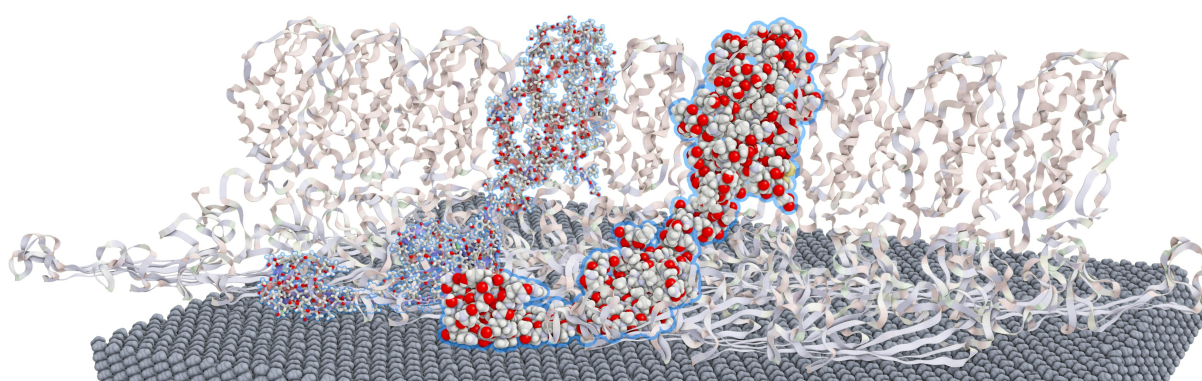# MolFind – Integrated Multi-Selection Schemes for Complex Molecular Structures

Robin Skånberg[1,2], Mathieu Linares[1,2], Martin Falk[1,2], Ingrid Hotz[1,2], and Anders Ynnerman[1,2]

[1]Scientific Visualization Group, Linköpings University, Sweden
[2]Swedish e-Science Research Centre (SeRC)

Figure 1: *Growing an initial selection of two residues along covalent bonds in a protein fibril.*

Abstract
Selecting components and observing changes of properties and configurations over time is an important step in the analysis of molecular dynamics (MD) data. In this paper, we present a selection tool combining text-based queries with spatial selection and filtering. Morphological operations facilitate refinement of the selection by growth operators, e.g. across covalent bonds. The combination of different selection paradigms enables flexible and intuitive analysis on different levels of detail and visual depiction of molecular events. Immediate visual feedback during interactions ensures a smooth exploration of the data. We demonstrate the utility of our selection framework by analyzing temporal changes in the secondary structure of poly-alanine and the binding of aspirin to phospholipase A2.

1 Introduction

"It is through the interactive manipulation of a visual interface – the analytic discourse – that knowledge is constructed, tested, refined and shared." [PSCO09]

This is also true for the analysis of molecular dynamics (MD) data and the generation of expressive visualization. Detailed study of microscopic properties based on molecular simulations, e.g. the binding between ligands and molecules or the assembly of macromolecular structures, is an essential part in nanotechnologies and life sciences. Size and complexity of the data require interactive exploration methods to fully leverage the content of the data. This includes easy-to-use methods for selection that cope with dense visual representations containing hidden atoms and structures and dynamically changing features only visible within a few time steps.

In this paper, we present a selection framework for molecular substructures embedded in the visual analytics tool VIA-MD [SLK*18, KSH*18]. The work is mostly targeted toward expert users from the molecular simulation domain. The user-driven design of the framework combines a large set of selection paradigms in a flexible way while maintaining a simple and intuitive interface with immediate visual feedback. This includes text-based queries, temporal and property-based filtering, diverse brushing functionality, morphological and structure-based growth operations combined using Boolean operators. A typical use case is the selection for an analysis based on properties and structures which are tracked over time. In contrast to the typical property summaries for the entire molecule this allows for a much more detailed analysis. An example is the calculation of root-mean-square deviation of atomic positions (RMSD) for a defined selection or the distance between

the center of mass of a selection and a specific residue. Another important use case is the specification of visual representation styles to highlight events for various representation and communication purposes.

Our main contributions are the assembly of a flexible and easy to use selection toolbox seamlessly integrating

- different selection types: text-based queries, spatial selections, and attribute-based selections;
- selection operators for context-based hierarchical growing or shrinking of the selections;
- immediate visual feedback through selection previews;

into an exploration framework supporting storing and loading of selections for dynamic property calculation and the generation of expressive visual representations.

2 Related Work

Most common Molecular Visualization tools provide some selection and interaction possibilities. The text-based query exposed in VMD [HDS96] serves mainly to apply representations to parts of the dataset. Spatial selection is limited to picking individual atoms. Jmol [Her06] and NGL [HR15] offers a text-based selection language similar to VMD but with a different keyword syntax. MolQL [RSB*18] offers a comprehensive declarative text-based query language that exposes convertions from aforementioned languages. Caver analyst [POB*06] exposes a similar text-based search function that allows the user to match text input against residues and atoms and append these to stored selections. It also exposes functionality to select regions in screen space using a rectangle, circle, or lasso tool. Chimera [PGH*04] presents virtually the same functionality as VMD and Caver analyst, but exposes the keywords as options in a drop-down field which is specific to the dataset. Chimera also offers operators to grow and shrink a selection, called *broaden* and *narrow*, respectively. The operation seems to be context-dependent based on which representation is used. For example, if an atom, shown in a van der Waals (VDW) representation, is selected, then the selection will grow to encompass all connected atoms, i.e. the entire molecule. But if a segment within a secondary structure is selected, e.g. a Ribbon in cartoon representation, the result will only grow to encompass just that structure. Furthermore, selections can be expanded radially and stored along with a label for later retrieval Driven by a more specific application Duran et al. [DHR*19], e.g., provide spatial selection of cavities or pockets by means of a sphere, whose position and radius can be adjusted interactively.

3 Selection

In this section we describe the components of the user interface for selections in MD data. This includes the supported types of selection, operators modifying the selection, and lastly the visual representation of a preview of the current selection. Once the user is satisfied with the active selection, it can be stored and labeled. Stored selections can be used in other parts of the application, for example to define properties such as Root Mean Squared Deviation (RMSD) and Center of Mass (COM) or to adjust the visual representations of the selected atoms. While text-based queries can be
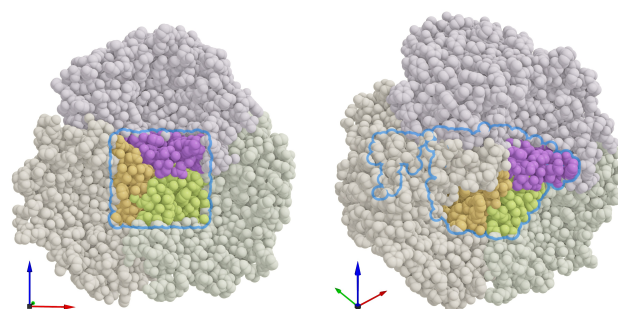
Figure 2: *Rectangular selection in screen space of residues in maltoporin, PDB-ID: 1AF6, (left). When rotated, the internal structure of the selection becomes apparent (right). Saturated colors indicate visible structures of the selection while occluded structures are outlined by a blue halo.*

applied temporally [HDS96], spatial selection is limited to a single time step as it requires a fixed reference location. In principle, spatio-temporal selection is feasible [DHR*19], however, it necessitates a re-centering of the dataset to fixate locations of structures of interest.

Type of Selection – Generally we can distinguish (i) text-based property queries, (ii) direct spatial selection by interaction, (iii) filtering in spatial and temporal plots. For all selection types there are hierarchical selection modes which define the atomic entities of the selection. These are either *atom*, *residue*, or *chain*. This enables the user to automatically select higher-ranking components matching the query and/or selection. For example when querying a calcium atom with selection level *atom* only matching atoms are selected while choosing the *residue* level would result in selecting all residues containing calcium atoms.

(i) Query-based selection: This enables selection based on keywords such as `resname`, `label`, `element`, `chain`, etc., in conjunction with logical constraints. Query-based selection is an efficient tool to select specific structures and is especially useful for an experienced user or domain scientist. For examples the query

    resname ALA and not label CA

selects all atoms within residues named `ALA` while excluding atoms labeled `CA`. Likewise,

    backbone and within 10 of residue 100

will select all atoms of the backbone of the protein within 10 units of residue 100.

(ii) Interactive spatial selection: Spatial selections provide a means to refine the selection with respect to spatial constraints, to include or exclude regions of the data. Our application support three different types of spatial interactions: picking individual atomic items, brushing regions on the surface with atomic brushes and spatial selection penetrating the entire molecule. This is done by dragging the mouse to form rectangular regions in screen space. These regions are then extruded into the scene either including or excluding all structures which fall inside (Fig. 2).

(iii) Filter based selection: In addition to textual selection our
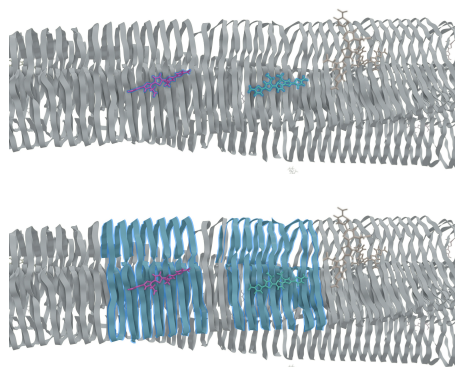
Figure 3: *Radial growth operator. The initial selection of two p-FTAA molecules (blue and purple) is grown by a radius of 3 units under a hierarchical selection mode of chains (data: amyloid fibril structure [KSH*18] based on PDB-ID: 5OQV).*

Figure 4: *Growth by covalent bonds. The active selection is depicted with a blue outline while a preview of the result is shown with an orange outline (data: β-sheet of an alanine-rich polypeptide).*

tool provides selections based on property plots, similar to the idea of brushing and linking. A typical example is the use of the Ramachandran plot [Ram63] for structure-based selections. The Ramachandran plot is spanned by the dihedral angles of the backbone in protein structure. It allows to distinguish e.g. β-sheet and α-helix structures. Every amino acid in the backbone chain is represented as a point in the plot. Selections are temporally persistent and can be tracked over time to analyze changes of key properties (cf. Fig. 5).

Selection Operators – In order to manipulate an active selection, we provide Boolean operations for adding and subtracting subsets as well as growth operators acting on covalent bonds or radially. The Boolean operations apply to all types of selection, i.e. queries as well as spatial selection.

A selection can be expanded radially by including atoms within a specified radius from the atoms within the active selection. Fig. 3 depicts an initial selection of two luminescent probes on top of a β-sheet (top). Growing the selection radially with the hierarchical selection level set to *chain* results in adding close-by β-sheets (bottom).

Growing by covalent bonds expands the selection by adding atoms which share a covalent bond with the active selection. The user can interactively adjust the number of edges taken into account while the updated selection is highlighted. In Fig. 4, the initial selection of four alanine amino acids (blue outline) is expanded along covalent bonds.

Visual Representation – The current selection is visualized with fully saturated colors and the extent of the selected structure is outlined by a silhouette (Fig. 2). The silhouette also includes occluded structures for maintaining spatial context. Any structure not being part of the selection is desaturated in order to direct the user's attention toward the selected regions.

We distinguish between the current, active selection and a selection in progress. This enables us to provide interactive feedback
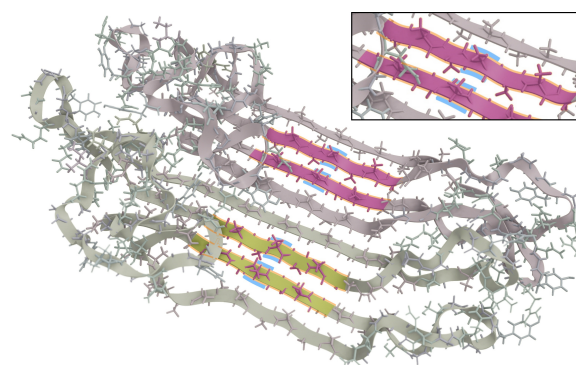
when modifying a selection, e.g. while applying a growth operation (cf. Fig. 4).

4 Use Cases

In the following we describe two examples how the selection tool can be used for data exploration and visualization.

4.1 Poly-alanine secondary structure

Amino acid chains might undergo changes in their secondary structure passing from α-helix to β-sheet during the MD simulation. The Ramachandran plot is widely used to track those changes. However, the Ramachandran plot becomes extremely cluttered when visualized for the entire protein hiding selected areas of interest of the protein. To illustrate this case, we follow the dynamic of an alanine-rich poly-peptide in a β-sheet structure on a hydrophilic surface. The user can perform a selection by clicking and dragging either on the molecule or in the Ramachandran plot and the selection will be highlighted accordingly. In the beginning of the dynamic, all the selected alanine residues are found in the β-sheet region (Fig. 5 left). The position of the selected amino acids can also be tracked in the Ramachandran plot. Due to hydrogen bonds forming with the surface, the β-sheet weakens during the dynamic and a transition to the α-helix region is observed (Fig. 5 right).

4.2 Aspirin

As a second case study, we use a molecular dynamics trajectory of the specific binding of aspirin to phospholipase A2. In this binding mode, the carboxylate group of the aspirin is bound to a calcium atom. The starting geometry was obtained from the crystal structure (1OXR) [SEJ*05] and the MD simulation has been performed with the gromos54a7 force field [SEC*11] using the Gromacs software [AMS*15]. During the trajectory the aspirin molecule is pulled out from the pocket. For a first visual inspection of the trajectory, the user displays the protein with Ribbons while representing the aspirin molecule and calcium atom with licorice and van der Waals representations, respectively. But to study in greater details the interaction between the aspirin and the amino acids composing the
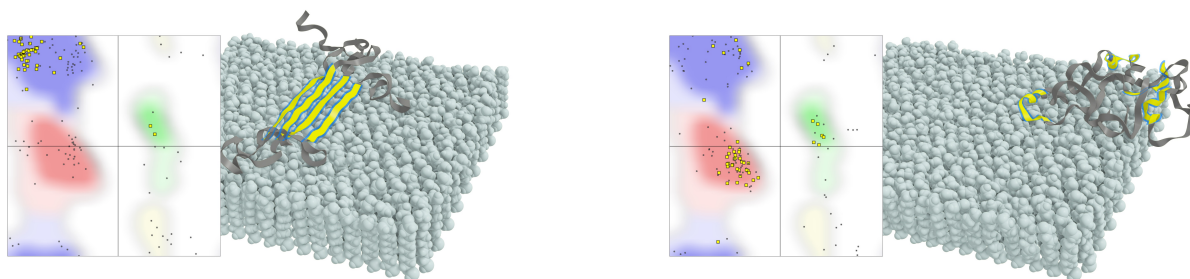
Figure 5: *Following the conformational change of an alanine-rich polypeptide on a hydrophilic surface. At the beginning of the dynamic (left) the selected alanine region is organized in a beta sheet as highlighted in the spatial visualization and the Ramachandran plot (inset). At the end of the dynamic (right) the system has undergone a conformational change toward α-helices. In the Ramachandran plot, the β-sheet region is shown as blue and the α-helix region is shown as red.*

(a)   (b)   (c)

(d)
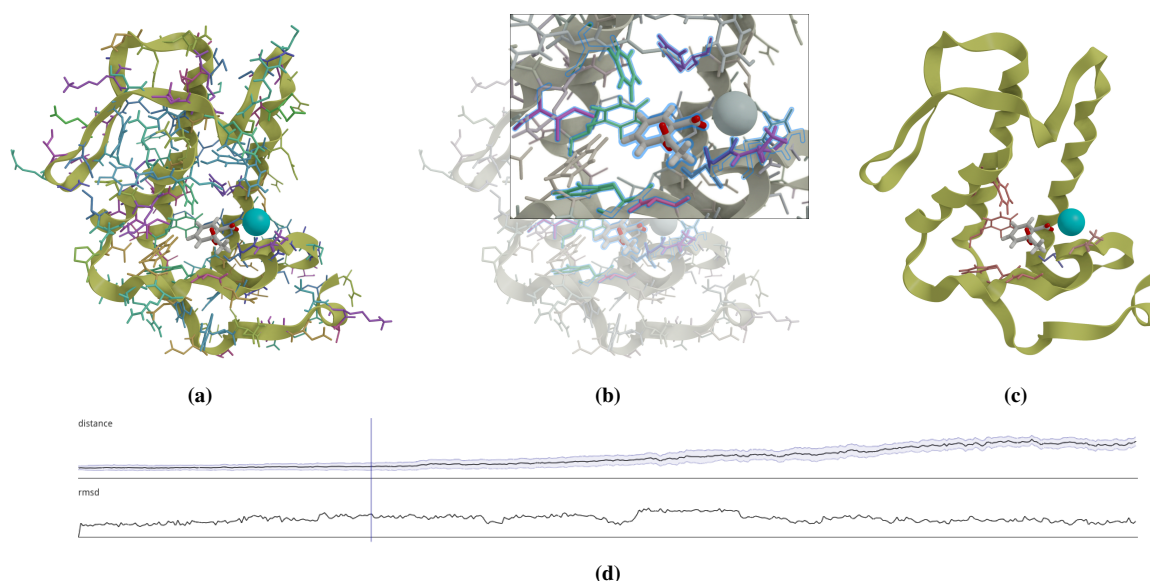
Figure 6: *Workflow showing the filtering process for the aspirin dataset. The backbone of the protein is shown in ribbons and its residues in licorice, the aspirin is shown in licorice with CPK coloring and the calcium is shown as a cyan VDW sphere. (a) Initial unfiltered representation where the aspirin is residing in the pocket. (b) radial growth operator applied to aspirin which selects the neighborhood. (c) filtered representation showing only structures in the vicinity of the aspirin. (d) Distance from the aspirin molecule to the selection (pocket) and RMSD of the pocket.*

pocket, there is a need to display their atomic representation. By doing so for the full protein, the view becomes cluttered (Fig. 6a). With a radial growth operation around the aspirin residue, we can select the amino acids in close interaction with the aspirin (Fig. 6b). This selection can then be used to either filter the representation (Fig. 6c) or to calculate and observe the change in the RMSD of the pocket while the aspirin is leaving it (Fig. 6d).

4.3 Conclusion

We propose in the present contribution a simple and intuitive implementation combining text-based query and spatial selection for molecular dynamic simulation. The initial combination can be expanded via growth operator either by covalent bonds or radial distance. The selection can then be stored and used later to filter representations or to calculate properties such as RMSD. As per-

spective, adding transparency could help in seeing internal parts of a selection which may be occluded depending on the current representations, but mouse interaction and transparency poses many challenges and pitfalls. Another potential improvement would be to render a continuous change in the representation between the selected and not selected part resulting in smooth transitions [vdZLBI11]. As future work, we propose to explore other types of growth operators by allowing growth over for example hydrogen bonds.

References

[AMS*15] ABRAHAM M. J., MURTOLA T., SCHULZ R., PÁLL S., SMITH J. C., HESS B., LINDAHL E.: GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1-2* (2015), 19–25. `doi:10.1016/j.softx.2015.06.001`. 3

[DHR*19] DURAN D., HERMOSILLA P., ROPINSKI T., KOZLÍKOVÁ B., VINACUA Á., VÁZQUEZ P.-P.: Visualization of large molecular trajectories. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (Jan 2019), 987–996. `doi:10.1109/TVCG.2018.2864851`. 2

[HDS96] HUMPHREY W., DALKE A., SCHULTEN K.: VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics 14* (1996), 33–38. `doi:10.1016/0263-7855(96)00018-5`. 2

[Her06] HERRAEZ A.: Biomolecules in the computer: Jmol to the rescue. *Biochemistry and Molecular Biology Education 34*, 4 (2006), 255–261. 2

[HR15] HILDEBRAND P. W., ROSE A. S.: NGL Viewer: a web application for molecular visualization. *Nucleic Acids Research 43*, W1 (04 2015), W576–W579. `doi:10.1093/nar/gkv402`. 2

[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). `doi:10.1039/c8cc00105g`. 1, 3

[PGH*04] PETTERSEN E. F., GODDARD T. D., HUANG C. C., COUCH G. S., GREENBLATT D. M., MENG E. C., FERRIN T. E.: UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of Computational Chemistry 25*, 13 (Oct 2004), 1605–1612. `doi:10.1002/jcc.20084`. 2

[POB*06] PETŘEK M., OTYEPKA M., BANAS P., KOŠINOVÁ P., KOCA J., DAMBORSKY J.: CAVER: A new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics 7*, 1 (2006), 316. `doi:10.1186/1471-2105-7-316`. 2

[PSCO09] PIKE W. A., STASKO J., CHANG R., O'CONNELL T. A.: The science of interaction. *Information Visualization* (2009). `doi:10.1057/ivs.2009.22`. 1

[Ram63] RAMACHANDRAN G. N.: Stereochemistry of polypeptide chain configurations. *J. Mol. Biol. 7* (1963), 95–99. 3

[RSB*18] ROSE A. S., SEHNAL D., BLIVEN S., BURLEY S. K., VELANKAR S.: Molql: Towards a common general purpose molecular query language. *Biophysical Journal 114*, 3 (2018), 342a. 2

[SEC*11] SCHMID N., EICHENBERGER A. P., CHOUTKO A., RINIKER S., WINGER M., MARK A. E., VAN GUNSTEREN W. F.: Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal 40*, 7 (Apr 2011), 843. `doi:10.1007/s00249-011-0700-9`. 3

[SEJ*05] SINGH R. K., ETHAYATHULLA A., JABEEN T., SHARMA S., KAUR P., SINGH T. P.: Aspirin induces its anti-inflammatory effects through its specific binding to phospholipase A2: Crystal structure of the complex formed between phospholipase A2 and aspirin at 1.9 Å resolution. *Journal of Drug Targeting 13*, 2 (2005), 113–119. `doi:10.1080/10611860400024078`. 3

[SLK*18] SKÅNBERG R., LINARES M., KÖNIG C., NORMAN P., JÖNSSON D., HOTZ I., YNNERMAN A.: VIA-MD: Visual interactive analysis of molecular dynamics. In *Workshop on Molecular Graphics and Visual Analysis of Molecular Data* (2018), The Eurographics Association. `doi:10.2312/molva.20181102`. 1

[vdZLBI11] 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`. 4