DOI: 10.1111/cgf.70155
EUROVIS 2025
M. Angellini, C. Garth, and A. Kerren
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 44 (2025), Number 3
STAR – State of The Art Report

Uncertainty-Aware Visualization of Biomolecular Structures

A. Sterzik[1], C. Gillmann[2], M. Krone[3], and K. Lawonn[1],

[1]Institute of Computer Science, University of Jena, Germany
[2]Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany
[3]Department of Computer Science, Stuttgart University of Applied Sciences, Germany

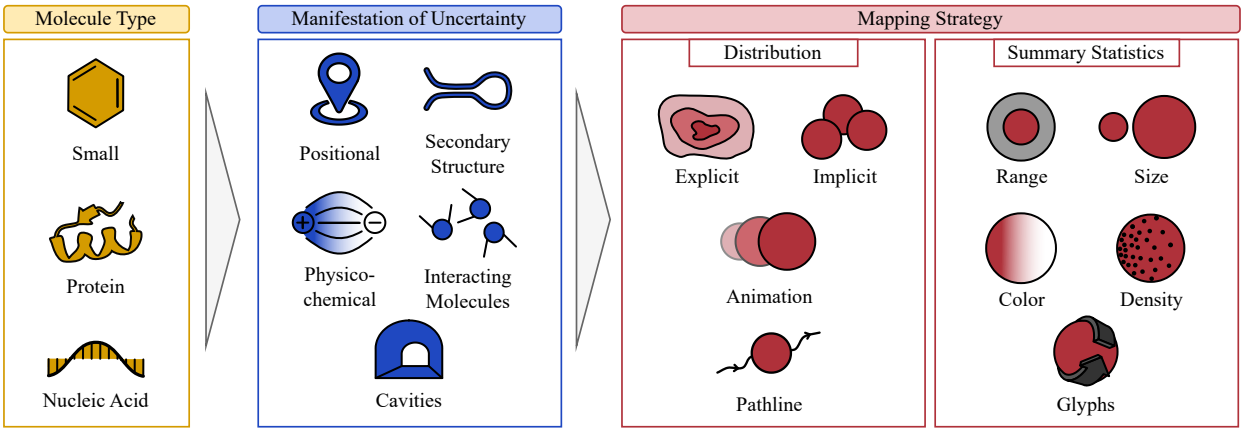


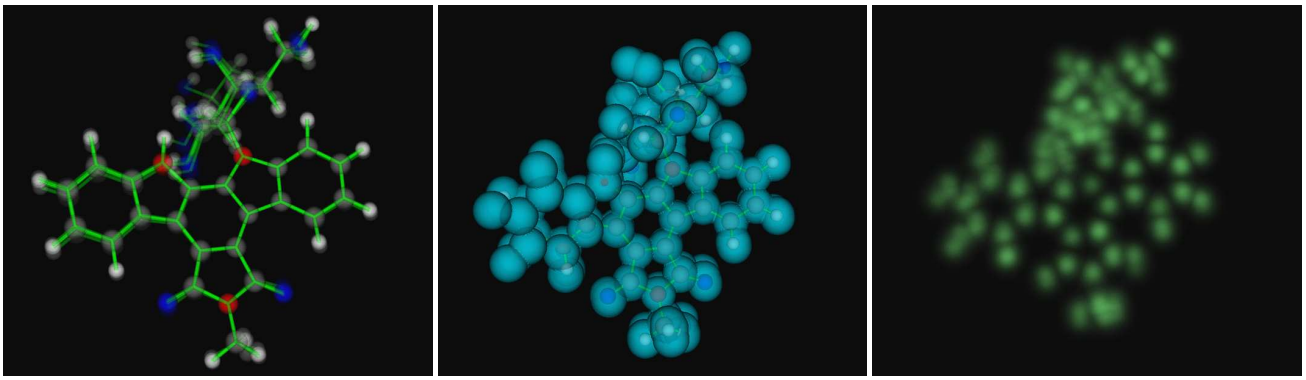
Figure 1: *Schema of the classification dimensions of uncertainty-aware visualization (UAV) of biomolecular structures. We categorize the relevant literature along three major dimensions: primary molecule type considered, manifestation of the uncertainty, and uncertainty mapping strategy. The mapping strategy is further divided into directly displaying the uncertainty distributions and using visual variables to encode summary statistics. The attributes listed for the dimensions are derived from the literature. They are not exhaustive and can be expanded as new visualization methods emerge.*

Molecular structure visualization is fundamental to molecular biology, aiding in understanding complex biological processes. While advancements in molecular visualization have greatly improved the representation of these structures, inherent uncertainties—such as inaccuracies in atomic positions or variability in secondary structure classifications—impact the accuracy of the visualizations. Uncertainty-aware visualization (UAV) emerged as a response to these challenges, integrating uncertainty into visual representations to improve data interpretation and decision-making. Despite extensive work on both molecular and uncertainty visualization (UV), there is a lack of comprehensive surveys addressing the intersection of these two fields. This paper provides a state-of-the-art review of UAV approaches for biomolecular structures. We propose a classification schema that organizes existing methods based on the type of molecule visualized, the manifestation of uncertainty, and the mapping of uncertainty to a visual representation. Using this framework, we identified research gaps and areas for future exploration in uncertainty-aware biomolecular structure visualization.

1. Introduction

The visualization of molecular structures has long been an integral part of structural molecular biology and continues to develop as a critical tool in scientific and industrial research. Since the early days of the field, molecular visualization has enabled researchers to gain an intuitive understanding of complex molecular systems that would otherwise be inaccessible through raw numerical data alone. Structural representations provide insight into the form and function of molecules, which are inherently linked. A well-known example is the lock-and-key metaphor for molecular docking, which, while simplified, illustrates the importance of molecular shape in determining biological interactions. By transforming abstract data into interpretable visual models, visualization allows scientists to identify relationships, mechanisms, and patterns essential for advancing molecular biology, drug design, and bioinformatics.

Levinthal [Lev66] was the first to use known positions of atoms, obtained by early X-Ray diffraction studies [Ken61, Per64], to interactively visualize 3D structural images of small proteins. Over



(a) *Semi-transparent ensemble visualization.* **(b)** *Ranges visualization.* **(c)** *Direct volume rendering of the distribution.*

Figure 2: *Some of the first uncertainty visualizations for biomolecular structures. In a) and b), the uncertainty is mapped to the ball-and-stick representation, whereas c) is a direct volume rendering of the structure [RJ99]. Reproduced from Visualization of Molecules with Positional Uncertainty, P. Rheingans and S. Joshi, Data Visualization '99, Eurographics and Springer, 1999. © Springer, reproduced with permission from SNCSC. Not covered by the article's Creative Commons license.*

the years, molecular visualization has advanced significantly. The field has seen the development of diverse visualization and visual analysis techniques, driven by growing computational capabilities and the increasing complexity of biological data. This progress is reflected in comprehensive surveys such as the one by Kozlíková et al. [KKF*17], which provides a detailed overview of the state-of-the-art in molecular structure visualization.

One of the inherent challenges with visualization is the potential introduction of uncertainty at any stage of the visualization pipeline, from data acquisition to the final visualization and its interpretation by the viewer [BAOL12]. There are many potential sources of uncertainty. For instance, the obtained positions of atoms can be inaccurate due to uncertainty in the data acquisition process. Another important example is the variability in molecular conformations, as represented by ensembles. Ensembles obtained from techniques such as nuclear magnetic resonance (NMR) spectroscopy illustrate a range of possible structures rather than a single, static conformation. Similarly, derived properties like the root mean square fluctuation (RMSF), which depends on atom positions during a dynamics simulation, can also be susceptible to uncertainties. Other forms of uncertainty come from computational methods such as assignments of secondary structure elements: Different classifiers that assign structural elements to parts of the amino acid sequence yield different results. Understanding these uncertainties is essential as they are directly impacting the reliability and interpretation of molecular visualizations. Coming back to the earlier example of molecular docking, uncertainties in atomic positions can severely affect the docking process. In such cases, it is not only the most probable structure that matters but also other plausible conformations, as they may influence the outcome and provide a more comprehensive understanding of molecular interactions.

To address these issues and communicate associated uncertainties, the field of UAV has emerged. It focuses on integrating data and its uncertainty into a unified computation and visualization pipeline, ultimately helping users make more informed decisions. This is particularly important because visualization tends to suggest a level of accuracy rarely present in the data [JS03]. However, incorporating uncertainty information into a visualization presents its own challenges. It tends to increase the visual complexity, potentially hindering the perception of the core data [BHJ*14]. In the context of biomolecular structures, UV has been an active area of research for over 30 years, with some of the earliest examples shown in Figure 2. Despite this, existing surveys focus on either specific aspects of molecular visualization or UV in general, with no survey addressing the unique intersection of these two topics.

Thus, this report provides an overview of existing UAV approaches for biomolecular structures. While this is particularly relevant to the study of biomolecules, it also holds mutual benefits for the broader visualization community. Molecular structure data is versatile, complex, and derived from a variety of sources. The techniques developed for visualizing uncertainty can enhance molecular visualization and provide valuable insights for other fields that handle complex and uncertain data. Additionally, innovations in these fields could inspire new methods and improvements in the visualization of uncertainty for biomolecular structures.

To structure and provide an overview of works on UAV of biomolecular structures, we introduce a classification schema (Figure 1) that categorizes approaches by molecule type, uncertainty manifestation, and uncertainty mapping. This classification allows for a better understanding of the diverse techniques and helps identify trends and gaps in current methodologies. For example, there is a significant lack of UAV techniques for large-scale data, despite the rapid growth of structural datasets. By identifying such gaps, we highlight promising opportunities for advancing the field.

Our survey's main contributions can be summarized as follows:

- A taxonomy for UAV of biomolecular structures.
- A thorough overview of existing research.
- The identification of patterns and gaps in current methods.
- Recommendations for promising areas of future research.

This report is organized as follows: First, we discuss related surveys (Section 2). Next, we introduce the necessary background

on molecular structure visualization (Section 3), and UAV (Section 4). Then, we discuss the scope of this paper in more detail and present our proposed taxonomy for uncertainty-aware visualization approaches for biomolecular structures (Section 5). Section 6, Section 7, and Section 8 then describes the approaches for small molecules, proteins, and nucleic acids, respectively. Afterward, we discuss the existing approaches (Section 9) and future research directions (Section 10) before we conclude this report in Section 11.

2. Related Work

While surveys on UV and biomolecular visualization individually are manifold, a comprehensive survey of the intersection of both areas is still missing in the literature. Surveys on biological or biomolecular visualization frequently acknowledge uncertainty but typically address it in a limited scope, often focusing solely on individual aspects. Kozlíková et al. [KKF*17] provide a broad overview of state-of-the-art techniques in biomolecular structure visualization. Their report includes a section on molecular dynamics visualization, covering methods for illustrating biomolecular flexibility and volumetric representations.

Krone et al. [KKL*16] review techniques for visualizing biomolecular cavities, also discussing several sources of uncertainty inherent to this field. They highlight the challenges of visualizing dynamic cavities, which fluctuate over time as molecular structures are inherently dynamic. Additionally, the authors emphasize the need for visualization tools that can compare and evaluate different cavity detection methods, noting significant discrepancies—up to 200%—in volume measurements across tools. Despite these challenges, only a few tools can currently calculate uncertainties in their measurements. The increasing size and complexity of data drive the need for multiscale visualization approaches. The 2019 survey by Miao et al. [MKK*19] explores these multiscale molecular visualization methods and emphasizes the importance of overview visualizations. For this purpose, highly abstracted and/or aggregated UV techniques are featured in their review. The recent review of biomolecular visualization by Li and Wei [LW24] includes a brief section on positional UV, outlining common encodings and some newer techniques.

Since molecular dynamics (MD) simulation computes the dynamic properties of molecules, this topic is closely related to molecular flexibility and positional uncertainty. Belghit et al. [BSD*24] reviewed the visualization of MD trajectories, while Corey et al. [CBC23] specifically reviewed the visualization of MD simulations for membrane systems.

In 1997, Pang et al. [PWL97] developed a classification scheme for early UV approaches. It incorporates properties of the data and visualization themselves, such as its dimensionality and discrete or continuous nature. Additionally, they categorize approaches according to how they add uncertainty information to the visualizations. They identified the methods: glyphs, adding/modifying geometry, modifying attributes, animation, sonification, and psychovisual. Brodlie et al. [BAOL12] discussed reasons for UV being a difficult problem. One major reason is the complexity of uncertainty itself, with even the terminology of uncertainty often being unclear. For biomolecular UV, for example, uncertainty is often used interchangeably with error, precision, or flexibility. Another difficulty is that uncertainty adds another dimension to a visualization. For one- or two-dimensional data, this makes the visualization more complex but is relatively easily solvable. Most biomolecular structure visualizations are inherently three-dimensional. Thus, adding another dimension—the uncertainty—becomes more challenging. Additionally, they specifically emphasize the distinction of *visualization of uncertainty* and *uncertainty of visualization*: Research usually considers the visualization of uncertainty in a dataset. However, the visualization itself—through filtering, mapping, and rendering—adds another layer of uncertainty. Bonneau et al. [BHJ*14] formally describe uncertainty and discuss UV in several fields like medical visualization or weather and climate.

Ensemble visualization is closely related to the visualization of uncertainty. In fact, a common metaphor for indicating uncertainty in biomolecular structures is ensemble visualization. The ensemble visualization survey by Wang et al. [WHLS19] does not explicitly address biomolecular structures; however, several techniques for visualizing ensembles of surfaces or volumes are either currently applied to biomolecular structures or are adaptable for this purpose.

Padilla et al. [PKH21] recently outlined best practices for UV, discussing cognitive theories that explain how these methods affect viewers' judgments. They primarily examined lower-dimensional (1D, 2D) UV, including error bars, icon arrays, and 2D ensemble plots. In their 2006 survey on UV, Griethe and Schumann [GS06] highlighted the scarcity of usability studies as a key gap in the field. Over a decade later, Hullman et al. [HQC*19] conducted an in-depth survey on evaluating UV, identifying several unresolved issues. Notably, they observed a prevalent research bias toward assessing performance accuracy over decision quality. They proposed guidelines for adopting transparent and valid evaluation methods.

Weiskopf [Wei22] discusses UV and illustrates general concepts with examples of biological data visualization. However, his paper is not a systematic survey and because it focuses on biological data in general, only a few molecular *structure* visualizations are featured. Therefore, a comprehensive overview of uncertainty in biomolecular structure visualization is still missing in the literature.

3. Molecular Structure Visualization

In this section, we provide a brief overview of molecular structure visualization. We begin with a summary of biomolecules (Section 3.1), followed by the description of various methods for acquiring biomolecular structure data (Section 3.2). Finally, we discuss molecular representation models (Section 3.3). We provide only brief overviews of these topics; for more in-depth information on biomolecules and biomolecular data acquisition, please see Nelson and Cox [NC21], Berg et al. [BTS10], and Alberts et al. [AHJ*22]. Further information on representation models and visualization can be found in the report by Kozlíková et al. [KKF*17].

3.1. A Brief Summary of Biomolecules

Biomolecules are essential to the structure, function, and regulation of living organisms. They are the building blocks of life and play critical roles in various biological processes. The main ingredient

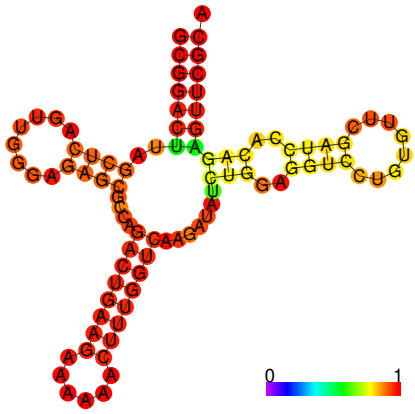
of biomolecules is carbon, a highly versatile element that can form a variety of bonds with itself and other elements such as hydrogen, oxygen, and nitrogen. Biomolecules can be classified into two major groups: small molecules and macromolecules.

Small molecules have a relatively low molecular weight, are highly diverse, and serve numerous functions. For example, they can serve as energy sources, signaling molecules, and building blocks for larger macromolecules. Although there are many types of small molecules, the four main classes are sugars, fatty acids, nucleotides, and amino acids. These molecules form the foundation for larger macromolecules: Larger carbohydrates such as polysaccharides and oligosaccharides are formed from smaller sugar molecules; fatty acids form fats and membrane lipids; nucleotides are the building blocks of nucleic acids; and amino acids combine to form proteins.

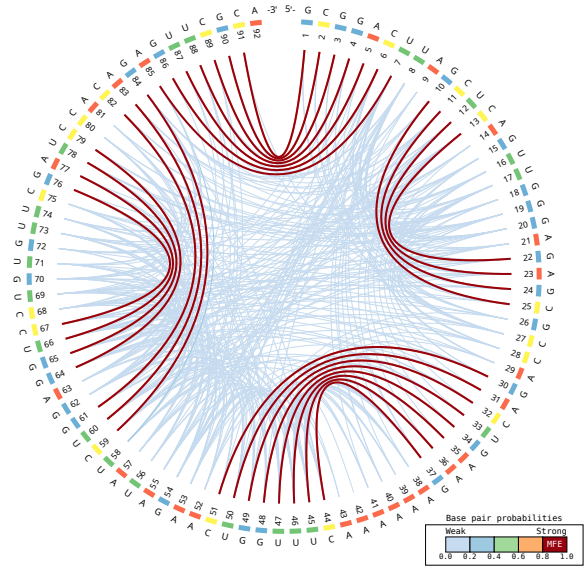
Carbohydrates serve as primary energy sources, particularly in the form of sugars such as glucose. They provide crucial structural support, particularly in plant cell walls, where cellulose plays a vital role. In addition, carbohydrates play a role in cellular signaling, acting as recognition molecules on cell surfaces, and influencing interactions between cells and their environment.

Lipids are water-insoluble, essential for energy storage, and form biological membranes. The membrane lipids have a hydrophilic head and a hydrophobic tail, which allows them to form bilayers. Their hydrophilic heads face outward toward the water, while their hydrophobic tails face inward, away from the water. This bilayer structure forms the foundation of biological membranes, acting as a selective barrier that separates the cell from its surroundings. In addition to their structural role, some lipids, such as hormones or enzyme cofactors, also actively participate in metabolism.

Nucleic Acids such as DNA and RNA, are central to storing and transmitting genetic information. DNA contains the genetic instructions determining an organism's characteristics. RNA, on the other hand, plays a critical role in translating these genetic instructions into functional proteins. In addition to this, certain types of RNA molecules function in various regulatory and catalytic roles, acting as messengers or catalysts in biological processes. The structure of nucleic acids is commonly described at four levels: primary, secondary, tertiary, and quaternary. The primary structure of nucleic acids refers to the linear sequence of nucleotides connected by phosphodiester bonds. Each nucleotide consists of a sugar, a phosphate group, and a nitrogenous base. The secondary structure of nucleic acids refers to the local folding patterns of the polynucleotide chain. In DNA, this is famously seen in the double helix, a spiral structure where two complementary strands of nucleotides form hydrogen bonds between complementary base pairs. In RNA, the secondary structure includes single- and double-stranded regions. In double-stranded regions, complementary base pairs form hydrogen bonds. A characteristic feature of RNA secondary structure is the hairpin loop, where a single-stranded RNA segment folds back on itself to form a double-stranded stem. This stem is held together by complementary base pairs, while the fold creates a single-stranded loop at the other end of the stem. Figure 3a shows examples of hairpin loops. The tertiary structure of nucleic acids is defined by the atomic coordinates of the molecules and refers to the overall



(a) *Secondary Structure visualization with base pairing probabilities encoded in the color map. This ribonucleic acid (RNA) molecule forms three hairpin loops, typical for single-stranded deoxyribonucleic acid (DNA) and RNA. Created with ViennaRNA Web Services [GLB*08].*



(b) *Circular adaptation of rainbow diagrams by Léger et al. [LCT19], created using their webserver.*

Figure 3: *Two uncertainty visualizations for RNA base pairing probabilities are demonstrated using a tRNA-Phe sequence.*

three-dimensional shape that results from further folding and twisting of the secondary structure. The quaternary structure refers to the higher-level organization of nucleic acid molecules, such as the arrangement of multiple RNA subunits into a functional complex.

Proteins carry out a wide range of essential biological functions. Enzymes catalyze chemical reactions such as metabolism and DNA replication. Other proteins provide structural support in cells and tissues. In molecular transport, they ensure the movement of molecules across membranes. In addition, proteins play a key role in immune responses by recognizing and neutralizing pathogens. Their diversity enables them to perform these varied roles. Pro-

teins are composed of long chains of amino acids, and the specific sequence of these amino acids determines how the protein folds and functions. Like nucleic acids, proteins are described by their primary, secondary, tertiary, and quaternary structure. The primary structure refers to the linear sequence of amino acids in the polypeptide chain. The secondary structure involves localized folding of the polypeptide chain into regular patterns, which are stabilized by hydrogen bonds. Alpha helices and beta sheets are the most common patterns. The tertiary structure describes how the entire polypeptide chain folds into a three-dimensional shape. When several polypeptides combine to form a protein complex, their arrangement is described by the quaternary structure. The spatial arrangement of proteins is critical to their function. Even small changes in the structure can lead to dysfunction and disease.

3.2. Data Acquisition

Structural molecular data is most often acquired *in vitro* using one of three techniques: X-ray crystallography, NMR, or cryogenic electron microscopy (cryo-EM).

Most of the structural data available today have been obtained by X-ray crystallography. This technique provides high-resolution detail and is effective for small molecules and macromolecules. In this process, X-ray beams pass through molecular crystals, and the resulting diffraction pattern is used to generate electron density maps. These electron density maps represent a probability distribution of electron positions. To generate atomic coordinate data, the maps are analyzed, and atoms are positioned within them. A key requirement for this method is creating molecular crystals. The crystallization conditions need to be carefully adjusted for each molecule, and large amounts of purified samples are required.

NMR is used to study molecules in solution. A sample is placed in a strong magnetic field, where pulses of electromagnetic radiation alter the spin of the atomic nuclei. The resulting signals, known as chemical shifts, are analyzed to assign probable 3D structures. NMR is primarily used to resolve smaller molecular structures.

Cryo-EM is used to analyze very large structures, such as large protein complexes, which are often difficult to crystallize and too large for NMR. A thin layer of sample is rapidly frozen to preserve the molecular shape. Beams of electrons are directed through the layer, and detectors capture the transmitted electrons to produce 2D images. Sets of similarly oriented molecules are superposed to improve the signal-to-noise ratio, and these views are combined into a 3D electron density map. For relatively stable molecules, cryo-EM resolution can approach the resolution of X-ray crystallography.

These *in vitro* techniques are complemented by *in silico* techniques such as molecular modeling and simulation. While *in vitro* techniques are indispensable for validating findings and providing direct experimental evidence *in silico* techniques fill in the gaps by offering atomic-level, continuous, and environmentally controlled insights that are difficult or impossible to achieve solely in vitro. MD simulations are commonly used to model molecular motions, simulating the movement of atoms over time. Force fields describe the interactions between individual atoms, guiding how they move. The atomic positions are updated at each time step, generating trajectory data that tracks the system's evolution. This enables the refinement of experimentally generated structural data, models molecular stability, and allows for the simulation of interactions between multiple molecules.

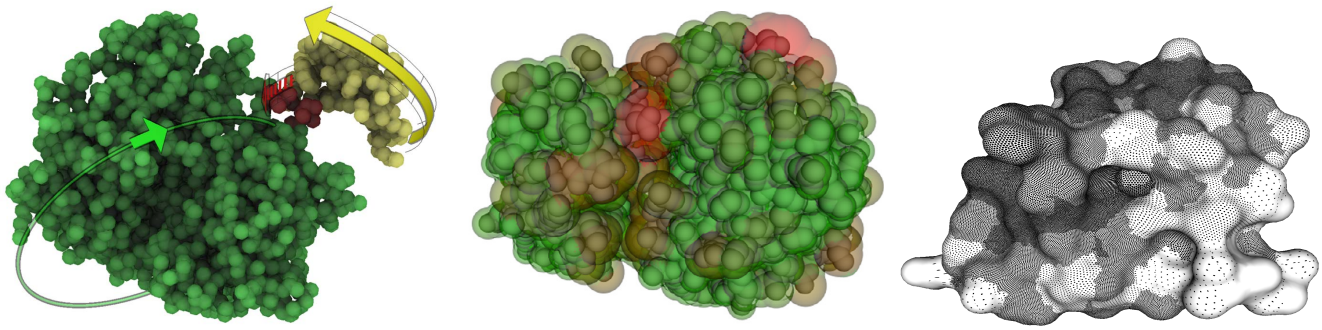
MD simulations can be extended in several ways to enhance their utility in studying biomolecular processes. For example, specific simulated forces, like those used in umbrella sampling or metadynamics, can be applied to study complex processes such as protein folding and unfolding by promoting the exploration of rare conformations and transition states. Additionally, coarse-grained techniques reduce the complexity of large systems by simplifying atomic detail, enabling the simulation of larger biomolecular systems over longer timescales. Another powerful extension is the integration with quantum mechanics (QM), where quantum mechanics/molecular mechanics (QM/MM) approaches allow for the accurate modeling of reactive sites or chemical reactions within a larger molecular environment, providing more precise insights into enzymatic mechanisms or drug interactions.

An alternative approach are Monte Carlo (MC) methods. MC methods are computational techniques that explore molecular configurations through random sampling rather than continuous time evolution. Instead of calculating forces and integrating equations of motion as in MD, MC randomly selects new configurations, which are accepted or rejected based on criteria like the Metropolis algorithm, ensuring that low-energy states are more likely to be sampled. MC is especially useful for studying equilibrium properties, conformational sampling, and thermodynamic properties of biomolecules. These methods can be applied to structures represented at atomic detail or with coarse-grained techniques, depending on the level of detail needed. Additionally, MC is often combined with enhanced sampling techniques, such as umbrella sampling, to overcome energy barriers and explore rare configurations that might otherwise be inaccessible.

Another less common technique is normal mode analysis (NMA). NMA provides a method for examining large, collective movements in biomolecules. By modeling the molecule as a set of point masses connected by harmonic springs, NMA identifies low-frequency movements, which typically correspond to functional changes in protein structures. This approach is computationally efficient, making it a useful tool for studying conformational dynamics on a broad scale, often in combination with more detailed methods like MD and MC simulations.

3.3. Molecular Representation Models

There are several options for creating a molecular representation from the acquired data. In this section, we follow the classification of Kozlíková et al. [KKF*17] which identifies three main categories: Atomistic models, abstract or illustrative models, and level of detail (LOD) models. Atomistic models give a fine-grained view of the model by displaying individual atoms or bonds. They can be further separated into bond-centric and surface models. Bond-centric models are common in chemistry and for smaller molecules. They emphasize the connections between individual atoms. The most basic representation uses lines to depict bonds, while a more sophisticated variant, the licorice or stick model, employs cylinders instead of lines. When spheres are additionally used to represent



(a) *Glyphs for UV added to space-filling representation. The arrow glyphs depict the motion of larger groups of atoms. Error-bars around the glyphs show the error of the fitting model [BPG12]. Reprinted with permission from IEEE Transactions on Visualization and Computer Graphics, © 2012 IEEE. Not covered by the article's Creative Commons license.*

(b) *Semi-transparent hulls represent the positional uncertainty on the van der Waals (vdW) surface. The b-factor is additionally encoded in the hulls color [MRW*21]. Reprinted with permission from Computers & Graphics, © 2021 Elsevier. Not covered by the article's Creative Commons license.*

(c) *The density of the stippling texture on the solvent excluded surface (SES) encodes the positional uncertainty of a protein [SLW*24]. Reprinted with permission from IEEE Transactions on Visualization and Computer Graphics, © 2023 IEEE. Not covered by the article's Creative Commons license.*

Figure 4: *Several metaphors for encoding summary statistics of the uncertainty distribution.*

atoms, the resulting depiction is called the ball-and-stick model. Bond-centric models can be efficiently computed on the GPU using raycasting and imposter-based techniques. Figure 2a and Figure 2b are examples for bond-centric representations.

Surface models represent the spatial arrangement of atoms as continuous surfaces. The simplest molecular representation is the space-filling, or calotte model, where atoms are depicted as spheres with sizes proportional to their atomic radii (see Figure 4a). While by itself, this representation is not a continuous surface, a surface can be generated by using the outer surface of the union of all spheres. This is done, for example, for the van der Waals (vdW) surface [Ric77], where the sphere sizes correspond to the vdW radii of the atoms (see Figure 4b). This surface is the basis for several other representations. The solvent accessible surface (SAS) [LR71] is computed by rolling a probe sphere (typically the size of a water molecule) over the vdW surface. The center of the probe sphere is tracked and marks the SAS, resulting in a smoother surface with fewer discontinuities than the vdW surface. It highlights regions accessible to interacting molecules. However, this approach inflates the molecular surface, as the SAS includes the radius of the probe sphere. The solvent excluded surface (SES) [Ric77] addresses this issue by tracking the point of the probe sphere closest to the atoms (see Figure 4c). This method more faithfully represents the molecular volume. Less common surface representations include the molecular skin surface (MSS) [Ede99], which is continuously differentiable but lacks a biophysical basis and is computationally complex. Another example is the ligand excluded surface (LES) [LBH14], which extends the SES concept by incorporating the geometry of the ligand's vdW surface. Convolution surface models [Bli82] offer an alternative visualization approach by blending atomic potentials using convolution kernels. They are often computationally efficient, with some providing among the fastest surface extraction algorithms available.

Abstract and illustrative models do not depict individual atoms directly. These models are particularly useful for representing large macromolecules. They are often derived from the surface models discussed earlier. Instead of focusing on individual atoms, they treat groups of atoms—such as molecular building blocks like amino acids—as the smallest unit for surface generation. Convolution surfaces can even be directly adapted by adjusting their parameters to produce coarser, more abstract visualizations. Such abstraction methods improve rendering efficiency and reduce visual clutter.

Illustrative models highlight key features in the data that are hard to discern with atomistic displays. They are often used to depict protein secondary structures. For example, the cartoon representation [Ric81] shows alpha helices as ribbons, beta sheets as arrows, and loops as lines. Several variants of the cartoon representation exist, such as using cylinders to represent beta sheets. The structure visualization in Figure 5 depicts a variant of the cartoon representation. Other illustrative methods represent the backbone as a continuous tube. The well-known double helix representation of DNA is another example of an illustrative visualization.

Similar to abstract models, LOD models are often used to represent large macromolecules. Within this category, a distinction can be made between semantic and spatial LOD. Semantic LOD methods involve visualizing molecular systems with varying levels of structural abstraction, allowing different parts of the system to be represented using different models. This approach, for example, enables the creation of focus-and-context visualizations [vLBI11]. Spatial LOD methods, on the other hand, focus on the spatial arrangement of molecules. Distant or occluded parts of the molecule can, for instance, be represented with lower resolution. Parulek et al. [PJR*14], for example, vary the abstractness of the surface based on its distance from the camera.

4. Uncertainty-Aware Visualization

Uncertainty visualization (UV) focuses on explicitly representing errors or uncertainty in the data. Uncertainty-aware visualization (UAV) builds on this by integrating the visualization of uncertainty

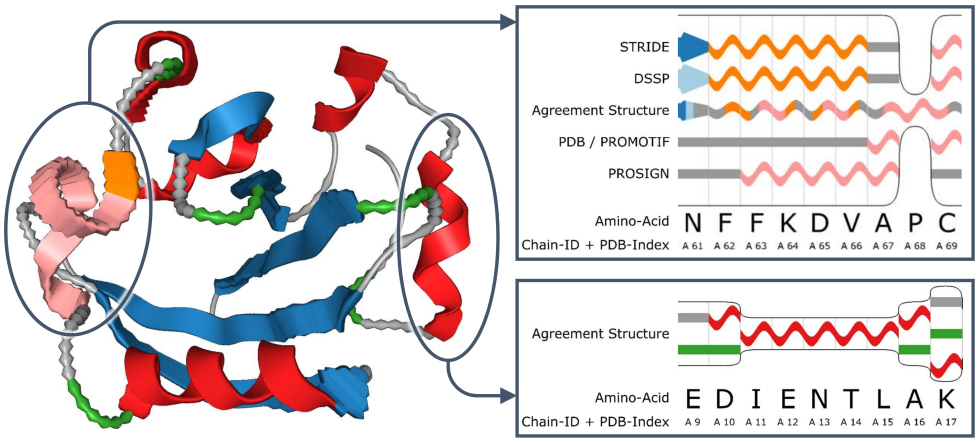


Figure 5: *Uncertainty visualization using geometric distortion of the cartoon representation of a protein to show the uncertainty of the secondary structure. The shape shows the most probable secondary structure over time, whereas the amplitude and frequency of the distortion illustrate the uncertainty. The cutout allows a more detailed investigation of the uncertainty by displaying the results of individual structure assignment methods [SSK*18]. Reprinted with permission from IEEE Proceedings, © 2018 IEEE. Not covered by the article's Creative Commons license.*

directly with the data itself, ensuring both are effectively represented [GSWS21]. Here, we focus on the latter approach, where the data and its uncertainty are visualized together. However, such visualizations are often still referred to as UV. In this section, we outline potential sources of uncertainty (Section 4.1), discuss how these map to biomolecular data (Section 4.2), introduce common mapping strategies for visualizing uncertainties (Section 4.3), and discuss specifics of uncertainty visualization for biomolecular data and commonalities with other domains (Section 4.4).

4.1. Sources of Uncertainty

In the context of visualization, sources of uncertainty are typically discussed in relation to the individual stages of the visualization pipeline [PWL97, BAOL12, BHJ*14]. For the data acquisition stage, a primary source of uncertainty is statistical variation. Experimental measurements often exhibit fluctuations; repeated measurements of the same quantity under similar conditions yield slightly different results [GMR*23]. There are also sampling uncertainties: in statistical surveys, even when samples are representative, different samples may produce slightly different results due to natural variability in the population [KF10]. Beyond statistical variation, uncertainty also stems from systematic biases, e.g., non-representative sampling can result in biased or incomplete datasets when improper sampling techniques are employed [GMR*23]. Another source of uncertainty is data incompleteness [GMR*23] (i.e., missing or unavailable data points). Moreover, finite instrument resolution introduces uncertainty due to the inherent limitations of the sensors capturing physical phenomena [GMR*23]. Each sensor has a specific resolution and range, which restricts the precision of the measurements. In data obtained from simulations or modeling, uncertainty arises due to limited computational power and knowledge. Model incompleteness is a common issue, as no model can fully represent real-world phenomena, and simplifications or assumptions are often introduced during model construction [WSC21]. Parameter uncertainty also affects computational models, as identifying optimal parameters is challenging and often requires experimental determination [GMR*23].

These uncertainties introduced during data acquisition are usually the primary uncertainties visualized. However, additional uncertainties can be introduced during the processing, modeling, and rendering stages. We only discuss them briefly here and refer to Brodlie et al. [BAOL12] for a more detailed discussion. During data processing, for example, the data is often smoothed or interpolated. By interpolating data points (such as drawing curves through discrete data points), the exact values at any given point can only be estimated. In the modeling stage, the chosen representation of the data is subject to uncertainty. Uncertainties might arise due to approximations or simplifications, such as representing continuous phenomena with discrete models (e.g., using a mesh to represent a smooth surface). During the rendering stage, further uncertainties are introduced. Rasterization, for example, is inherently discrete and has limited resolution; consequently, small details may be lost. Alternatively, different ray tracing techniques may yield different results depending on the algorithm or settings, adding more uncertainty to the final visualization. Finally, the viewers' perception introduces another source of uncertainty [GMR*23]. Factors such as prior experiences, vision deficiencies, and visualization literacy can influence how individuals interpret visualizations. Visual limitations, cognitive biases, and contextual factors can all contribute to varying interpretations, making it challenging to ensure accurate communication of uncertainty.

4.2. Uncertainty in Biomolecules

In biomolecular structure data, as in many areas of UV, the primary focus is on uncertainties arising during the data acquisition

stage. This does not imply that uncertainties are absent or irrelevant in other stages of the visualization pipeline. However, to the best of our knowledge, uncertainty quantification for these later stages has not yet been explored. The uncertainties introduced during later stages are likely similar to those encountered in general UV.

Within the data acquisition stage, one of the most critical uncertainties is positional uncertainty. Atomic positions or molecular shapes play a key role in many analyses, such as docking simulations or structural alignment. Even when the quantity of interest (QOI) is not explicitly positional, uncertainties in atomic positions still propagate to and influence the QOI [RCBB19]. The exact sources of positional uncertainty vary with the data acquisition method. Below, we outline several important sources of positional uncertainty for biomolecular structures.

For in vitro data, uncertainties such as statistical fluctuations, environmental variability, and limitations in instrumental resolution and range affect all structure acquisition techniques. Additionally, each acquisition method introduces specific sources of uncertainty: In X-ray crystallography, molecular crystals must be produced for the imaging process; thus, the structure's conformation in a more natural state—such as in solution—remains uncertain. Modeling uncertainties arise during the generation of electron density maps from diffraction patterns. Afterward, the structural model is fitted into the electron density map. This fitting process typically selects structures based on the best density fit and additional constraints, such as biophysical interactions, amino acid sequence, and secondary or tertiary structure [RCBB19]. To assess the quality of the structural fit, several metrics are available. A common metric is the b-factor, which measures the displacement or vibration of individual atoms around their average position [MDC24]. It is derived from the attenuation of X-ray scattering. In NMR spectroscopy, the amino acid sequence must be known to derive structures from the NMR spectrum. Uncertainties in these input data propagate throughout the process. Furthermore, several structures often satisfy the constraints equally well, resulting in an ensemble of structures that implicitly encodes the distribution of possible conformations. For cryo-EM, structural modeling involves a long pipeline. Thousands of molecular images must be acquired, filtered according to their orientation, and superimposed to improve the signal-to-noise ratio. These processed images are then combined into a 3D density map. At each stage of this pipeline, uncertainties propagate from the initial data, and additional uncertainties are introduced. Finally, as with X-ray crystallography, the structures must be fitted into the resulting density map. These various sources of uncertainty are often readily available. Structural databases such as the RCSB Protein Data Bank (`rcsb.org`) (PDB) [BWF*00] provide b-factors, alternative conformations, and electron density maps, which serve as direct indicators of measurement uncertainty.

For in silico data acquisition, experimentally resolved structures are usually used as input data, meaning uncertainties from structure acquisition propagate through the pipeline. Additional uncertainties arise during the simulation process. For instance, MD simulations are highly sensitive to initial conditions. They exhibit chaotic behavior, producing apparently random deviations that make single observations largely irreproducible. Generating ensembles instead of relying on single simulations can improve reliability and allow for quantifiable uncertainty estimates [WSC21]. Parameter choices are crucial not only in simulations but also in other computational tasks. For instance, in multiple sequence alignment, predictions can vary widely depending on parameters like gap costs [Ham14].

A common measure for analyzing the distance between conformations in an ensemble is the root mean square deviation (RMSD) [LGV*20]. It calculates the differences between atomic positions in individual structures and a reference structure (often the average structure of the ensemble). If the RMSD is calculated over a time series centered around a relatively stable average structure, it is referred to as RMSF. Simulation data are also affected by systematic errors, such as imperfections in design, execution, or analysis. A specific example of imperfect design in MD is the use of certain protein force fields that tend to favor specific secondary structure types. Recognizing these biases has led to efforts to develop improved force fields aimed at reducing or eliminating these systematic errors [WSC21]. Other errors include simplified model formulations, algorithmic approximations such as a discrete representation of continuous models, numerical integrators, and accumulation of rounding errors [RCBB19, WSC21].

As discussed earlier, generating biomolecular data for structure visualization is a lengthy process, inherently making it susceptible to the accumulation of uncertainty at each step of the progress. Uncertainties introduced in one step propagate throughout the progress, with additional uncertainties potentially being introduced at every subsequent step. This issue is even more severe if additional QOI are calculated based on the structure, which extends the pipeline and amplifies the uncertainty [RCBB19]. A prime example of this is physicochemical uncertainties. Physicochemical uncertainties are uncertainties in the physical and chemical properties of molecules, such as charge distributions, hydrophobicity, and vdW forces. These properties are crucial for understanding how molecules interact with each other and their environment. Rasheed et al. [RCBB19] found that positional uncertainties—more specifically b-factors— seem to propagate to only low uncertainties for simple quantities (e.g., exposed surface area) but high uncertainties for complex quantities such as total energy.

Another challenge in further processing is estimation in high-dimensional discrete spaces, such as secondary structure assignments or sequence alignments. In these problems the number of solutions is vast—it scales exponentially with the length of the sequence—resulting in a very small probability for any single solution [Ham14]. Possible ways to make the computations more reliable are ensemble methods or marginal probabilities. Marginal probabilities are unconditional probabilities that are obtained by integrating the joint probabilities over all possible values of the other variables in the system. As a result, these marginal probabilities are much larger than the probability of any individual solution [Ham14]. Base pairing probabilities are an example of marginal probabilities and serve as a common uncertainty measure for secondary structures in RNA. It refers to the likelihood that two nucleotides will form a pair. It sums the probabilities of all conformations in which the two bases form a pair. A similar concept in sequence alignments is called aligned pairing probabilities.

While positional uncertainty is usually readily accessible for visualization, other manifestations of uncertainty often require indi-

rect inference, such as consensus models or confidence scores. Addressing these challenges remains an ongoing area of research in uncertainty quantification and visualization.

4.3. Mapping Strategies

Visualizing uncertainty in data is essential for accurately representing the inherent variability and limitations of the information. Different strategies can be used to map this uncertainty to the visualization, each with its strengths depending on the type of data and the intended audience and tasks. In this chapter, we discuss possible mapping strategies for uncertainty in biomolecular data found in the literature. As mentioned above, a major challenge in biomolecular UV is that the uncertainty usually needs to be visualized together with other biochemical attributes to allow for a comprehensive analysis. We follow the classification by Weiskopf et al. [Wei22] and primarily distinguish between directly displaying the uncertainty distribution and visualizing summary statistics.

4.3.1. Visualizing distributions

Direct visualizations of uncertainty distributions can give a comprehensive overview of the whole uncertainty space. These visualizations are particularly useful in scenarios where understanding the full range of possibilities is required. While they offer detailed insights into the nature of uncertainty, they also increase the complexity of the visualizations. Explicit distribution visualizations represent the distribution directly, with common examples including density plots, dot plots, histograms, and violin plots. These methods are particularly suited for one-dimensional datasets, as the uncertainty information is often mapped to an additional axis. For three-dimensional datasets, such as molecular structures, explicit visualizations typically involve volume rendering. Since an additional axis is infeasible for these datasets, the distribution is rendered by modifying visual properties, such as adjusting opacity based on the distribution's density.

Another approach is to visualize distributions implicitly by sampling and displaying individual data points, often through superposition or juxtaposition. Examples include scatterplots, spaghetti plots, and small multiples. For biomolecular structures, this is often done via ensemble visualizations that overlay multiple conformations using abstract forms like cartoon representations.

Closely related to implicit mapping is animation. Animation for UV leverages temporal changes to convey uncertainty dynamics. For instance, data points can blink or oscillate slightly within a constrained region, representing uncertainty of position or value. Similarly, surfaces might expand or contract over time. Hypothetical Outcome Plots [HRA15], which animate samples drawn from the uncertainty distribution, are a relatively recent example of UV. Animation relies on familiar visual dynamics without introducing new encodings. However, it requires time as an additional dimension, which can make it harder to gain an overview of the data and may limit the accurate representation of complex distributions. Pathlines can be employed to create a static representation of an animation. Although pathlines are not traditionally considered a mapping for uncertainty, they highlight dynamic behavior. By overlaying multiple alternative pathlines for the same feature, they are an alternative form of ensemble visualization. An example of this technique can be seen in visualizations of forecasting ensembles for tropical cyclone paths [LPCH19].

4.3.2. Visualizing summary statistics

An alternative to directly displaying uncertainty distributions is the visualization of summary statistics. These visualizations are more concise and focus on specific properties of the uncertainty distribution. While this approach sacrifices some details about the uncertainty, it is often easier to integrate into a visualization and useful if only certain aspects of the uncertainty are relevant to the task. A widespread method for representing uncertainty in summary statistics is through ranges. In one-dimensional data, this is typically done using error bars or confidence intervals. In line charts, confidence bands—regions around a central line—are often used to depict uncertainty intervals. For three-dimensional data, uncertainty ranges are frequently represented as uncertainty hulls.

Alternatively, visual variables such as size, color, and density can be modified to represent uncertainty. For example, in scatter plots, data points can be made larger when uncertainty is high and smaller when uncertainty is low. In biomolecular structure visualization, color often represents uncertainty measures, such as the b-factor, on molecular surface representations. Density variations can also illustrate uncertainty, for instance, by adjusting the density of primitives in texturing techniques like stippling or hatching based on uncertainty levels. In line plots, the frequency of dashes in a dashed line can be increased or decreased to reflect uncertainty. MacEachren et al. [MRO*12] discuss and evaluate several visual variables and their utility for representing uncertainty. Another approach involves using or modifying glyphs to indicate uncertainty. For instance, glyphs have been employed to visualize vector fields, with their attributes altered to reflect uncertainty [WPL96] or to represent uncertain tensors by adding transparent hulls to the glyphs [GRT19].

Finally, it is possible to use separate visualizations to represent uncertainty alongside the data visualization. This approach is common in UV systems. However, this report focuses on UAV and thus methods that integrate uncertainty information directly into the visual representation of the data.

4.4. Biomolecular specifics and commonalities

The visualization of uncertainty in biomolecular structural data is challenging due to its inherently three-dimensional nature and diverse representation possibilities. Molecular structures can be depicted using a variety of representations, including cartoon and surface renderings. Additionally, the data are investigated at different levels of granularity, ranging from quantum-physical to mesoscale models. Moreover, ensembles, which capture a range of conformational states, are crucial for biomolecular structure visualization. However, there is also a need to present scalar uncertainty values. Furthermore, visualizations frequently integrate uncertainty data with other physicochemical properties. While this combination of challenges is specific to biomolecular structures, there also exist commonalities with other domains. Here, we highlight two selected domains to showcase commonalities with biomolecular UV.

An example of a domain where uncertainty visualization is crucial is earth sciences. In climate science, UV is widely used to represent uncertainty and variability in climate model projections. Ensemble data is highly relevant to both earth sciences and biomolecular data. Techniques such as ensemble data visualization can reveal the spread of variable values in a sequence of model runs, resulting in an improved understanding of model uncertainties [PWB*09]. Furthermore, uncertainty visualization can influence decision-making in weather forecasting. Studies suggest that some visualization techniques enhance the interpretation and utilization of forecast uncertainty [NJT08]. In ocean simulations, Raith et al. [RSG21] identified and measured various uncertainties and presented a visualization approach that integrates these uncertainties for depicting ocean eddies. This approach enhances ocean dynamic perception by facilitating a general depiction of uncertainty in detecting eddies, as demonstrated in a case study of the Red Sea. In contrast to biomolecular data, the uncertainty visualizations for earth sciences often display the uncertainty in 2D visualizations such as maps.

In medical visualization, however, 3D representations are common: In multimodal medical data visualization, several techniques are employed to encode different types of information onto surfaces [LSBP18]. The organic surface shapes here have similar geometric properties to some molecular representations such as the SES. Furthermore, medical visualization fuses heterogeneous data to preserve user trust. In addition to color-based encoding, multimodal data visualization relies on illustrative techniques [LBSP14, LP16, LVPI18]. The challenge of displaying a diverse set of data is similar to biomolecular structure visualization, where multiple physicochemical features must be represented simultaneously. A specific example of illustrative visualization for multimodal medical visualization is PelVis [SLK*17], where illustrative visualization is applied for oncologic pelvic surgery to minimize the risk of autonomic nerve damage and subsequent postoperative complications. The technique incorporates patient-specific MRI data and an anatomical atlas. The uncertainty was coded and evaluated on the surface using various techniques such as isolines or grid-based encodings. Subsequently, similar techniques were applied to biomolecular surface data [SLK*22, SMCL24]. Similarly, frequently used techniques for biomolecular structures are sometimes also applied to medical data. Uncertainty hulls—one of the earliest forms of uncertainty visualization for biomolecules [RJ99]—have also been applied to medical data [GWHA18]. Meanwhile, some techniques lack direct counterparts in biomolecular visualization, such as probabilistic animation methods developed for illustrating uncertainty in medical volume rendering [LLPY07]. These methods use probabilistic transfer function models to highlight uncertain regions, aiding in tasks like tissue classification and diagnosis. Given their application in medical visualization, it could be valuable to investigate their potential for biomolecular visualization.

5. Taxonomy

Next, we specify our scope and selection process (Section 5.1) and detail our classification schema (Section 5.2).

5.1. Scope

We focus on publications that introduce or adapt UAV in the field of biomolecular structure visualization. The online libraries that we searched were the *ACM Digital Library*, *IEEEXplore*, *Computers & Graphics (ScienceDirect)* and the *Eurographics Digital Library*, as well as *PubMed*. Searches in the individual libraries were conducted using combinations of the search terms:

- Biomolecule | Molecule | Protein
- Uncertainty | Ensemble | Flexibility | Probability | Density
- Visualization

The first part of the search was intended to find works that consider molecular data. In the second part, we searched for approaches that deal with uncertainty. As uncertainty is not a unique term and comes with many synonyms, we also search for the related terms ensemble, flexibility, probability, and density. The purpose of the last part was to find approaches that include a visual component. To be considered in our report, a publication needs to deal with UAV of spatial or sequential molecular data. With this restriction, we exclude more general UV approaches such as graph visualization or heatmaps, where biomolecular data is only one possible application area. The publication also needs at least one UAV image.

Even though our search terms include related terms like flexibility, publications need to visualize uncertainty (see Section 4). For example, concepts like allosteric response—the change of a protein shape or activity as a result of the binding of a molecule at a site other than the active site—are also often referred to as protein flexibility, but the visualization of allosteric response is not a visualization of uncertainty and thus not included in our survey. In addition to our primary search methods, we employed a recursive citation analysis strategy. More specifically, we identified relevant papers referenced within the sources we had already reviewed. By tracing citations backward, we aimed to uncover additional relevant but previously unconsidered publications. In total, we identified 71 publications that met our selection criteria, including 10 follow-up works, resulting in 61 distinct projects.

5.2. Classification schema

Our classification schema consists of three main dimensions: molecule type, manifestation of uncertainty, and mapping strategy, see Figure 1. For the molecule type dimension, we categorize molecules into small molecules, proteins, and nucleic acids. In Section 3.1, we addressed carbohydrates and lipids as additional, distinct molecule types. While our schema could be expanded to include these categories, to the best of our knowledge, no research on UAV specific to these molecule types exists. We discuss reasons for this gap in Section 9. In visualizations where multiple molecules interact, we classified the visualization based on the type of the primary molecule—the molecule that serves as the focus of the visualization.

There are many potential manifestations of uncertainty in biomolecular data. For this report, we classify the manifestations of uncertainty into five major categories: positional uncertainty, uncertainty in secondary structure, uncertainty in

Table 1: *Overview of relevant literature grouped according to molecule type. The following properties are encoded in the table:*
Molecule Type (Mol.): *Small Molecule*, *protein*, *nucleic acid*.
Manifestation of Uncertainty (Manif.): *Positional*, *secondary structure*, *physicochemical*, *interacting molecules*, *cavities*.
Mapping Strategy (Mapping): *Explicit*, *implicit*, *animation*, *pathline*, *range*, *size*, *color*, *density*, *glyphs*.
Task: *Analyze*, *compare*, *integrate*, *overview*.
Software (Soft.): *Available* ✓, *not available* ✗, *not available anymore*.
Evaluation (Eval.): *None, case study (Case), expert interview (EI), Prelim. survey (Prelim.), Percep. study (Percep.), user study (User).*
∗ *If authors published follow-up work to their papers, we treated them as one method and only cited the first publication in the table.*

Reference	Mol.	Manif.	Mapping	Task	Soft.	Eval.
[RJ99]					✗	None
[BH02]					✗	Case
[SBH02]					✓	Case
[KCL*13]					✗	Case
[WH17]					✓	Case
[SUS*21]					✓	EI, Case
[GMO89]					✗	Case
[FTPG90]					✗	Case
[ZKB95]					✗	Case
[KBW96]					✓	None
[KS97]						None
[TLR*01]					✗	Case
[LV02]					✗	None
[LKCW04]						Case
[HCZP06]					✗	Case
[BGB*08]					✓	EI
[DMR08]*						Prelim.
[DLB08]					✗	Case
[KBE09]					✓	None
[MWT10]					✓	None
[AMA*12]					✗	Case
[BPG12]					✗	EI
[LBBH12]					✗	Case
[HKOW14]						Case
[BJG*15]					✓	EI, Case
[OSK*15]*					✓	Percep.
[SS15]					✗	Case
[BLMG*16]					✓	Case
[MS16]					✗	Case
[RCBB16b]*					✗	Case
[KFS*17]					✓	EI, Case
[KJB*17]					✗	Case
[VBJ*17]					✗	EI, Case
[AKCL18]*					✗	EI
[BGB*18]*					✗	Case
[SSK*18]					✓	EI, Case
[VHG*18]					✗	EI
[JFB*19]					✗	EI, Case
[BRHB20]					✗	Case
[DCS20]					✗	Case
[FJK*20]					✗	EI, Case
[KRH20]					✗	EI, User
[MRW*21]					✗	Case
[SFS*21]					✓	EI, Case
[ST21]					✓	Case
[SLK*22]*					✗	Percep.
[BHP*23]					✗	Case
[CP23]					✓	Case
[PF23]					✓	Case
[FTB*24]					✗	Case
[SMCL24]					✗	Percep.
[SBB*24]					✓	EI, Case
[GLB*08]					✓	None
[SHAM09]					✓	None
[AJ13]						Case
[EKK*14]					✓	Case
[LCT19]					✓	Case
[XZF20]					✓	Case
[IW24]					✓	Case
[BJ96]					✓	Case
[SFL*21]*					✓	Case

physicochemical properties, uncertainties concerning interacting molecules, and uncertainties concerning biomolecular cavities. We addressed the sources of positional uncertainty in Section 4. Many sources of biomolecular structure data, such as structural files from the PDB [BWF*00], include positional uncertainty data such as the b-factor or a set of possible molecular conformations. As a result, positional uncertainty is generally easy to access, and many uncertainty visualizations focus on this manifestation. As discussed in Section 4, positional uncertainties can propagate through subsequent modeling and computations, often compounding with other types of uncertainty to form more complex manifestations. Although some manifestations—such as uncertainties in secondary structure—ultimately derive from positional uncertainties, it is important to distinguish between them. Simply presenting raw positional uncertainty data does not sufficiently address many of the specific questions related to secondary structure. Given the need for distinction, we have identified our categories to capture different aspects of uncertainty in biomolecular structures. We selected key categories with properties central to UV research that are particularly relevant to scientists studying biomolecular structures. The chosen categories are not exhaustive but offer a practical framework for understanding and analyzing uncertainty. Additional categories could be introduced if the need arises in future research.

Not all types of uncertainty apply universally across biomolecule types. Secondary structure uncertainties are relevant to proteins and nucleic acids, as these larger biomolecules form repetitive struc-

tural elements or motifs. Small molecules are too small and compact to exhibit secondary structures. Cavities are predominantly relevant to proteins, where they are critical for functions such as ligand binding. While nucleic acids can form cavities in certain contexts, they are less common and less studied than in proteins. Small molecules are generally too small to form proper cavities, though they may exhibit indentations.

The third dimension of our classification schema, mapping strategy, categorizes approaches primarily into two groups: visualizing uncertainty distributions and visualizing summary statistics. For a more detailed breakdown, we subdivide the visualization of distributions into four categories: explicit and implicit displays, animation, and pathline visualization. Summary statistics can be further divided into displays of ranges, glyphs, and the visual attributes size, color, and density. These finer-grained categories are not exhaustive and may evolve as research in biomolecular structure visualization progresses. We especially anticipate additional visual attributes as the field advances.

In addition to these main classification dimensions, we documented common tasks supported by the visualizations, the availability of an implementation of the methods, and the evaluation strategy of the proposed methods. We defined four abstract task types to generalize the often highly specific purposes of these visualizations and help readers identify suitable methods. The analyze task encompasses cases where details of the uncertainty data need to be clearly visible. This may involve identifying regions where uncertainty falls within a specific value range or comparing uncertainty values at different positions within a molecule. The compare task focuses on comparing different molecules, explicitly excluding ensemble visualizations of multiple molecular conformations. The integrate task applies when uncertainty needs to be considered alongside other information. For example, this could mean simultaneously examining electrostatic potential and positional uncertainty, requiring both properties to be displayed together to facilitate an assessment of their combined effects. Finally, the overview task enables a broad understanding of uncertainty at a glance, ensuring that the molecule and its associated uncertainty are visible in a static form without user interaction.

To assist readers in choosing an appropriate UAV, we established a clear requirement: a visualization method qualifies for a specific task attribute only if the molecule's UAV directly supports that task. Auxiliary visualizations—such as heat plots within larger frameworks—are insufficient. Additionally, a method must either present an example visualization demonstrating the task or explain how it supports it.

In the following chapters, we review publications that discuss molecular UV. Table 1 lists all the considered methods and classifies them according to our schema.

6. Small Molecules

While the methods discussed in this section are illustrated with small molecules, they are not inherently restricted to them. Certain techniques, particularly those that rely on ball-and-stick representations, may not be optimal for larger molecules, though theoretically, they could still be used. On the other hand, methods employing volume visualization should be more readily applicable to various types of biomolecules.

Explicit In 1999, Rheingans and Joshi were the first to write a publication explicitly about molecular UV [RJ99]. They proposed three options for visualizing molecules with positional uncertainty (Figure 2). Two methods are based on a likelihood volume computed from the ensemble. The likelihood volume stores the probability of an atom being at each voxel. They gave two options for visualizing this volume. The first volume visualization extracts transparent isosurfaces from the volume and renders them on top of the opaque ball-and-stick visualization (Figure 2b). The other volume visualization method renders the volume directly, resulting in a cloud-like appearance (Figure 2c). The third method superposes each conformer in the ensemble using its ball-and-stick representation. Each conformation can either be rendered opaquely, or semi-transparently (Figure 2a) so that more certain regions appear opaque while uncertain regions appear more transparent.

Schmidt-Ehrenberg et al. [SBH02] used direct volume rendering (DVR) or opaque isosurfaces to display likelihood volumes. They focused on metastable conformations, which are states in which the molecule stays for a long time. Therefore, they visualized several metastable conformations of a molecule at once, while also displaying the likelihood volume for each of the metastable conformations separately.

Implicit Best and Hege [BH02] let users cluster groups of atoms into ellipsoids in their application to reduce clutter and thus ease the detection of conformational changes. Additionally, they blended several time frames of a molecular simulation to facilitate detecting large changes. Similar to the implicit mapping proposed by Rheingans and Joshi [RJ99], Sabando et al. [SUS*21] visualized chemical compound similarity by layering semi-transparent visualizations of individual conformations. They added an option to invert the opacity. As a result, more uncertain parts of the molecule can be highlighted by appearing more opaque.

Ranges Knoll et al. [KCL*13] defined the interval between the chemical bond radii and the vdW radii as an uncertainty interval over a charge density distribution. They rendered the molecules in black using the ball-and-stick representation. Then, they overlaid a visualization of an electron density volume. They used DVR with a custom transfer function. In most of their examples, they used two peaks for displaying the isolevel of chemical bonds in blue and the isolevel corresponding to the vdW radii in red. In their paper, they focused on molecular interfaces rather than biomolecules. Due to using the ball-and-stick representation, their visualization is suited to small molecules rather than larger biomolecules.

Size Wagner and Himmel [WH17] used color and size to encode the RMSD between two molecular structures in a modified ball-and-stick representation. The sphere color indicates the atom pair's absolute RMSD. The sphere size is proportional to the relative contribution of the atom pair to the total RMSD.

The methods developed for small molecules generally provide overview visualizations of uncertainty. They indicate the presence of uncertainty in the molecule and give a rough sense of its magnitude. However, apart from the combined color and size coding by Wagner and Himmel [WH17], making more detailed judgments is

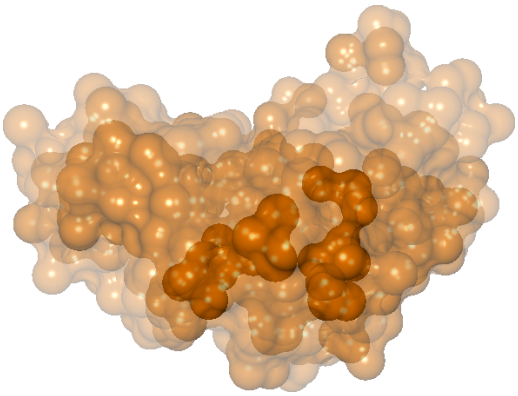


Figure 6: *Stacked semi-transparent surfaces display the spatial probability density of the molecule [KBE09]. Reprinted with permission from IEEE Transactions on Visualization and Computer Graphics, © 2009 IEEE. Not covered by the article's Creative Commons license.*

likely challenging. These methods also do not facilitate compare or integrate tasks.

7. Proteins

Proteins are the most extensively studied biomolecules for UV, with visualizations available for all types of uncertainty manifestations and mappings according to our schema. Similar to small molecules, many methods for UV apply to other biomolecule types. Nonetheless, certain techniques are specific to proteins, such as the visualization of secondary structures or cavities.

7.1. Positional

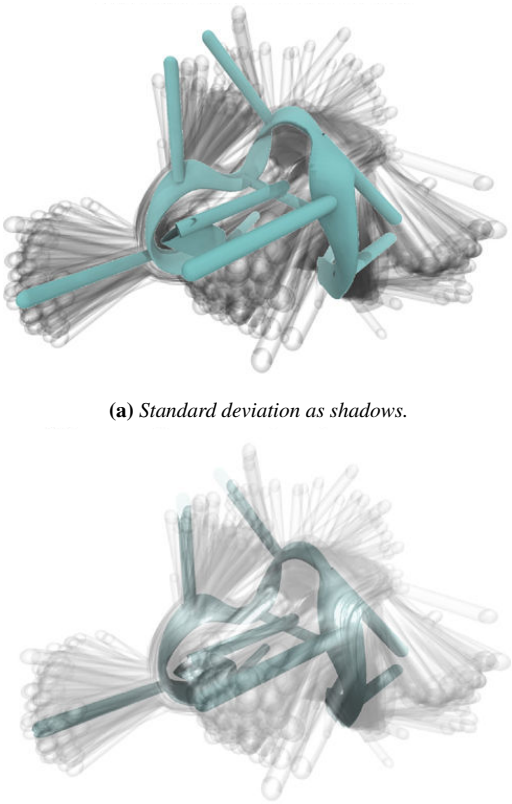
Explicit As early as 1989, Goodsell et al. [GMO89] started visualizing distributions of molecules rather than exact atom positions. They employed volume rendering with various transfer functions to visualize volumetric molecular data, including electron density and electrostatic potential. Specifically, they proposed three transfer functions. One generates a cloud representation, and the other two generate surface representations. The cloud representation assigns higher opacities to higher densities, resulting in more transparent uncertain areas. One of the surface representations combines several opaque isosurfaces with different layers becoming visible when using a clipping plane. The other representation consists of stacked transparent isosurfaces that become more opaque as the densities in the volume increase. Lee and Varshney [LV02] proposed another isosurface method: the fuzzy molecular surface, an uncertainty-aware version of the SES. They rendered several stacked semi-transparent surfaces to create this fuzzy surface. In contrast to most other methods, Lee and Varshney did not compute their surfaces from volume data. Given the coordinate data of the molecule, they computed *p-probability spheres* for each atom. These *p-probability spheres* are the smallest spheres that contain the atom center with probability p. They selected several levels of p for rendering the fuzzy surface. For each p, they extended all *p-probability spheres* by the probe radius. Then, for each p, they computed and rendered a semi-transparent SES-style surface. The fuzzy molecular surface appears blurry compared to the molecular surface without uncertainty representation. Krone et al. [KBE09] also used nested shells of semi-transparent molecular surfaces to show the positional uncertainty of proteins (Figure 6). They calculated the spatial probability density from simulation data on a regular grid. Next, they generated and rendered surfaces corresponding to different probability ranges by placing virtual atom spheres within the probability grid and computing a SES for each range. Layers with higher uncertainty were rendered more transparently than those with lower uncertainty. Hu et al. [HCZP06] used DVR to visualize 3D volume data generated by quantum mechanics simulations of proteins. Punjani and Fleet [PF23] developed 3DFlex, a motion-based neural network for determining the structure and motion of proteins from cryo-EM data. The models' latent variables describe the possible deformations of the molecule. They visualized several isosurfaces along one latent dimension ranging from minus to plus one standard deviation.

Similar to methods developed for small molecules, explicit distribution visualizations provide overviews of the uncertainty. Lee and Varshney [LV02] additionally demonstrate that additional data attributes can be visualized on the molecular surface using color.

Implicit Closely related to the direct visualization of the distribution are ensemble visualizations. In this implicit representation, individual molecular conformations are superposed to represent the continuous distribution. In 1997, Kelley and Sutcliffe. [KS97] proposed the online database OLDERADO that facilitates viewing an ensemble of molecular conformations and the most representative conformation in separate windows. The conformations were displayed using the stick representation. Lai et al. [LKCW04] visualized ensembles using sticks, ribbons, or cartoon representations. The web server MOBI [MWT10] visualizes NMR ensembles of proteins and encodes the residue mobility in color. Heinrich et al. [HKOW14] presented a visualization system for visualizing intrinsically disordered regions (IDRs). They combined parallel coordinates with ensemble visualization. Brushing and linking between the parallel coordinates and the ensembles enables better inspection and analysis of the proteins. Melvin and Salsbury [MS16] specifically selected conformations to be superimposed to create more manageable ensemble visualizations. One of their proposed selection criteria was frames lying within one standard deviation of the median. To additionally minimize visual clutter, they used a focus-and-context technique: They displayed one representative conformation opaque, the others transparently, resulting in a cloud or shadow-like visualization (Figure 7).

While all ensemble visualizations provide overviews of the data, the methods often appear cluttered. To reduce clutter, Melvin and Salsbury [MS16] introduced initial improvements. The additional use of color [MWT10] and the brushing capabilities introduced by Heinrich et al. [HKOW14] further support analysis tasks. Most ensemble visualization methods can also use color to incorporate additional data into the visualization.

Pathlines Dabdoub et al. [DMR08, DRSR15] drew pathlines that trace the movement of individual atoms in the molecule over time (Figure 8). Their preliminary study showed that pathline visualizations can help to understand molecular motion. However, their



(a) *Standard deviation as shadows.*

(b) *Standard deviation as clouds.*

Figure 7: *Ensemble visualization using focus and context to reduce visual clutter [MS16]. Reprinted with permission from Journal of Molecular Graphics and Modelling, © 2016 Elsevier. Not covered by the article's Creative Commons license.*

examples focused on relatively small molecules. Using this technique for larger molecules or over many time frames might lead to a cluttered appearance.

Ranges An alternative option for displaying uncertainty information is summary statistics. Here, instead of visualizing the distribution directly, uncertainty statistics—most often scalar uncertainty measures—are derived and subsequently displayed using a variety of metaphors. Most similar to the explicit visualization of the uncertainty distribution is displaying ranges. Many of the direct distribution visualizations mentioned earlier feature multiple layers of stacked isosurfaces. The techniques we classify as range visualizations are similar but only display one or two isosurface layers. If two isosurfaces are displayed, the inner surface usually corresponds to the surface as it would be displayed without uncertainty information. The outer layer then visualizes the uncertainty measure. However, the boundaries between the *explicit* and the *ranges* mapping categories are fluid, and the distinction based on a fixed number of layers is inherently flexible. An example of a method that uses only one layer is the work of Burnett and Johnson [BJ96]. They developed ORTEP-III, a program displaying thermal ellipsoids for crystal structures. These ellipsoids represent the anisotropic thermal motion of atoms, with their size and shape reflecting the average atomic displacement in different directions. This approach enables the visualization to support both analysis and overview tasks. However, for larger molecules—because of the atomic-level depictions—the visualizations can become increasingly difficult to interpret. In contrast, Kniesel et al. [KRH20] used transparency to visualize an inner and outer surface. For a density dataset of amyloid-beta fibrils, they rendered two isosurfaces with a semi-transparent outer hull and an opaque inner surface. This enables overview visualizations even for these larger molecules. Maack et al. [MRW*21] also used a semi-transparent hull in their visual analytics approach to visualize positional uncertainty in molecules (Figure 4b). The hull's position can be adjusted by a scaling factor accessible to the user. In addition, the system includes a previously developed visualization view [MGH19] that indicates the potential changes in the dihedral angles along the backbone of the molecule, considering potential changes in the position of the underlying molecule—an uncertainty-aware version of the Ramachandran plot. In addition to supporting the overview task, the use of color enables analysis and the integration of additional data. The authors also demonstrated a side-by-side comparison of a molecule with and without mutations. The differences in this visualization are primarily highlighted through color-coded surfaces.

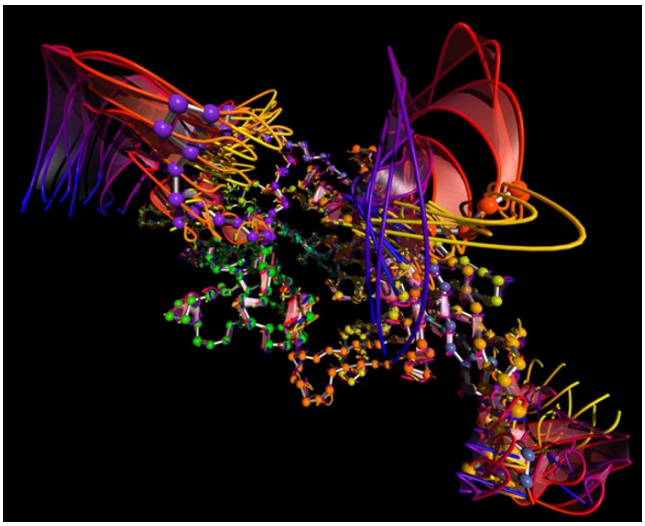
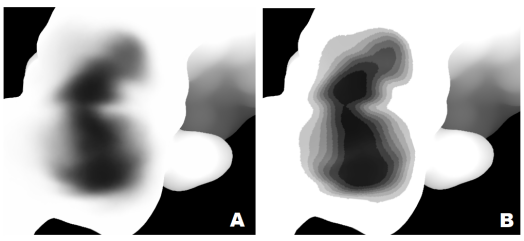
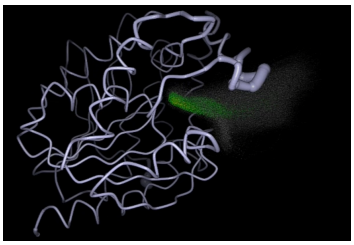


Figure 8: *Pathline visualization for MD. Reproduced from Dabdoub et al. [DRSR15], licensed under CC BY 4.0. No changes were made.*

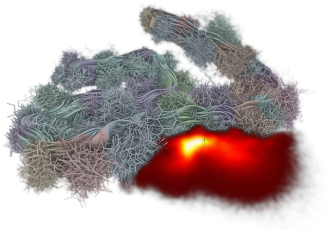
Animation Fisher et al. [FTPG90] visualized positional uncertainty by rendering a consensus surface of the two maximally displaced protein surfaces. The color of the surface varied over time, according to changes in the electrostatic field. As a second visualization option, the electrostatic potential was color-coded on the protein surface, and the electrostatic field was indicated with arrows. With this method, animation makes changes visible over time. This method inherently integrates multiple types of data and enables some level of analysis. However, it is the only method in the collected literature that does not offer an overview at a glance.



(a) *Two types of explicit distribution visualization highlight the potential impact of glycans on a protein surface. Reproduced from Besançon et al. [BWR*20], licensed under CC BY 4.0.*



(b) *Distribution of ligand trajectories with the backbone width of the protein encoding positional uncertainty [JFB*19]. Image courtesy of Katarína Furmanová, used with permission. Not covered by the article's Creative Commons license.*



(c) *DVR of the distribution of small molecules around a reference structure [SFL*21]. Image courtesy of Robin Skånberg, used with permission. Not covered by the article's Creative Commons license.*

Figure 9: *Distribution visualizations for molecules interacting with a protein.*

Size In contrast to range mappings, which indicate parameters of the distribution, methods that use size mapping are less restricted. Rather than directly mapping distribution attributes, size-based methods may, for instance, use size to convey that larger values correspond to higher uncertainty. Koradi et al. [KBW96] developed the visualization software MOLMOL. Positional uncertainty can be either displayed by the superposition of several conformers or by varying the tube radius of a backbone presentation. Wider tubes have higher positional uncertainty. This particular size encoding is often called sausage or putty representation. An example of the sausage representation is the depiction of the protein backbone in Figure 9b. Scott and Straus [SS15] used ensemble visualization and the sausage metaphor in a combined visualization. Since the color channel remains available for encoding different properties, these techniques support the integrate task alongside the overview task. Their suitability for analysis or comparison tasks is unclear, as no evaluations or example applications are demonstrating their effectiveness in these contexts.

Color A very popular option for displaying summary statistics is the use of color. There is no standard color mapping for uncertainty. Relatively common are blue-white-red colormaps. However, this colormap is also strongly associated with electrostatic potential. Another popular option, the red-green color map, is problematic for people with color-vision deficiencies. Some of the methods discussed previously use color as an additional mapping method. Here, we discuss methods that primarily use the color attribute. In 1995, Zachmann et al. [ZKB95] used texture mapping to map molecular flexibility to the SES. Thorpe et al. [TLR*01] used constraint theory to compute protein flexibility and colored their proteins according to this flexibility measure. Falk et al. [FTB*24] proposed a visual framework to support model building from cryo-EM density maps. It facilitates the verification and validation of 3D atomic models constructed from these maps. Several quality measures are displayed as a heat map, and for spatial analysis, a user-selected measure can be mapped onto a 3D representation of the model. While the visual mapping of uncertainty data is relatively simplistic, verification of 3D models has traditionally relied on tabular views or simple line graphs. This approach represents an important step forward in supporting model building.

Color encodings offer the advantage of providing detailed depictions of uncertainty, making them generally well-suited for analysis and comparison tasks. While the two methods specifically mentioned in this section did not explicitly demonstrate a comparison example, side-by-side displays are typically feasible for such tasks [MRW*21]. Additionally, color helps provide an overview of the data. However, since the color encodes uncertainty information, it cannot represent other attributes. As a result, the methods are not suitable for integrate tasks.

Density Similar to color mapping, different textures can encode uncertainty. One advantage of using non-color-based textures is that they can be combined with color to represent multiple attributes. For example, color could be used for a primary attribute like electrostatic potential, while texture could represent a secondary attribute such as uncertainty. Sterzik et al. [SMCL24] evaluated the perception of three illustrative texture maps (stippling, hatching, and triangles) with varying primitive density (Figure 4c). They conducted crowdsourced perceptual evaluations and provided approximately perceptual uniform reparameterizations for the density values. Sterzik et al. [SLK*22, SLK*23] proposed the use of stylized feature lines to encode uncertainty in biomolecular visualizations (Figure 10). They modified the line attributes blur, sketchiness, dashing, width, and grayscale (color value) to represent different levels of uncertainty. Additionally, they conducted two crowdsourced studies to evaluate the effectiveness of these line variables in distinguishing uncertainty values. Among the attributes tested, they identified line width as particularly suitable for effectively distinguishing between different levels of uncertainty.

In contrast to color-based methods, these illustrative techniques inherently allow for the integration of additional data. Otherwise, they should support the same tasks as color encodings. However, their effectiveness for specific tasks compared to other encodings has yet to be evaluated.

Glyphs Glyphs are another less common mapping metaphor. Bryden et al. [BPG12] used arrow glyphs to display larger motions of molecular groups. They visualized the results of normal mode analysis to illustrate molecular flexibility by summarizing groups of atoms into larger groups and visualizing the groups' mo-

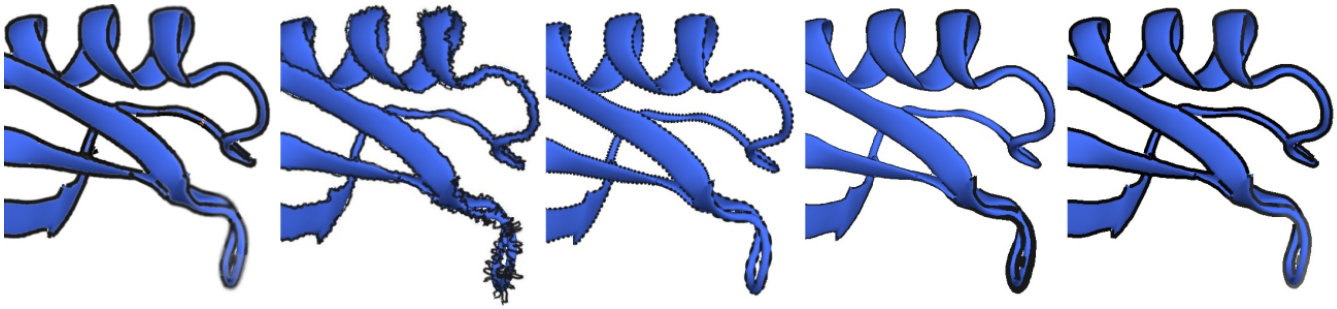


Figure 10: *Line attributes encode positional uncertainty. From left to right: blur, sketchiness, dashing, width, grayscale (color value). Adapted from Sterzik et al. [SLK*23], licensed under CC BY 4.0.*

tions with arrow glyphs. Although these larger motions may not strictly represent uncertainty, the glyphs' attributes can encode the fitting model's error and the group's rigidity. These properties help in judging the reliability of the motion model. They visualized the error in their fitting model with error bars around the glyphs. The error levels were discretized into three levels: small errors, medium, and larger errors. The respective encodings were no error bars, small error bars, and large error bars. Additionally, they added a stripe pattern to the arrows to encode the rigidity of the molecular group. The stripe pattern also had three discrete levels: completely colored, dense stripes, and sparse stripes, from low to high nonrigid energy (Figure 4a). Bedoucha et al. [BRHB20] incorporated these glyphs into their visual analytics tool for normal mode analysis. Additionally, atomic fluctuations were indicated by color mapping. In addition to supporting overview and analysis tasks, glyphs enable the integration of additional data by adding separate visual elements to the representation. However, since they are placed at discrete positions—and in the methods discussed here, rather sparsely—they can only convey detailed information at specific points. This makes them unsuitable for tasks that require a continuous representation of uncertainty across all positions.

7.2. Secondary structure

Implicit Three papers use ensembles to compare secondary structures. Kocincová et al. [KJB*17] introduced a system for comparing the secondary structures of protein ensembles. This system integrates a traditional 3D ensemble view with two views directly comparing secondary structure elements. The comparison views consist of superimposed and juxtaposed sequences. The secondary structure elements are depicted using conventional symbols: arrows for beta-sheets, spirals for alpha helices, and lines for coils. Although the superimposed and juxtaposed views are presented in a flattened 2D format, they retain and encode some information about the mutual orientation of corresponding elements through angles relative to the horizontal line. Chen and Porter [CP23] also displayed juxtaposed alignments of protein sequences with symbols encoding the secondary structure elements. The sequences are colored by conservation score, b-factor, or a user-defined measure.

Both methods explicitly support comparison. The approach by Kocincová et al. [KJB*17] preserves some orientation information in the 2D view, whereas Chen and Porter [CP23] completely flatten the sequences. However, Chen and Porter's color coding enables quantitative judgments of scalar uncertainty measures.

Schulz et al. [SSK*18] utilized juxtaposed sequences to compare secondary structure assignments generated by different methods. If a consensus sequence is available, it is displayed as a single sequence. In cases of assignment uncertainty, multiple sequences are stacked over each other. To further represent uncertainty, they introduced a combination of geometric distortions and transparency in their cartoon representations. The ribbons and tubes used to depict secondary structure elements are distorted using sine or triangle functions, with the frequency and amplitude reflecting the degree of uncertainty (Figure 5). Additionally, different structural predictions (e.g., helix, sheet, or coil) can be overlaid using screen-door transparency, where more probable structures appear more opaque. They also demonstrated the application of their distortion technique to illustrate positional uncertainty in a backbone tube representation. Their tool is the only existing one in the collected literature for comparing different secondary structure assignment methods, but they also demonstrate use cases for other types of uncertainty, particularly positional uncertainty.

Color The tool Aquaria [OSK*15] uses homology modeling, a method that predicts a 3D structure of a protein based on its similarity to known structures (templates). The tool generates a structural model by aligning the input sequence with these templates, assuming that similar sequences will adopt similar structures. The model is colored based on two attributes: The secondary structure element determines the hue, while the degree of conversation between the most relevant and the input structure is indicated by muting the colors. Highly conserved regions are brightly colored, unconserved regions are drawn in black, and inserted parts are colored gray. All other related structures are displayed in a 2D view as sequences of rectangles, colored according to the above-mentioned scheme, giving an overview of all sequences. Heinrich et al. [HKO15] evaluated this color mapping in a crowdsourced study. They found a significant correlation between the perceived quality of images and the alignment quality.

Vázquez et al. [VHG*18] developed a visual analysis tool for protein-ligand interactions that represents both positional and secondary structure uncertainties. The protein is depicted as a circular arrangement of its residues, with the ligand optionally displayed at the center. Positional uncertainty can be incorporated by adding

bars to the residues, where height and color encode the RMSF. Additionally, intra-molecular hydrogen bonds are visualized as ribbons connecting the respective residues, with ribbon color indicating bond persistence. While this is a highly abstracted structural visualization, we include this edge case in our survey due to its similarity to some of the few existing examples of UAV for RNA (see Section 8.1). To mitigate the limitations of such an abstracted representation, the authors incorporated a separate 3D structure view; however, this view does not include an uncertainty visualization.

7.3. Physicochemical

Rasheed et al. [RCBB16b, RCBB19] developed a statistical framework to quantify uncertainties for user-defined QOI, such as the Poisson-Boltzmann potential. They proposed visualizing these uncertainties through DVR or color mapping. Furthermore, ligand binding probabilities were visualized by mapping colors onto the protein surface and the ligand. In a preprint [RCBB16a], they also introduced a method to visualize positional uncertainty using vector glyphs. At each atomic position, three arrow glyphs represent the variability in the atom's location, with the glyph lengths indicating the magnitude of the uncertainty. While some other methods take physicochemical properties into account, this is the only one with a clear focus on them. The proposed visualizations support analysis, comparison, and overview tasks.

7.4. Cavities

While the previously discussed methods visualized manifestations of uncertainty that concerned the whole molecule, the following paragraph discusses methods developed for displaying uncertainty for biomolecular cavities.

Color Ashford et al. [AMA*12] visualized uncertainties concerning molecular pockets. The probability of an atom or residue belonging to a pocket is color-mapped to the surface, thereby supporting analysis and overview tasks.

Implicit Lindow et al. [LBBH12] proposed several visualizations for the dynamics of protein cavities and tunnels. They computed residence probabilities for every point on a regular grid and visualized these probabilities as iso-surfaces. The residence probability was defined as "the proportion of time in which the point is inside a cavity". Additionally, they used two types of 2D ensemble visualizations. A line graph showed topological changes in channels and cavities. Each component—the network of paths through the molecule that defines a single cavity or channel—was rendered as a single polyline. If splits or merges of the components occurred during the simulation, the lines split or merged respectively. A second line graph showed the extension of the components over time. In this graph, the positions and widths of the lines represent the location and extent of the path component along a user-defined direction. As the lines in this graph could overlap, they were drawn semi-transparently and blended. While the spatial uncertainty visualization only gives an overview of the data, the additional tools can be used for analysis tasks.

Byška et al. [BJG*15] provide an overview visualization of the time-dependency of tunnel bottlenecks. The bottleneck shape is visualized as a contour, and several time steps are overlaid to generate an ensemble visualization. The contours are colored according to the time step at which they occurred. They also indicated the surrounding amino acids with bars arranged radially around the contours based on their positions. The color of the bars indicates the amino acids' physio-chemical properties. Because it is common for amino acids to be replaced by other amino acids over time, individual bars for each amino acid were placed next to each other. The bar height indicates the number of time steps an amino acid was part of the tunnel boundary. The bottlenecks were aligned such that the surrounding amino acids overlapped the most. The AnimoAminoMiner [BLMG*16], a follow-up work, visualizes the tunnel width along its length, presenting multiple time steps as an ensemble. A second view enables exploration of the amino acids surrounding the tunnel. While both methods by Byška et al. provide highly abstracted structural representations, we included them in this review due to the limited number of works on uncertainty visualizations for tunnels.

Both methods are integrated into the analysis tool Caver Analyst [JBB*18]. The first method by Byška et al. is especially useful when the primary focus is on the tunnel's bottleneck. Otherwise, the time dependency of the entire tunnel can be examined using the AnimoAminoMiner [BLMG*16] or the method by Lindow et al. [LBBH12]. All three methods employ abstracted 2D structural representations; however, the approach by Lindow et al. [LBBH12] additionally incorporates a 3D visualization.

7.5. Interacting molecules

Next, we discuss methods that incorporate uncertainty measures for molecules interacting with a visualized protein. Most methods described below focus on the distribution of smaller molecules surrounding the main molecule.

Color Alharbi et al. [AKCL18, AKCL19a, AKCL19b] visualized protein-protein and protein-lipid interactions. They aggregated the number of contacts with another protein or a lipid during a simulation for each protein and showed the results using color-coding on an abstract representation of the protein (a cylinder). Similarly, Duncan et al. [DCS20] mapped contact frequencies of different lipids onto the surface of membrane proteins using color. The framework ProLint [ST21] uses a similar technique but allows for the selection of different metrics for the distribution, such as total contacts or longest contacts. Additionally, ProLint allows visualization of the lipid distribution via isolevels. Schatz et al. [SFS*21] computed the contact probability for a ligand across the molecular surface by aggregating how often a ligand of a certain type is in contact with a point on the molecular surface. A color-coding shows likely contact points. In addition, they presented the Compressed Ligand Interaction Sequence Diagram, which shows the contact probabilities, the RMSF, and other values per residue. Krone et al. [KFS*17] also used color-coding of the surface to show uncertainty. They introduced Molecular Surface Maps: map projections of molecular surfaces that give an occlusion-free overview of colored contact points on the surface. They demonstrated their approach for several properties, including the protein's b-factor and the contact probabilities for water. The methods in this section ap-

ply to different scenarios but share the same underlying technique. Color is used to map a scalar uncertainty value related to interacting molecules onto the molecular surface. This provides an overview of the data and supports analysis tasks.

Ranges In the collected literature, only one paper specifically focused on protein-protein interactions. Furmanová et al. [FJK*20] developed a visual analysis tool for large ensembles of multi-body protein complexes. The tool features a 3D view in which the distribution of secondary proteins around a primary protein is represented using an isolevel hull. This approach helps reduce visual clutter that would arise from directly displaying the full ensemble of possible configurations. Additionally, individual protein complexes can be explored in an exploded view, where contact frequencies are mapped to the surface using color. Beyond structural representations, the tool includes non-structural views that provide further uncertainty mappings, such as interaction frequency heatmaps.

Bhati et al. [BHP*23] displayed ligand occupancy using a wireframe isovalue surface. The method is similar to some of the explicit distribution visualizations discussed below; however, the wireframe representation can sometimes obscure the underlying surface, and only a single isolevel is provided, making a detailed assessment of the uncertainty distribution difficult. Nevertheless, the method offers an initial overview of the ligand distribution.

Distribution visualizations Besançon et al. [BGB*18, BWR*20, BGB*20] focused on a specific class of carbohydrates: n-glycans. They mapped the carbohydrates' distribution to the surface of the protein, either directly by adding up the opacities of individual glycan conformations or by abstracting and drawing contour lines for the distribution. Both versions of the visualization can be seen in Figure 9a. This method falls somewhere between scalar and distribution visualizations. Like color-based approaches, it maps uncertainty to color on the surface. At the same time, similar to distribution visualizations, it represents the uncertainty distribution without first converting it into scalar values for each point on the molecular surface. As with most other distribution visualizations, it is well suited for gaining an overview of the data, but finer details may not be sufficiently visible for analysis purposes.

Bidmon et al. [BGB*08] presented a method to extract and visualize the most likely paths of solvent molecules near a protein cavity. They first collected the paths of all solvent molecules that enter a user-defined area of interest during the simulation. Similar paths are clustered, and the results are shown as tubes, with the tube's thickness indicating the number of clustered paths. That is, paths with a higher probability were more prominent than less likely ones. The tubes were color-coded to show the direction of movement, and users could filter clusters with a low probability. Vad et al. [VBJ*17] developed a system for exploring water trajectories for MD simulations of proteins. Their system also visualizes trajectories that can optionally be smoothed. Additionally, the water molecule distribution can be visualized explicitly with density isosurfaces. Since directly visualizing all trajectories of the interacting molecules leads to a highly cluttered display, both methods [BGB*08, VBJ*17] rely on simplified representations. Further evaluation is necessary to determine their drawbacks and benefits. Durrieu et al. [DLB08] explicitly visualized the water molecule distribution in a protein complex through DVR. Jurčík et al. [JFB*19] developed a visual analytics tool for the analysis of ligand trajectories for MD simulations. The tool offers a 3D view with two uncertainty encodings (Figure 9b). Firstly, they visualize the distribution of ligand trajectories based on a voxel grid. The voxel grid is computed by counting the number of trajectories for each voxel. Each voxel is visualized by a sphere whose size and color correspond to the local density value. Secondly, the protein's flexibility is encoded in the backbone width (sausage plot). Skånberg et al. [SFL*21] developed a similar visualization for exploring the spatial relations between atomic structures. The distribution of the small molecules is determined by providing an internal frame of reference for a meaningful aggregation of results. For this purpose, the reference structure (large molecule) is assumed to be a rigid object. The small molecules are aggregated over time and molecular category. The method includes two options for displaying the distribution of the small molecules. The distribution could be displayed with two isosurfaces: a high-density opaque surface and a low-density transparent surface. Alternatively, the distribution was visualized with DVR and color coding. For this type of visualization, a cutting plane was used to reveal the core of the distribution (Figure 9c). Because molecules are not rigid bodies, Skånberg et al. introduced a shape variation plot to assess whether the rigid-body assumption is approximately true. The plot's barycentric shape space includes three dimensions: linear, planar, and spherical. In this shape-plot, rigid molecules only cover a small area, while flexible molecules cover a larger area. The method applies to broader applications, and the reference structure is not limited to a particular molecule type, e.g., reference structures in the paper included a protein, DNA, and small molecules. In VIAMD [SHYL23], these visualizations are embedded in a more comprehensive visual analysis tool for MD.

An alternative representation of uncertainties was chosen in the visual analytics application InVADo by Schäfer et al. [SBB*24], which allows docking data to be analyzed and explored. To investigate the docking probability at a binding site, docked ligand molecules were first clustered. Parts of the molecular surface that are close to one of these cluster positions are highlighted by color and opacity to indicate probable binding sites. The number of possible interactions can also be color-mapped onto the protein surface to show the spatial probability of an interaction. This can be done for different interaction types (i.e., physico-chemical properties). The possible distributions of hydrogen bonds are shown as cone glyphs connected to the participating protein atom. InVADo also offers non-spatial visualizations to show the uncertainty of the docking results: The distribution of docking scores (which indicates the probability of the corresponding docking event) per cluster is shown as a box plot and by color in a segmented heatmap. The frequency of different interactions within a cluster is shown as a bar chart to allow users to analyze the most likely interaction.

8. Nucleic Acids

Most papers about nucleic acids focus on visualizing RNA. Only two papers specifically visualized DNA.

8.1. RNA

All papers concerning RNA encode secondary structure uncertainties by using color. The secondary structure of RNA

molecules—specifically, the alignment of bases—is commonly visualized by displaying only the most probable alignments. The bases are represented by their corresponding letters and arranged so that paired bases are positioned close to each other. Additionally, paired bases are often connected by lines to indicate their interactions. ViennaRNA [GLB*08]—a toolkit for prediction and analysis of RNA secondary structures—incorporates basic UV by providing the option to color the bases according to their base pairing probabilities (Figure 3a). CentroidFold [SHAM09] is a web server that offers a similar visualization option. Beyond these simple uncertainty visualizations, which support basic overview and analysis tasks, there are also methods based on rainbow diagrams that facilitate comparison tasks. Rainbow diagrams consist of a sequence of base pairs written in a single row with bows connecting paired bases. RNABows [AJ13] enhance rainbow diagrams with uncertainty information by encoding the base pair probabilities in the bows' width and lightness. Improbable pairings are connected with thinner and lighter lines and are thus less visible than probable pairings. Additionally, RNABows offers the possibility to compare two sequences by plotting them back to back. Léger et al. [LCT19] adapted RNABows by using a circle as the foundation for plotting the sequence (see Figure 3b). Several sequences are compared by plotting their respective circles next to each other. This allows for the comparison of more than two sequences, but the association between them is not as clear as in RNABows [AJ13]. Irving and Weeks [IW24] proposed the tool RNAvigate to compare base pair predictions from several studies. They adapted rainbow plots and colored the bows according to the number of studies that predicted a specific pairing. Additionally, minimum free energy (MFE) pairings were indicated below the sequence in gray.

8.2. DNA

Ertl et al. [EKK*14] visualized the ion density near a strand of DNA in a nanopore using nested semi-transparent surfaces. Since the DNA is negatively charged, positive ions are likely to be close to the DNA (see Figure 11). That is, the different isosurfaces illustrate different density ranges or ion probabilities. Different sequential colors (blue, green, yellow) are assigned for better discriminability of the nested surfaces. In addition, users can visualize the pathlines of the ions. Xie et al. [XZF20] proposed a technique for uncertainty-aware DNA visualization. Their tool PyShifts facilitates comparisons between experimentally measured and computed chemical shifts. These shifts provide information on the conformational states of biomolecules and can be used to model their secondary and tertiary structures. The magnitude of the difference in measured and computed shifts is encoded in the size of the sphere drawn at a nucleus position. The color of the sphere indicates the sign of the difference. The methods for visualizing DNA are more generic than for the proposed methods RNA and could also be applied to other types of biomolecules.

9. Discussion

In the following, we will discuss patterns and trends of UV techniques for biomolecular structures. Figure 12 shows a parallel sets plot [KBH06] of the distribution and correlation of the collected literature's attributes along our schemas main dimensions. In the visualization, the dimensions—molecule type, manifestation, and mapping—are represented with consistent total contributions across layers. However, papers often vary in the number of applicable categories per dimension. We distribute each paper's contribution equally among the applicable categories within a given dimension. For example, if a paper employs three mapping strategies, each mapping receives a proportional width of 1/3 of the paper's total contribution. Conversely, a paper with only one mapping allocates its full contribution to that mapping. This strategy ensures that each paper contributes equally to the total width per dimension.

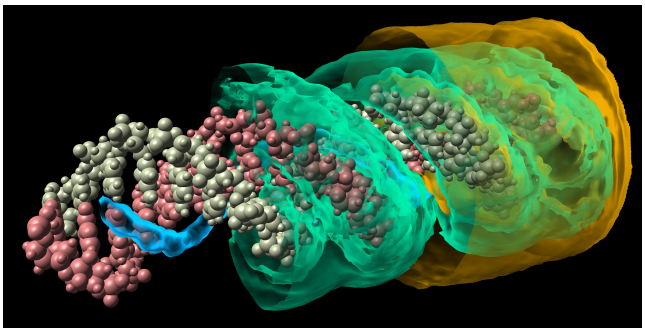


Figure 11: *An example for uncertainty visualization showing the probability density of ions around DNA [EKK*14].*

9.1. Molecule Type

The field of uncertainty-aware biomolecular structure visualization clearly emphasizes proteins, with the vast majority of published papers focusing on this type of biomolecule. This is not a major concern for small molecules, as most alternative methods are readily applicable. Small molecules lack unique uncertainty manifestations, ensuring that existing approaches are generally applicable without significant adaptation. However, the representation of nucleic acids is also limited, especially for DNA. This is a more significant issue since nucleic acids have unique visualization needs—for example, for base pairing probabilities—that necessitate the development of specific uncertainty visualizations. Interestingly, except for the paper by Ertl et al. [EKK*14] on DNA, and the more general framework applicable to a broad range of molecules proposed by Skanberg et al. [SFL*21, SHYL23], the majority of papers on nucleic acids stem from domain experts in bioinformatics and related fields, rather than from the visualization research community. Carbohydrates and lipids are similarly underrepresented. There are some works concerning carbohydrates [BGB*18, BWR*20, BGB*20] and lipids [DCS20, ST21]. However, they are primarily concerned with protein interactions and thus sorted into the protein category of our schema. This focus is somewhat expected, as the visualization of individual carbohydrates or lipids tends to hold limited interest. Instead, the primary relevance lies in their interactions with other molecules, such as proteins, or, in the case of lipids, their collective behavior within larger molecular assemblies like membranes.

The focus on proteins aligns with the emphasis in the domain sciences. Historically, and to a lesser extent even today, proteins have been a primary focus of research [Cam02, BBD*22].

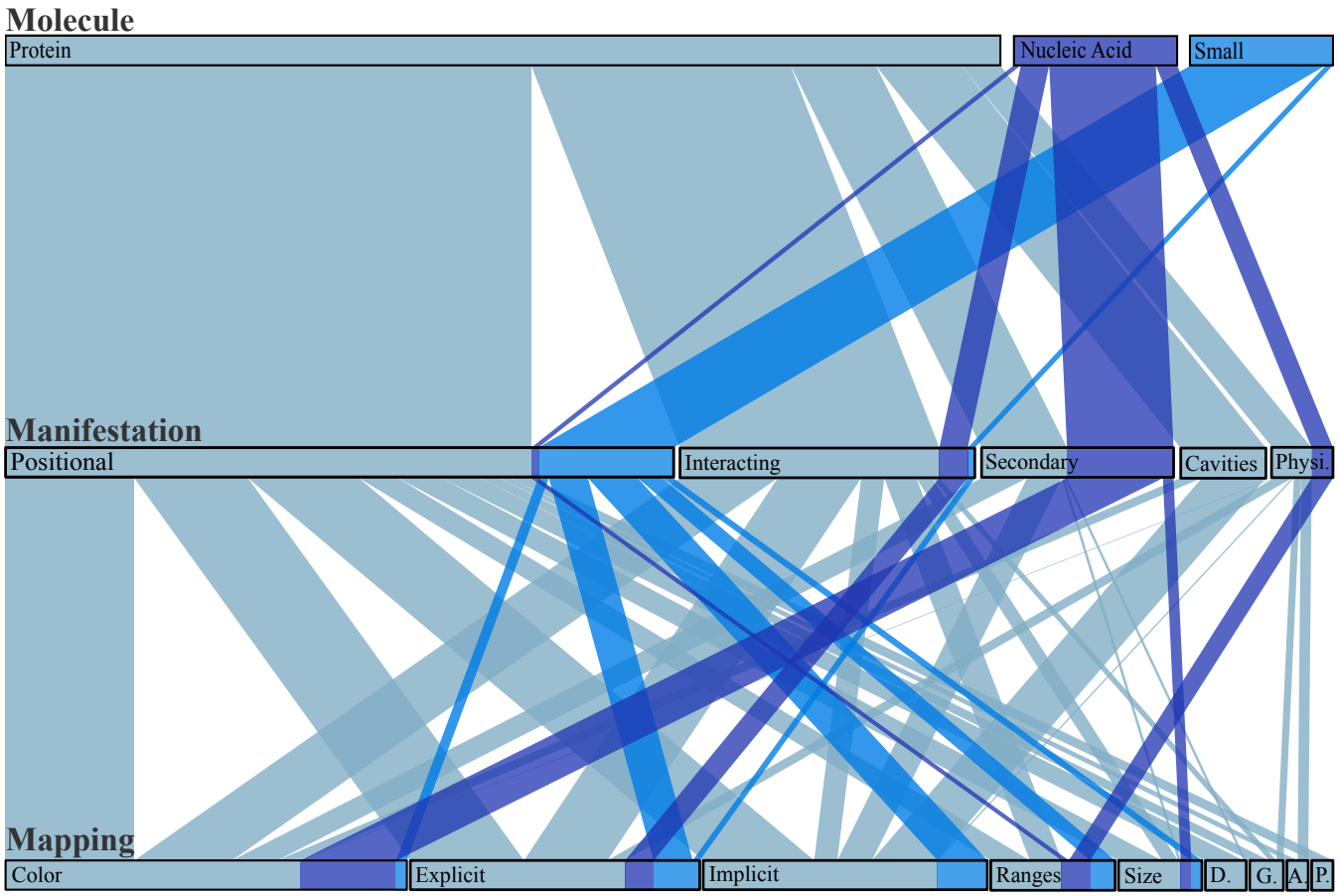


Figure 12: *Parallel sets plot illustrating the relationships between molecule types, manifestations, and mappings in the collected literature. For proteins, all manifestations and mappings are represented, while far fewer combinations have been explored for nucleic acids and small molecules. Abbreviations: Physi. (Physicochemical), D. (Density), G. (Glyphs), A. (Animation), P. (Pathline).*

This focus can be attributed to several factors, such as the availability of structural data. Proteins, for example, are generally easier to crystallize than RNA because RNA is more flexible [MSAC15]. Additionally, proteins play a central role as functional molecules within cells [AHJ*22]. However, this focus has gradually shifted over time, particularly as RNA is no longer viewed primarily as a messenger molecule [MM14]. Moreover, advancements in techniques like cryo-EM have enabled the visualization of larger molecular systems, such as membranes. This has expanded the scope of research beyond small, pharmaceutically relevant molecules and proteins to include a broader range of biomolecules and larger structural scales, such as entire membranes or large molecular complexes [RCC*18]. However, uncertainty-aware approaches for larger-scale (e.g. mesoscale) and multiscale molecular visualizations remain largely unexplored. Some progress has been made in related areas, such as the dynamics of membrane systems. For example, Chavent et al. [CRG*14] employed pathlines, arrow glyphs, and streamlines to visualize lipid flow in membrane systems. As structural biology has advanced in recent decades, the visualization research community needs to address the growing complexity of larger datasets and more intricate molecular systems.

9.2. Manifestation of Uncertainty

Uncertainty in molecular visualization appears in several forms, depending on the molecule or system studied. A primary focus has been on positional uncertainty. As discussed in Section 4.2, positional uncertainty information is readily available and typically one of the first manifestations of uncertainty considered. Unsurprisingly, the majority of visualization methods focus on positional uncertainty. Additionally, methods developed for one molecule type are often transferable to others when investigating positional uncertainty. Furthermore, positional uncertainty influences most other properties investigated for biomolecules and propagates, affecting uncertainty measures for other manifestations as well. Physicochemical properties are another area of interest, with visualizations typically using generic methods like direct distribution representations. While these techniques are broadly applicable, they are not tailored to specific properties, which may limit their effectiveness in some contexts. Interactions with other molecules are an important but underexplored topic in UV. If interactions were addressed, existing work primarily focused on proteins and mainly addressed interactions with small surrounding molecules, such as water. However, many other types of interactions require

appropriate uncertainty visualizations, including but not limited to protein-protein, protein-ligand, nucleic acid-protein, ligand-nucleic acid, and membrane interactions. Progress is being made in visualizing these other types of interactions, as shown by the visual analysis tool developed by Furmanová et al. [FJK*20] for large ensembles of multi-body protein complexes. However, the topic remains largely underexplored. As computational power increases and imaging techniques improve, research has shifted towards studying more dynamic and complex molecular systems, highlighting the need for further development in this area. For secondary structure uncertainties, visualization efforts for RNA have primarily focused on base-pairing probabilities. These visualizations are either quite basic—such as color mapping applied to existing 2D structure representations—or limited to linear sequence visualizations. Comprehensive 3D visualizations for RNA secondary structure uncertainties are entirely missing. Additionally, visualizations of conformational spaces are absent, highlighting a significant gap in capturing the full scope of structural variability. UAV for biomolecular cavities in proteins has also received little attention. Part of it is likely due to cavities being a relatively specialized manifestation of uncertainties. Existing tools for visualizing cavities often rely on rather abstract 2D representations for uncertainty. This is unsurprising, as cavities are inherently complex due to their dependence on positional uncertainty and physicochemical properties. Thus, it is crucial to properly evaluate existing methods to determine whether they successfully aid in making these complex phenomena understandable and whether the proposed abstract mappings are appropriate.

9.3. Mapping

In visualizing molecular structure uncertainty, explicit and implicit mapping of distributions, along with the use of attributes like color, are the most commonly employed techniques. Ensemble visualizations are particularly popular, which is unsurprising given that many imaging methods and simulations produce a set of conformations. This makes ensemble visualizations a natural choice for representing uncertainty. Similarly, explicit distribution visualizations often directly depict electron density maps or the distribution of physicochemical properties, providing another straightforward approach. Additionally, color mapping is very common because it is simple, implemented in standard visualization software, and easy to apply.

Despite these established approaches, other techniques remain relatively unexplored. For example, animation is not used to visualize uncertainty in biomolecules, except for one of the earliest papers on the subject [FTPG90]. A possible reason for this is that biomolecules are inherently dynamic structures, and animations are typically expected to represent molecular motion rather than uncertainty. While animations that incorporate subtle motion, such as wiggling, are sometimes used in molecular science communication, a proper and systematic investigation of the possibilities of using animation to visualize uncertainty is still lacking. Another underexplored mapping strategy is glyphs. While Bryden et al. [BPG12] introduced glyph-based approaches, these techniques have not been widely adopted. This leaves considerable potential for glyphs to be utilized, especially in static visualizations or for science communication. Similarly, other static representations of dynamics like pathlines, motion lines, or motion blur are rarely used and remain largely unexplored for depicting uncertainty in molecular systems.

An example of motion blur for molecular dynamics visualization can be found in the paper by Eschner et al. [EMW23], who used motion blur to guide attention in molecular dynamics visualizations. Spalvieri et al. [SML*22] illustrated molecular motion using motion blur for a DNA strand in their paper on collaborations with designers for improving molecular visualization. Another example result of such a collaboration was the temporal visualization of a membrane protein that serves as an ion channel through the membrane barrier; abstracted, semi-transparent 2D visualizations of the channel's key constriction were superimposed to form a single 2D image, similar to the contour line approach of Byška et al. [BJG*15]. This type of collaboration can inspire new ideas for molecular UV. Recently, research has explored alternative visual attributes for uncertainty mapping, such as density in illustrative textures [SMCL24] and different line attributes [SLK*23]. These methods can display summary statistics while leaving the color channel open for other properties. However, we believe there is still significant potential for further exploration of alternative attributes, particularly through collaboration with design experts.

We classified the mapping strategies into two main categories: visual representations of the uncertainty distributions and visual representations of summary statistics. The distribution visualizations are particularly useful for comprehending the uncertainty inherent to biomolecules, which often translates to understanding the potential conformational space that a given molecule can occupy. Summary statistics can typically be easily integrated into existing visualizations, offering uncertainty information in a less obtrusive way while providing only aggregated data. However, currently available distribution visualizations are often very cluttered and difficult to comprehend.

Efforts in related areas could serve as inspiration for more advanced techniques. Liu et al. [LLBP12], for example, used Gaussian mixture models to represent a data distribution per voxel in a volume. Probabilistic rendering could then be used to visualize uncertainty. For dynamic rendering, the volume is displayed by sampling the distributions at each frame, where areas of higher uncertainty appear as increased flickering compared to regions of lower uncertainty. For static images, samples can be integrated in screen space, producing more blurring in areas with higher uncertainty.

Another area that has received little attention is the uncertainty-aware comparison of similar molecules. Comparisons of similar structures—such as wild types and mutated structures— are highly relevant. Another potential application for such comparative visualizations is ensembles of MD trajectories, which Belghit et al. [BSD*24] identified as one of the main challenges for MD visualization. Nevertheless, UV for such comparisons remain scarce. In existing work, several structures were either displayed by direct overlay—that is, as an ensemble—or as side-by-side visualization. Kocincová et al. [KJB*17] proposed a more advanced approach by combining the typical ensemble view with juxtaposed or superposed 1D representations of protein secondary structures.

Uncertainty mapping in biomolecules could benefit from more

abstract or illustrative visualization techniques, an area that remains largely untapped. While abstraction approaches do not allow for precise judgments of the level of uncertainty, they can represent broad categories of uncertainty, as illustrated by Strothotte et al. [SMI99] for ancient architectural reconstructions. To our knowledge, Sorger et al. [SMK*16] are the only authors to have presented a paper using abstraction in an uncertainty-related context for biomolecules. They interpolate between different states of biological mesoscale models at varying levels of detail by applying an abstraction technique. The model with the higher level of detail is abstracted until it has a resolution similar to that of the less detailed model. At this level of detail, the models can be interpolated and subsequently reverse-transformed to the original level of abstraction. Thus, while this approach does not utilize abstraction to highlight regions of higher uncertainty within the same image, the concept could also be applied to UV. Additional approaches for creating abstract or illustrative visualizations of biomolecules exist for other purposes. Parulek et al. [PJR*14] use seamless visual abstraction for increasing rendering performance. Van der Zwan et al. [vLBI11] use continuous illustrative abstraction for focus and context and to guide attention. Thus, while methods for illustration and abstraction exist, their utility for UV is still unexplored.

9.4. Software Availability

In general, the availability of software for uncertainty-aware biomolecular structure visualizations is poor. For only half of the projects (30 out of 61), software was provided to reproduce the visualizations. Only 25 projects are still available. This presents a significant issue, as the importance of UV is frequently emphasized in the literature [Joh04, BHJ*14, PKH21]. However, the mere development of these methods is insufficient for their adoption. To ensure usability, practitioners must have access to these tools. Ideally, new approaches should be integrated into widely used visualization frameworks such as PyMOL [Sch], VMD [HDS96], and ChimeraX [PGH*21]. At a minimum, reproducible source code should be provided to support broader use and validation.

However, some mappings are available in standard molecular renderers and widely used. Color mapping of scalar uncertainty values, such as the b-factor, is generally accessible and frequently applied. Simple ensemble visualizations are also common. The sausage metaphor, which varies the tube radius of the backbone based on uncertainty, is available in many renderers. DVR and isolevel extraction of probability distributions are also widely supported. Additionally, some renderers include options for displaying glyphs, such as motion vectors for NMA. However, the availability of more advanced methods or solutions for less common manifestations of uncertainty is often poor.

9.5. Evaluation

A significant gap is the evaluation of proposed methods. Most methods are assessed primarily through case studies. When the authors are not domain experts, expert interviews often supplement the evaluations. While case studies are valuable for demonstrating that a method can offer advantages over traditional visualization techniques, they are rarely sufficient for a comprehensive evaluation. Only five methods were subjected to more rigorous quantitative evaluations, such as user studies or perceptual studies. This lack of empirical evaluation complicates informed decision-making about UV methods. A deeper understanding of the strengths and weaknesses of different approaches is generally necessary for choosing an appropriate method for a specific problem.

9.6. Techniques Across Domains

Taxonomies for UV often classify methods based on the dimensionality of the data [PWL97, BAOL12, BHJ*14]. Most methods discussed in reports on general UV focus on 1D or 2D data. Biomolecular structure data is inherently 3D. In visualizations of uncertain 1D or 2D data, uncertainty is often represented by incorporating additional dimensions. However, this approach is typically impractical for 3D, as all available spatial dimensions are already utilized. Therefore, alternative uncertainty mappings must be used. That being said, many other metaphors used for uncertainty in lower-dimensional data can be adapted for 3D data. For example, error bars for point estimates in 1D become error bands in 2D and hulls in 3D. This adaptability means metaphors from general taxonomies can often inform biomolecular structure UV. Some metaphors proposed in the literature, though not yet applied to biomolecular structures, could also be explored for biomolecular structure data. Examples include blinking or oscillation [PWL97, BHJ*14], sonification [PWL97, BAOL12], and surface displacement or bumps [PWL97, BHJ*14]. For example, Grigoryan and Rheingans [GR04] render noisy surfaces by representing a surface as a collection of points displaced based on the uncertainty of the structure. This technique could be directly adapted for molecular surfaces. Thus, biomolecular structure visualization can benefit from techniques and insights from other disciplines. Conversely, the reverse holds also true: Many methods developed for uncertainty-aware biomolecular structure visualization are not limited to this domain and can be applied to related fields, such as material science, chemical engineering, and nanotechnology. Furthermore, disciplines that rely on 3D models with inherent uncertainty, such as medical imaging, face similar challenges and could benefit from shared approaches. Ensemble data, one of the most common forms of uncertain biomolecular structure data, are an essential part of many other fields such as meteorology, climatology, astrophysics, high-energy physics, and oceanography [WHLS19], suggesting opportunities for cross-disciplinary collaboration and method transfer. UV for biomolecular structures is an active area of research, supported by the extensive availability of data and its importance to the life sciences. Developing visualization techniques for this domain enhances the interpretation of biomolecular data and provides insights applicable to more general UV challenges, including the visualization of ensembles and uncertain 3D data.

10. Research Directions

In the previous chapter, we explored trends and outlined various challenges in the field. Here, we narrow our focus and highlight a few areas we consider the most important for future research.

Complex and Multiscale Systems Advances in imaging technologies enable the visualization of increasingly complex systems.

While some approaches exist for visualizing systems of interacting molecules, such as protein-ligand interactions, there is significant room for improvement. UAV for even larger systems, such as mesoscale structures or visualizations spanning multiple scales, remain largely unexplored. Methods that work well for small to medium-sized molecular structures often become cluttered or impractical when applied to more complex systems. This highlights the need for approaches that can effectively manage the scale and complexity of these data. A potential solution is adopting more abstract and illustrative visualization techniques. These approaches could offer clearer, more interpretable visualizations while addressing the unique challenges of complex molecular systems.

Conformational Space While there is already a substantial amount of literature visualizing the conformational space, the problem remains challenging, particularly for larger molecules, as existing techniques often become cluttered and difficult to interpret. Initial efforts to enhance these visualizations were undertaken by Melvin et al. [MS16], who employed focus and context, as well as a deliberately selected subset of conformations, to reduce visual complexity. Nevertheless, easily comprehensible visualizations of the entire conformational space remain scarce. Similar to the visualization of complex systems, the volume and complexity of the data can be overwhelming. However, the multidimensional nature of conformational spaces introduces additional challenges. As with complex systems, abstraction and illustrative techniques offer promising solutions. Simplified representations that emphasize key states or transitions could enhance clarity, while interactive tools for exploring specific regions of the conformational landscape could improve usability. Alternatively, animations—similar to HOPs [Hul16]—could dynamically represent conformational changes, reducing the need to convey all information within a single, static visualization.

Comparison Techniques Comparing molecular configurations and their uncertainties typically involves side-by-side visualizations or simple overlays. These methods are effective for small datasets but become impractical as the number of molecules increases. Overlays face challenges similar to visualizing conformational space, such as becoming cluttered and difficult to interpret. This issue is especially problematic when comparing dynamic molecules. Ensembles of molecular dynamics (MD) simulations have become crucial for addressing the lack of reproducibility in individual simulations [WSC21]. However, MD ensembles generate vast amounts of data, requiring new visualization techniques capable of handling this complexity without overwhelming users. This challenge was also highlighted by Belghit et al. [BSD*24]. This scalability problem is not limited to biomolecular systems; it also affects broader ensemble visualization methods [WHLS19]. Advancements in both fields could offer mutual benefits, leading to better solutions for visualizing large, complex datasets.

Nucleic Acids Despite being the focus of numerous current research topics [WEFC21, BJZ*22, WD23], nucleic acids are often underrepresented in terms of specialized UV techniques. Current visualizations primarily focus on base pair probabilities and rely on relatively simple methods. While these approaches provide some insight into nucleic acid structure, they are limited in scope and fail to capture the full complexity of these molecules. Comprehensive 3D visualizations, which could provide a more complete representation of nucleic acid structures, are largely absent. Additionally, visualizing the conformational space of nucleic acids remains a significant gap despite its importance for understanding molecular dynamics, interactions, and function.

Evaluation UV techniques for biomolecular structures are not being evaluated rigorously enough. For example, it is crucial to determine which methods are most intuitive for viewers, how effectively they support the inference of quantitative values, and whether they accurately represent the conformational space of biomolecules. Since humans often reason in non-intuitive ways when interpreting uncertainty [PKH21], it is essential to rigorously evaluate uncertainty visualizations to ensure they effectively support accurate understanding and decision-making. Evaluating UV methods is inherently challenging [Hul16], even for simple 1D and 2D visualizations. Incorporating and evaluating uncertainty in the context of 3D biomolecular structure visualizations is even more complex. However, addressing this gap is essential. It will enable practitioners to choose the most suitable methods for specific problems, identify shortcomings in existing techniques, and drive the development of new methods that better meet the demands of UAV.

11. Conclusion

In this report, we presented UAV approaches for biomolecular structures, classifying the methods along three primary dimensions: molecule type, manifestation of uncertainty, and mapping strategy. While these methods demonstrate strategies for representing uncertainty for individual molecules, they often fall short of addressing the demands of modern biomolecular research. The constantly increasing size and complexity of biomolecular datasets require visualization techniques that scale effectively and remain interpretable. Current approaches frequently appear cluttered even for small to medium-sized molecules, making them unsuitable for larger systems such as mesoscale data or ensembles of dynamic structures. These gaps highlight the need for innovative visualization strategies that can handle the complexity and scale of modern biomolecular data. Advancing UAV methods is essential for improving how researchers interpret and interact with complex biomolecular data, enabling clearer insights and more informed decision-making. Given the parallels to other domains, such as general uncertainty and ensemble visualization, advancements in UAV for biomolecular structures are likely to have a broader impact, benefiting the entire field of UAV.

Acknowledgments

We want to thank Frank Cordes, Ute Hellmich, Philipp Schnee, Philipp Thiel, Markus Weber, and Christoph Wiedemann for the insightful discussions on UV for biomolecular structures. We would also like to thank Katarína Furmanová, Barbora Kozlíková, and Robin Skånberg for sharing images with us for this report. This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project-ID 437702916. Open Access funding enabled and organized by Projekt DEAL.

References

[AHJ*22] ALBERTS B., HEALD R., JOHNSON A., MORGAN D., RAFF M., ROBERTS K., WALTER P.: *Molecular Biology of the Cell (Seventh Edition)*. W. W. Norton, Incorporated, 2022. 3, 20

[AJ13] AALBERTS D. P., JANNEN W. K.: Visualizing RNA base-pairing probabilities with RNABow diagrams. *RNA (New York, N.Y.) 19*, 4 (2013), 475–478. doi:10.1261/rna.033365.112. 11, 19

[AKCL18] ALHARBI N., KRONE M., CHAVENT M., LARAMEE R. S.: VAPLI: Novel visual abstraction for protein-lipid interactions. In *2018 IEEE Scientific Visualization Conference (SciVis)* (2018), pp. 6–10. doi:10.1109/SciVis.2018.8823785. 11, 17

[AKCL19a] ALHARBI N., KRONE M., CHAVENT M., LARAMEE R. S.: Hybrid visualization of protein-lipid and protein-protein interaction. *Eurographics Workshop on Visual Computing for Biology and Medicine* (2019), 11 pages. doi:10.2312/VCBM.20191247. 17

[AKCL19b] ALHARBI N., KRONE M., CHAVENT M., LARAMEE R. S.: LoD PLI: Level of detail for visualizing time-dependent, protein-lipid interaction. In *10th International Conference on Information Visualization Theory and Applications (IVAPP)* (2019), pp. 164–174. 17

[AMA*12] ASHFORD P., MOSS D. S., ALEX A., YEAP S. K., POVIA A., NOBELI I., WILLIAMS M. A.: Visualisation of variable binding pockets on protein surfaces by probabilistic analysis of related structure sets. *BMC bioinformatics 13* (2012), 39. doi:10.1186/1471-2105-13-39. 11, 17

[BAOL12] BRODLIE K., ALLENDES OSORIO R., LOPES A.: A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*. Springer London,, 2012, pp. 81–109. doi:10.1007/978-1-4471-2804-5_6. 2, 3, 7, 22

[BBD*22] BURLEY S. K., BERMAN H. M., DUARTE J. M., FENG Z., FLATT J. W., HUDSON B. P., LOWE R., PEISACH E., PIEHL D. W., ROSE Y., SALI A., SEKHARAN M., SHAO C., VALLAT B., VOIGT M., WESTBROOK J. D., YOUNG J. Y., ZARDECKI C.: Protein Data Bank: A comprehensive review of 3D structure holdings and worldwide utilization by researchers, educators, and students. *Biomolecules 12*, 10 (2022), 1425. doi:10.3390/biom12101425. 19

[BGB*08] BIDMON K., GROTTEL S., BÖS F., PLEISS J., ERTL T.: Visual abstractions of solvent pathlines near protein cavities. *Computer Graphics Forum 27*, 3 (2008), 935–942. doi:10.1111/j.1467-8659.2008.01227.x. 11, 18

[BGB*18] BESANÇON C., GUILLOT A., BLAISE S., DAUCHEZ M., BELLOY N., PRÉVOTEAU-JONQUET J., BAUD S.: New visualization of dynamical flexibility of N-glycans: Umbrella visualization in UnityMol. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (2018), pp. 291–298. doi:10.1109/BIBM.2018.8621256. 11, 18, 19

[BGB*20] BESANÇON C., GUILLOT A., BLAISE S., DAUCHEZ M., BELLOY N., PRÉVOTEAU-JONQUET J., BAUD S.: Umbrella visualization: A method of analysis dedicated to glycan flexibility with UnityMol. *Methods 173* (2020), 94–104. doi:10.1016/j.ymeth.2019.07.010. 18, 19

[BH02] BEST C., HEGE H.-C.: Visualizing and identifying conformational ensembles in molecular dynamics trajectories. *Computing in Science & Engineering 4*, 3 (2002), 68–75. doi:10.1109/5992.998642. 11, 12

[BHJ*14] BONNEAU G.-P., HEGE H.-C., JOHNSON C. R., OLIVEIRA M. M., POTTER K., RHEINGANS P., SCHULTZ T.: Overview and state-of-the-art of uncertainty visualization. In *Scientific Visualization*. Springer, London, 2014, pp. 3–27. doi:10.1007/978-1-4471-6497-5_1. 2, 3, 7, 22

[BHP*23] BHATI A. P., HOTI A., POTTERTON A., BIENIEK M. K., COVENEY P. V.: Long time scale ensemble methods in molecular dynamics: Ligand–protein interactions and allostery in SARS-CoV-2 targets. *Journal of Chemical Theory and Computation 19*, 11 (2023), 3359–3378. doi:10.1021/acs.jctc.3c00020. 11, 18

[BJ96] BURNETT M., JOHNSON C.: *ORTEP-III: Oak Ridge Thermal Ellipsoid Plot Program for Crystal Structure Illustrations*. Tech. Rep. ORNL–6895, 369685, 1996. doi:10.2172/369685. 11, 14

[BJG*15] BYŠKA J., JURČÍK A., GRÖLLER M. E., VIOLA I., KOZLÍKOVÁ B.: MoleCollar and tunnel heat map visualizations for conveying spatio-temporo-chemical properties across and along protein voids. *Computer Graphics Forum 34*, 3 (2015), 1–10. doi:10.1111/cgf.12612. 11, 17, 21

[BJZ*22] BARBIER A. J., JIANG A. Y., ZHANG P., WOOSTER R., ANDERSON D. G.: The clinical progress of mRNA vaccines and immunotherapies. *Nature Biotechnology 40*, 6 (2022), 840–854. doi:10.1038/s41587-022-01294-2. 23

[Bli82] BLINN J. F.: A generalization of algebraic surface drawing. *ACM Transactions on Graphics 1*, 3 (1982), 235–256. doi:10.1145/357306.357310. 6

[BLMG*16] BYŠKA J., LE MUZIC M., GRÖLLER M. E., VIOLA I., KOZLÍKOVÁ B.: AnimoAminoMiner: Exploration of protein tunnels and their properties in molecular dynamics. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (2016), 747–756. doi:10.1109/TVCG.2015.2467434. 11, 17

[BPG12] BRYDEN A., PHILLIPS G. N., GLEICHER M.: Automated illustration of molecular flexibility. *IEEE Transactions on Visualization and Computer Graphics 18*, 1 (2012), 132–145. doi:10.1109/TVCG.2010.250. 6, 11, 15, 21

[BRHB20] BEDOUCHA P., REUTER N., HAUSER H., BYŠKA J.: Visual exploration of large normal mode spaces to study protein flexibility. *Computers & Graphics 90* (2020), 73–83. doi:10.1016/j.cag.2020.05.025. 11, 16

[BSD*24] BELGHIT H., SPIVAK M., DAUCHEZ M., BAADEN M., JONQUET-PREVOTEAU J.: From complex data to clear insights: Visualizing molecular dynamics trajectories. *Frontiers in Bioinformatics 4* (2024). doi:10.3389/fbinf.2024.1356659. 3, 21, 23

[BTS10] BERG J., TYMOCZKO J., STRYER L.: *Biochemistry*. W. H. Freeman, 2010. 3

[BWF*00] BERMAN H. M., WESTBROOK J., FENG Z., GILLILAND G., BHAT T. N., WEISSIG H., SHINDYALOV I. N., BOURNE P. E.: The Protein Data Bank. *Nucleic Acids Research 28*, 1 (2000), 235–242. doi:10.1093/nar/28.1.235. 8, 11

[BWR*20] BESANÇON C., WONG H., RAO R., DAUCHEZ M., BELLOY N., PRÉVOTEAU-JONQUET J., BAUD S.: Improved umbrella visualization implemented in UnityMol gives valuable insight on sugar/protein interplay. In *Workshop on Molecular Graphics and Visual Analysis of Molecular Data (MolVA) (2020)* (2020). doi:10.2312/molva.20201097. 15, 18, 19

[Cam02] CAMPBELL I. D.: The march of structural biology. *Nature Reviews Molecular Cell Biology 3*, 5 (2002), 377–381. doi:10.1038/nrm800. 19

[CBC23] COREY R. A., BAADEN M., CHAVENT M.: A brief history of visualizing membrane systems in molecular dynamics simulations. *Frontiers in Bioinformatics 3* (2023). doi:10.3389/fbinf.2023.1149744. 3

[CP23] CHEN E. A., PORTER L. L.: SSDraw: Software for generating comparative protein secondary structure diagrams. *Protein Science 32*, 12 (2023), e4836. doi:10.1002/pro.4836. 11, 16

[CRG*14] CHAVENT M., REDDY T., GOOSE J., E. DAHL A. C., E. STONE J., JOBARD B., P. SANSOM M. S.: Methodologies for the analysis of instantaneous lipid diffusion in MD simulations of large membrane systems. *Faraday Discussions 169*, 0 (2014), 455–475. doi:10.1039/C3FD00145H. 20

[DCS20] DUNCAN A. L., COREY R. A., SANSOM M. S. P.: Defining how multiple lipid species interact with inward rectifier potassium (Kir2) channels. *Proceedings of the National Academy of Sciences 117*, 14 (2020), 7803–7813. doi:10.1073/pnas.1918387117. 11, 17, 19

[DLB08] DURRIEU M.-P., LAVERY R., BAADEN M.: Interactions between neuronal fusion proteins explored by molecular dynamics. *Biophysical Journal 94*, 9 (2008), 3436–3446. doi:10.1529/biophysj.107.123117. 11, 18

[DMR08] DABDOUB S., MOHAN A., RAY W. C.: Visualizing molecular uncertainty: A path to the path. In *ACM SIGGRAPH 2008 Posters* (2008), p. 1. doi:10.1145/1400885.1401022. 11, 13

[DRSR15] DABDOUB S. M., RUMPF R., SHINDHELM A. D., RAY W. C.: MoFlow: Visualizing conformational changes in molecules as molecular flow improves understanding. *BMC Proceedings 9* (2015). doi:10.1186/1753-6561-9-S6-S5. 13, 14

[Ede99] EDELSBRUNNER H.: Deformable smooth surface design. *Discrete & Computational Geometry 21*, 1 (1999), 87–115. doi:10.1007/PL00009412. 6

[EKK*14] ERTL T., KRONE M., KESSELHEIM S., SCHARNOWSKI K., REINA G., HOLM C.: Visual analysis for space–time aggregation of biomolecular simulations. *Faraday Discussions 169* (2014), 167–178. doi:10.1039/C3FD00156C. 11, 19

[EMW23] ESCHNER J., MINDEK P., WALDNER M.: Illustrative motion smoothing for attention guidance in dynamic visualizations. *Computer Graphics Forum 42*, 3 (2023), 361–372. doi:10.1111/cgf.14836. 21

[FJK*20] FURMANOVÁ K., JURČÍK A., KOZLÍKOVÁ B., HAUSER H., BYŠKA J.: Multiscale visual drilldown for the analysis of large ensembles of multi-body protein complexes. *IEEE Transactions on Visualization and Computer Graphics 26*, 1 (2020), 843–852. doi:10.1109/TVCG.2019.2934333. 11, 18, 21

[FTB*24] FALK M., TOBIASSON V., BOCK A., HANSEN C., YNNERMAN A.: A visual environment for data driven protein modeling and validation. *IEEE Transactions on Visualization and Computer Graphics 30*, 8 (2024), 5063–5073. doi:10.1109/TVCG.2023.3286582. 11, 15

[FTPG90] FISHER C. L., TAINER J. A., PIQUE M. E., GETZOFF E. D.: Visualization of molecular flexibility and its effects on electrostatic recognition. *Journal of Molecular Graphics 8*, 3 (1990), 125–132, 145. doi:10.1016/0263-7855(90)80052-h. 11, 14, 21

[GLB*08] GRUBER A. R., LORENZ R., BERNHART S. H., NEUBÖCK R., HOFACKER I. L.: The Vienna RNA websuite. *Nucleic Acids Research 36*, suppl_2 (2008), W70–W74. doi:10.1093/nar/gkn188. 4, 11, 19

[GMO89] GOODSELL D. S., MIAN I., OLSON A. J.: Rendering volumetric data in molecular systems. *Journal of Molecular Graphics 7*, 1 (1989), 41–47. doi:10.1016/0263-7855(89)80055-4. 11, 13

[GMR*23] GILLMANN C., MAACK R. G. C., RAITH F., PÉREZ J. F., SCHEUERMANN G.: A Taxonomy of Uncertainty Events in Visual Analytics. *IEEE Computer Graphics and Applications 43*, 5 (2023), 62–71. doi:10.1109/MCG.2023.3299297. 7

[GR04] GRIGORYAN G., RHEINGANS P.: Point-based probabilistic surfaces to show surface uncertainty. *IEEE Transactions on Visualization and Computer Graphics 10*, 5 (2004), 564–573. doi:10.1109/TVCG.2004.30. 22

[GRT19] GERRITS T., RÖSSL C., THEISEL H.: Towards glyphs for uncertain symmetric second-order tensors. *Computer Graphics Forum 38*, 3 (2019), 325–336. doi:10.1111/cgf.13692. 9

[GS06] GRIETHE H., SCHUMANN H.: The visualization of uncertain data: Methods and problems. In *Simulation and Visualization* (2006), vol. 2006. 3

[GSWS21] GILLMANN C., SAUR D., WISCHGOLL T., SCHEUERMANN G.: Uncertainty-aware visualization in medical imaging - a survey. *Computer Graphics Forum 40*, 3 (2021), 665–689. doi:10.1111/cgf.14333. 7

[GWHA18] GILLMANN C., WISCHGOLL T., HAMANN B., AHRENS J.: Modeling and visualization of uncertainty-aware geometry using multi-variate normal distributions. In *2018 IEEE Pacific Visualization Symposium (PacificVis)* (2018), IEEE, pp. 106–110. doi:10.1109/PacificVis.2018.00021. 10

[Ham14] HAMADA M.: Fighting against uncertainty: An essential issue in bioinformatics. *Briefings in Bioinformatics 15*, 5 (2014), 748–767. doi:10.1093/bib/bbt038. 8

[HCZP06] HU M., CHEN W., ZHANG T., PENG Q.: Direct volume rendering of volumetric protein data. In *Advances in Computer Graphics* (2006), Lecture Notes in Computer Science, pp. 397–403. doi:10.1007/11784203_34. 11, 13

[HDS96] HUMPHREY W., DALKE A., SCHULTEN K.: VMD – visual molecular dynamics. *Journal of Molecular Graphics 14* (1996), 33–38. 22

[HKO15] HEINRICH J., KAUR S., O'DONOGHUE S.: Evaluating the effectiveness of color to convey alignment quality in macromolecular structures. In *2015 Big Data Visual Analytics (BDVA)* (2015), pp. 1–8. doi:10.1109/BDVA.2015.7314292. 16

[HKOW14] HEINRICH J., KRONE M., O'DONOGHUE S. I., WEISKOPF D.: Visualising intrinsic disorder and conformational variation in protein ensembles. *Faraday Discussions 169* (2014), 179–193. doi:10.1039/C3FD00138E. 11, 13

[HQC*19] HULLMAN J., QIAO X., CORRELL M., KALE A., KAY M.: In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 903–913. doi:10.1109/TVCG.2018.2864889. 3

[HRA15] HULLMAN J., RESNICK P., ADAR E.: Hypothetical Outcome Plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLOS ONE 10*, 11 (2015). doi:10.1371/journal.pone.0142444. 9

[Hul16] HULLMAN J.: Why evaluating uncertainty visualization is error prone. *ACM International Conference Proceeding Series 24-October-2016* (2016), 143–151. doi:10.1145/2993901.2993919. 23

[IW24] IRVING P. S., WEEKS K. M.: RNAvigate: Efficient exploration of RNA chemical probing datasets. *Nucleic Acids Research 52*, 5 (2024), 2231–2241. doi:10.1093/nar/gkae089. 11, 19

[JBB*18] JURCIK A., BEDNAR D., BYSKA J., MARQUES S. M., FURMANOVA K., DANIEL L., KOKKONEN P., BREZOVSKY J., STRNAD O., STOURAC J., PAVELKA A., MANAK M., DAMBORSKY J., KOZLIKOVA B.: CAVER Analyst 2.0: Analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics 34*, 20 (2018), 3586–3588. doi:10.1093/bioinformatics/bty386. 17

[JFB*19] JURČÍK A., FURMANOVÁ K., BYŠKA J., VONÁSEK V., VÁVRA O., ULBRICH P., HAUSER H., KOZLÍKOVÁ B.: Visual analysis of ligand trajectories in molecular dynamics. In *2019 IEEE Pacific Visualization Symposium (PacificVis)* (2019), pp. 212–221. doi:10.1109/PacificVis.2019.00032. 11, 15, 18

[Joh04] JOHNSON C.: Top scientific visualization research problems. *IEEE Computer Graphics and Applications 24*, 4 (2004), 13–17. doi:10.1109/MCG.2004.20. 22

[JS03] JOHNSON C., SANDERSON A.: A next step: Visualizing errors and uncertainty. *IEEE Computer Graphics and Applications 23*, 5 (2003), 6–10. doi:10.1109/MCG.2003.1231171. 2

[KBE09] KRONE M., BIDMON K., ERTL T.: Interactive visualization of molecular surface dynamics. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 1391–1398. doi:10.1109/TVCG.2009.157. 11, 13

[KBH06] KOSARA R., BENDIX F., HAUSER H.: Parallel Sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics 12*, 4 (2006), 558–568. doi:10.1109/TVCG.2006.76. 19

[KBW96] KORADI R., BILLETER M., WÜTHRICH K.: MOLMOL: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics 14*, 1 (1996), 51–55. doi:10.1016/0263-7855(96)00009-4. 11, 15

[KCL*13] KNOLL A., CHAN M. K. Y., LAU K. C., LIU B., GREELEY J., CURTISS L., HERELD M., PAPKA M. E.: Uncertainty classification and visualization of molecular interfaces. *International Journal for Uncertainty Quantification 3*, 2 (2013), 157–169. doi:10.1615/Int.J.UncertaintyQuantification.2012003950. 11, 12

[Ken61] KENDREW J. C.: The three-dimensional structure of a protein molecule. *Scientific American 205*, 6 (1961), 96–111. 1

[KF10] KONOLD T. R., FAN X.: Hypothesis testing and confidence intervals. In *International Encyclopedia of Education (Third Edition)*. Elsevier, Oxford, 2010, pp. 216–222. doi:10.1016/B978-0-08-044894-7.01337-3. 7

[KFS*17] KRONE M., FRIESS F., SCHARNOWSKI K., REINA G., FADEMRECHT S., KULSCHEWSKI T., PLEISS J., ERTL T.: Molecular surface maps. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 701–710. doi:10.1109/TVCG.2016.2598824. 11, 17

[KJB*17] KOCINCOVÁ L., JAREŠOVÁ M., BYŠKA J., PARULEK J., HAUSER H., KOZLÍKOVÁ B.: Comparative visualization of protein secondary structures. *BMC Bioinformatics 18*, S2 (2017), 23. doi:10.1186/s12859-016-1449-z. 11, 16, 21

[KKF*17] KOZLÍKOVÁ B., KRONE M., FALK M., LINDOW N., BAADEN M., BAUM D., VIOLA I., PARULEK J., HEGE H. C.: Visualization of biomolecular structures: State of the art revisited. *Computer Graphics Forum 36*, 8 (2017), 178–204. doi:10.1111/cgf.13072. 2, 3, 5

[KKL*16] KRONE M., KOZLÍKOVÁ B., LINDOW N., BAADEN M., BAUM D., PARULEK J., HEGE H.-C., VIOLA I.: Visual analysis of biomolecular cavities: State of the art. *Computer Graphics Forum 35*, 3 (2016), 527–551. doi:10.1111/cgf.12928. 3

[KRH20] KNIESEL H., ROPINSKI T., HERMOSILLA P.: Real-time visualization of 3D amyloid-beta fibrils from 2D cryo-EM density maps. In *Eurographics Workshop on Visual Computing for Biology and Medicine* (2020), p. 11 pages. doi:10.2312/VCBM.20201178. 11, 14

[KS97] KELLEY L. A., SUTCLIFFE M. J.: OLDERADO: On-line database of ensemble representatives and domains. *Protein Science: A Publication of the Protein Society 6*, 12 (1997), 2628–2630. doi:10.1002/pro.5560061215. 11, 13

[LBBH12] LINDOW N., BAUM D., BONDAR A.-N., HEGE H.-C.: Dynamic channels in biomolecular systems: Path analysis and visualization. In *2012 IEEE Symposium on Biological Data Visualization (BioVis)* (2012), pp. 99–106. doi:10.1109/BioVis.2012.6378599. 11, 17

[LBH14] LINDOW N., BAUM D., HEGE H.-C.: Ligand Excluded Surface: A new type of molecular surface. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2486–2495. doi:10.1109/TVCG.2014.2346404. 6

[LBSP14] LAWONN K., BAER A., SAALFELD P., PREIM B.: Comparative evaluation of feature line techniques for shape depiction. In *Proc. of Vision, Modeling and Visualization* (2014), pp. 31–38. doi:10.2312/vmv.20141273. 10

[LCT19] LÉGER S., COSTA M. B. W., TULPAN D.: Pairwise visual comparison of small RNA secondary structures with base pair probabilities. *BMC Bioinformatics 20*, 1 (2019), 293. doi:10.1186/s12859-019-2902-6. 4, 11, 19

[Lev66] LEVINTHAL C.: Molecular model-building by computer. *Scientific American 214*, 6 (1966), 42–52. doi:10.1038/scientificamerican0666-42. 1

[LGV*20] LAZAR T., GUHAROY M., VRANKEN W., RAUSCHER S., WODAK S. J., TOMPA P.: Distance-based metrics for comparing conformational ensembles of intrinsically disordered proteins. *Biophysical Journal 118*, 12 (2020), 2952–2965. doi:10.1016/j.bpj.2020.05.015. 8

[LKCW04] LAI P., KAPLAN W., CHURCH W. B., WONG R. K.: Informative 3D visualization of multiple protein structures. In *Proceedings of the Second Conference on Asia-Pacific Bioinformatics - Volume 29* (2004), APBC '04, pp. 201–208. doi:10.5555/976520.976547. 11, 13

[LLBP12] LIU S., LEVINE J. A., BREMER P.-T., PASCUCCI V.: Gaussian mixture model based volume visualization. In *IEEE Symposium on Large Data Analysis and Visualization (LDAV)* (2012), pp. 73–77. doi:10.1109/LDAV.2012.6378978. 21

[LLPY07] LUNDSTRÖM C., LJUNG P., PERSSON A., YNNERMAN A.: Uncertainty visualization in medical volume rendering using probabilistic animation. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1648–1655. doi:10.1109/TVCG.2007.70518. 10

[LP16] LAWONN K., PREIM B.: *Feature Lines for Illustrating Medical Surface Models: Mathematical Background and Survey*. Springer Verlag, 2016, ch. Visualization in Medicine in Life Sciences III, pp. 93–132. doi:10.1007/978-3-319-24523-2_5. 10

[LPCH19] LIU L., PADILLA L., CREEM-REGEHR S. H., HOUSE D. H.: Visualizing uncertain tropical cyclone predictions using representative samples from ensembles of forecast tracks. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 882–891. doi:10.1109/TVCG.2018.2865193. 9

[LR71] LEE B., RICHARDS F. M.: The interpretation of protein structures: Estimation of static accessibility. *Journal of molecular biology 55*, 3 (1971), 379–IN4. 6

[LSBP18] LAWONN K., SMIT N. N., BÜHLER K., PREIM B.: A survey on multimodal medical data visualization. *Computer Graphics Forum 37*, 1 (2018), 413–438. doi:10.1111/cgf.13306. 10

[LV02] LEE C. H., VARSHNEY A.: Representing thermal vibrations and uncertainty in molecular surfaces. In *Electronic Imaging 2002* (2002), pp. 80–90. doi:10.1117/12.458813. 11, 13

[LVPI18] LAWONN K., VIOLA I., PREIM B., ISENBERG T.: A survey of surface-based illustrative rendering for visualization. *Computer Graphics Forum 37*, 6 (2018), 205–234. doi:10.1111/cgf.13322. 10

[LW24] LI H., WEI X.: A concise review of biomolecule visualization. *Current Issues in Molecular Biology 46*, 2 (2024), 1318–1334. doi:10.3390/cimb46020084. 3

[MDC24] MLYNEK G., DJINOVIĆ-CARUGO K., CARUGO O.: B-factor rescaling for protein crystal structure analyses. *Crystals 14*, 5 (2024), 443. doi:10.3390/cryst14050443. 8

[MGH19] MAACK R. G. C., GILLMANN C., HAGEN H.: Uncertainty-aware ramachandran plots. In *2019 IEEE Pacific Visualization Symposium (PacificVis)* (2019), pp. 227–231. doi:10.1109/PacificVis.2019.00034. 14

[MKK*19] MIAO H., KLEIN T., KOUŘIL D., MINDEK P., SCHATZ K., GRÖLLER M. E., KOZLÍKOVÁ B., ISENBERG T., VIOLA I.: Multiscale molecular visualization. *Journal of Molecular Biology 431*, 6 (2019), 1049–1070. doi:10.1016/j.jmb.2018.09.004. 3

[MM14] MORRIS K. V., MATTICK J. S.: The rise of regulatory RNA. *Nature Reviews Genetics 15*, 6 (2014), 423–437. doi:10.1038/nrg3722. 20

[MRO*12] MACEACHREN A. M., ROTH R. E., O'BRIEN J., LI B., SWINGLEY D., GAHEGAN M.: Visual semiotics & uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2496–2505. doi:10.1109/TVCG.2012.279. 9

[MRW*21] MAACK R. G., RAYMER M. L., WISCHGOLL T., HAGEN H., GILLMANN C.: A framework for uncertainty-aware visual analytics of proteins. *Computers & Graphics* (2021), 293–305. doi:10.1016/j.cag.2021.05.011. 6, 11, 14, 15

[MS16] MELVIN R. L., SALSBURY F. R.: Visualizing ensembles in structural biology. *Journal of Molecular Graphics and Modelling 67* (2016), 44–53. doi:10.1016/j.jmgm.2016.05.001. 11, 13, 14, 23

[MSAC15] MARCHANKA A., SIMON B., ALTHOFF-OSPELT G., CARLOMAGNO T.: RNA structure determination by solid-state NMR spectroscopy. *Nature Communications 6*, 1 (2015), 7024. doi:10.1038/ncomms8024. 20

[MWT10] MARTIN A. J. M., WALSH I., TOSATTO S. C. E.: MOBI: A web server to define and visualize structural mobility in NMR protein ensembles. *Bioinformatics (Oxford, England) 26*, 22 (2010), 2916–2917. doi:10.1093/bioinformatics/btq537. 11, 13

[NC21] NELSON D., COX M.: *Lehninger Principles of Biochemistry: International Edition*. Macmillan Learning, 2021. 3

[NJT08] NADAV-GREENBERG L., JOSLYN S., TAING M.: The effect of uncertainty visualizations on decision making in weather forecasting. *Journal of Cognitive Engineering and Decision Making 2* (2008), 24–47. doi:10.1518/155534308X284354. 10

[OSK*15] O'DONOGHUE S. I., SABIR K. S., KALEMANOV M., STOLTE C., WELLMANN B., HO V., ROOS M., PERDIGÃO N., BUSKE F. A., HEINRICH J., ROST B., SCHAFFERHANS A.: Aquaria: Simplifying discovery and insight from protein structures. *Nature Methods 12*, 2 (2015), 98–99. doi:10.1038/nmeth.3258. 11, 16

[Per64] PERUTZ M. F.: The hemoglobin molecule. *Scientific American 211*, 5 (1964), 64–79. 1

[PF23] PUNJANI A., FLEET D. J.: 3DFlex: Determining structure and motion of flexible proteins from cryo-EM. *Nature Methods 20*, 6 (2023), 860–870. doi:10.1038/s41592-023-01853-8. 11, 13

[PGH*21] PETTERSEN E. F., GODDARD T. D., HUANG C. C., MENG E. C., COUCH G. S., CROLL T. I., MORRIS J. H., FERRIN T. E.: UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science: A Publication of the Protein Society 30*, 1 (2021), 70–82. doi:10.1002/pro.3943. 22

[PJR*14] PARULEK J., JÖNSSON D., ROPINSKI T., BRUCKNER S., YNNERMAN A., VIOLA I.: Continuous levels-of-detail and visual abstraction for seamless molecular visualization. *Computer Graphics Forum 33*, 6 (2014), 276–287. doi:10.1111/cgf.12349. 6, 22

[PKH21] PADILLA L., KAY M., HULLMAN J.: Uncertainty visualization. In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd, 2021. doi:9781118445112.stat08296. 3, 22, 23

[PWB*09] POTTER K., WILSON A., BREMER P.-T., WILLIAMS D., DOUTRIAUX C., PASCUCCI V., JOHNSON C.: Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. *Journal of Physics: Conference Series 180*, 1 (2009), 012089. 10

[PWL97] PANG A. T., WITTENBRINK C. M., LODHA S. K.: Approaches to uncertainty visualization. *The Visual Computer 13*, 8 (1997), 370–390. doi:10.1007/s003710050111. 3, 7, 22

[RCBB16a] RASHEED M., CLEMENT N., BHOWMICK A., BAJAJ C.: Quantifying and visualizing uncertainties in molecular models, 2016. arXiv:1508.03882. 17

[RCBB16b] RASHEED M., CLEMENT N., BHOWMICK A., BAJAJ C.: Statistical framework for uncertainty quantification in computational molecular modeling. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (2016), BCB '16, pp. 146–155. doi:10.1145/2975167.2975182. 11, 17

[RCBB19] RASHEED M., CLEMENT N., BHOWMICK A., BAJAJ C. L.: Statistical framework for uncertainty quantification in computational molecular modeling. *IEEE/ACM Transactions on Computational Biology and Bioinformatics 16*, 4 (2019), 1154–1167. doi:10.1109/TCBB.2017.2771240. 8, 17

[RCC*18] RENAUD J.-P., CHARI A., CIFERRI C., LIU W.-T., RÉMIGY H.-W., STARK H., WIESMANN C.: Cryo-EM in drug discovery: Achievements, limitations and prospects. *Nature Reviews Drug Discovery 17*, 7 (2018), 471–492. doi:10.1038/nrd.2018.77. 20

[Ric77] RICHARDS F. M.: Areas, volumes, packing and protein structure. *Annual Review of Biophysics and Bioengineering 6* (1977), 151–176. doi:10.1146/annurev.bb.06.060177.001055. 6

[Ric81] RICHARDSON J. S.: The anatomy and taxonomy of protein structure. In *Advances in Protein Chemistry*, vol. 34. Elsevier, 1981, pp. 167–339. doi:10.1016/S0065-3233(08)60520-3. 6

[RJ99] RHEINGANS P., JOSHI S.: Visualization of molecules with positional uncertainty. In *Data Visualization '99* (Vienna, 1999), Springer Vienna, pp. 299–306. doi:https://doi.org/10.1007/978-3-7091-6803-5_30. 2, 10, 11, 12

[RSG21] RAITH F., SCHEUERMANN G., GILLMANN C.: Uncertainty-aware detection and visualization of ocean eddies in ensemble flow fields - a case study of the red sea. In *Workshop on Visualisation in Environmental Sciences (EnvirVis)* (2021), The Eurographics Association. doi:10.2312/envirvis.20211080. 10

[SBB*24] SCHÄFER M., BRICH N., BYSKA J., MARQUES S. M., BEDNAR D., THIEL P., KOZLIKOVA B., KRONE M.: InVADo: Interactive visual analysis of molecular docking data. *IEEE Transactions on Visualization and Computer Graphics 30*, 4 (2024), 1984–1997. doi:10.1109/TVCG.2023.3337642. 11, 18

[SBH02] SCHMIDT-EHRENBERG J., BAUM D., HEGE H.-C.: Visualizing dynamic molecular conformations. In *IEEE Visualization* (2002), pp. 235–242. doi:10.1109/VISUAL.2002.1183780. 11, 12

[Sch] SCHRÖDINGER, LLC: The PyMOL molecular graphics system, version 3.0. 22

[SFL*21] SKÅNBERG R., FALK M., LINARES M., YNNERMAN A., HOTZ I.: Tracking internal frames of reference for consistent molecular distribution functions. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. doi:10.1109/TVCG.2021.3051632. 11, 15, 18, 19

[SFS*21] SCHATZ K., FRANCO-MORENO J. J., SCHÄFER M., ROSE A. S., FERRARIO V., PLEISS J., VÁZQUEZ P.-P., ERTL T., KRONE M.: Visual analysis of large-scale protein-ligand interaction data. *Computer Graphics Forum 40*, 6 (2021), 394–408. doi:10.1111/cgf.14386. 11, 17

[SHAM09] SATO K., HAMADA M., ASAI K., MITUYAMA T.: CentroidFold: A web server for RNA secondary structure prediction. *Nucleic Acids Research 37*, Web Server issue (2009), W277–W280. doi:10.1093/nar/gkp367. 11, 19

[SHYL23] SKÅNBERG R., HOTZ I., YNNERMAN A., LINARES M.: VIAMD: A software for visual interactive analysis of molecular dynamics. *Journal of Chemical Information and Modeling 63*, 23 (2023), 7382–7391. doi:10.1021/acs.jcim.3c01033. 18, 19

[SLK*17] SMIT N., LAWONN K., KRAIMA A., DERUITER M., SOKOOTI H., BRUCKNER S., EISEMANN E., VILANOVA A.: PelVis: Atlas-based surgical planning for oncological pelvic surgery. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 741–750. doi:10.1109/TVCG.2016.2598826. 10

[SLK*22] STERZIK A., LICHTENBERG N., KRONE M., CUNNINGHAM D. W., LAWONN K.: Perceptual evaluation of common line variables for displaying uncertainty on molecular surfaces. In *Eurographics Workshop on Visual Computing for Biology and Medicine* (2022). doi:10.2312/vcbm.20221186. 10, 11, 15

[SLK*23] STERZIK A., LICHTENBERG N., KRONE M., BAUM D., CUNNINGHAM D. W., LAWONN K.: Enhancing molecular visualization: Perceptual evaluation of line variables with application to uncertainty visualization. *Computers & Graphics* (2023), 401–413. doi:10.1016/j.cag.2023.06.006. 15, 16, 21

[SLW*24] STERZIK A., LICHTENBERG N., WILMS J., KRONE M., CUNNINGHAM D. W., LAWONN K.: Perception of line attributes for visualization. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (2024), 1041–1051. doi:10.1109/TVCG.2023.3326523. 6

[SMCL24] STERZIK A., MEUSCHKE M., CUNNINGHAM D. W., LAWONN K.: Perceptually uniform construction of illustrative textures. *IEEE Transactions on Visualization and Computer Graphics 30*, 1 (2024), 1052–1062. doi:10.1109/TVCG.2023.3326574. 10, 11, 15, 21

[SMI99] STROTHOTTE T., MASUCH M., ISENBERG T.: Visualizing knowledge about virtual reconstructions of ancient architecture. In *1999 Proceedings Computer Graphics International* (1999), pp. 36–43. doi:10.1109/CGI.1999.777901. 22

[SMK*16] SORGER J., MINDEK P., KLEIN T., JOHNSON G., VIOLA I.: Illustrative transitions in molecular visualization via forward and inverse abstraction transform. *Eurographics Workshop on Visual Computing for Biology and Medicine* (2016), 10 pages. doi:10.2312/VCBM.20161267. 22

[SML*22] SPALVIERI D., MAUVIEL A.-M., LAMBERT M., FÉREY N., SACQUIN-MORA S., CHAVENT M., BAADEN M.: Design – a new way to look at old molecules. *Journal of Integrative Bioinformatics 19*, 2 (2022). doi:10.1515/jib-2022-0020. 21

[SS15] SCOTT W. R. P., STRAUS S. K.: Determining and visualizing flexibility in protein structures. *Proteins: Structure, Function, and Bioinformatics 83*, 5 (2015), 820–826. doi:10.1002/prot.24776. 11, 15

[SSK*18] SCHULZ C., SCHATZ K., KRONE M., BRAUN M., ERTL T., WEISKOPF D.: Uncertainty visualization for secondary structures of proteins. In *2018 IEEE Pacific Visualization Symposium (PacificVis)* (2018), pp. 96–105. doi:10.1109/PacificVis.2018.00020. 7, 11, 16

[ST21] SEJDIU B. I., TIELEMAN D. P.: ProLint: A web-based framework for the automated data analysis and visualization of lipid–protein interactions. *Nucleic Acids Research 49*, W1 (2021), W544–W550. doi:10.1093/nar/gkab409. 11, 17, 19

[SUS*21] SABANDO M. V., ULBRICH P., SELZER M., BYŠKA J., MIČAN J., PONZONI I., SOTO A. J., GANUZA M. L., KOZLÍKOVÁ B.: ChemVA: Interactive visual analysis of chemical compound similarity in virtual screening. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 891–901. doi:10.1109/TVCG.2020.3030438. 11, 12

[TLR*01] THORPE M. F., LEI M., RADER A. J., JACOBS D. J., KUHN L. A.: Protein flexibility and dynamics using constraint theory. *Journal of Molecular Graphics and Modelling 19*, 1 (2001), 60–69. doi:10.1016/S1093-3263(00)00122-4. 11, 15

[VBJ*17] VAD V., BYŠKA J., JURCÍK A., VIOLA I., GRÖLLER E., HAUSER H., MARQUES S. M., DAMBORSKÝ J., KOZLÍKOVÁ B.: Watergate: Visual exploration of water trajectories in protein dynamics, 2017. doi:10.2312/VCBM.20171235. 11, 18

[VHG*18] VÁZQUEZ P., HERMOSILLA P., GUALLAR V., ESTRADA J., VINACUA A.: Visual analysis of protein-ligand interactions. *Computer Graphics Forum 37*, 3 (2018), 391–402. doi:10.1111/cgf.13428. 11, 16

[vLBI11] VAN DER ZWAN M., LUEKS W., BEKKER H., ISENBERG T.: Illustrative molecular visualization with continuous abstraction. *Computer Graphics Forum 30*, 3 (2011), 683–690. doi:10.1111/j.1467-8659.2011.01917.x. 6, 22

[WD23] WANG J. Y., DOUDNA J. A.: CRISPR technology: A decade of genome editing is only the beginning. *Science 379*, 6629 (2023), eadd8643. doi:10.1126/science.add8643. 23

[WEFC21] WINKLE M., EL-DALY S. M., FABBRI M., CALIN G. A.: Noncoding RNA therapeutics — challenges and potential solutions. *Nature Reviews Drug Discovery 20*, 8 (2021), 629–651. doi:10.1038/s41573-021-00219-z. 23

[Wei22] WEISKOPF D.: Uncertainty visualization: Concepts, methods, and applications in biological data visualization. *Frontiers in Bioinformatics 2* (2022). doi:10.3389/fbinf.2022.793819. 3, 9

[WH17] WAGNER A., HIMMEL H.-J.: aRMSD: A comprehensive tool for structural analysis. *Journal of Chemical Information and Modeling 57*, 3 (2017), 428–438. doi:10.1021/acs.jcim.6b00516. 11, 12

[WHLS19] WANG J., HAZARIKA S., LI C., SHEN H.-W.: Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics 25*, 9 (2019), 2853–2872. doi:10.1109/TVCG.2018.2853721. 3, 22, 23

[WPL96] WITTENBRINK C., PANG A., LODHA S.: Glyphs for visualizing uncertainty in vector fields. *IEEE Transactions on Visualization and Computer Graphics 2*, 3 (1996), 266–279. doi:10.1109/2945.537309. 9

[WSC21] WAN S., SINCLAIR R. C., COVENEY P. V.: Uncertainty quantification in classical molecular dynamics. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering sciences 379*, 2197 (2021), 20200082. doi:10.1098/rsta.2020.0082. 7, 8, 23

[XZF20] XIE J., ZHANG K., FRANK A. T.: PyShifts: A PyMOL plugin for chemical shift-based analysis of biomolecular ensembles. *Journal of Chemical Information and Modeling 60*, 3 (2020), 1073–1078. doi:10.1021/acs.jcim.9b01039. 11, 19

[ZKB95] ZACHMANN C.-D., KAST S. M., BRICKMANN J.: Quantification and visualization of molecular surface flexibility. *Journal of Molecular Graphics 13*, 2 (1995), 89–97. doi:10.1016/0263-7855(94)00015-K. 11, 15