explicitly depict secondary structures [5] or to concentrate on molecular surfaces as more encompassing bounding structures [6].

In the early days of computer graphics, it was a computational challenge to interactively display a single macromolecule. For quite some time, simple approaches were in use to represent a molecular structure as a mesh geometry that is displayed with accelerated graphics routines. Such a representation was suitable to display only a few hundreds up to thousands of atoms interactively, which was sufficient for most of the analytical tasks back then. The mentioned approaches are inappropriate for dynamic molecular representations. The mesh extraction would hardly be feasible, and meshing the entire dynamics sequence would lead to excessive memory requirements. Until recently, molecular visualization thus predominantly conveyed using single-scale structures and dynamics.

Scale and scalability are recurring topics in visualization and visual computing. Multiscale techniques have been efficient strategies to cope with complexity in various areas of our field. In the simulation community, multigrid and scale-space approaches have been, for a long time, an appropriate tool to accommodate accuracy requirements and performance constraints. Hierarchical methods and hierarchical data structures are used extensively in all areas of (computer) science. Recent developments like Big and Smart Data, machine learning with vast training data, and advances in various vibrant application domains pose novel challenges to scalability and use of scale way beyond the typical multiscale techniques. In the biological domain, phenomena are now modeled across many scales, from individual atoms to proteins, to entire cells, and even beyond. In another domain, for example, smart cities require the representation of many diverse networks, entire communities, neighborhoods, individual buildings, down to specific rooms. Scalability has always been an important issue in information visualization and visual analytics as well. Recently, increased data sizes and data complexity led to innovative strategies in incremental analytics and prediction, for example. Cartography has been an area where several scales have traditionally been handled concurrently. In the past, islands of data were living in relative isolation. Today, this information is increasingly growing together to fuel the digital transformation of society. The need for data integration poses interesting challenges with respect to scales, multiscale representations, and scalability. Interesting topics in this respect are: crossscale visualization and interaction, massive multiscale techniques, scale integration,

crossscale labeling and annotation, cross scales on time and space, cross scales on structure and dynamics (e.g., physiology), and continuous scales.

2. Scalable Molecular Visualization Fundamentals

Structural biology as a scientific discipline drives the need for interactive molecular visualization and has dramatically developed over the last decades [7]. Currently, small molecules with only thousands of atoms or short molecular dynamics simulations with only thousands of frames are rarely interesting for researchers anymore. The analysis rather focuses on very long simulations of structural models, where several molecules can mutually interact with a macromolecular structure. A ligand, which is an interacting chemical compound, is often simulated to interact with the studied macromolecule. Furthermore, the solvent molecules are also present in the simulation, raising new challenges for the visualization. Nowadays, it is no longer an issue to render several thousands of atoms interactively, even if they change over time. Now the challenge is to understand the dynamic behavior, captured in several millions of timesteps. A direct playback of such long molecular dynamics sequences is unsuitable for a visual analysis and more advanced techniques are required that convey several scales of dynamics.

In the past, biologists have developed a visual language [8] to communicate various features inside of molecules, such as bond types, secondary structures, or boundary representations. All these visual representations are designed for a single scale and are limited to a few thousands of atoms. In the example of DNA-nanotechnology, however, tens of thousands of base pairs are potentially needed to construct a particular nano-machine. The designer would construct a complex model piece-by-piece where all structures are formed by a very long DNA strand or a set of DNA strands. Even if an all-atom model could be displayed at interactive framerates, the representation would not be effective in communicating information necessary for a multiscale structural design. For this purpose, new visual metaphors are needed to tackle the visual complexity and allow the designer to access and modify the model on various levels of structural organization.

Previously, researchers in the field of molecular visualization focused on interactive visualization of a single molecule or a few molecules at most. Integrative structural biology research creates all-atom models of entire organisms, bacteria, cell organelles, up to entire cells. It is a considerable challenge to interactively display such huge datasets and only achievable by a smart combination of various real-time rendering algorithms. Care has to be taken that the resulting image contains expressive and interpretable information. If, for example, structural details of a particular scale are conveyed, but the projection- or zoom-level settings produce views too close or too far away, the resulting image will be either cluttered with details that are impossible to comprehend or too little information will be visually accessible. Therefore, visualization designers need to consider which scale to visually promote for a given viewpoint setting and which interaction metaphors are available for that particular magnification level.

2.1. Tasks Influencing Visualization Decisions
Overall, data coming from simulations, modeling, or imaging is nowadays substantially bigger, longer, and more complex than datasets from just a decade ago. The suitability

of a particular visualization technique, whether multiscale or not, highly depends on the involved goals and tasks. Tasks are categorized into analytical, modeling, and explanatory ones.

Analytical tasks. Analytical tasks typically concern the interactive *analysis* of molecular dynamics simulations. This requires the selection of certain data subsets and to check their relevant physical quantities. In case of multiple simulation runs, comparison is an important task to assess whether certain conformations remain stable with respect to slightly changing initial conditions. In particular, the analyst needs mechanisms to cope with the temporal emergence of features. For shorter molecular dynamics sequences, the movements of individual components are captured in trajectory data, which can simply be re-played, to visually identify interesting events. Nowadays, molecular dynamics simulations contain millions of timesteps and direct browsing through the trajectory-data timesteps is no longer feasible. In this case, selective visualization [9], visual guidance [10], overview and detail [11], focus+context [12], visual aggregation [13], reformation, deformation, or warping [14] can be strategies to convey trajectory information on various levels of scale in an effective way (see Sect. 3).

Modeling tasks. Modeling tasks occur in the computer-aided design of DNA-nanotechnological objects where the primary goal is to interactively *model* a nanostructure and to assess its various qualities. Previously, this modeling process has been possible on a single scale and with a static schematic layout. With increasing model complexity, however, the result is less and less understandable to the modeler and even more so to the experts that were not involved in the modeling process. Common tasks are the modeling of a structure from scratch, or the modeling of a 3D DNA structure that is shaped analogous to a given macromolecule. This is usually a protein whose function shall be replicated. During this process, researchers need to access simple visual representations that characterize various spatial organizations. A typical approach is to develop a seamless transition between these spatial scale organizations so that the modeler can at any time access a desired level he or she is interested to inspect. Understanding the correspondences between a schematic two-dimensional layout and the corresponding three-dimensional conformation is another important aspect during the modeling task. Here, linked views can be used, connected through interaction schemes, such as selection or brushing [15, 16, 17]. Another strategy are integrated views where the two-dimensional schematics can be smoothly transformed into the three-dimensional representation and vice versa. Such morphing through animation establishes visual correspondences between two different structural arrangements in the same visualization space. We provide more details about these techniques in Sect. 4.

Explanatory tasks. Our example for an explanatory task comes from integrative structural biology, which is concerned with the assembly of larger biological arrangements. These consist of known elementary pieces and the distribution of these pieces in the particular biological system. Complex structures can be assembled, like viruses, bacteria, cell organelles, and hopefully soon even entire eukaryotic cells. In the future, dynamic models will clearly be of special interest, but currently rather static multiscale models of complex biological organizations are available. These models contain the macromolecular composition of proteins, lipid membranes, and fibers of sugar polymers or nucleic

4

acid strands. The total atom count varies between several tens of millions up to several billions for models with roughly one micrometer in diameter. While the integrated models are relevant for the domain science itself, they are also invaluable for scientific outreach [18]. They present tangible, understandable, and fascinating depictions to the broader audience. Typical visualization tasks comprise: the interactive 3D visualization of the structures, where the observer can freely navigate in the scene; comprehensive communication of structural details on all scales of spatial organization; and visibility management tasks [19] to overcome the occlusion of internal structures. The first group of visualization tasks is related to problems addressed in multiscale visualization for 3D virtual globe technologies [20]. Compared to these approaches there are two significant aspects that increase the complexity for the visual *explanation* of biological structures. Virtual globes deal with multiscale models that are organized on 2D manifold structures, while biological organisms are organized as dense three-dimensional collections of building blocks. Each building block occurs up to several millions of times in the structural composition of the modeled organism (e.g., hemoglobin molecules in a red blood cell). Understanding the guiding principles behind the hierarchical biological organization is the key for designing effective rendering acceleration algorithms as well as visual encodings that expressively communicate details of the structural composition on all spatial scales. We discuss details about explanatory visualization tasks in Sect. 5.

2.2. Challenges for Visualization

The three examples presented above illustrate how multiscale visualization supports a spectrum of analytical, modeling, and explanatory tasks [21]. While there are certain specifics, on a general level the visualization tasks have many commonalities. First, a shared challenge is to efficiently handle large data for interactive rendering. Second, multiscale visualization often visually abstracts [22] to navigate the viewer across levels of structural or dynamic details. Sometimes the visual abstraction creates interactive overview maps, to access desired detail information quickly and effortlessly. In another case, visual abstraction controls the complexity of the illumination scheme or the geometric detail. Various types of visual abstractions imply the existence of a multidimensional visual abstraction space that can be effectively utilized in multiscale visualization tasks. Such a spanned space allows the user to seamlessly transit between visual representations for the purpose of integrated visualizations. Visual abstractions together with spatio-temporal deformations, reformations, and warping schemes [23, 24] effectively support visibility management [19]. Visibility management alleviates spatial or temporal occlusions or reduces the overall visual complexity to the essential visual information that is necessary to complete a particular task. Navigational guides, such as textual labels, are designed to present information at the best visible scale of the current 3D visualization. Access to multiple scales implies the necessity of a scale-dependent user interactions [25]. These interactions are either intended for 3D navigations or modifications of structural details during modeling tasks.

The interplay of domain requirements for handling complex molecular data and multiscale visualization techniques is depicted in Fig. 1. We now present particular domain scenarios and concrete multiscale visualization techniques that give access to the complex information in an effective way. This review is by no means exhaustive,
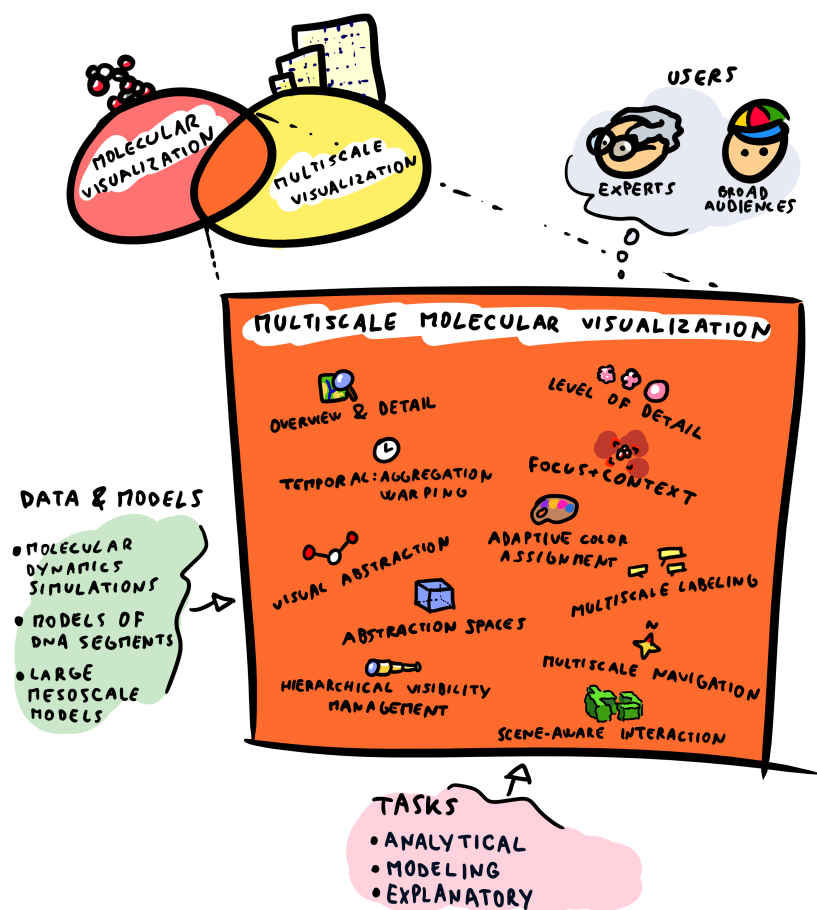
Figure 1: Multiscale molecular visualization in a nutshell: The long, complex, and large models pose problems for a comprehensive representation in the context of analytical, modeling, or explanatory tasks. Several visualization approaches, often in combination, can effectively address these challenges.

there exist many other examples, which could be illustrative showcases the same way as those presented below.

3. Visual Analysis of Molecular Dynamics Simulations

The exploration of molecular dynamics simulations is a prominent domain where multiscale molecular visualization has been intensively applied. Here it facilitates our understanding of the behavior of molecular complexes and the interactions between molecules [26, 27].

Molecular dynamics is currently one of the most powerful computational methods to simulate conformational biomolecular transitions that induce chemical reactions. Molecular dynamics is a very popular and commonly used technique, although it cannot

model complex processes such as creating and breaking chemical bonds. For this we would need to use quantum mechanics or hybrid simulations. In these simulations, particles shift according to their potential and kinetic energies and their re-locations result in permanent movements of the whole molecule. The particle movements influence different properties of the molecule, such as its stability or reactivity. Protein structures fold to recurring structural motifs and they can also carry a limited set of recurring mobility patterns [28]. The simulation of molecular dynamics basically results in a set of motifs, forming fingerprints of the dynamics. The subsequent comparison of these fingerprints is a possible option to identify similarities and differences between the simulations. A study by Hensen et al. [28] concluded that for most proteins there is a strong correlation between protein structure and dynamics, i.e., similar protein structures carry out similar dynamics.

Chemical processes are caused by the underlying physical phenomena, which can be expressed by systems of equations. Molecular dynamics simulations are derived from various computational models, differing in scale and approximation level. They range from quantum-based to coarse-grained approaches. The applicability and usefulness of these models have been shown in many studies [29, 30].

Computational demands for molecular dynamics simulations depend on three main factors: the complexity of the model (complexity of interactions), the size of the simulated system (number of particles), and the simulation length. Nowadays, the simulation length scales up significantly due to the increased computational capabilities. The possibility of simulating longer dynamics opens entirely new opportunities for the biochemists and biologists. For example, the passing of a virus through a cell's membrane is a lengthy process, which requires a long simulation period. This is a typical example of phenomena, which could not be studied and explored without sufficient computational power. In general, large systems require more time to express significant changes. The simulation of a whole cell on an atomic scale, observed for a picosecond, is meaningless as within this time interval typically no significant event occurs.

Currently, simulations are capable of capturing timescales of microseconds or even milliseconds [31]. Beyond increasing simulation lengths, another trend is to capture many simulations of the same process to reveal variations in the behavioral patterns of a molecule. The main task is to compare the content of multiple simulations and reveal similarities and differences. The increasing data scales pose many challenges for the exploration process, tightly related to the multiscale visual representations. The multiscale aspect requires to explore the simulations at different spatio-temporal scales. Specialized techniques, like temporal aggregation, reveal interesting features of the simulations that are not detected by investigating individual timesteps. Such strategies support the user to understand the main trends and patterns in the simulation as well as the detailed behavior of selected parts on an atomic level.

3.1. Tasks Associated with Molecular Dynamics Simulations

Molecular dynamics simulations focus not only on movements of the molecule itself, but also on its interactions with other molecules. Studying interactions between molecules is at the core of many research fields, such as drug design or protein engineering. Researchers aim at the correct combination of a protein and a ligand in order to design a new drug or to change protein properties and its function. The interactions between
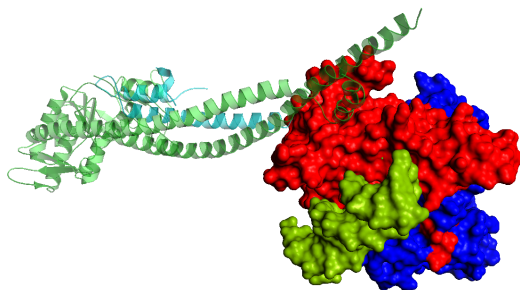
Figure 2: Part of the structurepreservation in the chromosome SMC5-6 complex of proteins, governing the organization of the DNA in the cell nucleus.

the ligand and the protein happen on the protein surface or deeply within the protein structure. In the latter case, the ligand traverses a void space of sufficient width that leads from the protein outer environment to its inner active site. This void space is called tunnel. A tunnel is categorized according to its properties and is visualized using one of several visualization techniques [32]. Other void space structures include channels and cavities.

The dynamic behavior of molecules is also influenced by their environment, i. e., the solvent that surrounds the molecules [33]. The molecules of the solvent interact with the protein in a similar manner as ligands, but the number of interacting molecules is by orders of magnitude larger. In this case, molecular simulation produces the trajectories of thousands of solvent molecules, floating around the protein and eventually moving through the protein's inner structures, i. e., through the cavities. This opens plenty of opportunities for visualization to explore the solvent behavior on different scales, with a focus on trends as well as individual molecules of the solvent.

Proteins often mutually interact also on their surfaces, forming protein-protein complexes, which are responsible for important functions in living cells (see Fig. 2). Several computational tools predict possible spatial configurations of mutually interacting proteins. A comprehensive survey in this respect was published by Huang [34]. The tools can produce an extensive number of possible configurations, which are further explored concerning biological relevance. This is again a task for multiscale visualization.

3.2. Visualization of Molecular Dynamics Simulations

The exploration process of large molecular dynamics simulations and their ensembles intrinsically requires a proper visual representation. Biologists and biochemists are trained and well accustomed to understand the static structure of molecules by using traditional visualization models and their combinations. They can easily perceive and understand the dynamic behavior by observing the animated movements of these models. Therefore, animating the 3D spatial representation of molecules has been the most common approach for decades. In essence this representation counts as the lowest-level temporal abstraction [22]. With the possibility to calculate very long molecular dynamics simulations, this traditional method has become tedious and cumbersome.
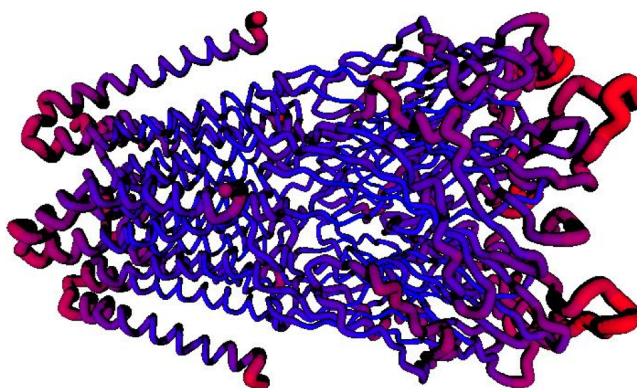
8

Figure 3: Visualization of the backbone of the GLIC ion channel augmented with its dynamic flexibility through a modulated tube [36]. Flexible parts (red) occupy more space than the rigid parts (blue).

Biologists therefore usually subsample their simulated trajectories, making the overall sequence possible to analyze in limited time. Generally speaking, temporal abstraction techniques [22] from the visualization literature, might be good candidates for supporting visual analysis of molecular dynamics trajectories. It has also been proven that human perception can process and analyze static images better than animations [35]. With the current simulation lengths, domain experts would have to spend hours and even days to observe the molecular movements. Except for this obvious disadvantage, the 3D representation suffers from occlusion issues, so even when observing the animation, important events can be easily overlooked. These problems lead to completely new approaches to visualize simulations, which utilize several hierarchical levels of visual abstraction. On the highest level, the basic statistical and physico-chemical information about the simulation is derived and presented to the user. This information can be depicted using simple chart representations, which help the user to select possibly interesting sequences of the simulation. The sequences are then explored in more detail on a lower level. The multiscale nature of the data is tightly connected with the dynamic behavior in the simulations. Proper visualization techniques help the user to explore details of individual timesteps or get a valuable summary overview by aggregating (e. g., summarizing, averaging) the information from the individual timesteps. Using such aggregated views is one of the most common techniques for creating an overview of a certain simulation property. An illustrative example is the visualization of the molecular chain flexibility. It can be mapped on a static 3D model, which depicts the molecular backbone as a tube. By modulating the width of the tube, we can naturally highlight the most flexible and stable parts of the backbone over time (see Fig. 3).

Visual Abstraction Techniques. Many existing techniques for the visual exploration of long molecular dynamics simulations on different scales are based on abstracted representations. They are often tailored to the exploration of specific properties, interactions, or inner cavities. The interactive exploration of a ligand trajectory inside a protein structure has been introduced by Furmanová et al. [37]. The visual analysis tool consists of a multiscale simplification model of the ligand trajectory. The user gets the information
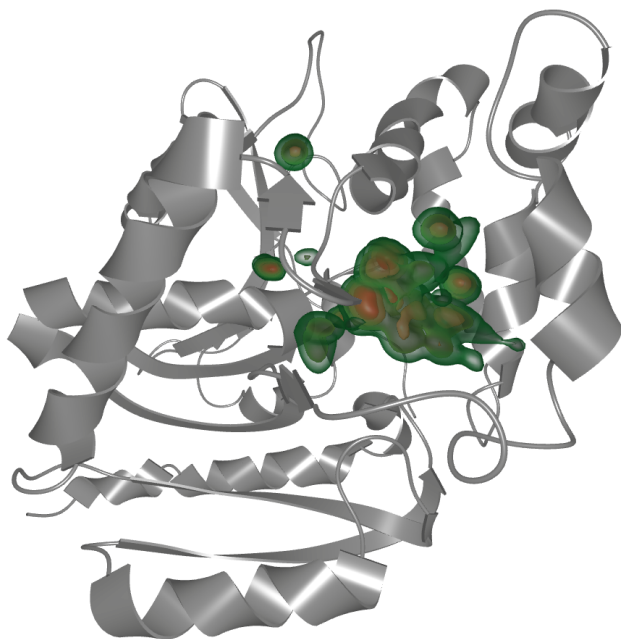
9

Figure 4: Isosurface representation of the inner protein space (in green) occupied by solvent molecules during a molecular dynamics simulation. The red isosurfaces represent the most densely occupied parts.

about the overall trend of the ligand movement within the protein. On demand, selected parts of the trajectory are augmented with details of the ligand movement. In this way, the user can observe several scales in one representation. The visual exploration of interactions between a protein and solvent molecules, using multiple scales, has been studied and presented by Vad et al. [38]. They propose a set of visualization techniques for the interactive exploration and filtering of solvent molecules entering the protein inner space. The techniques consist of highly abstracted 2D views, showing the overall behavior of solvent molecules and enabling their filtering according to different criteria. The selected trajectories can be further explored in more detail in 3D through individual, simplified paths. Another representation aggregates the information on the coverage of the protein by solvent molecules in the whole simulation. The occupied space is visualized in a static view, where the boundaries of densely occupied areas are represented by red surfaces and sparsely occupied regions are shown in green (see Fig. 4). This representation shows not only the most densely occupied parts of the protein, but also suggests the main entrance paths taken by the solvent.

Another set of techniques for visualizing the dynamic behavior of proteins is tightly connected to the entrance paths of proteins through voids, i. e., tunnels, channels, or cavities. The visual exploration of tunnels, leading from the protein outer environment to the active site, has been proposed in two publications by Byška et al. [39, 40]. Both techniques visualize the dynamic behavior of a single tunnel. The AnimoAminoMiner technique [40] enables researchers to visualize the tunnel profile over time, along with
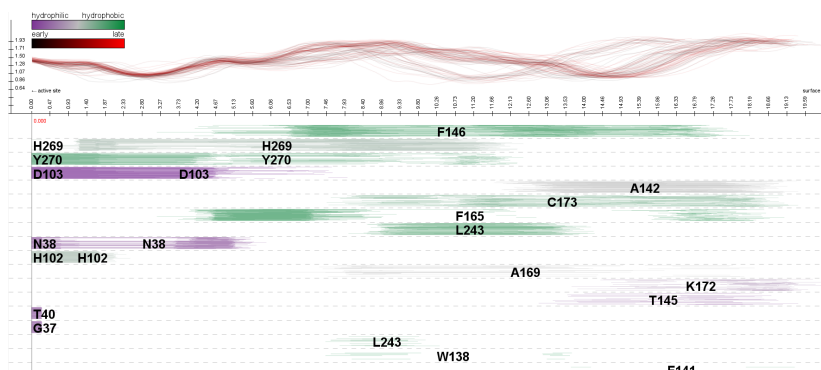
Figure 5: The AminoAnimoMinertechnique showing the tunnel profile and theset of amino acids forming the tunnel boundary [40].
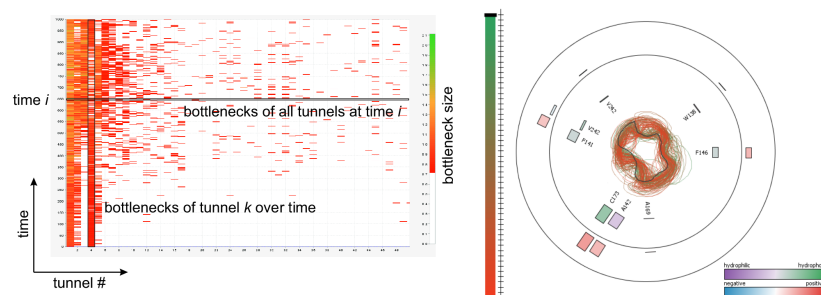

Figure 6: The MoleCollar technique showing (left) a heat plot representation to select a tunnel for exploration. The selection is based on the tunnel bottleneck. Temporal evolution (right) of the bottleneck shape in a selected tunnel [39].

the associated set of amino acids, which form the tunnel boundary (see Fig. 5). It clearly conveys information about the most stable and flexible parts of the tunnel with respect to its width. The set of amino acids, along with filtering and sorting according to different criteria, helps to find those amino acids, which influence the tunnel significantly. Mutating these amino acids leads to substantial changes in the protein function. The MoleCollar technique [39] focuses on the exploration of the most crucial part of a tunnel, i. e., its bottleneck. The bottleneck represents the narrowest location of the tunnel, limiting the size of a ligand, which can pass through. The visualization consists of two basic methods. The first method is based on heat plots (see Fig. 6 (left)) and enables the users to select the most appropriate tunnel for further exploring its bottleneck.

Appropriateness is determined by the tunnel width over time. A selected tunnel is then scrutinized using an abstracted view whose central region shows the contour of the tunnel bottleneck and its evolution over time (see Fig. 6 (right)). The central region is surrounded by rectangular representations of amino acids forming the boundary of the bottleneck.

11

(a) Scale 2: Atoms

(b) Scale 4: Nucleotides

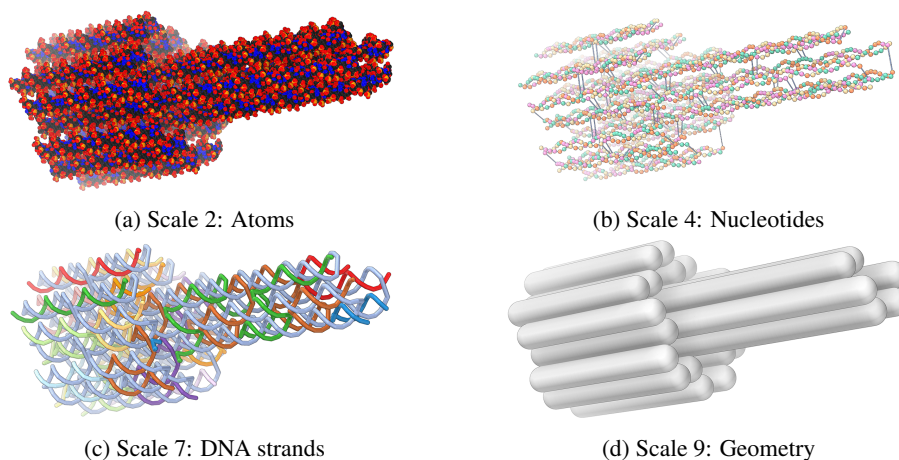(c) Scale 7: DNA strands

(d) Scale 9: Geometry

Figure 7: Different semantic scales of a DNA nanotube, starting with (a) a highly-detailed atom-based representation and ending with (d) a representation of the modeled higher-level nanostructure geometry. Users can seamlessly transition between these stages.

Temporal Aggregation Techniques. A more general tunnel visualization has been proposed by Krone et al. [41]. It supports the multiscale aspect by showing general statistics of the detected tunnels over time. As a starting point, the tunnels of the surface mesh are determined using the ambient occlusion illumination model [42], a widely used technique for approximating shadows. Their method identifies larger shadowed areas as tunnels. Properties of the detected tunnels are tracked over time, including covered surface areas as well as split and merge events. The results are either presented per timestep or as abstracted graphs of the computed values. While this method identifies interesting tunnels, a further analysis can be performed using other methods, such as the previously mentioned MoleCollar technique.

4. Visual Modelitics for DNA-Nanotechnology Design

Visual Modelitics combines automated, smart, and intelligent modeling and synthesis techniques with interactive visualizations for an effective specification of large, complex, dense, multi-instance, and multiscale environments. A key challenge in DNA nanotechnology is to represent complexity across several hierarchical scales, and also to facilitate the design and modification of these structures. The design of nanostructures such as produced by DNA origami [43] consists of many interwoven DNA strands, which in total comprise several hundred thousands of atoms. Structures of this size and complexity require visualization techniques that incorporate high-level views, which enable researchers to discard low-level distracting details. *Visual Modelitics* combines the visual analysis and synthesis of properties across scales with using scales for the interactive modeling of the underlying data. It supports complex modeling tasks through integrated analyses and recommendations. An important requirement in visual modelitics is that the visualizations consist of simple geometries, which, in turn, facilitate
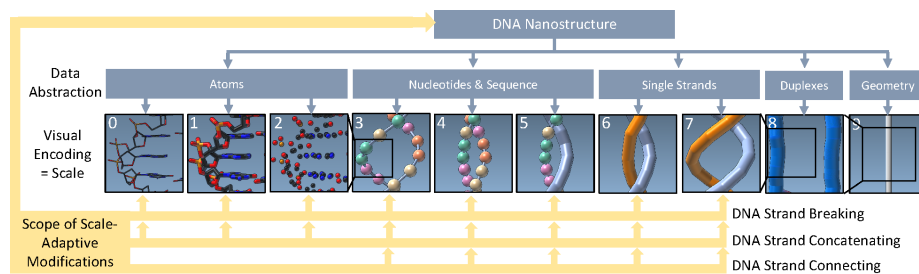
Figure 8: Miao et al. [44] describe an approach for visualizing and modifying DNA nanostructures on multiple scales. The DNA's atomic details are increasingly abstracted and seamlessly transitioned across ten scales. Based on this multiscale visualization, designers can apply operations such as DNA strand breaking, concatenating, and connecting at different scales.

interactive modeling. Crafting representations across scales and organizing them in a meaningful multiscale arrangement is a key challenge in visual modelitics. With the advancement of DNA nanotechnology, all-atom representations are no longer adequate to model the emerging large macromolecular structures. It becomes crucial for designers to interactively model not only with a single low-level representation, but rather on several high-level abstract representations that support scale-varying tasks. In the visual modelitics approach for DNA nanotechnology by Miao et al. [44], the authors developed ten scales for examination and interaction purposes that enable the designer to apply targeted operations. An example of such a multiscale approach is shown in Fig. 7, where a DNA nanotube is depicted at several scales.

4.1. Designing the Scales

Visual modelitics facilitates the multi-faceted visual presentation of the data and also its automation-supported interactive specification and manipulation. DNA nanostructure data is organized hierarchically and can be represented on several levels, as shown in Fig. 8. Miao et al. [44] created their approach in close collaboration with experts in the field. The complexity of the object is distributed among several representations, which enables the user to interact with the structure through task-dependent scales. Each scale highlights certain characteristics of the data, without a visual distraction from the other scales. This is demonstrated on the example of the DNA nanotube in Fig. 7.

With a multiscale visualization approach, it is possible to combine atom-based representations with abstract, non-atomistic scales. The problem with all-atom representations is that they are visually cluttered and it is tedious to modify a large number of atoms. Nonetheless, all-atom representations are needed as there are many situations where precise modifications of single atoms are required. Multiscale representations show the needed details in dependence of the tasks, and it is thus possible to satisfy both, initially incompatible constraints.

The complexity of nanostructure representations arises from the semantics of the scales, and also from their spatial layout. Modelers of DNA nanostructures, for example, often use specific layouts to model structural properties. caDNAno [46] is a popular tool that employs 2D DNA diagrams for the modeling of DNA origami structures. However, it does not provide the designer with mechanisms to understand and directly model the
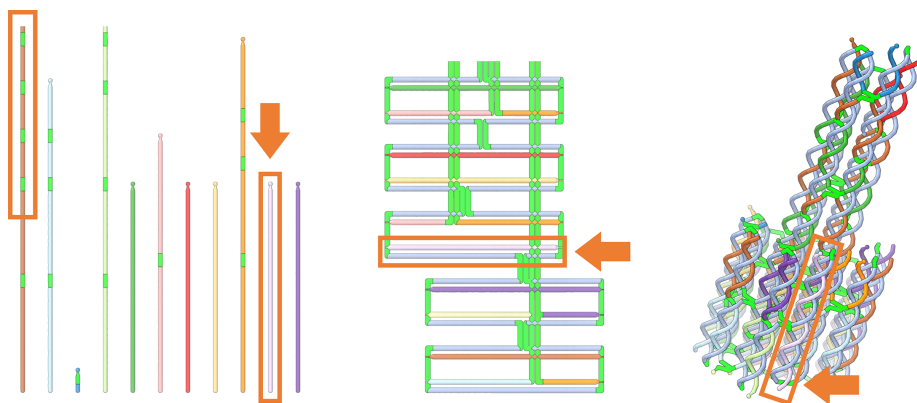
13

Figure 9: DNA nanotube in 1D, 2D and 3D [45]. Crossovers are highlighted in green.
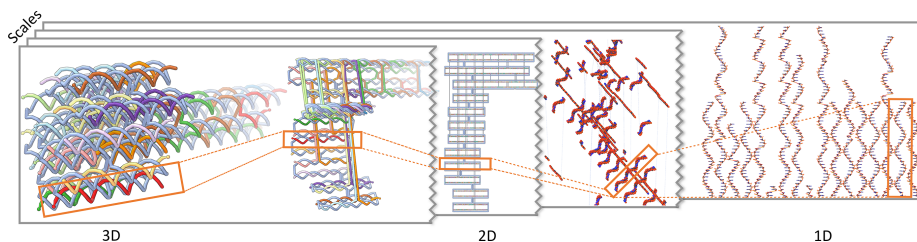

Figure 10: Through multiple scales and dimensions, a DNA nanotube transitions from a 3D shape to a diagrammatic 2D representation, and finally to a clean 1D alignment of single strands [45]. The orange rectangles indicate an individual strand through the dimension and scale transition.

resulting three-dimensional layout of the final structure. To further ease the modeling of these complex structures, Miao et al. [45] employed the spatial dimension as another axis of visual abstraction. As shown in Fig. 9, the DNA nanotube can now be depicted in 1D, 2D, and 3D layouts. Inspired by the primary, secondary, and tertiary structures of the DNA, the 3D spatial layout enables the estimation of distances, while the 2D diagrammatic layout is free of spatial occlusion. The 1D sequence layout finally enables the scientists to inspect the individual strands in an uncluttered way.

4.2. Seamless Transitions

The design across multiple scales advantageously reduces the overall complexity by utilizing different data representations, which highlight specific features. However, many different representations of the same data burdens the observer with a certain cognitive load when mentally linking those. For instance, in traditional molecular visualization tools, such as VMD [47] or Maestro [48], the user has to mentally link different representations. In order to ease the cognitive load, the relationship between the representations can be externalized and smoothly controlled [45]. Morphing between scales is an effective way to relate representations. For the transitioning between dimensions, the intermediate positions between two layouts have to be interpolated. For

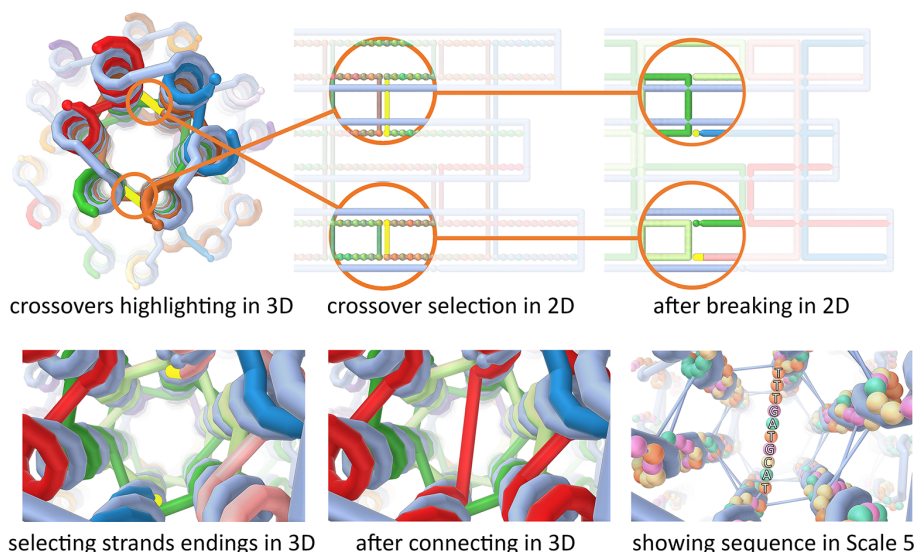| crossovers highlighting in 3D | crossover selection in 2D | after breaking in 2D |
| selecting strands endings in 3D | after connecting in 3D | showing sequence in Scale 5 |

Figure 11: A bridging strand is added to the DNA nanotube by connecting two strands [45]. The examination and operations are carried out at several scales and dimensions.

example, in Fig. 10 we show such transitions in action when scale and dimension are changed at the same time.

In addition to temporal transitions between different representations, it is also beneficial to support spatial transitions. This is in particular the case for DNA-based models, which are elongated and exhibit a high level of similarity at higher levels of abstraction. It is possible to change to a lower level based on the position along the DNA model as demonstrated by Lueks et al. [49]. This is possible under the assumption that a seamless transition between the different geometric and representational features can be designed. In both transition types, the resulting animation typically does not have a physically correct spatial interpretation, yet it eases the mental integration between different scales or dimensions.

4.3. Multiscale Modeling

An integral requirement of designing DNA nanostructures is is to facilitate modeling on muliptle scales. This enhances the structural stability as well as the design of structures with increasing complexity. With multiscale modeling [45, 44], Miao et al. support designers to apply operations at different scales so that the changes are automatically adjusted to the given semantic granularity. For instance, modeling operations, like DNA concatenating, connecting, breaking, and deleting, make up the basic operations that can be applied on all scales with analogous effects. The choice of granularity usually depends on the user task. In Fig. 11, the user adds a bridging strand into the cavity of a DNA nanotube. The chosen scale enables the modeler to select the nucleotide at the appropriate granularity. Instead of selecting groups of atoms in the backbone and side-chain of a nucleotide, only a single sphere is selected on a higher scale. The single

sphere represents all the atoms belonging to the nucleotide. As a multiscale operation, its effect is automatically propagated to other scales as well.

4.4. Abstraction Spaces and Overview Maps

Using different semantic scales and depicting data at different dimensional layouts are often independent from each other. Similar semantic scales can be shown for 3D, 2D, as well as 1D representations. These types of abstraction define an abstraction space in which practitioners can select the most suitable representation for a given task or goal. In addition to the axes of semantic scale and dimensionality, other types of abstraction are possible as well. For example, van der Zwan et al. [50] discussed the two abstraction axes "illustrativeness" of the depiction and support of spatial perception.

In an important application of these abstraction spaces they serve as mental models for the selection of appropriate representations. Miao et al. [45] use a two-dimensional abstraction space to assist designers in selecting appropriate levels of abstraction, to transition between different representations, and thus to better understand the whole visualization space and landscape.

5. Explanatory Visualization in Integrative Structural Biology

Over the years, structural databases [51] have grown in the number of identified structures. This includes larger macromolecular complexes containing several tens of thousands of atoms, or even entire viral capsids with several millions of atoms. Recently, biologists procedurally generated complex structures of viruses and bacteria into all-atom integrated models, where the atom count explodes as compared to previously available structural data [52] (http://www.cellpack.org/use). To specify an entire macromolecular composition of the *E.coli* bacterium, the model would need to contain approximately 15 billions of carbon, nitrogen, and oxygen atoms, abstracting away all hydrogens. A model of an entire eukaryotic cell would consist of around 100 trillions of atoms. This relatively new requirement of domain science raises an urgent challenge on how to effectively handle large data and how to display these at interactive framerates. The large range of spatial scales from a single atom ($\sim$0.1nm) up to a complex eukaryotic cell ($\sim$50μm)clearly requires data handling and rendering to be multiscale in nature.

The key task of explanatory visualization is to comprehensively communicate the essentials of a particular structure or phenomenon. In a multiscale context, the structural organization of each scale and the interaction between scales are among these essentials. First and foremost, only an interactive visualization can engage the audience to explore the essentials by themselves. Then, *staging* is performed, i. e., the visualization is split into several discrete stages, consecutively demonstrating the high-level details and the low-level individual deep structural or physiological details. Visualizations need to incorporate level-of-detail schemes for all visual information, whether this concerns geometric details, illumination models [53], coloring, visibility management [53], or labeling [54]. Furthermore, interaction and 3D navigation are also adapted to the currently visible scale detail. Next, we review exemplary cases of required technologies.

5.1. Multiscale Molecular Rendering

Rendering of molecules in multiple scales requires the usage of a variety of computer graphics techniques and algorithms. We discuss the most prominent ones below.

Procedural Billboarding. Triangle meshs are commonly used in many computer graphics applications today, but are unsuitable for rendering of large molecular scenes. Each atom, represented as a sphere, would require many triangles to smoothly approximate a spherical surface. The approximation of a surface by triangles or planar polygons is called tessellation. Tessellation artifacts are especially visible and disturbing on the contour of the approximating polyhedron, where a curved contour is represented by an edgy polyline. Interactively rendering tens of thousands of atoms, each represented by a spherical mesh, easily becomes intractable. The need for high triangle rates has been effectively eliminated by the procedural billboarding technique by Tarini et al. [55]. A sphere is thereby represented by a billboard consisting of just one triangle or one quad of geometry and the rendering of an atom is done through ray casting. Billboards are simple 2D patches that mimic the visual impression of much more complex 3D spherical meshes. They are placed in the scene instead of tessellated spheres and are oriented to always face the viewer. The size of a billboard ensures that it tightly encloses the projection of an atomic sphere on the image plane. Billboards are positioned either just in front or at the center of the atomic sphere. The last stage of the graphics pipeline produces the final image. Billboards are rasterized and through ray casting the correct depth value, i. e., distance from the viewer, is determined. The ray from the viewer intersects the billboard and the corresponding sphere. Here the distance to the sphere and not to the billboard is required for further calculations. As an example, for shading the necessary normal vector is determined from the sphere center and the sphere-ray intersection.

Image-Space Metaballs. Metaballs [56], also known as Gaussian surface models, are an implicit representation where atoms define point densities and the desired 3D structure is an iso-surface in the underlying density field. With metaballs, individual atoms are still distinguishable, while macromolecules appear as cohesive structures. Metaballs produce better results than billboarding, for example. In computer graphics rendering algorithms are either object-space (exact, costly) or image-space (approximative, cheap) approaches. Rendering metaballs in object space either by polygonization or by ray casting is computationally too expensive for large datasets. Therefore, in molecular rendering, image-space techniques are frequently used. Image-space approaches replace a complex 3D geometry by a simplified 2D patch with additional textural detail information. Billboarding is one example of such an approach. Müller et al. [57] proposed two methods to render large point datasets as metaballs in image space. The first approach uses a texture to store vicinity information. The performance is limited by the high number of necessary texture accesses. The second method approximates the metaballs per pixel by sliding a view-aligned depth image. This avoids the vicinity texture. Without pre-computations the approach is suitable for displaying molecular dynamics simulations.

Occlusion Culling. The above techniques discuss the rendering of individual atoms with efficient drawing routines. A different strategy to ensure interactive performance is occlusion culling. Here, objects are not rendered if they are either outside the visible region or hidden behind other objects. This technique is particularly efficient for molecular environments since they are usually densely packed with molecules, of which only a small part is actually visible. There are several ways to realize occlusion culling, the two most common types are early-depth rejection and techniques that make use of the GPU's z-buffer, which stores the depth values of the scene.

Early depth rejection is a feature supported by most GPUs to reduce the overall number of rendered pixels. A pixel is rendered multiple times if multiple objects occupy the same pixel position. Early-depth rejection is capable to reduce the number of times a pixel is rendered. It compares the depth of the previously rendered pixel at this position with the current one. In case the depth of the current pixel is greater (further away from the camera), it is occluded by the previous pixel and thus does not need to be rendered. This saves computation time and improves the performance of the rendering. However, this technique does not directly work together with procedural billboarding. Early-depth rejection relies on fixed depth values. Procedural billboarding renders a billboard instead of the actual geometry where the depth reflects the one of the billboard and not the desired geometry. In order to adapt the depth so that it corresponds to the actual geometry it needs to be changed. This contradicts the requirement for early-depth rejection. Modern GPUs offer a solution through a feature called 'conservative depth output.' This feature does not require the depth value to be fixed. Nevertheless, to make this feature work the depth value cannot be arbitrarily changed but rather only increased or only decreased. For instance, if we only want to increase the depth of a pixel, we can simply render the billboards not at the center of the desired geometry (e. g., atoms) but in front of them instead.

Early-depth rejection has the potential to improve performance though also has its limitations. It is applied on the latest stage of the rendering process. There is no performance improvement of the processing that happens before. Algorithms that exploit the GPU's z-buffer can exclude whole objects from the rendering altogether when they are occluded. Such techniques initially compute only the depth values of current frame, or utilize the depth values of the previous frame. With the help of the now existing depth values (GPU's z-buffer) the visibility of the objects in the scene is determined. This technique can be improved by using the hierarchical z-buffer solution [58, 59]. The hierarchical z-buffer improves the performance of the depth comparison by creating multiple resolutions of the depth buffer. A further extension of the hierarchical z-buffer is presented by Le Muzic et al. [60] making use of the temporal coherency of the scene.

Instant Construction of Biological Structures. With larger and more complex macro-molecular structures, new visualization challenges emerge for the rendering approach and also for the memory management. Macromolecular structures may be too large to fit into GPU memory and data streaming from the CPU memory is potentially too slow for real-time purposes. In 2007, Lampe et al. [62] introduced a two-level method to illustrate slow dynamics of large protein assemblies. Instead of transferring the positions of all atoms to the GPU, they just transfer the much fewer positions and rotations of all amino acid residues. This approach exploits the natural hierarchical
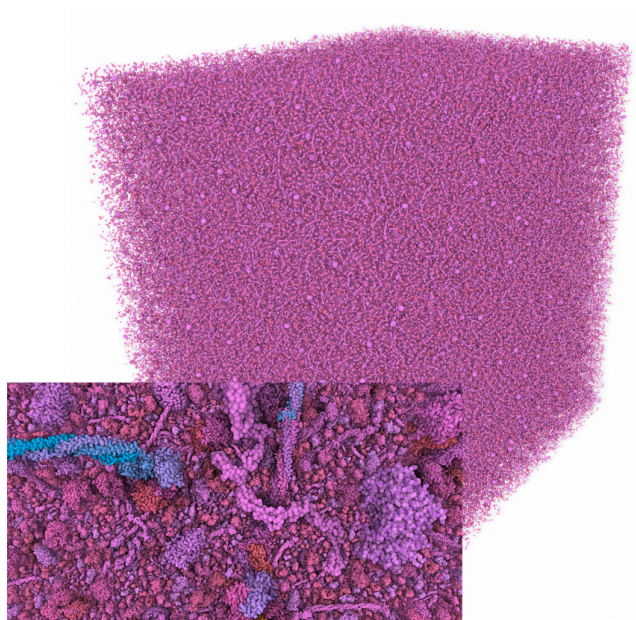
Figure 12: Two cubic microns filled with blood plasma, with over 4.6 million soluble protein ingredients, populated with the instant construction algorithm by Klein et al. [61] in 600 ms. This work demonstrates the potential of on-the-fly generation and visualization of large-scale molecular scenes.

structure of proteins. The geometries of the residues are constructed on-the-fly in the geometry shader stage on the GPU. Le Muzic et al. [63] later extended this concept beyond composing proteins to generate entire molecules directly on the GPU. They use a single vertex as input per molecule and utilize the tessellation shader to emit, i. e., generate, its individual atoms. These approaches improve the rendering performance drastically, however, they still require the transfer of positions and orientations of individual molecules to the GPU. Biological mesoscale models, like viruses and bacteria, are often computationally generated from experimentally obtained building rules. A more recent instant-construction approach by Klein et al. [61] incorporates such rules to unite the modeling and visualization phases. It is able to generate large biological models, as depicted in Fig. 12, on-the-fly. Within one framework, soluble components, membranes, and fibrous structures are interactively generated and visualized. This provides the possibility to render large-scale scenes by generating and visualizing only the currently visible parts of the scene, which considerably reduces the computational and memory requirements.

5.2. Multiscale Visualization Techniques

Visualizing molecular biological structures for presentation purposes often requires an enhanced visual appearance of the displayed scene. For this purpose, visibility management concerning the displayed objects, intuitive navigation, and guidance can be applied.
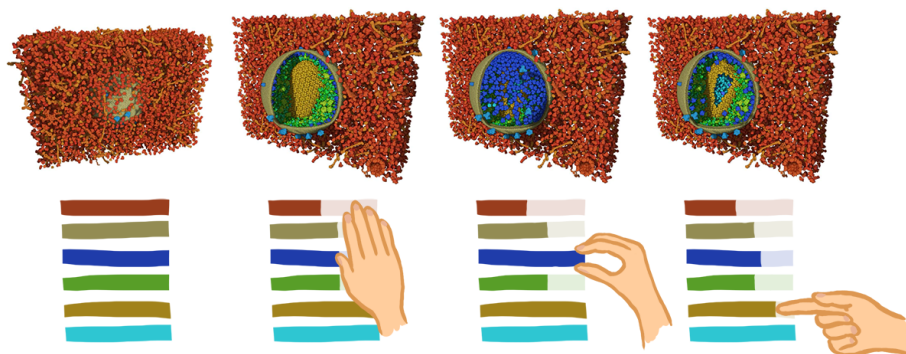
19

Figure 13: Visibility Equalizer to selectively visualize a subset of molecular structures. The result depends on the chosen clipping geometry and the desired number of molecules to display [64].

Visibility Management. Complex mesoscale models are densely packed, i. e., outer structures heavily include inner structures. Cutting planes are often used to expressively visualize interior parts. However, with cutting planes or other cutting objects, the illustrator does not have full control on the removal of structures, like proteins, and on how many of them are still visible. Le Muzic et al. [64] proposed the Visibility Equalizer to control the number of molecular structures removed by various cutting objects. Initially, the Visibility Equalizer conveys through stacked bars an overview of the number of removed as well as visible structures according to type (e. g., differing protein macromolecules). Subsequently, the user can modify some of the stacked bars to increase or decrease the desired number of visible structures. This either leads to the addition of previously removed structures or the removal of structures not yet cut away. The user controls the visibility of the various protein types, which together with the cutting objects reveals inner structures. The Visibility Equalizer also exploits the hierarchy intrinsically given in many mesoscale models. The stacked bars are displayed for every level of the hierarchy, where intermediate levels correspond to compartments in the mesoscale models. The visibility can thus be controlled for entire compartments, and not just for individual protein types only.

Visual Guidance. In multiscale situations, typical visualization steps, like color or textual label assignment, require quite sophisticated solutions. In an all-atom molecular visualization, the coloring is given by conventions from chemistry. The CPK model, for example, provides default colors like white for hydrogen, red for oxygen, gray for carbon, etc. On another hierarchy level, a different coloring scheme is needed to distinguish amino acids, secondary structures, or compartments. The most suitable coloring depends on the scale the model is observed. In interactive 3D visualizations, assignments need to happen dynamically. Waldin et al. [65] proposed dynamic coloring in multiscale environments. They consider available color as a resource that needs to be distributed. Color assignment is based on the number of distinct structures, which are visible in the current view, and on the particular magnification or hierarchy level. Various levels of spatial detail and the corresponding colorings are shown in Fig. 14.
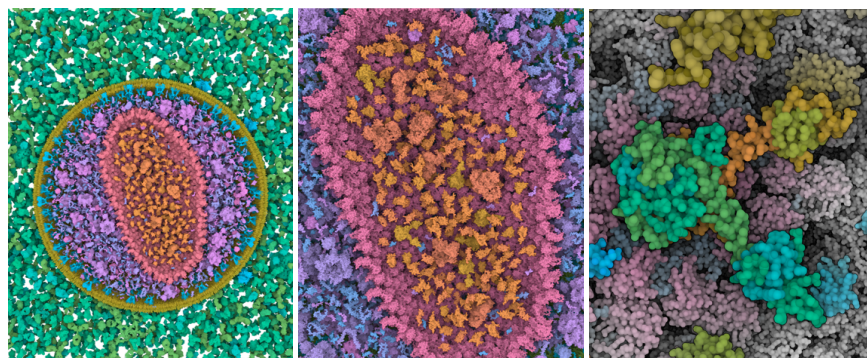
Figure 14: Dynamic adaptive multiscale coloring: The camera zooms into details and the coloring adapts to support discrimination of structures at the current scale [65].

To establish structural and hierarchical relationships between elements, colors can be used, but textual labels as well. Labeling has already received much attention in cartography and visualization research [66]. Kouřil et al. [67] proposed dynamic labeling of dense multiscale 3D molecular environments. The scene hierarchy is matched by an associated label hierarchy. For each image region the currently given detail level determines the corresponding label level. Labels of only clearly visible structures are displayed. An example is given in Fig. 15.

3D Navigation. Complementary and equally important to multiscale visualization is multiscale interaction. The DNA-nanotechnology scenario of Sect. 4 exemplifies the importance of scale-aware interactions for modeling. For 3D navigation McCrea et al. [68] proposed a cubemap that guides camera settings in support of multiscale interaction. A cubemap is an image-space approach where the 3D environment is projected/mapped onto a cube enclosing the camera. Costly processing of 3D geometry is again traded for handling the six 2D images on the cubemap. It is important to adjust the camera speed when traversing the scene across scales. It is also essential to select the appropriate scale and object the interaction should operate on. When moving through the scene different parts are of varying interest. The cubemap efficiently informs the camera about relevant objects in its vicinity. In addition, it facilitates the determination of paths so that the camera will not collide with objects in the scene. These interaction techniques have been showcased on a virtual 3D globe, but are equally relevant in the context of mesoscale visualizations.

Multiscale Camera. The rendering of hierarchical mesoscale models requires a camera that can accommodate the various scale levels in the scenery. Continuously zooming from an entire cell down to a single atom with a traditional camera based on a single transformation matrix would result in numerical issues. These are addressed by Axelsson at al. [69]. Their work focuses on navigation in scenes depicting the universe. The solution is relevant for multiscale molecular visualization as well, since the range of scales in the universe (galaxies vs. planets) is comparable to the one in the molecular world (cells vs. atoms). The authors propose a dynamically assigned frame of reference.
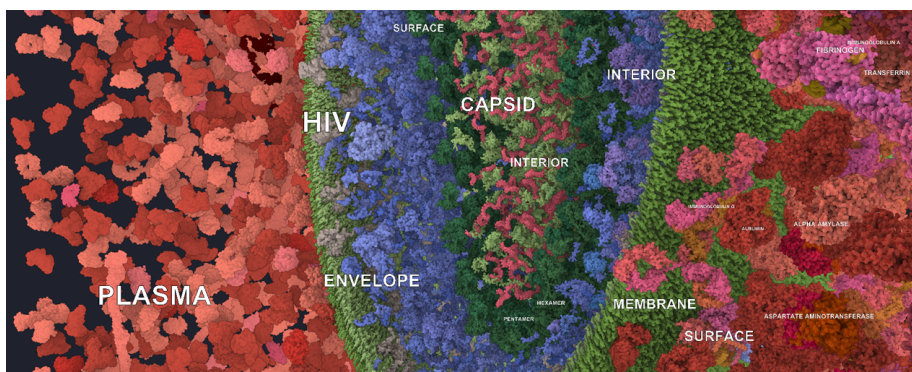
Figure 15: Multiscale labeling of an HIV in blood plasma [67]. For each image region the label is displayed that annotates the currently best visible structure.

The hierarchical scene is organized in an acyclic directed graph, which is called *scene graph* [70] in computer graphics. Depending on the current object of interest, the camera is attached to individual nodes of the scene graph. The scene graph is then traversed from this node, ensuring a higher precision as compared to the positional precision of the scene elements. Animations are supported through dynamic camera reattachments in the scene graph.

Illustrative purposes sometimes require the simultaneous display of multiple scales in a special oblique view. In the illustration of certain processes this might be helpful to also communicate the context. Hsu et al. [71] propose a method to seamlessly integrate into a single image several renderings of a scene from multiple viewpoints. The method utilizes image masks for individual viewpoints. The seamless integration of the individual renderings is based on non-linearly bent camera rays, which traverse the scene. Properly setting up the cameras, produces an integrated multiscale rendering of the scene. Similar techniques have been proposed for 3D virtual globe technologies.

Depth Perception Enhancement. It is difficult to perceive intricate molecular shapes in 3D space with traditional shading models from computer graphics. Ambient occlusion is a widespread technique in molecular visualization to emphasize hollows and protrusions on a molecular surface, even down to an atomic level (see Fig. 16 and Sect. 3). The technique increases depth perception by simulating the transport of diffuse light between objects, i.e., atoms, leading to localized shadowing. Ambient occlusion is computationally very expensive, thus several accelerations have been proposed. Ambient occlusion can be approximated in a postprocessing step, as an image-based technique. This variation is called Screen-Space Ambient Occlusion (SSAO) [23, 72]. SSAO is a fairly realistic approximation of ambient occlusion, while the computational requirements are independent from the 3D scene complexity. SSAO is significantly faster than objects-space ambient occlusion for very complex scenes, such as molecular landscapes.

Another possibility is to simulate ambient occlusion in object space. Grottel et al. [73] proposed an interactive technique, based on local neighborhood information.
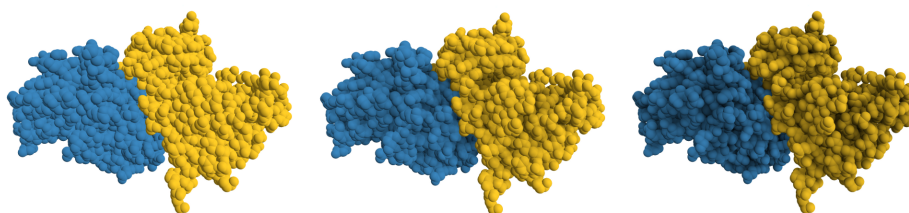
Figure 16: Comparison between local illumination (left), Screen-Space Ambient Occlusion (middle), and Object-Space Ambient Occlusion (right) [73].

The interactive performance is reached by the aggregation of the particle dataset into a coarse resolution density volume. The density value directly influences the geometry neighborhood taken into account when calculating the ambient occlusion.

In multiscale environments, shading is done on several levels. A relatively inexpensive approach performs several consecutive SSAO passes with different parameters. Le Muzic et al. [60] used two passes with varying the impact radius in the SSAO algorithm, effectively implementing a two-level shading. The first pass determines the shading of atoms inside a molecule, while the second pass approximates shading caused by other molecules. Depending on the distance to the viewer, these two results are either combined or turned off.

Depth perception in complex scenes can be assisted by other approaches as well. The method by Parulek et al. [74] visualizes simulations of molecular dynamics in real-time, incorporating full shading and contours to enhance depth perception. The method utilizes three representations of the molecular surface, which based on the distance to the camera are seamlessly combined into one visualization. The three levels of representation are associated with specific shading levels. On the level closest to the camera, local diffuse shading in combination with constant shading is employed. The middle and far levels use constant shading. Their method incorporates silhouettes and contours as well.

The method by Lindow et al. [75] interactively visualizes very complex scenes, which contain biological structures with several billions of atoms. The method introduces an efficient GPU rendering scheme and also deferred shading for visualizing the depth in the scene. Deferred shading is a two-pass rendering concept, postponing most of the complex rendering tasks, e. g., lighting, to a later stage. This stage is typically carried out in image space and operates of series of buffers, such as depth and normal buffers, as well as an object ID buffer.Molecular structures are visualized according to their distance where the authors propose a shading that is based on three different normal calculations. For a position close to the camera, they compute the analytically correct normal. With growing distance from the camera, this normal is linearly interpolated with an approximated normal derived from surrounding positions. For positions far from the camera, the approximated normal is interpolated with the reversed view direction.

Geometric Level of Detail. Photometric abstraction [22] concerns different levels of illumination, which are varied acoording to the scale that is shown. Orthogonal to abstractions based on illumination and color scheme, there is also a structural or *geometric*
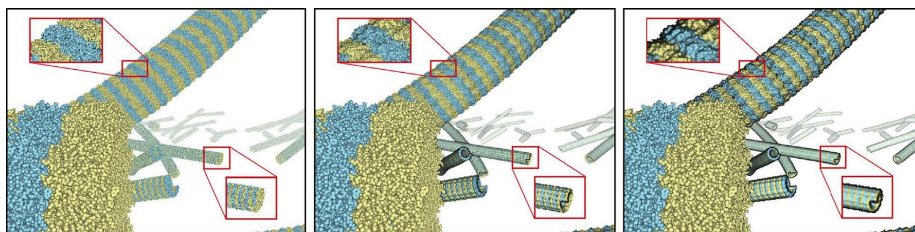
23

Figure 17: Three types of normal calculations, depending on the distance d of a position from the camera: analytically correct normal (a), approximated normal (b), reversed view direction (c) [75].

abstraction. DNA nanotechnology (Sect. 4) utilizes many geometric abstractions to facilitate efficient design workflows. The display of several scales from atomistic detail up to the level of entire organisms employs geometric abstraction as one of the most essential concepts to enable cross scale traversals.

The current way of handling geometric abstraction for atomistic data is to gradually abstract from the geometry, dictated by the van der Waals representation, into clusters of atoms. The ultimate coarse detail leads to representation, where one or few spheres approximate the macromolecular shape instead of the full atomistic detail. This can be done either by heuristics such as skipping every k-th atom of the structure and increasing radii of the remaining atoms to preserve the overall volume of the macromolecule [63] as shown in Fig. 18. Another approach uses a clustering algorithm for merging the individual atoms into higher level geometric representations. The clustering approach will come with a slight memory overhead of representing the intermediate nodes of the hierarchy, but will generally better preserve anisotropically structured macromolecules. For clustering, the affinity propagation approach seemed to be the most suitable approach [76].

Bridging Temporal Scales. We have discussed temporal abstraction mostly for visualizing molecular dynamics simulations. The diffusion process of entire molecules, however, happens on a different temporal scale, typically modeled as a random walk simulation. Random walks cause chemical compounds to come close together from time to time and under favorable conditions to react with each other. This occurs on yet another
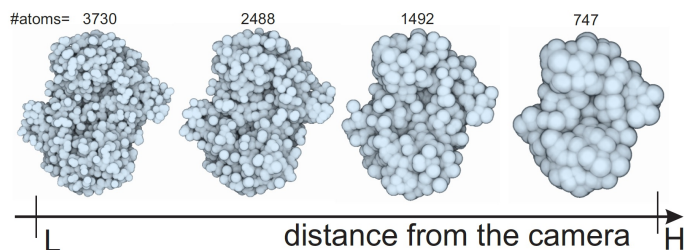

Figure 18: Geometric level of detail from atomistic resolution down to shape approximation with few overlapping spheres [63]

.

temporal scale. On average a reaction happens only once every thousand diffusion steps. Either the diffusion steps or the reaction events can be observed. The simultaneous perception of both phenomena requires dedicated visualization techniques. Le Muzic et al. [24] have proposed a technique to bring the two timescales together. The physical time is mapped to visual times, which are different for the diffusion and the reactions. Diffusion trajectories are smoothed to accommodate an accelerated visual diffusion time. The visual reaction time is decelerated to allocate more temporal emphasis to the fast-paced events. The unequal time warp allows the viewer to perceive both phenomena simultaneously.

6. Tools

Multiscale visualization is an emerging topic of research. Some of the mentioned techniques have already been integrated into software tools. Initially we mention systems for molecular-dynamics simulations, then frameworks for nanotechnology, and finally tools for explanatory molecular visualization are discussed. This is not a comprehensive review of all available tools, but rather of selected applications concerning multiscale techniques, which have been presented in this paper.

VMD – Visual Molecular Dynamics [47] is a multiplatform robust software tool for the simulation, visualization, and analysis of large biomolecular systems. It supports a wide range of molecular representations, coloring styles, transparency, and material properties. The NAMD [77] software results from the same research environment as VMD. NAMD is a parallel molecular dynamics simulator that is capable of scaling to hundreds or thousands of processors. This enables the large scale modeling of cellular systems.

MegaMol [78] is an open-source cross-platform visualization prototype. It quickly renders large particle-based structures and serves basically as a development framework for fast prototyping of new algorithms and visualization approaches. MegaMol is suitable for large static as well as dynamic scenes.

CAVER Analyst [79] is a multiplatform tool for the visualization and analysis of protein structures and their molecular dynamics simulations. The major focus is to explore tunnels in proteins and their dynamic behavior. The latest version supports advanced visual analysis by integrating multiscale representations as described above [39, 40], for example showing an overview of tunnel behavior.

SAMSON [80] is a software for the adaptive modeling and simulation of nanosystems. It is a platform for computational nanoscience and offers various features through modules, which are called *Elements*. Elements include Apps, Editors, Visual Models, File Loaders and can be extended through the supported open architecture. This creates a rather comprehensive software toolkit for various domains.

Adenita is a specialized toolkit for the visualization and modeling of DNA nanostructures.

It provides modelers with a modelitics concept, as described in Sect. 4 that includes: multiscale and multilayout inspection functionalities; from-scratch design capabilities; and adaptive modeling of large DNA nanostructures with realistic spatial arrangements. It provides some tools for a multicomponent design of nanorobots. Adenita is still in an alpha-phase, but a test version will soon be available from the project page [81].

CellVIEW [60] is a tool for illustrative multiscale rendering of large biomolecular datasets. It is integrated into the Unity 3D game engine. CellVIEW renders molecular scenes that comprise up to 15 billion atoms. Interactive frame rates are achieved through the incorporation of advanced GPU algorithms from computer graphics and visualization, like occlusion culling, instancing, and dynamic tessellation. CellVIEW implements innovative level-of-details that enable a seamless visual transition between different levels of abstraction. CellVIEW showcases mesoscale models, which are generated by cellPACK [52]. This is a publicly available software tool to assemble large scale models from molecular building-blocks. It integrates data from multiple sources into a 'recipe',i. e., a declarative specification of the desired environment. With the recipe and a packing algorithm cellPACK generates comprehensive 3D models of cell-scale structures in molecular detail. CellVIEW has also been implemented as a plug-in in the *Marion* [82] rendering framework.

7. Conclusions

The omnipresence of Big Data has inspired and driven recent developments in data sciences. Some data-processing areas have been dealing with multiscale representations already for several decades. In visualization, this topic certainly requires further research. This is especially the case for molecular visualization as the generated data have become very complex in recent years. In this review, we have discussed three distinct scenarios in molecular visualization, where multiscale solutions are meanwhile imperative. We have presented approaches: to handle very long simulation sequences; visually very complex arrangements of DNA in the context of nanotechnology; and very large spatial data from integrative structural biology innovations.

An upcoming challenge for visualization is to cope with molecular data that are very long, large, and complex simultaneously. The visualization solutions in this survey have addressed these challenges in isolation. Soon this will no longer be sufficient. For example, interactive simulations of an entire bacterial life cycle result in data that are both very long temporally and very large spatially. This raises several research questions that require the attention of the visualization research community. Which data and visual abstraction methods can help us to cope with such data? How many spatio-temporal scales can be displayed interactively? How can complex data be presented to communicate them comprehensively to the target viewer. The many fascinating advances in the various areas of biology are generating equally many challenges for visualization research.

Acknowledgements

References

[1] S. K. Card, J. D. Mackinlay, B. Shneiderman, Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann, 1999.

[2] T. Munzner, Visualization Analysis and Design, A K Peters/CRC Press, New York, 2014. `doi:10.1201/b17511`.

[3] C. Ware, Information Visualization: Perception for Design, 3rd Edition, Morgan Kaufmann Publishers Inc., San Francisco, 2012. `doi:10.1016/C2009-0-62432-6`.

[4] A. C. Telea, Data Visualization: Principles and Practice, A K Peters/CRC Press, New York, 2014. `doi:10.1201/b17217`.

[5] J. S. Richardson, Schematic drawings of protein structures, in: Diffraction Methods for Biological Macromolecules Part B, Vol. 115 of Methods in Enzymology, Academic Press, 1985, pp. 359 – 380. `doi:10.1016/0076-6879(85)15026-3`.

[6] C. L. Bajaj, V. Pascucci, A. Shamir, R. J. Holt, A. N. Netravali, Dynamic maintenance and visualization of molecular surfaces, Discrete Applied Mathematics 127 (1) (2003) 23–51. `doi:10.1016/S0166-218X(02)00283-4`.

[7] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, Sixth Edition, Taylor & Francis Group, 2014.

[8] M. Erwig, K. Smeltzer, X. Wang, What is a visual language?, Journal of Visual Languages & Computing 38 (2017) 9–17. `doi:10.1016/j.jvlc.2016.10.005`.

[9] T. Walsum, F. H. Post, Selective visualization of vector fields, Computer Graphics Forum 13 (3) (1994) 339–347. `doi:10.1111/1467-8659.1330339`.

[10] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, C. Tominski, Characterizing guidance in visual analytics, IEEE Transactions on Visualization and Computer Graphics 23 (1) (2017) 111–120. `doi:10.1109/TVCG.2016.2598468`.

[11] A. Cockburn, A. Karlson, B. B. Bederson, A review of overview+detail, zooming, and focus+context interfaces, ACM Computing Surveys 41 (1) (2009) 2:1–2:31. `doi:10.1145/1456650.1456652`.

[12] H. Hauser, Generalizing focus+context visualization, in: G.-P. Bonneau, T. Ertl, G. M. Nielson (Eds.), Scientific Visualization: The Visual Extraction of Knowledge from Data, Springer, Berlin, Heidelberg, 2005, pp. 305–327. `doi:10.1007/3-540-30790-7_18`.

[13] D. Albers, M. Correll, M. Gleicher, Task-driven evaluation of aggregation in time series visualization, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2014, pp. 551–560. `doi:10.1145/2556288.2557200`.

[14] J. Kreiser, M. Meuschke, G. Mistelbauer, B. Preim, T. Ropinski, A survey of flattening-based medical visualization techniques, Computer Graphics Forum 37 (3) (2018) 597–624. `doi:10.1111/cgf.13445`.

[15] R. A. Becker, W. S. Cleveland, Brushing scatterplots, Technometrics 29 (2) (1987) 127–142. `doi:10.1080/00401706.1987.10488204`.

[16] A. Buja, J. A. McDonald, J. Michalak, W. Stuetzle, Interactive data visualization using focusing and linking, in: Proc. Visualization, IEEE Computer Society, Los Alamitos, 1991, pp. 156–163. `doi:10.1109/VISUAL.1991.175794`.

[17] G. Wills, Linked data views, in: C.-h. Chen, W. Härdle, A. Unwin (Eds.), Handbook of Data Visualization, Springer, Berlin/Heidelberg, 2008, Ch. II.9, pp. 217–241. `doi:10.1007/978-3-540-33037-0_10`.

[18] P. A. Craig, L. V. Michel, R. C. Bateman, A survey of educational uses of molecular visualization freeware, Biochemistry and Molecular Biology Education 41 (3) (2013) 193–205. `doi:10.1002/bmb.20693`.

[19] N. Elmqvist, P. Tsigas, A taxonomy of 3d occlusion management for visualization, IEEE Transactions on Visualization and Computer Graphics 14 (5) (2008) 1095–1109. `doi:10.1109/TVCG.2008.59`.

[20] A. Mahdavi-Amiri, T. Alderson, F. Samavati, A survey of digital earth, Computers & Graphics 53 (2015) 95–117. `doi:10.1016/j.cag.2015.08.005`.

[21] H. Schumann, W. Müller, Visualisierung: Grundlagen und allgemeine Methoden, Springer, Berlin, Heidelberg, 2000. `doi:10.1007/978-3-642-57193-0`.

[22] I. Viola, T. Isenberg, Pondering the concept of abstraction in (illustrative) visualization, IEEE Transactions on Visualization and Computer Graphics 24 (9) (2018) 2573–2588. `doi:10.1109/TVCG.2017.2747545`.

[23] M. Mittring, Finding next gen: Cryengine 2, in: ACM SIGGRAPH 2007 Courses, SIGGRAPH '07, 2007, pp. 97–121. `doi:10.1145/1281500.1281671`.

[24] M. Le Muzic, M. Waldner, J. Parulek, I. Viola, Illustrative timelapse: A technique for illustrative visualization of particle-based simulations, in: Proc. IEEE Pacific Visualization Symposium (PacificVis), IEEE Computer Society, Los Alamitos, 2015, pp. 247–254. `doi:10.1109/PACIFICVIS.2015.7156384`.

[25] G. W. Furnas, B. B. Bederson, Space-scale diagrams: Understanding multiscale interfaces, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1995, pp. 234–241. `doi:10.1145/223904.223934`.

[26] M. Karplus, J. A. McCammon, Molecular dynamics simulations of biomolecules, Nature Structural & Molecular Biology 9 (9) (2002) 646–652. `doi:10.1038/nsb0902-646`.

[27] M. Karplus, Molecular dynamics of biological macromolecules: A brief history and perspective, Biopolymers 68 (3) (2003) 350–358. `doi:10.1002/bip.10266`.

[28] U. Hensen, T. Meyer, J. Haas, R. Rex, G. Vriend, H. Grubmüller, Exploring protein dynamics space: The dynasome as the missing link between protein structure and function, PLoS ONE 7 (5) (2012) e33931:1–16. `doi:10.1371/journal.pone.0033931`.

[29] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, P. Zhang, Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics, Nature 497 (7451) (2013) 643–646. `doi:10.1038/nature12162`.

[30] M. Baram, Y. Atsmon-Raz, B. Ma, R. Nussinov, Y. Miller, Amylin-Aβ oligomers at atomic resolution using molecular dynamics simulations: A link between Type 2 diabetes and Alzheimer's disease, Physical Chemistry Chemical Physics 18 (4) (2016) 2330–2338. `doi:10.1039/c5cp03338a`.

[31] F. Noe, Beating the millisecond barrier in molecular dynamics simulations, Biophysical Journal 108 (2) (2015) 228–229. `doi:10.1016/j.bpj.2014.11.3477`.

[32] M. Krone, B. Kozlíková, N. Lindow, M. Baaden, D. Baum, J. Parulek, H. Hege, I. Viola, Visual analysis of biomolecular cavities: State of the art, Computer Graphics Forum 35 (3) (2016) 527–551. `doi:10.1111/cgf.12928`.

[33] N. Prabhu, K. Sharp, Protein-solvent interactions, Chemical Reviews 106 (5) (2006) 1616–1623. `doi:10.1021/cr040437f`.

[34] S.-Y. Huang, Search strategies and evaluation in protein-protein docking: Principles, advances and challenges, Drug Discovery Today 19 (8) (2014) 1081–1096. `doi:10.1016/j.drudis.2014.02.005`.

[35] B. Tversky, J. B. Morrison, M. Betrancourt, Animation: Can it facilitate?, International Journal of Human-Computer Studies 57 (4) (2002) 247–262. `doi: 10.1006/ijhc.2002.1017.`

[36] B. Kozlíková, M. Krone, M. Falk, N. Lindow, M. Baaden, D. Baum, I. Viola, J. Parulek, H. Hege, Visualization of biomolecular structures: State of the art revisited, Computer Graphics Forum 36 (8) (2017) 178–204. `doi:10.1111/ cgf.13072.`

[37] K. Furmanová, M. Jarešová, J. Byška, A. Jurčík, J. Parulek, H. Hauser, B. Kozlíková, Interactive exploration of ligand transportation through protein tunnels, BMC Bioinformatics 18 (Suppl. 2) (2017) 22:1–16. `doi:10.1186/ s12859-016-1448-0.`

[38] V. Vad, J. Byška, A. Jurčík, I. Viola, E. Gröller, H. Hauser, S. M. Marques, J. Damborský, B. Kozlíková, Watergate: Visual exploration of water trajectories in protein dynamics, in: Proc. Eurographics Workshop on Visual Computing for Biology and Medicine, The Eurographics Association, Goslar, Germany, 2017, pp. 33–42. `doi:10.2312/vcbm.20171235.`

[39] J. Byška, A. Jurčik, M. E. Gröller, I. Viola, B. Kozlíková, MoleCollar and Tunnel Heat Map visualizations for conveying spatio-temporo-chemical properties across and along protein voids, Computer Graphics Forum 34 (3) (2015) 1–10. `doi: 10.1111/cgf.12612.`

[40] J. Byška, M. Le Muzic, M. E. Gröller, I. Viola, B. Kozlíková, AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics, IEEE Transactions on Visualization and Computer Graphics 22 (1) (2016) 747–756. `doi:10.1109/TVCG.2015.2467434.`

[41] M. Krone, G. Reina, C. Schulz, T. Kulschewski, J. Pleiss, T. Ertl, Interactive extraction and tracking of biomolecular surface features, Computer Graphics Forum 32 (3) (2013) 331–340. `doi:10.1111/cgf.12120.`

[42] S. Zhukov, A. Iones, G. Kronin, An ambient light illumination model, in: Rendering Techniques'98, 1998, pp. 45–56. `doi:10.1007/ 978-3-7091-6453-2_5.`

[43] P. W. Rothemund, Folding DNA to create nanoscale shapes and patterns, Nature 440 (7082) (2006) 297–302. `doi:10.1038/nature04586.`

[44] H. Miao, E. De Llano, J. Sorger, Y. Ahmadi, T. Kekic, T. Isenberg, M. E. Gröller, I. Barišić, I. Viola, Multiscale visualization and scale-adaptive modification of DNA nanostructures, IEEE Transactions on Visualization and Computer Graphics 24 (1) (2018) 1014–1024. `doi:10.1109/TVCG.2017.2743981.`

[45] H. Miao, E. De Llano, T. Isenberg, M. E. Gröller, I. Barišić, I. Viola, DimSUM: Dimension and scale unifying map for visual abstraction of DNA origami structures, Computer Graphics Forum 37 (3) (2018) 403–413. `doi:10.1111/cgf. 13429.`

[46] S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, W. M. Shih, Rapid prototyping of 3D DNA-origami shapes with caDNAno, Nucleic Acids Research 37 (15) (2009) 5001–5006. `doi:10.1093/nar/gkp436`.

[47] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, Journal of Molecular Graphics 14 (1) (1996) 33–38. `doi:10.1016/0263-7855(96)00018-5`.

[48] Schrödinger release 2017-1: Maestro, Schrödinger, LLC, New York, NY (2017). URL `https://www.schrodinger.com/`

[49] W. Lueks, I. Viola, M. van der Zwan, H. Bekker, T. Isenberg, Spatially continuous change of abstraction in molecular visualization, in: Abstracts of BioVis, IEEE Computer Society, Los Alamitos, 2011. URL `https://hal.inria.fr/hal-00781520`

[50] M. van der Zwan, W. Lueks, H. Bekker, T. Isenberg, Illustrative molecular visualization with continuous abstraction, Computer Graphics Forum 30 (3) (2011) 683–690. `doi:10.1111/j.1467-8659.2011.01917.x`.

[51] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The protein data bank, Nucleic Acids Research 28 (1) (2000) 235–242. `doi:10.1093/nar/28.1.235`.

[52] G. T. Johnson, L. Autin, M. Al-Alusi, D. S. Goodsell, M. F. Sanner, A. J. Olson, cellPACK: A virtual mesoscope to model and visualize structural systems biology, Nature Methods 12 (1) (2015) 85–91. `doi:10.1038/nmeth.3204`.

[53] P. Shirley, M. Ashikhmin, S. Marschner, Fundamentals of Computer Graphics, 3rd Edition, AK Peters / CRC Press, 2010. URL `https://www.taylorfrancis.com/books/e/9781482229417`

[54] S. Oeltze-Jafra, B. Preim, Survey of labeling techniques in medical visualizations, in: Proc. of VCBM, Eurographics Association, Goslar, germany, 2014, pp. 199–208. `doi:10.2312/vcbm.20141192`.

[55] M. Tarini, P. Cignoni, C. Montani, Ambient occlusion and edge cueing for enhancing real time molecular visualization, IEEE Transactions on Visualization and Computer Graphics 12 (5) (2006) 1237–1244. `doi:10.1109/TVCG.2006.115`.

[56] J. F. Blinn, A generalization of algebraic surface drawing, ACM Transactions on Graphics 1 (3) (1982) 235–256. `doi:10.1145/357306.357310`.

[57] C. Müller, S. Grottel, T. Ertl, Image-space GPU metaballs for time-dependent particle data sets, in: Proceedings of VMV, Aka GmbH, 2007, pp. 31–40.

[58] W. Straßer, Schnelle Kurven- und Flächendarstellung auf grafischen Sichtgeräten, Ph.D. thesis, Germany (1974). URL `https://diglib.eg.org:443/handle/10.2312/2631196`

[59] N. Greene, M. Kass, G. Miller, Hierarchical z-buffer visibility, in: Proc. SIG-GRAPH, ACM, New York, 1993, pp. 231–238. `doi:10.1145/166117.166147`.

[60] M. Le Muzic, L. Autin, J. Parulek, I. Viola, cellVIEW: A tool for illustrative and multi-scale rendering of large biomolecular datasets, in: Proc. Eurographics Workshop on Visual Computing for Biology and Medicine, The Eurographics Association, Goslar, Germany, 2015, pp. 61–70. `doi:10.2312/vcbm.20151209`.

[61] T. Klein, L. Autin, B. Kozlíková, D. S. Goodsell, A. Olson, M. E. Gröller, I. Viola, Instant construction and visualization of crowded biological environments, IEEE Transactions on Visualization and Computer Graphics 24 (1) (2018) 862–872. `doi:10.1109/TVCG.2017.2744258`.

[62] O. D. Lampe, I. Viola, N. Reuter, H. Hauser, Two-level approach to efficient visualization of protein dynamics, IEEE Transactions on Visualization and Computer Graphics 13 (6) (2007) 1616–1623. `doi:10.1109/TVCG.2007.70517`.

[63] M. Le Muzic, J. Parulek, A. Stavrum, I. Viola, Illustrative visualization of molecular reactions using omniscient intelligence and passive agents, Computer Graphics Forum 33 (3) (2014) 141–150. `doi:10.1111/cgf.12370`.

[64] M. Le Muzic, P. Mindek, J. Sorger, L. Autin, D. Goodsell, I. Viola, Visibility equalizer: Cutaway visualization of mesoscopic biological models, Computer Graphics Forum 35 (3) (2016) 161–170. `doi:10.1111/cgf.12892`.

[65] N. Waldin, M. Le Muzic, M. Waldner, M. E. Gröller, D. Goodsell, L. Autin, I. Viola, Chameleon: Dynamic color mapping for multi-scale structural biology models, in: Proc. Eurographics Workshop on Visual Computing for Biology and Medicine, The Eurographics Association, Goslar, Germany, 2016, pp. 11–20. `doi:10.2312/vcbm.20161266`.

[66] K. Been, M. Nöllenburg, S.-H. Poon, A. Wolff, Optimizing active ranges for consistent dynamic map labeling, Computational Geometry 43 (3) (2010) 312–328. `doi:10.1016/j.comgeo.2009.03.006`.

[67] D. Kouřil, L. Čmolík, B. Kozlíková, H. Wu, G. Johnson, D. S. Goodsell, A. Olson, M. E. Gröller, I. Viola, Labels on levels: Labeling of multi-scale multi-instance and crowded 3D biological environments, IEEE Transactions on Visualization and Computer Graphics 25 (1), to appear. `doi:10.1109/TVCG.2018.2864491`.

[68] J. McCrae, I. Mordatch, M. Glueck, A. Khan, Multiscale 3D navigation, in: Proc. Symposium on Interactive 3D Graphics and Games, ACM, New York, 2009, pp. 7–14. `doi:10.1145/1507149.1507151`.

[69] E. Axelsson, J. Costa, C. Silva, C. Emmart, A. Bock, A. Ynnerman, Dynamic scene graph: Enabling scaling, positioning, and navigation in the universe, Computer Graphics Forum 36 (3) (2017) 459–468. `doi:10.1111/cgf.13202`.

[70] J. H. Clark, Hierarchical geometric models for visible surface algorithms, Communications of the ACM 19 (10) (1976) 547–554. `doi:10.1145/360349.360354`.

[71] W.-H. Hsu, K.-L. Ma, C. Correa, A rendering framework for multiscale views of 3D models, ACM Transactions on Graphics 30 (6) (2011) 131:1–131:10. `doi:10.1145/2070781.2024165`.

[72] T. Luft, C. Colditz, O. Deussen, Image enhancement by unsharp masking the depth buffer, ACM Transactions on Graphics 25 (3) (2006) 1206–1213. `doi:10.1145/1141911.1142016`.

[73] S. Grottel, M. Krone, K. Scharnowski, T. Ertl, Object-space ambient occlusion for molecular dynamics, in: Proc. IEEE Pacific Visualization Symposium, IEEE Computer Society, Los Alamitos, 2012, pp. 209–216. `doi:10.1109/PacificVis.2012.6183593`.

[74] J. Parulek, T. Ropinski, I. Viola, Seamless abstraction of molecular surfaces, in: Proc. Spring Conference on Computer Graphics, ACM, New York, 2013, pp. 107–114. `doi:10.1145/2508244.2508258`.

[75] N. Lindow, D. Baum, H. Hege, Interactive rendering of materials and biological structures on atomic and nanoscopic scale, Computer Graphics Forum 31 (3pt4) (2012) 1325–1334. `doi:10.1111/j.1467-8659.2012.03128.x`.

[76] J. Parulek, D. Jönsson, T. Ropinski, S. Bruckner, A. Ynnerman, I. Viola, Continuous levels-of-detail and visual abstraction for seamless molecular visualization, Computer Graphics Forum 33 (6) (2014) 276–287. `doi:10.1111/cgf.12349`.

[77] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, Journal of Computational Chemistry 26 (16) (2005) 1781–1802. `doi:10.1002/jcc.20289`.

[78] S. Grottel, M. Krone, C. Müller, G. Reina, T. Ertl, MegaMol – A prototyping framework for particle-based visualization, IEEE Transactions on Visualization and Computer Graphics 21 (2) (2015) 201–214. `doi:10.1109/TVCG.2014.2350479`.

[79] B. Kozlíková, E. Šebestová, V. Šustr, J. Brezovský, O. Strnad, L. Daniel, D. Bednář, A. Pavelka, M. Manak, M. Bezděka, P. Beneš, M. Kotry, A. Gora, J. Damborský, J. Sochor, CAVER Analyst 1.0: Graphic tool for interactive visualization and analysis of tunnels and channels in protein structures, Bioinformatics 30 (18) (2014) 2684–2685. `doi:10.1093/bioinformatics/btu364`.

[80] NANO-D, Inria, SAMSON – Software for adaptive modeling and simulation of nanosystems, Website: `https://samson-connect.net`, visited May 2018 (2018).

[81] Austrian Institute of Technology, MARA – Molecular analytical robotics assays, Website: `http://maraproject.eu/`, visited in May 2018 (2018).

[82] P. Mindek, D. Kouřil, J. Sorger, D. Toloudis, B. Lyons, G. Johnson, M. E. Gröller, I. Viola, Visualization multi-pipeline for communicating biology, IEEE Transactions on Visualization and Computer Graphics 24 (1) (2018) 883–892. `doi:10.1109/TVCG.2017.2744518`.