Workshop on Molecular Graphics and Visual Analysis of Molecular Data (2018)
J. Byška, M. Krone, and B. Sommer (Editors)

VIA-MD: Visual Interactive Analysis of Molecular Dynamics

R. Skånberg[1,3], C. König[2], P. Norman[2,3], M. Linares[2,3], D. Jönsson[1,3], I. Hotz[1,3] and A. Ynnerman[1,3]

[1]Media and Information Technology, Linköping University, Sweden
[2]Division of Theoretical Chemistry and Biology, KTH Royal Institute of Technology, Sweden
[3]Swedish e-Science Research Centre (SeRC)

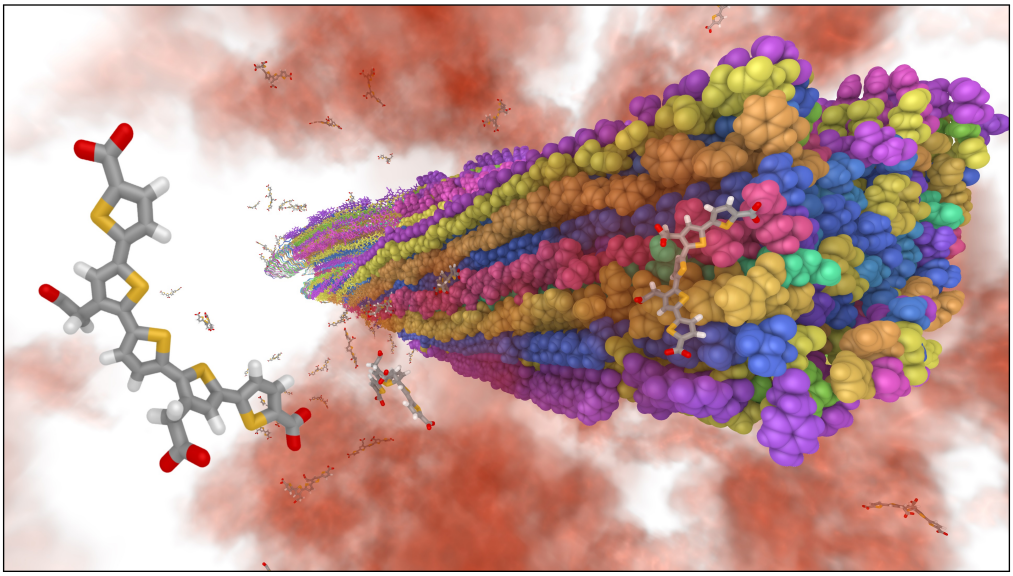


Figure 1: *The central twisted molecular structure is an amyloid fibril and the smaller molecules surrounding it are luminescent probes that bind to the amyloid fibril. The probe movement can be observed from the red density field, depicting the spatial distribution integrated over time. The rendering above is produced by the presented visual exploration environment, which enable query driven exploration of spatio-temporal molecular dynamics simulations.*

Abstract

We present a visual exploration environment tailored for large-scale spatio-temporal molecular dynamics simulation data. The environment is referred to as VIA-MD (visual interactive analysis of molecular dynamics) and has been developed in a participatory design process with domain experts on molecular dynamics simulations of complex molecular systems. A key feature of our approach is the support for linked interactive 3D exploration of geometry and statistical analysis using dynamic temporal windowing and animation. Based on semantic level descriptions and hierarchical aggregation of molecular properties we enable interactive filtering, which enables the user to effectively find spatial, temporal and statistical patterns. The VIA-MD environment provides an unprecedented tool for analysis of complex microscopic interactions hidden in large data volumes. We demonstrate the utility of the VIA-MD environment with four use cases. The first two deal with simulation of amyloid plaque associated with development of Alzheimer's, and we study an aqueous solution of 100 probes and an amyloid fibril. The identification of interaction "hotspots" is achieved with the use of combined filter parameters connected with probe molecular planarity and probe–fibril interaction energetics. The third and fourth examples show the wide applicability of the environment by applying it to analysis of molecular properties in material design.

DOI: 10.2312/molva.20181102

1. Introduction

Today, classical force field Molecular Dynamics (MD) simulations are routinely carried out in full atomistic detail for systems involving several million atoms over time scales ranging into the microsecond region and it has become an indispensable tool in nanotechnologies and life sciences alike [DFZ16]. MD simulation is arguably one of the most efficient ways to sample the phase-space of large-scale systems and it will gradually replace the use of more approximate coarse-grain approaches in studies of biochemical macromolecules. With this development comes a need to effectively analyze the vast amount of available data. The generated MD data consist of representations of constantly moving atoms and molecules, in the space of possible molecular arrangements and conformations, and not converging to an equilibrium state. In such a highly dynamic setting events of interest, like probe docking, are hidden in long time series and large molecules. Software tools like 'Visual Molecular Dynamics' (VMD) [HDS96], provide advanced rendering possibilities of the resulting molecular structures. However, little effort has been made to support an interactive exploration, which is fundamental for understanding the data.

In this paper we introduce a novel visual environment for exploration of the simulated molecules and their descriptive properties. The environment is built on a tight integration of statistics, plots and visualization of the molecules. This enables global overview as well as in-detail inspection of selected events and features. The environment includes spatial, temporal and property-distribution windows, which are linked together to facilitate effective data filtering. Selection of individual molecules and their corresponding trajectories is supported to enable close-up analysis of docking events. The molecular properties are derived from geometrical relations of their constituent atoms for each frame of the simulation. They can be aggregated into temporal and spatial summaries on demand according to the filtering state to find patterns within the data. A special strength of the system is its flexibility, as it is not restricted to one type of simulation but can easily be configured to different molecular properties of interest. This is achieved by introducing a hierarchical structure of groups and instances for the properties.

The environment has been realized in a participatory design process with the domain experts to meet their specific requirements. This tight integration between visualization and domain experts has resulted in an iterative development cycle where features and ideas have been discussed, implemented and tested in short cycles. A key principle has been to identify science query based challenges specifying the users' needs driving the development agenda. The requirements that resulted from this process are provided in the following section. Apart from enabling research in the MD domain and applications, the work has led to a number of concepts that are also of relevance for other visualization applications dealing with dynamical trajectory data. They can be summarized as:

- A flexible software environment supporting interactive query driven exploration of large and complex MD simulations.
- A semantic level description approach for flexible complex and high dimensional molecular properties.
- User defined hierarchical clustering and aggregation of simulated and derived molecular properties.
- Spatio-temporal filtering and scrubbing of parameter spaces using time windows linked to statistical and geometrical views.

In Section 4 we describe in detail how our Visual Analysis of Molecular Dynamics (VIA-MD) environment is designed around these contributions, and in Section 6 we show how the MD analysis workflow can be supported in a specific scenario.

2. MD data analysis requirements

Analysis of MD simulations is challenging from both a data size and parameter space point of view. Detecting the relevant changes in a sea of constantly moving atoms and molecules over thousands of time steps require analysis of a multitude of parameters in both the spatial and the parameter space domain. Moreover, the sheer size of the entire simulated trajectory makes it infeasible to compute all geometric properties on the fly. A pre-determined set of properties cannot be used since it limits the types of questions that can be answered.

Available MD software packages such as Gromacs [BvdSvD95, LHvdS, VDSLH*05], Amber [SFCW] and LAMMPS [PLI95] provide several data analysis and visualization functionalities. These functionalities include statistical analyses of trajectories and extraction of properties, such as distribution of bond lengths, angles or dihedral angles. Plots of time evolution of single molecular structure and energetics properties are readily available in these tools. However, these approaches have focused on either rendering of geometry or statistical analysis independently, this correlated analysis has so far not been possible.

One of the one main requirements on our environment is thus to enable generic interactive property filtering and provide visual illustrations of its correlation with phase-space distributions. This will have powerful consequences for the analysis of MD data, making it possible to pose queries that are completely new in nature, as exemplified by

1. Where in space do we find system A when its set of geometric properties $\{a_1, a_2, ...\}$ are confined to certain values or regions?
2. How are systems A and B oriented when their interaction energies V_{AB} are confined to certain values or regions?

Effective support to answer queries of this nature has been the guiding principle in identifying required features and developing new approaches as well as tailoring existing methods for integration into the VIA-MD environment.

3. Related Work

The VIA-MD environment combines a wide range of visualization techniques for integrated analysis of MD simulation data. Here, we first discuss similar applications and systems for visualization of MD data before we move on to a brief discussion on the specific visualization techniques used in VIA-MD.

Much of the previous work on visualization of MD simulation data has been tailored towards specific problems, or focused on developing novel techniques targeting specific types of datasets. Among these techniques we see exploration of Polymer-Solvent interaction [TWK*11], abstractions of solvents

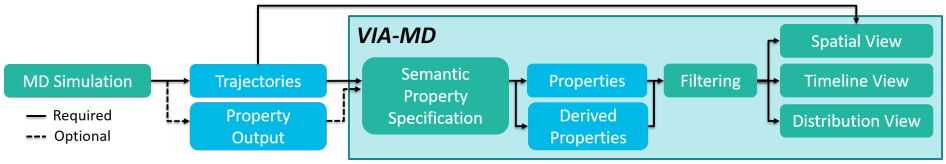


Figure 2: *Illustration of the VIA-MD data-flow, where MD simulation data is the basis for the visual environment. Query driven exploration is enabled by the property specification, in which the systems of interest in the simulation output, as well as derived properties, are specified. Spatio-temporal overview, filtering and navigation is tightly integrated to support interactive exploration of the MD simulation data.*

near protein cavities using path lines [BGB*08], space–time aggregation of biomolecules [EKK*14], variations in protein ensembles [HKOW14], temporal clustering of mixed lipid bilayers [TPRH11], visual exploration of water trajectories in proteins [VBJ*17], visual analysis of interaction forces in molecular dynamics [HEG*17] and fully developed interactive visualization environments for exploring protein tunnels [KSS*14] [FJB*17]. Though the specific applications in these papers vary a lot the concept of visual exploration of specific properties, often using linked views and some type of data aggregation, is common to most of the techniques. Similar concepts have also been the driving force behind the design of VIA-MD, exploiting them to build a generic tool for the exploring of a variety of user defined properties of Molecular Dynamics data.

In contrast to the specialized techniques, there also exists generic software with capabilities to render 3D illustrations of large-scale systems. These software packages include features such as interactive camera adjustments, multitude of residue representations (spheres, tubes, ribbons, etc.), identification and illustration of hydrogen bonds and much more. A selection of powerful tools that are particularly adept for MD simulation data include VMD [HDS96], MegaMol [GKM*15], Chimera [PGH*04] and NGL [RH15]. Most of these software packages focus on rendering the molecular data with rich set of graphical molecular representations and settings to choose from, however if any analysis of molecular properties is present, it is often implemented as ad-hoc plug-ins which only focus on particular properties and does not integrate well with other parts of the software. VIAMD aims to bridge the gap between the two groups (the specialized techniques and the generic rendering software).

4. The VIA-MD Environment

To meet the challenges presented in Section 2, we provide a system which combines spatial and temporal views of the data with dynamic statistics. We use a semantic description to allow the user to specify sets of molecular properties of interest. The property specification forms the basis for filtering, selection and statistics in the application. As such, the specification provides a flexible way for the user to adapt the visual environment to different use-cases and, as illustrated in Figure 2, can be seen as an abstraction layer between the MD simulation output and VIA-MD. Furthermore, a graphical user interface adapts to the specification by showing values and distributions of the specified property sets. The environment utilizes the following linked views for finding and analyzing events of interest in the entire molecular simulation (see Figure 3):

- **Spatial view:** shows the spatial geometry of the molecule(s) to reveal *where* events occur in relation to the molecular structure
- **Distribution view:** shows an overview of parameters over time to reveal the *distribution* and type of events occurring.
- **Timeline view:** shows the values of properties over time to reveal *when* events occur and for temporal navigation, zooming and filtering.

The combination of these linked views with the semantic parameter description form a powerful and flexible tool for answering the queries posed by domain experts. The following subsections provide a description of the VIA-MD components. Technical details are described later in Section 5.

4.1. Semantic Property Specification

The combination of properties of interest cannot in general be specified once and used for multiple scenarios. One scenario can for example include interaction between small molecular structures, often referred to as ligands, and proteins. Another scenario considers ligand–ligand interactions, which can aggregate into more complex structures. This involves different combinations of molecules and interactions affected by different molecular properties and parameters.

To handle all these configurations the system supports the specification of molecules and properties of interest using domain specific semantics, e.g. residue names or indices. The approach is sim-

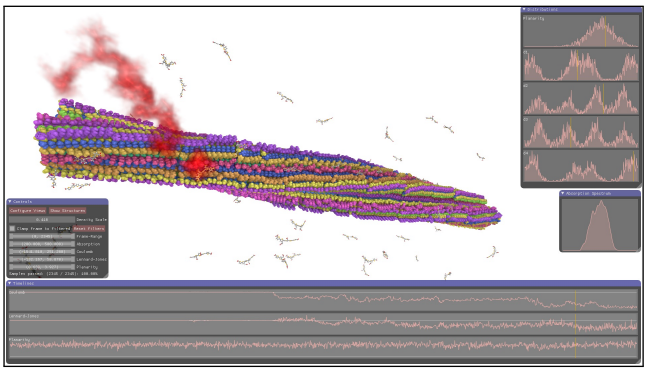


Figure 3: *Screenshot of the VIA-MD environment with the Amyloid-p-FTAA dataset. Spatio-temporal analysis is enabled by combining information from the spatial view (center), filtering (mid-left), timelines (bottom) and distributions (top right).*

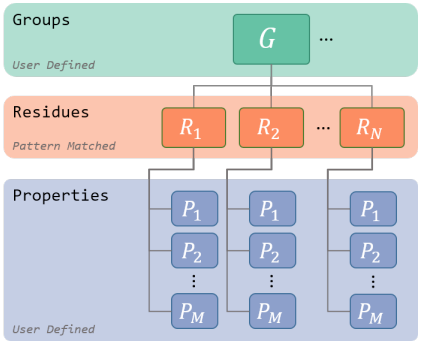


Figure 4: *Every use case require analysis of groups of specific residues and specific sets of properties. VIA-MD exposes a hierarchical model where the user specifies groups containing a pattern that matches residues within the dataset. The groups also contain recipes for computing properties for each of the matched residues. The figure illustrates the flexible hierarchical structure which is used by application to enable selection, filtering and aggregation.*

ilar to the approach used by Trellet et al. [TFBB16] but is more explicit as the user defines what properties should be computed and aggregated under what group. The core component is a hierarchical structure of groups, residues and properties. Groups are at top of the hierarchy and define collections of residues. Residues are sequences of atoms, where each residue can have a set of properties.Typically, properties of interest include bond length, angle between two atoms and dihedral angle. The semantic also supports definition of new derived properties, which were not generated by the MD simulation. The hierarchical relations between groups, residues and properties are depicted in Figure 4. It forms the core structure of the data managed by the application.

An example specification is included in Figure 5. Here, the first word (in green) is the identifier, which will be displayed in the application, the second word (in red) is the command to apply, and the following words (in black) are sent as arguments to the command. Arguments commonly include local indices of atoms in residues, or an external file containing a property. In this example, group G_A is created from residues with name A. Property d_1 is the dihedral angle computed from the specified atoms, given by their index lo-

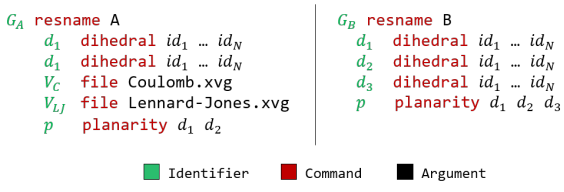


Figure 5: *An example of a property specification. Groups are selected from the residue together with a recipe to compute properties. The groups matched based on its criteria to the molecular structures within the simulation and for each match an instance of the group is generated and the properties are computed.*

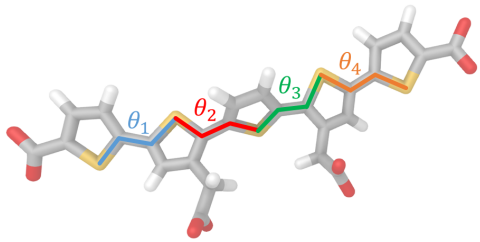


Figure 6: *An example of how planarity is computed from the four dihedral angles θ_i between the five aromatic Thiophene rings found in p-FTAA molecules.*

cation in the residue. Molecule and atom attributes necessary for computing derived properties are cached instead of the properties to reduce memory overhead.

Structural properties commonly included in the specification are interatomic and residue separation distances. An example of a more application specific property, related to molecular functionality, is the molecular planarity of π-conjugated systems shown in Figure 6. It consists of an oligomeric sequence of aromatic rings bonded together with carbon–carbon single bonds. For such a system, the molecular planarity P is defined as

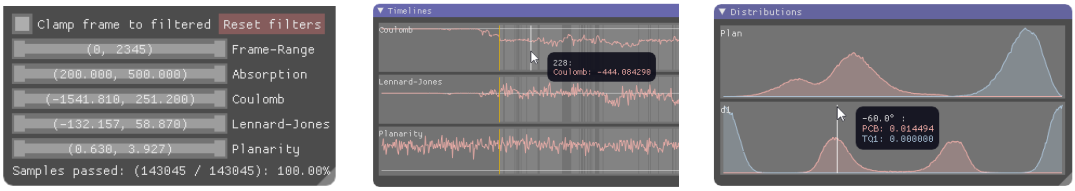
$$P = \sum_i \frac{||\theta_i| - 90|}{90}, \tag{1}$$

where θ_i is the dihedral angle formed between two adjacent aromatic rings, defined between -180° and 180°, and where the sum ranges over all such angles in the chromophore. Thus, for each bond between two aromatic rings, a value between 0 and 1 is added, where 0 indicates that the two rings are completely out of plane with each other and 1 indicates that the are completely planar. For this class of systems, the optical properties are very much governed by this planarity property [SMS*14], and to e.g. correlate planarity with spatial location can provide immediate information about chromophore–environment interactions. The support for derived properties in the semantic specification enables planarity to be included in the VIA-MD environment.

4.2. Interactive Filtering And Selection

The number of time steps in large-scale MD simulations can be overwhelming and, in combination with high-dimensional complex parameter spaces, it is difficult to find features of interest. Query 1 in Section 2 further stresses the importance of being able to find a system with specific properties.

Selecting a subset of the data through brushing can help reveal features hidden in large time, space and parameter domains. VIA-MD therefore supports frame filtering based on integrated properties and time window selection. Thus, the unique aspect of the filtering in VIA-MD is that it is based on the semantic specification (Section 4.1) and that filtering can be applied to each group in the specification. An example of filters exposed to the user is shown in Figure 7(a). Here, 'Clamp frame to filtered' limits the displayed



(a) (b) (c)

Figure 7: *(a) Example filters from the Amyloid–p-FTAA data set. Time (frame-range) is always present, due to the temporal nature of MD simulations, while the others are generated based on the parameter specification. (b) Automatically generated timelines for Coulomb, Lennard-Jones and planarity parameters. The timelines provide both temporal overview and navigation in the visual environment. (c) Displaying the distribution of the two specified groups in the Interface data set. One group is shown in red while the other is shown in blue.*

molecules to the subset of frames within the simulation that fulfill the conditions set by the filters. The number of samples that passes the current filter conditions is shown in the bottom. Throughout the article, the set of remaining frames after filtering will be referred to as the *filtered-frames*.

4.3. Atom Trajectory Overview

Analysis of the location of atoms and molecules provides essential information, such as how close a molecule is to the host system, or to which regions molecules are confined to for certain parameter settings.

Trajectories can be displayed by for example showing each individual path, or in an aggregated representation showing the spatial distribution of paths [FKRE10]. Together with the domain scientists, we have chosen to show the spatial distribution of locations integrated over time since the details of each individual path is generally not important and therefore clutter the view. The spatial distribution is represented by a density volume, see red areas in Figure 1. The density volume is computed based on the filtered frame set, meaning that it can provide both overview and details of atom trajectories depending on the current filters set by the user.

4.4. Temporal overview and navigation

The dynamic nature of MD simulations requires means for temporal navigation. Correlating time and property values are essential to discovering and understanding when certain events occur.

VIA-MD provides interactive time-lines of each property. An example is shown in Figure 7(b), with property values on the vertical axis and the simulation time on horizontal axis. Each time-line displays the value over time (red), the current time (orange vertical bar), and which temporal regions that are currently included in the filtered-frames (highlighted in light grey). The horizontal white line provides a reference for the zero-value of the property. Hovering shows the property values and clicking sets the current frame.

4.5. Distribution plots

Probabilities of conformations for specific filter settings are in many cases more important than the individual frames in which they occur. Distributions of properties allow ensembles of ligands or probes to be analyzed and their collective behavior to be studied.

To address this challenge, we supply a customizable distribution view, see Figure 7(c). It displays property-value histograms for each specified group, cf. Section 4.1. The distributions are computed from the filtered frame-set for each group individually. Each distribution is assigned a unique transparent color, making it possible to compare overlapping group distributions. Selection, by picking, of a structure in the spatial view displays the distribution of the selected structure. Aggregated distributions of each group are shown by default, i.e. when no specific structure is selected.

5. Implementation

VIA-MD has been developed using C++ and OpenGL to enable low-level access to the native capabilities of the target platforms, Windows, Linux and Mac OS. The bulk of the implementation of VIA-MD has been done in Inviwo [Inv], an open-source visualization framework designed for rapid prototyping released under BSD-license. This has allowed us to leverage existing state-of-the-art volumetric rendering techniques, one of the core rendering techniques used in VIA-MD. The Graphical User Interface (GUI) has been developed using Dear ImGui [Cor], an open-source immediate mode GUI for C++, released under MIT-License.

To test the limits in terms of performance of the application, a dataset of a p-FTAA probe solvated in water was simulated over the course of one Million frames. This resulted in 59 GB of compressed trajectory data and has served as a worst case test scenario.

5.1. Filtering

Properties are internally represented using arrays. They have the same length as the number of frames contained within the simulation, they do not change. Filtering operations can therefore use indices into these arrays. A single property filter operation returns the indices to values passing the filter requirements. The filtering result of all properties for residue i, $P_{i1} \dots P_{iN}$, are combined into one residue-specific index set R_i using set intersection: $R_i = P_{i1} \cap P_{i2} \cap P_{i3} \dots \cap P_{iN}$.

5.2. Atom Spatial Distribution Volume Generation

The system supports generation of density volumes for all user-defined groups. The computation uses the residues in the group passing the filter criteria, all if filtering is disabled. The volume

is constructed by binning the atomic positions of the trajectories integrated over time. This operation can be computationally expensive. In the worst case it requires binning each atomic position of each residue of each group in every frame of the simulation, which can take seconds for large datasets. The binning operation can be parallelized using OpenGL and atomic operations, but they are not supported in OpenGL 4.1, which is currently the latest version supported by the target platform Mac OS. Thus, we use a CPU-based approach and address the performance issue by applying the following steps. First, the binning operations are parallelized and performed in background threads to avoid stalling the main GUI thread.This improves interactivity but can still take a long time to finish. Therefore we add a second step, inspired by Monte-Carlo sampling and progressive estimate refinement. The sorted indices of the atomic trajectories R_i are first randomly permuted to get a more uniform distribution of indices U_i. This operation could kill the benefits of potential cache locality from close indices, but since R_i may be sparse good cache locality in never guaranteed. By utilizing the permuted set of uniformly distributed indices U_i, the density can be estimated by processing chunks of U_i over multiple frames. The result is a temporally smooth density volume, which is refined over time. The first frames provide a good estimate which over time converges to the true density.

Before display, the density volume is normalized based on the maximum value found in the volume. This normalization greatly simplifies the mapping from density to optical values required in the volume rendering, as a linear mapping between the densities and opacity can be applied with a single global scaling parameter. Exponential mapping was also tested but the added complexity of having to manipulate both scale and falloff did not pay off in practice.

5.3. Timelines and Histograms

As the number of frames within a simulation is typically much larger than the amount of pixels available, we show the mean property value of all frames projected into the same pixel. This gives a accurate and stable representation of the data, especially when zooming or resizing the window. This is computationally more expensive than picking one representative sample per pixel, but even with the dataset containing 10^6 frames and 5 separate timelines, this has not been an issue.

The histograms used to visualize the distributions are computed for a preset number of bins. The user can optionally change the number of bins. The fixed number of bins enables A to B comparisons between groups in the same plot.

5.4. Rendering

The rendering can be separated into three stages. In the first stage, the molecules are rendered using a user specified representation and color mapping. Currently we support three common visual representations: Space-Fill (van der Waals), Licorice and Ribbons. Space-Fill or van der Waals [Ric77] represents atoms as spheres with a radii that is determined by the element of the atom. Licorice visualizes covalent bonds explicitly as tubes connecting the atoms. Ribbons provide a more abstract view displaying the core structure or 'backbone' of the protein as a ribbon. For further details we refer to [Ric81]. To accelerate the rendering of geometric primitives such as spheres and tubes, impostor based rendering techniques are utilized as described by Grottel et al. [GRE09].

In the second stage, the density volume is rendered using a GPU-based direct volume raycaster, which uniformly samples the density volume from the direction of the viewpoint. Each voxel is considered as an emissive light source with a given color and opacity. For comprehensive overview of GPU-based volume rendering and transfer functions design we refer to the articles [BHP,LKG*]. The density value given in each voxel of the volume, see Section 4.3, is mapped to opacity such that a high density value yields a high opacity while the emissive part is fixed. To improve the performance of the volume rendering, which is computationally expensive, the opaque geometry are rendered first and used for early ray termination during raycasting.

As a third and last rendering stage, the GUI is rendered with all of its windows.

6. Use Cases

In this section, we examine four different types of use cases and the insights gained using the VIA-MD environment. All datasets used have been obtained using the Gromacs software [BvdSvD95, LHvdS, VDSLH*05] and touch upon different applications from Amyloid detection (Case 1), bio-templating of organic molecules (Case 2), organic solar cell (Case 3) and self-assembly of molecules (Case 4). We discuss the first case in detail, the second case is an extension of the first case with different types of probes. The last two cases do not exploit the full capacity of the system and are mainly used to illustrate the flexibility of the VIA-MD environment and its usefulness to different applications.

6.1. Amyloid-p-FTAA

Alzheimer's is one of the major diseases of our time and even though the effects of the disease are well understood, the underlying cause is still unknown. What is known as a hallmark indicator of Alzheimer's is that there is an unusual build up of Amyloid Plaque within the brain of the affected. Therefore, researchers have developed optical markers in the form of small luminescent molecules or probes that target the main component of the plaque, the Amyloid Fibrils, which consist of misfolded Beta-Amyloid proteins, and bind to these. As the optical markers bind to the fibrils, they seem to favor certain conformations, i.e. spatial arrangements, which determine the absorption and emission spectrum of the markers [SLS*]. The conformation is mainly defined by the dihedral angles between the inner structures of the molecules. In this use case, we study these angles over the course of the simulation for a group of identical molecules which provide an aggregated view on their typical behavior.

The amyloid fibril structure is based on the PDB entry 5OQV determined by cryo-EM [GSS*].The simulation extent is $63 \times 22 \times 22$ nm. 100 p-FTAA molecules have been placed randomly in the beginning of the simulation. Water molecule ions and Na+ ions have been added to charge-equilibrate the system. The simulation took

about 24 days of single-core computation time to finish. In total, the system consists of 2,983,792 atoms; 158,631 atoms for the protein, 5,100 atoms for the 100 p-FTAA molecules, 1,159 Na+ ions, and 2,818,902 atoms for the water. The dataset was simulated over a time span of 23.44 ns from which 2,344 snapshot frames containing only the protein and the p-FTAA molecules have been stored on disk, leaving a footprint of about 1.4 GB after compression.

The structure of the p-FTAA probe is depicted in Figure 1 and 6 and consists of five thiophene rings, each of them composed of five atoms: four carbons (grey) and one sulfur atom (yellow). The thiophene groups rotate with respect to each other. The angle between two planes defined by neighboring aromatic rings is known as a dihedral angles—it is this set of dihedral angles that is at focus in our study (d_i). The planarity (P) parameter defined in Eq. (1) can range from 0.0 up to 4.0 for a completely planar structure. In addition to these geometrics parameters, we also focus on the Coulombic interaction (V_C) and the Lennard-Jones interaction (V_{LJ}) between individual p-FTAA molecules and the amyloid fibril, as well as the absorption spectrum of individual p-FTAA molecules ($\sigma(\lambda)$).

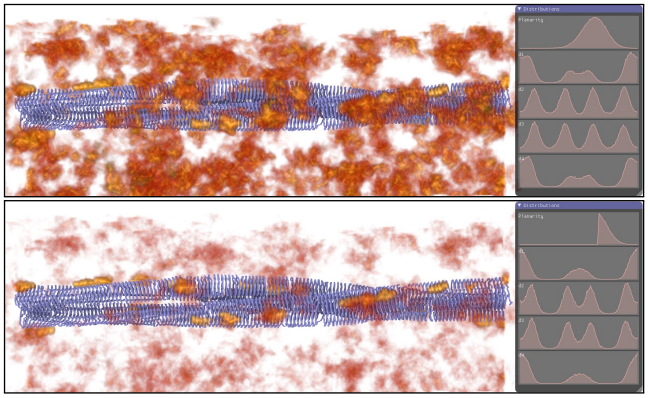


Figure 8: *Depicting the amyloid fibril, represented with ribbons, and the occurrences of probe atoms, represented using the density volume (red to yellow). No filters have been applied in the top figure, while the planarity range of interest has been identified from the distributions view and filtered in the bottom figure. The interesting spatial locations are revealed when observing the movement density field of the filtered occurrences.*

In a first analysis step the aim is to locate regions of high planarity for the p-FTAA molecules within the Amyloid–p-FTAA data set. Figure 8 represents the amyloid fibril by ribbons to avoid occlusion while still providing a structural context to the density volume. In the top view we see the trajectories of all atoms over the extent of the entire simulation. In the bottom view a high-pass filter has been applied to the planarity parameter and a cut-off value has been set at 2.6, which roughly corresponds to the peak of the planarity distribution (see top distribution in Figure 8). The resulting density volume shows a series of "hotspots" along the surface of the amyloid fibril from which we can deduce that the p-FTAA molecules have a *tendency* to exhibit a higher planarity close to the amyloid fibril. The density volume shows other traces of occurrences of high planarity further away from the amyloid fibril, which is expected

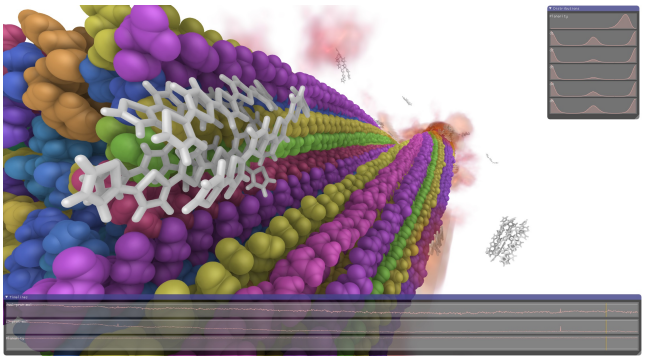


Figure 9: *Probe molecules consisting of 6T aggregating on the surface of the amyloid fibril structure. The density volume (Red) in the background highlights regions of strong Lennard-Jones interaction between the probes and the amyloid fibril.*

as these conformations can occur sporadically when the molecule is solvated in water. Moreover, we are able to identify a very stable binding site [KSH*18] where p-FTAA is locked in an all-trans conformation with a Coulombic binding energy of 1200 kJ/mol due to the interactions between the anionic carboxyl groups of the probe and the Lys 16 of the Amyloid fibril. Upon binding, the conformationally restricted probes show a pronounced increase in molecular planarity. This is in line with the observed changes in luminescence properties that serve as the foundation for their use as biomarkers.

6.2. Amyloid-6T

In the second data set, we study the interaction between the same amyloid fibril as in the Amyloid-p-FTAA data set with sexithiophene (6T). 6T is a chain of six thiophene moieties but in contrast to the p-FTAA molecule it does not carry any COO- groups. Therefore, we can anticipate that it will interact differently with the amyloid fibril surface. Experimentally, 6T molecules tend to aggregate in solution but the amyloid act as a dispersive agent and the molecules will tend to form smaller aggregate when in contact with the amyloid surface [BWW*14]. The VIA-MD environment allow to us to identify the specific region of the amyloid where the 6T molecules adsorb, i.e. the hydrophilic surface formed by Gly 9 and Tyr 10, as well as in the groove formed by Tyr 10, Val 12 and Hys 14. Lennard-Jones interactions are mainly responsible for the adsorption of 6T, and is accompanied by a more planar structure as illustrated in Figure 9.

6.3. Material Interface

This third data set [VL17] presents the interface between two materials, a donor polymer (TQ1) and a acceptor fullerene derivative (PCBM), used widely in the field of organic solar cell. Even though this particular dynamic is very short (100 frames) it contains many instances of the same type of molecules. This exemplifies the strength of aggregation as individual may not provide sufficient data to draw conclusions over such a short simulation. But by aggregating properties in groups, one can clearly see the patterns and tendencies of the system. VIA-MD allows to treat differ-

ent groups of molecules, aggregates properties and plot them separately as shown in Figure 10.

6.4. BTA

The fourth data set is a MD simulation of benzene-1,3,5-tricarboxamide (BTA) molecules. Those molecules have been widely used for self-assembly [DRP*11] and for ferroelectric materials [GMU*17]. Their assembly properties are governed by the conformation of the amide bond with respect to the benzene core which we can easily follow by looking at the distribution of dihedral angles. In Figure 11, we illustrate the aggregated distributions for the three dihedral angles, that correlates to the helicity of the stack.

7. Discussion & Conclusions

In this paper we present an environment for visual exploration of large scale MD simulation data. While originally motivated from a specific application its concepts are applicable to a much wider set of MD simulations. A key component is the query driven workflow interlinking visual data exploration and statistical tools. This enables intuitive property exploration and filtering in both space and time. Significant effort has been spent on efficient aggregation of data into meaningful representations. Interactivity and high quality rendering have been central throughout the design of the system.

A significant technical contribution integrated into the VIA-MD environment is the flexible interactive filtering of multiple structural, energetic, and spectral properties. Common analysis of MD simulation requires the use of multiple separate tools which make an interactive exploration unfeasible. Our implementation of this idea is open-ended in the sense that it does not set any restrictions on the imposed property correlations, e.g. in terms of type or number of properties. This current version of VIA-MD supplies an initial set of parameters motived by the showcased applications and will be expanded with further applications and a growing user group. A challenge arising along with this flexibility is, however, ensuring that the system is intuitive and easy to use, an aspect that will play an increasingly important role in the future development of VIA-MD.

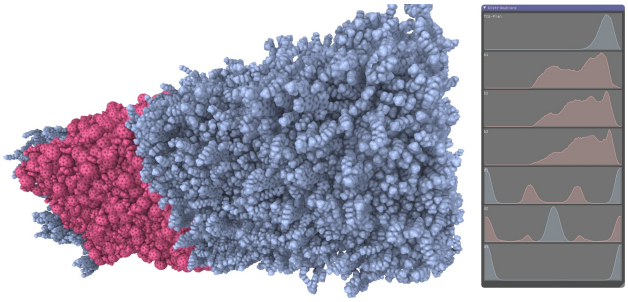


Figure 10: *Interface between TQ1 (in blue) and PCBM (pink). The statistics of the two groups are plotted in matching colors in the distribution view. In this example aggregation of properties greatly help when studying the overall characteristics of the system.*

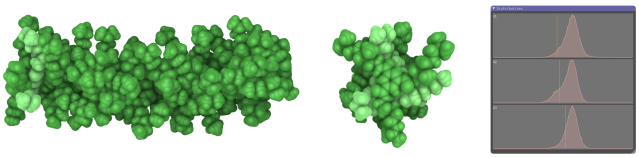


Figure 11: *The BTA dataset consisting of a stack of BTA Molecules where the second molecule in the stack has been selected (Light Green). From left to right: The stack as seen from the side, view down the axis of the stack and the distributions of dihedral angles d_1, d_2 and d_3 of the selected molecule.*

The use cases have demonstrated the unique usefulness of VIA-MD for the investigation of correlation of advanced molecular properties such as planarity. A future perspective is to employ such data directly into VIA-MD which would open many new options for analysis of molecular properties, e.g. correlating spectral response properties with molecular dynamics.

To conclude, the utility of the VIA-MD environment was demonstrated using examples of large-scale MD simulations of amyloid fibril protein with molecular probes, providing insights into molecular mechanisms involved in the development of neurodegenerative diseases. The insights gained were made possible by the flexible filtering and combined spatio-temporal visualization components. We also showed how MD simulations in material science can be analyzed using the same concepts and environment.

It is clear that the sum of all components in the VIA-MD environment made it possible to answer the types of questions posed in Section 2. Based on the success of the current work, we see potential usability of the VIA-MD environment in several other fields, as applications are numerous and range from the field of drug design, molecular probe and sensor technology, DNA technology, organic electronics, etc.

The source code and binaries for VIA-MD are freely available at https://github.com/scanberg/viamd.

Acknowledgements

The authors acknowledges funding by a Marie Skłodoswka–Curie International Fellowship “FreezeAlz” by the European Commission, the Swedish Research Council (Grant No. 621- 2014-4646) and SeRC (Swedish e-Science Research Center) for funding and providing a research environment enabling innovative research across e-science disciplines. The Swedish National Infrastructure for Computing (SNIC) at National Supercomputer Centre (NSC) are acknowledged for providing computer resources. The prototype of this project has been implemented using the visualization framework Inviwo [Inv] and immediate mode GUI Dear ImGui [Cor].

References

[BGB*08] BIDMON K., GROTTEL S., BÖS F., PLEISS J., ERTL T.: Visual abstractions of solvent pathlines near protein cavities. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 935–942. 3

[BHP] BEYER J., HADWIGER M., PFISTER H.: State-of-the-art in gpu-based large-scale volume visualization. *Computer Graphics Forum 34*, 8. 6

[BvdSvD95] BERENDSEN H. J., VAN DER SPOEL D., VAN DRUNEN R.: Gromacs: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications 91*, 1-3 (1995), 43–56. 2, 6

[BWW*14] BÄCKLUND F. G., WIGENIUS J., WESTERLUND F., INGANÄS O., SOLIN N.: Amyloid fibrils as dispersing agents for oligothiophenes: control of photophysical properties through nanoscale templating and flow induced fibril alignment. *Journal of Materials Chemistry C 2*, 37 (2014), 7811–7822. 7

[Cor] CORNUT O.: Dear imgui. URL: https://github.com/ocornut/imgui. 5, 8

[DFZ16] DUAN L. L., FENG G. Q., ZHANG Q. G.: Large-scale molecular dynamics simulation: Effect of polarization on thrombin-ligand binding energy. *Scientific reports 6* (2016), 31488. 2

[DRP*11] DANILA I., RIOBEÌĄ F., PIRON F., PUIGMARTÍ-LUIS J., WALLIS J. D., LINARES M., ÅGREN H., BELJONNE D., AMABILINO D. B., AVARVARI N.: Hierarchical chiral expression from the nano-to mesoscale in synthetic supramolecular helical fibers of a nonamphiphilic c 3-symmetrical π-functional molecule. *Journal of the American Chemical Society 133*, 21 (2011), 8344–8353. 8

[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. 3

[FJB*17] FURMANOVÁ K., JAREŠOVÁ M., BYŠKA J., JURČÍK A., PARULEK J., HAUSER H., KOZLÍKOVÁ B.: Interactive exploration of ligand transportation through protein tunnels. *BMC bioinformatics 18*, 2 (2017), 22. 3

[FKRE10] FALK M., KLANN M., REUSS M., ERTL T.: 3d visualization of concentrations from stochastic agent-based signal transduction simulations. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2010), pp. 1301–1304. 5

[GKM*15] GROTTEL S., KRONE M., MULLER C., REINA G., ERTL T.: Megamol – a prototyping framework for particle-based visualization. *Visualization and Computer Graphics, IEEE Transactions on 21*, 2 (Feb 2015), 201–214. 3

[GMU*17] GORBUNOV A., MENG X., URBANAVICIUTE I., PUTZEYS T., WÜBBENHORST M., SIJBESMA R., KEMERINK M.: Polarization loss in the organic ferroelectric trialkylbenzene-1, 3, 5-tricarboxamide (bta). *Physical Chemistry Chemical Physics 19*, 4 (2017), 3192–3200. 8

[GRE09] GROTTEL S., REINA G., ERTL T.: Optimized data transfer for time-dependent, gpu-based glyphs. In *Visualization Symposium, 2009. PacificVis' 09. IEEE Pacific* (2009), IEEE, pp. 65–72. 6

[GSS*] GREMER L., SCHÖLZEL D., SCHENK C., REINARTZ E., LABAHN J., RAVELLI R. B. G., TUSCHE M., LOPEZ-IGLESIAS C., HOYER W., HEISE H., WILLBOLD D., SCHRÖDER G. F.: Fibril structure of amyloid-β(1-42) by cryo-electron microscopy. *Science (80-.).*, 6359, 116–119. 6

[HDS96] HUMPHREY W., DALKE A., SCHULTEN K.: VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics 14* (1996), 33–38. 2, 3

[HEG*17] HERMOSILLA P., ESTRADA J., GUALLAR V., ROPINSKI T., VINACUA A., VÁZQUEZ P.-P.: Physics-based visual characterization of molecular interaction forces. *IEEE transactions on visualization and computer graphics 23*, 1 (2017), 731–740. 3

[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. 3

[Inv] Inviwo: Interactive visualization workshop. The Inviwo Foundation. URL: http://www.inviwo.com. 5, 8

[KSH*18] KÖNIG C., SKANBERG R., HOTZ I., YNNERMAN A., NORMAN P., LINARES M.: Binding sites for luminescent amyloid biomarkers from non-biased molecular dynamics simulations. *Chemical Communications* (2018). 7

[KSS*14] KOZLIKOVA B., SEBESTOVA E., SUSTR V., BREZOVSKY J., STRNAD O., DANIEL L., BEDNAR D., PAVELKA A., MANAK M., BEZDEKA M., ET AL.: Caver analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics 30*, 18 (2014), 2684–2685. 3

[LHvdS] LINDAHL E., HESS B., VAN DER SPOEL D.:. 2, 6

[LKG*] LJUNG P., KRÃIJGER J., GROLLER E., HADWIGER M., HANSEN C. D., YNNERMAN A.: State of the art in transfer functions for direct volume rendering. *Computer Graphics Forum 35*, 3. 6

[PGH*04] PETTERSEN E. F., GODDARD T. D., HUANG C. C., COUCH G. S., GREENBLATT D. M., MENG E. C., FERRIN T. E.: Ucsf chimeraâĂŤa visualization system for exploratory research and analysis. *Journal of computational chemistry 25*, 13 (2004), 1605–1612. 3

[PLI95] Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics 117*, 1 (1995), 1 – 19. 2

[RH15] ROSE A. S., HILDEBRAND P. W.: Ngl viewer: a web application for molecular visualization. *Nucleic acids research 43*, W1 (2015), W576–W579. 3

[Ric77] RICHARDS F. M.: Areas, volumes, packing, and protein structure. *Annual review of biophysics and bioengineering 6*, 1 (1977), 151–176. 6

[Ric81] RICHARDSON J. S.: The anatomy and taxonomy of protein structure. *Advances in protein chemistry 34* (1981), 167–339. 6

[SFCW] SALOMON-FERRER R., CASE D. A., WALKER R. C.: An overview of the amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science 3*, 2, 198–210. 2

[SLS*] SHIRANI H., LINARES M., SIGURDSON C. J., LINDGREN M., NORMAN P., NILSSON K. P. R.: A palette of fluorescent thiophene-based ligands for the identification of protein aggregates. *Chemistry - A European Journal 21*, 43. 6

[SMS*14] SJÖQVIST J., MARIA J., SIMON R. A., LINARES M., NORMAN P., NILSSON K. P. R., LINDGREN M.: Toward a molecular understanding of the detection of amyloid proteins with flexible conjugated oligothiophenes. *J. Phys. Chem. A 118* (2014), 9820–7. 4

[TFBB16] TRELLET M., FEREY N., BAADEN M., BOURDOT P.: Interactive visual analytics of molecular data in immersive environments via a semantic definition of the content and the context. In *Immersive Analytics (IA), 2016 Workshop on* (2016), IEEE, pp. 48–53. 4

[TPRH11] TURKAY C., PARULEK J., REUTER N., HAUSER H.: Interactive visual analysis of temporal cluster structures. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 711–720. 3

[TWK*11] THOMASS B., WALTER J., KRONE M., HASSE H., ERTL T.: Interactive exploration of polymer-solvent interactions. In *VMV* (2011), pp. 301–308. 2

[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. In *Proceedings of Eurographics Workshop on Visual Computing for Biology and Medicine (EG VCBM)* (2017), pp. 33–42. 3

[VDSLH*05] VAN DER SPOEL D., LINDAHL E., HESS B., GROENHOF G., MARK A. E., BERENDSEN H. J.: Gromacs: fast, flexible, and free. *Journal of computational chemistry 26*, 16 (2005), 1701–1718. 2, 6

[VL17] VOLPI R., LINARES M.: Study of the cold charge transfer state separation at the tq1/pc71bm interface. *Journal of computational chemistry 38*, 14 (2017), 1039–1048. 7