

# visualization of large molecular trajectories

---

D. Duran-Rossich, P. Hermosilla, T. Ropinski, B.  
Kozlikova, À. Vinacua, P. Vázquez

*VIS 2018*

[pere.pau@cs.upc.edu](mailto:pere.pau@cs.upc.edu)

*IEEE*

# motivation

## Molecular Simulations

- Use computers to simulate whether a small molecule (ligand) can bind to a larger biomolecule (the protein)
  - Activate or inhibit a certain biomolecule function
  - Provide a therapeutic benefit for the patient
- Great utility for drug design
- Two typical approaches: Monte Carlo and Molecular Dynamic methods
- Result is a trajectory
  - Atoms' positions
  - Energy of the system (aggregation of energies)

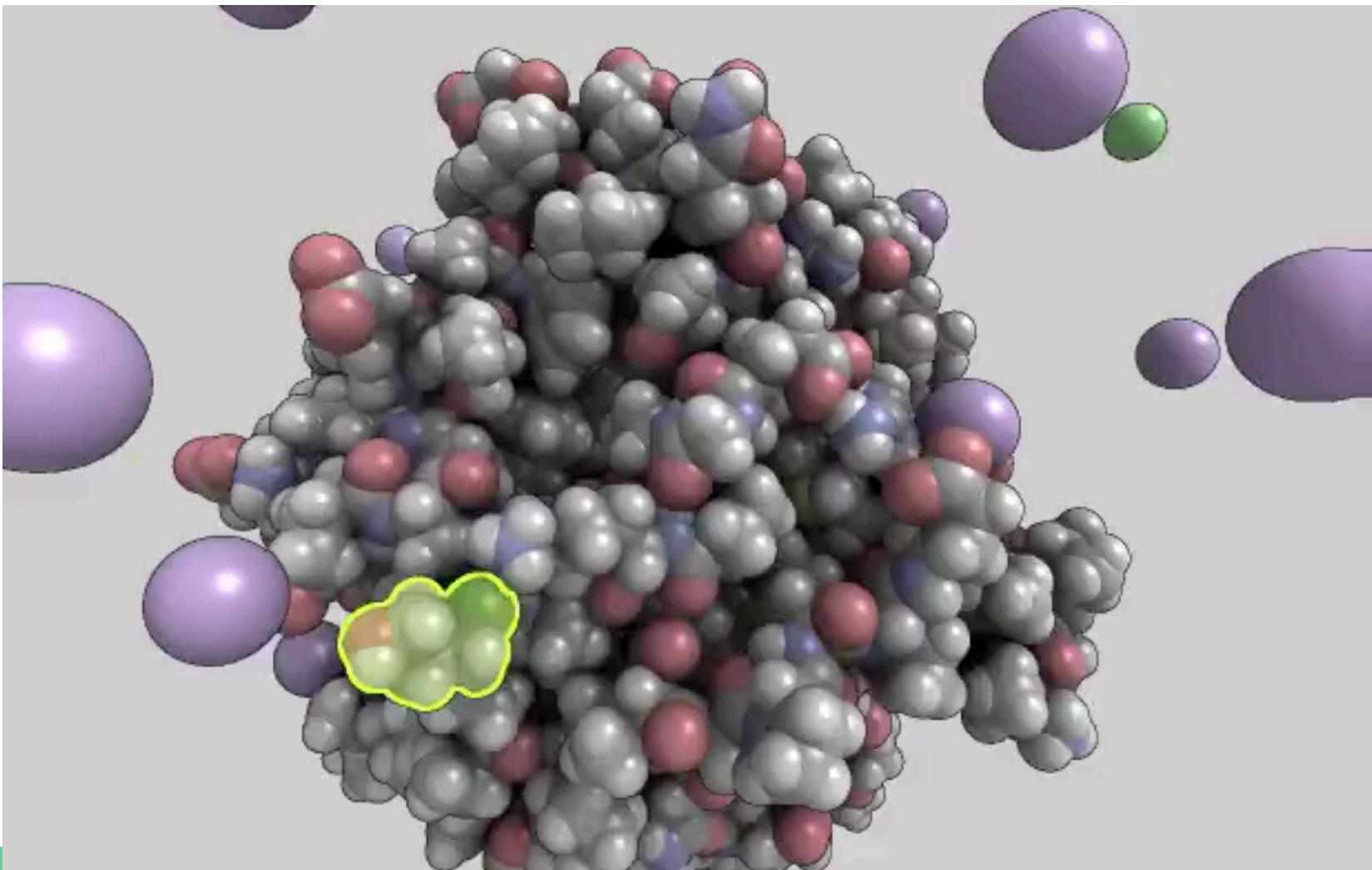
# motivation

## Simulation energy

- Strength of the binding (aka binding affinity)
- Can be approximated as the sum of the interaction energies between the protein residues and ligand
- Low level conformations are more stable
- Three main types (additive):
  - Van der Waals (short range)
  - Electrostatic energy in the vacuum (long-range)
  - Screening of that electrostatic energy due to the solvent

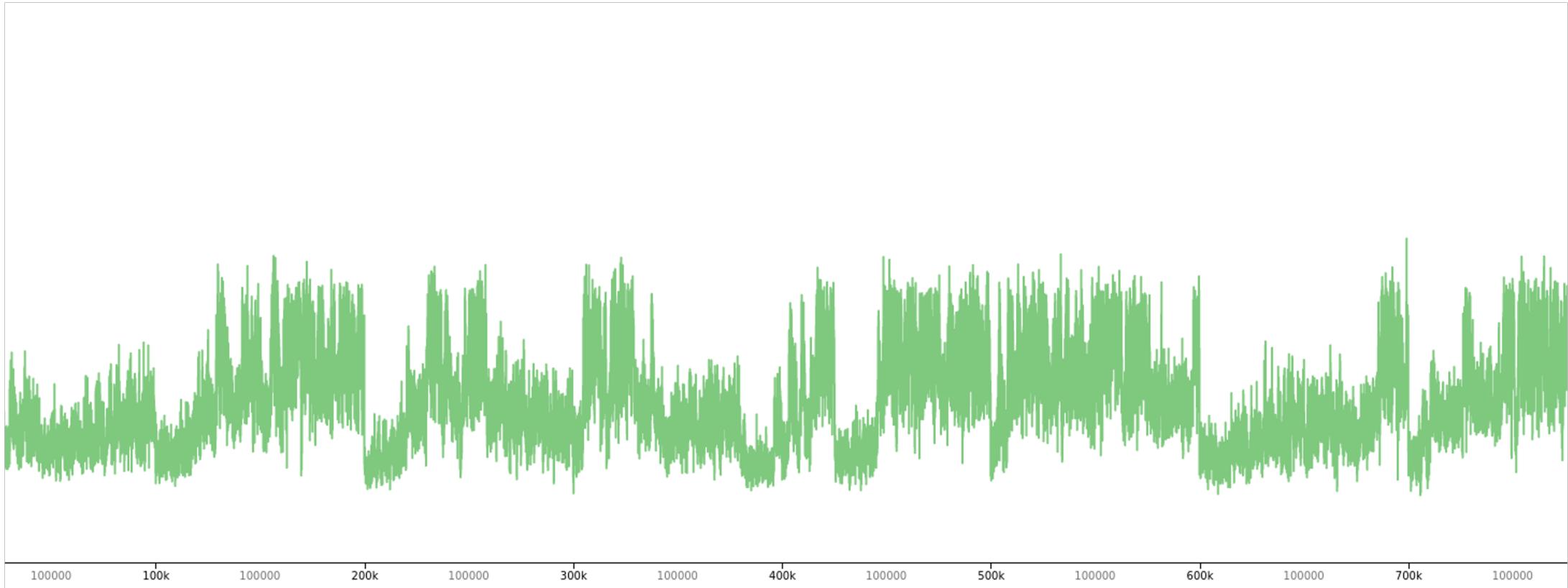
# motivation

Protein-ligand interactions may generate extremely large trajectories (esp. in MD)  
Large simulations difficult to explore interactively



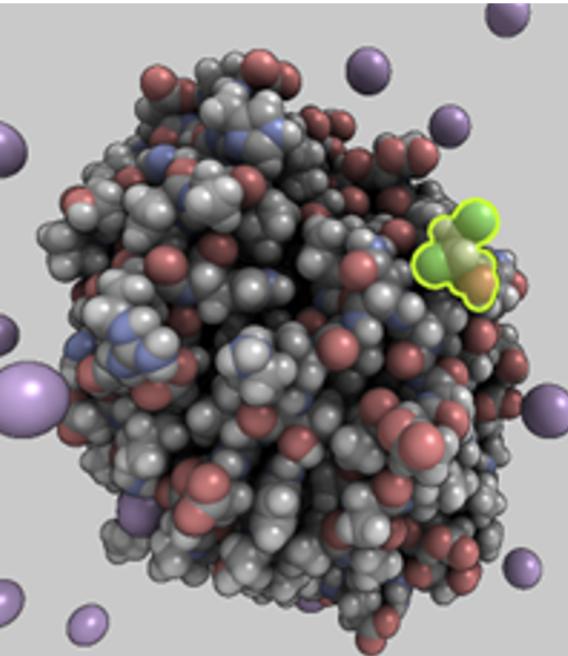
# motivation

Protein-ligand interactions may generate extremely large trajectories (esp. in MD)  
Large simulations difficult to explore interactively

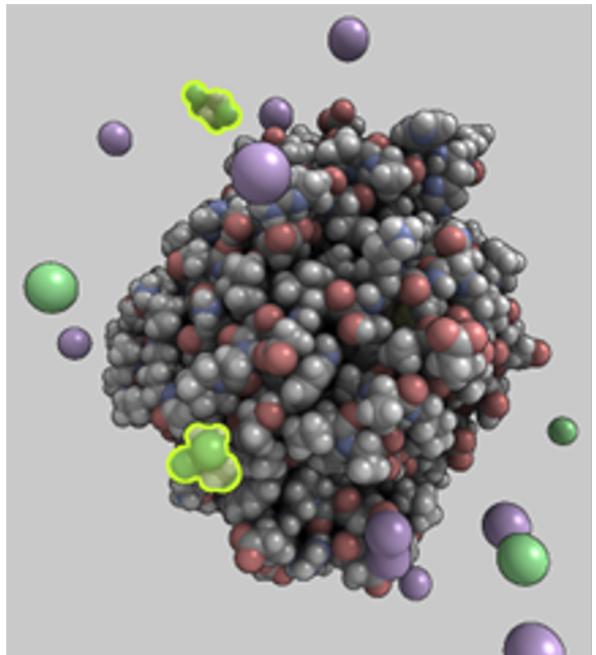


# challenges

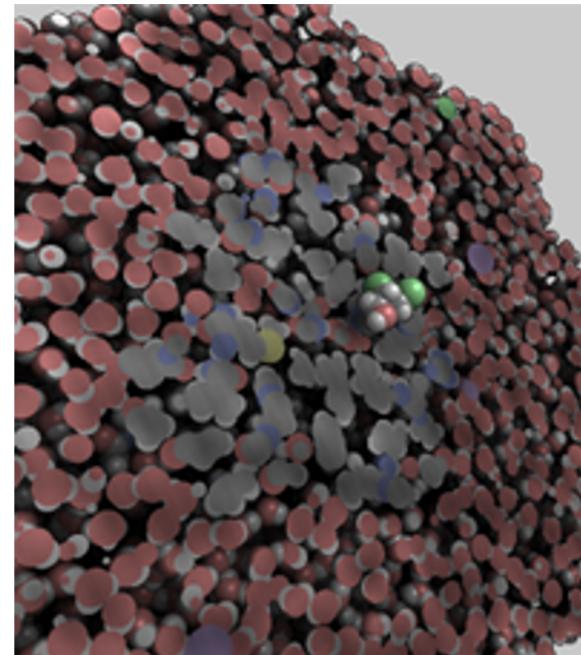
Large amounts of data



- One ligand
- 800K frames
- 45 GB



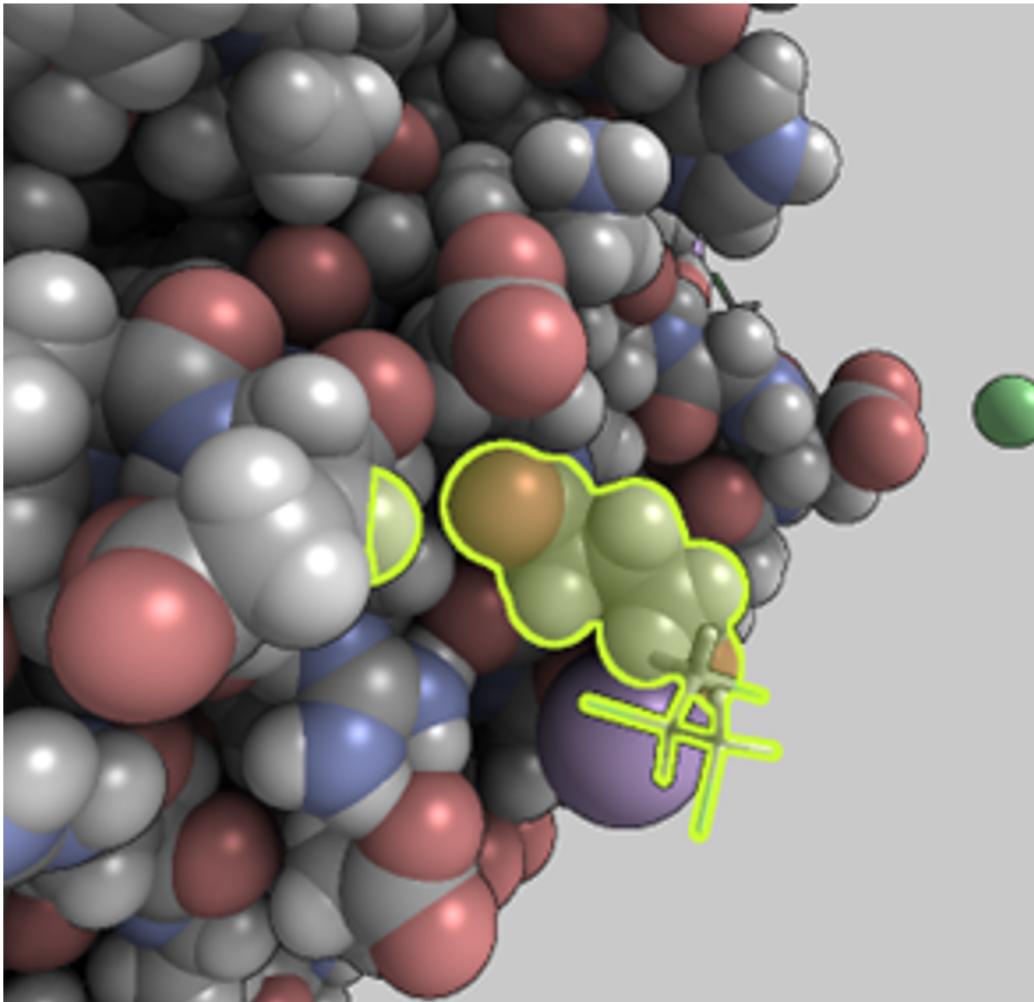
- Three ligands
- 50K frames
- 2.8 GB



- One ligand
- 100K frames
- 41.6 GB

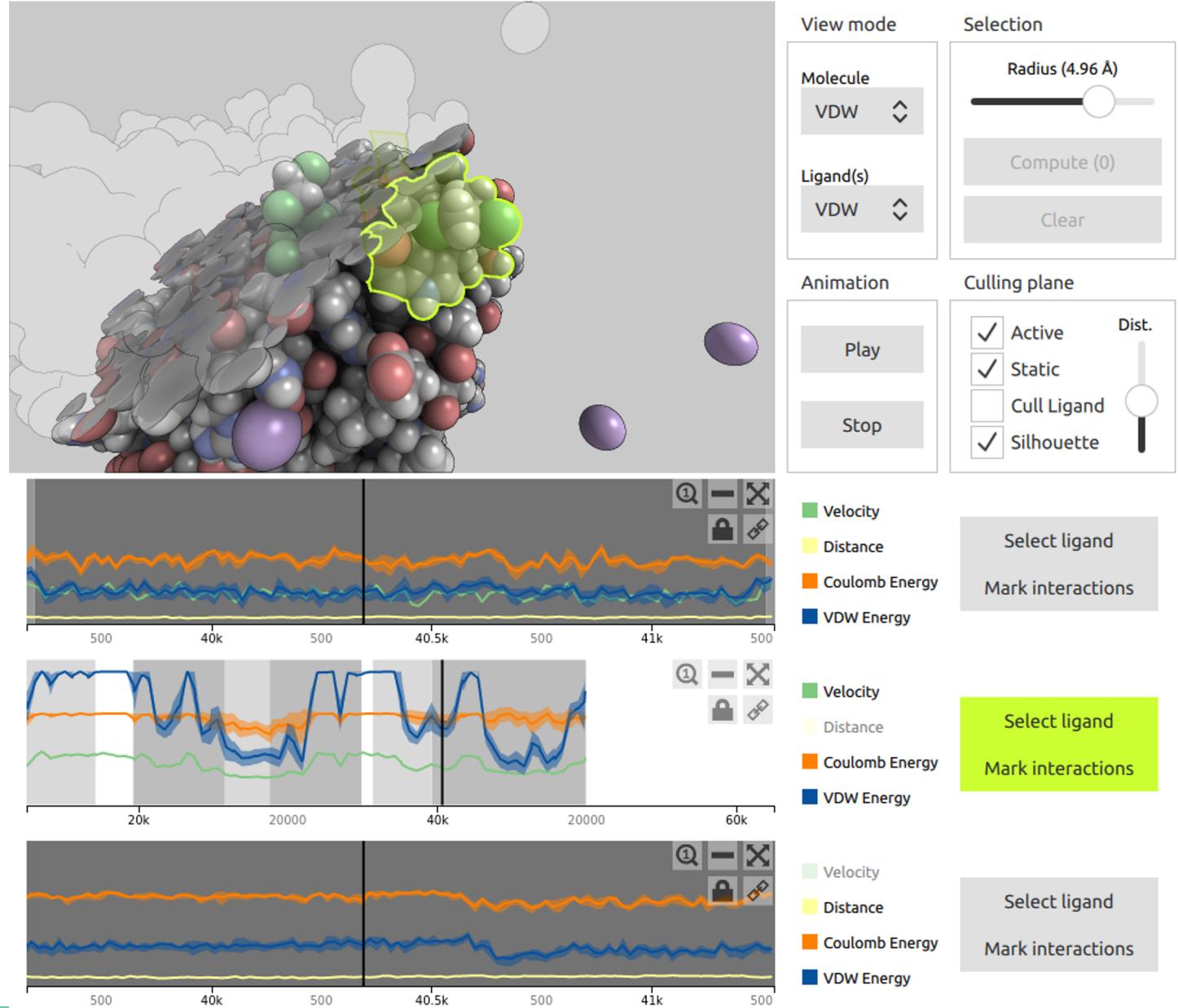
# challenges

Illustrate ligand-protein interactions



# challenges

## Exploratory visualization



## related work

- 3D rendering of large molecules or cells [Grott et al. 2015, Hermosilla et al., 2015, Le Muzic et al. 2016]
- Exploration of MS trajectories
  - Lindow et al. 2012, Voronoi diagrams for cavity exploration and time evolution
  - Byska et al. 2015, Byska et al., 2016 analysis of protein tunnels' properties over time
  - Furmanova et al. 2017, ligand behavior through tunnels
  - Hermosilla et al. 2017, ligands interactions with proteins over time
  - Vázquez et al. 2018, whole simulation exploration with ligands' interactions
    - **Do not address the problem of extremely large charts/trajectories**
- Visualization of large charts
  - Data aggregation, progressive exploration...
    - **Focus on pattern finding or abstract representations**
    - **MD energy data essentially consists of arrays of floating points**

# outline

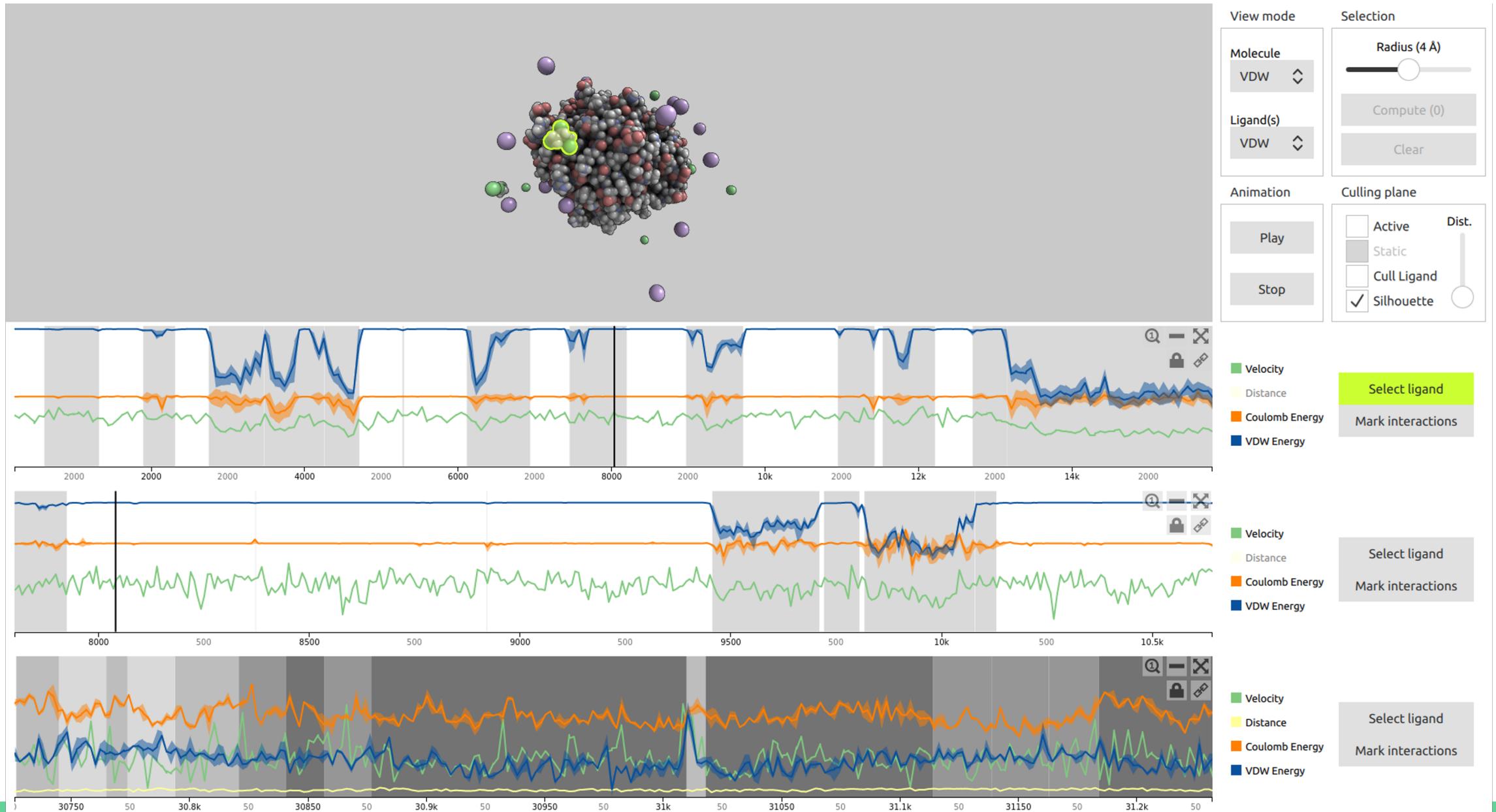
- System Overview
- Hierarchical trajectory exploration
- Coordinated views
- Results & Evaluation
- Conclusions & Future Work

# system overview

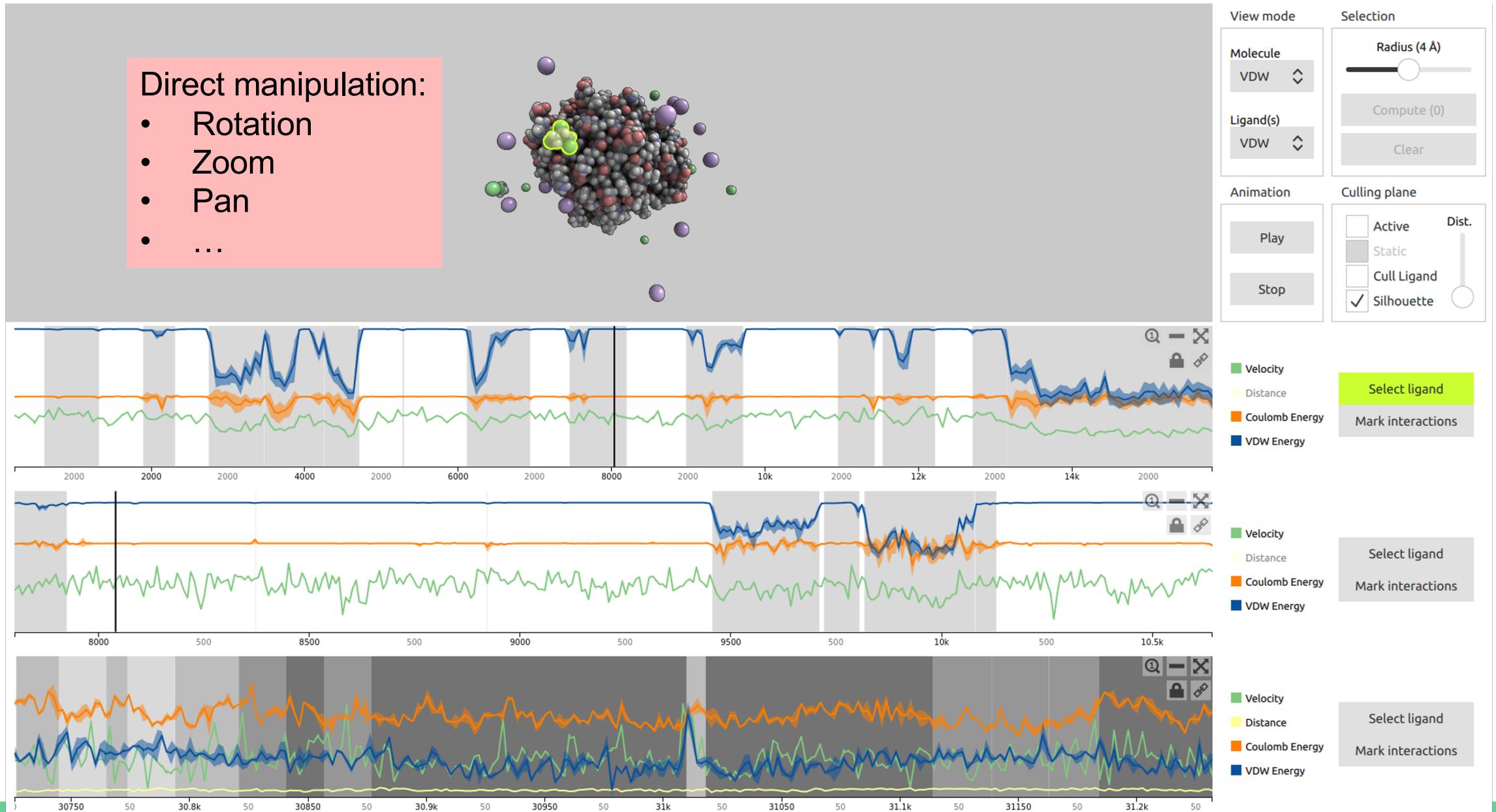
Main ideas:

- Progressive exploration of simulation data
  - Analyze input data to provide insights on interesting regions
- Interaction techniques to explore multiple simulations at once
- Combine (bidirectionally linked) 3D and 2D views
- Special 3D ROI selection with 3D → 2D highlighting

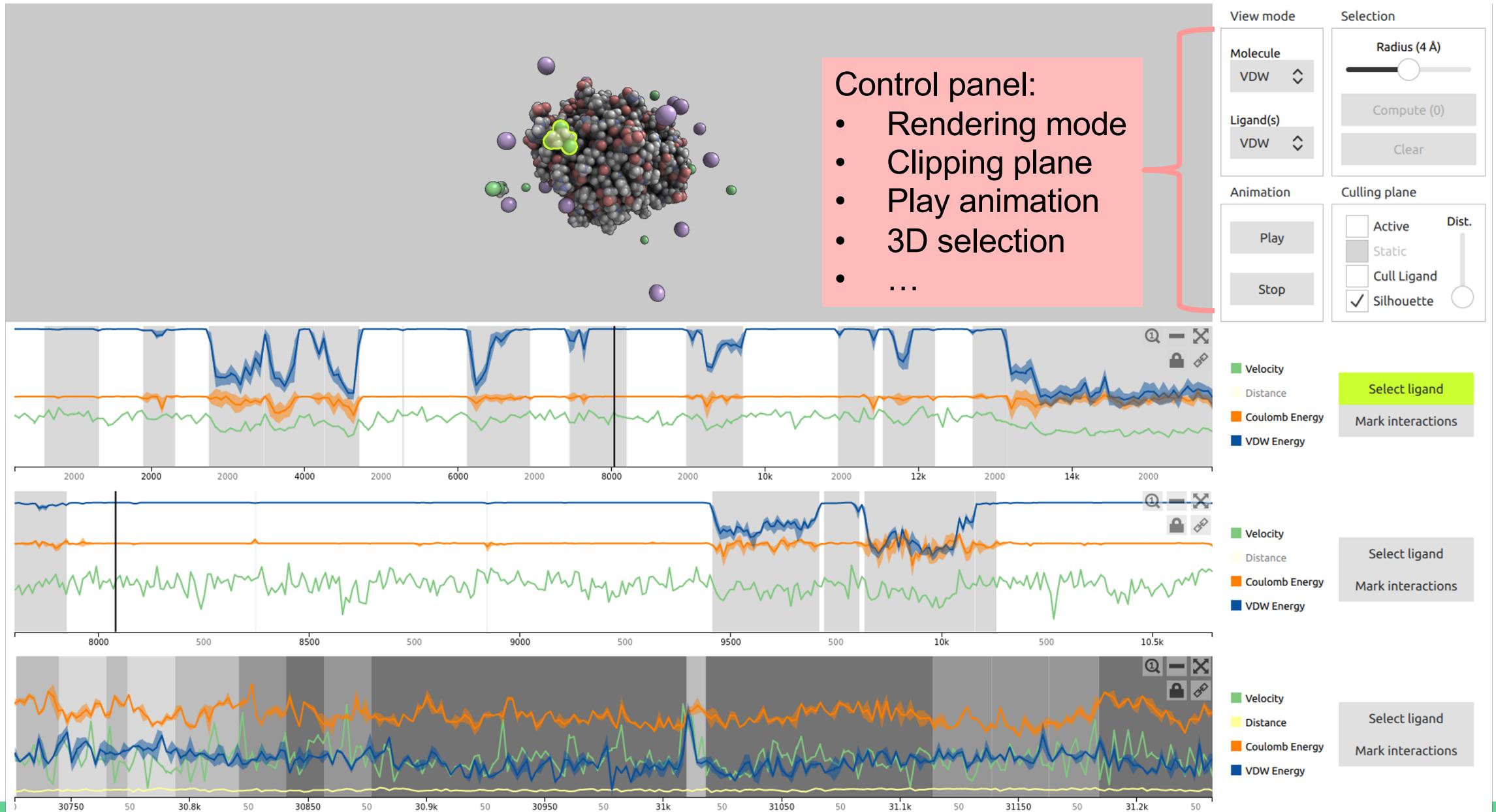
# system overview



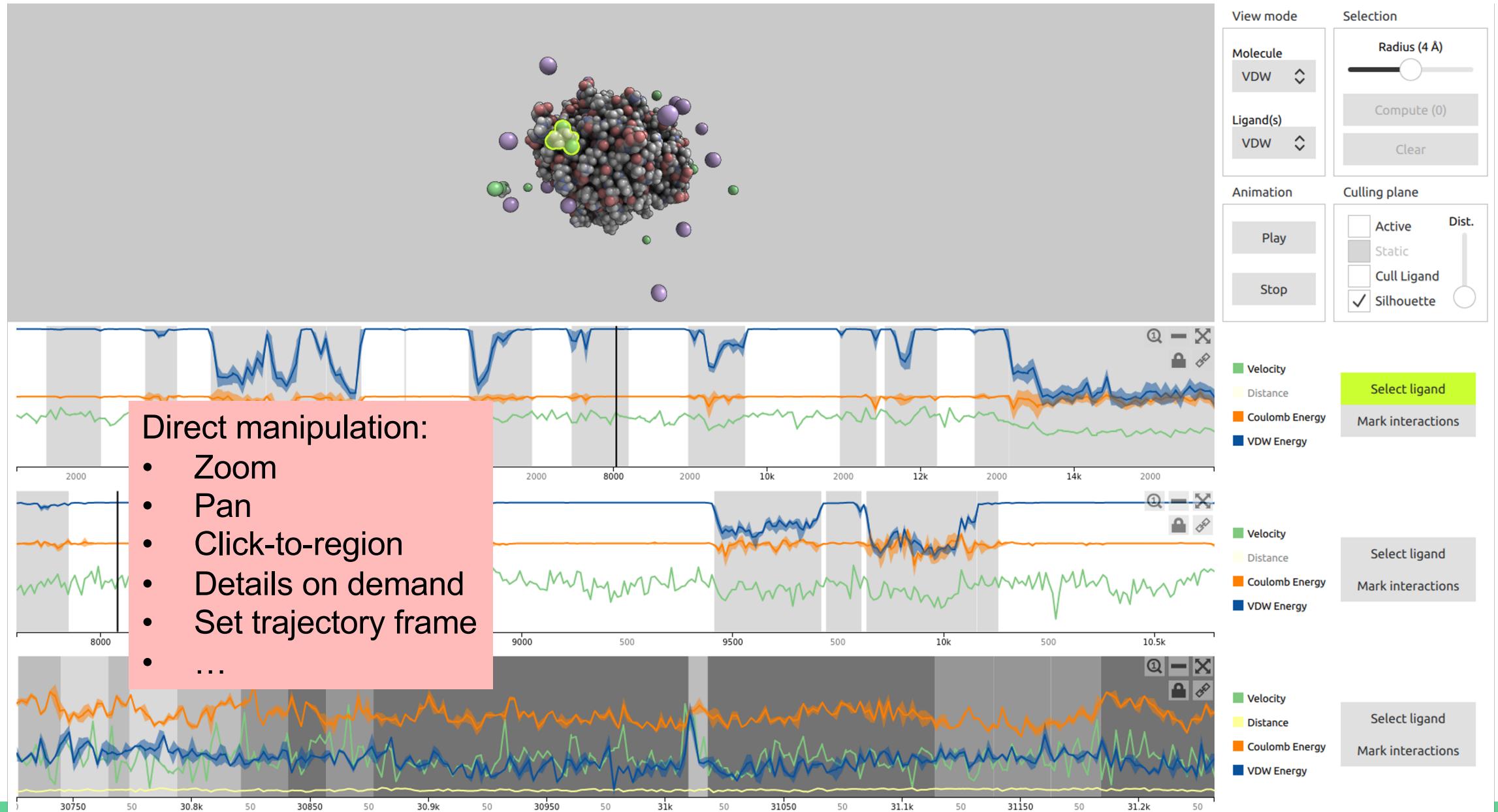
# system overview



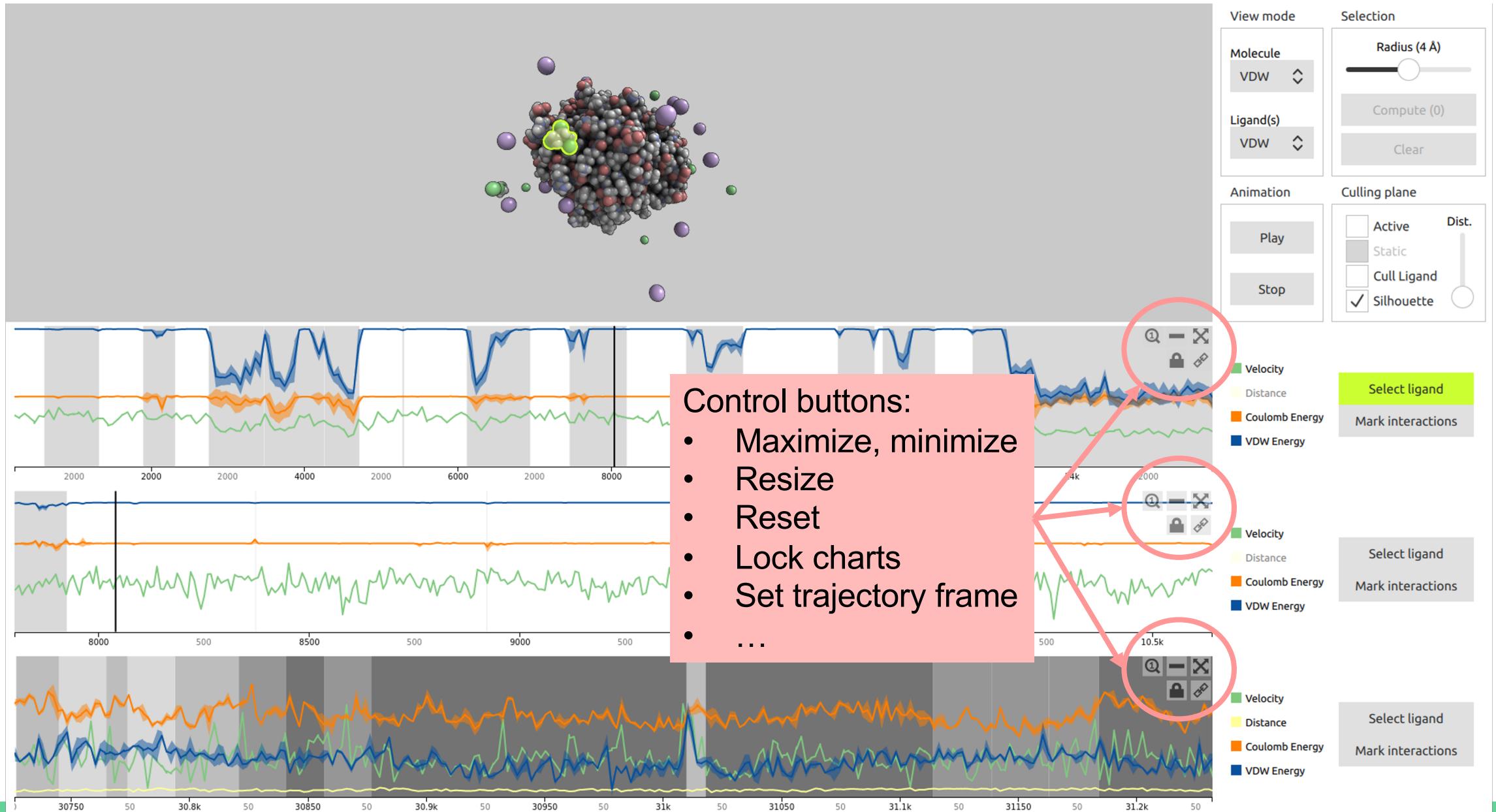
# system overview



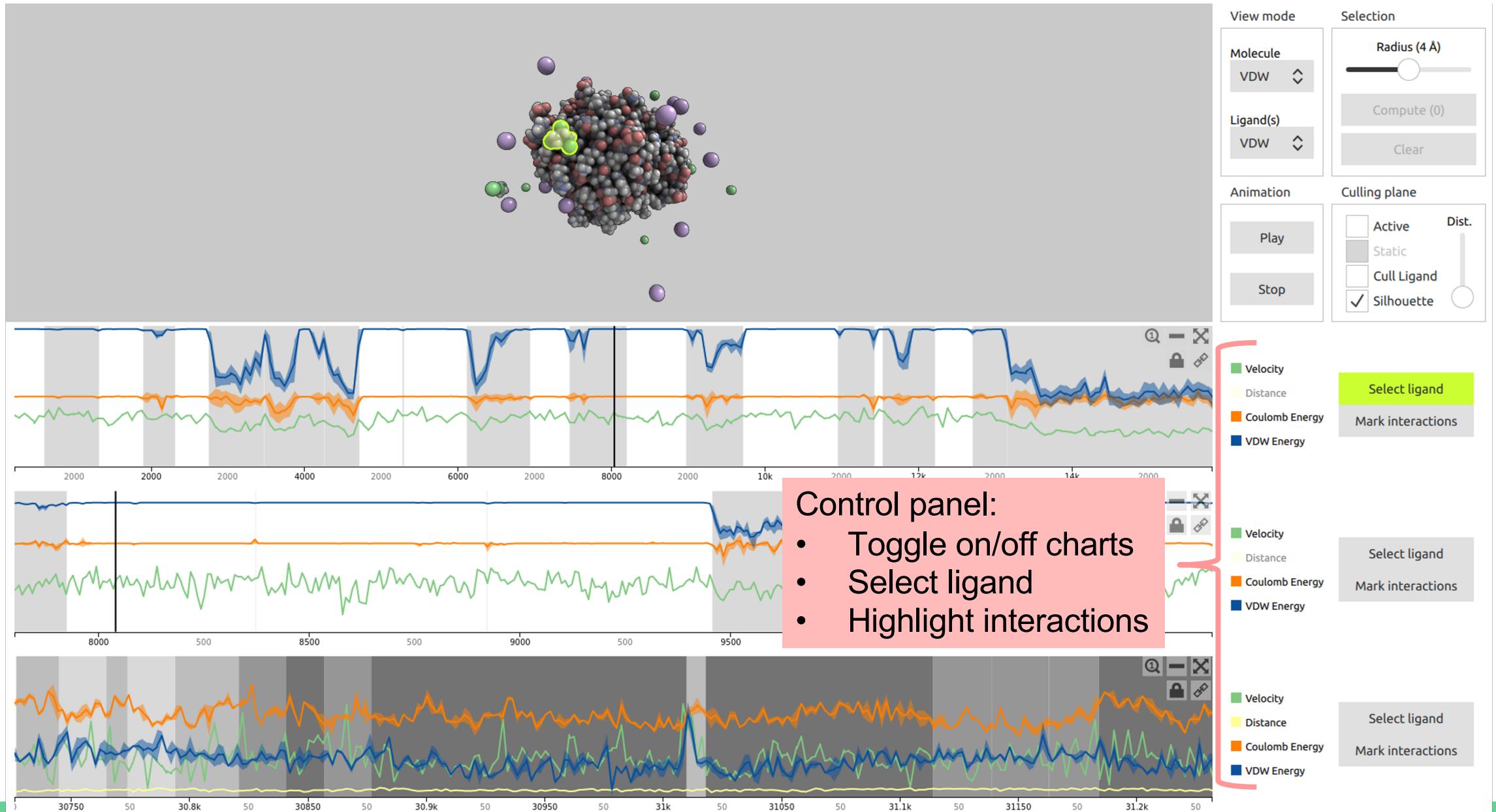
# system overview



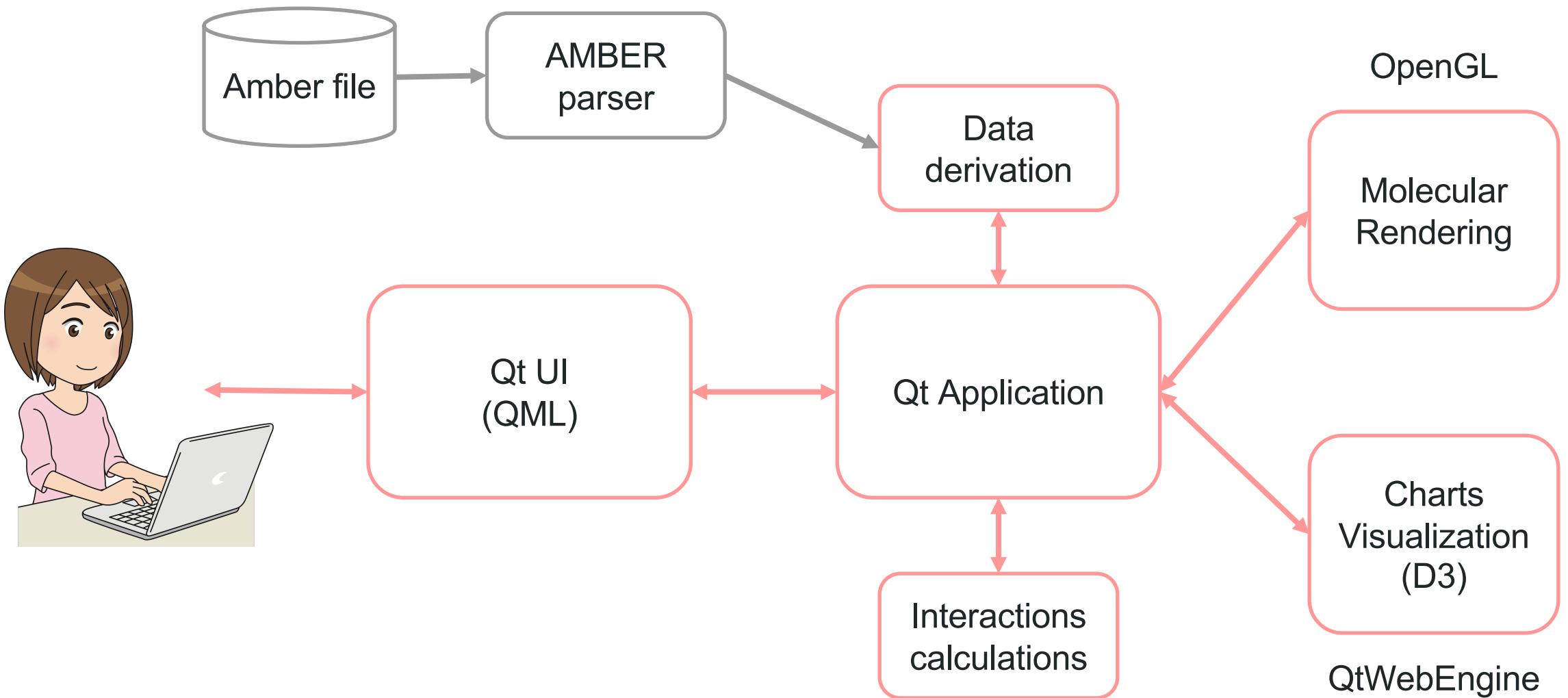
# system overview



# system overview



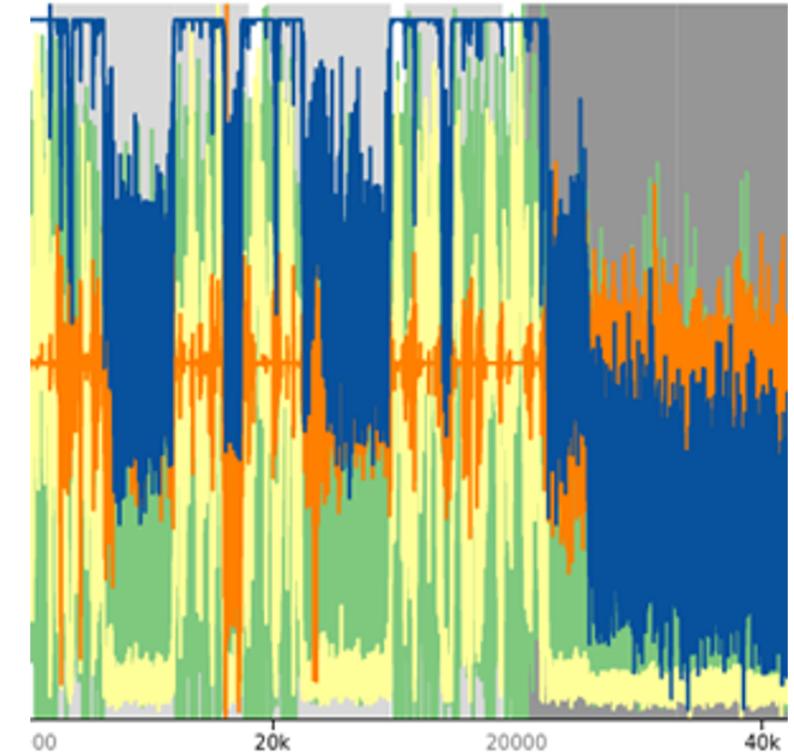
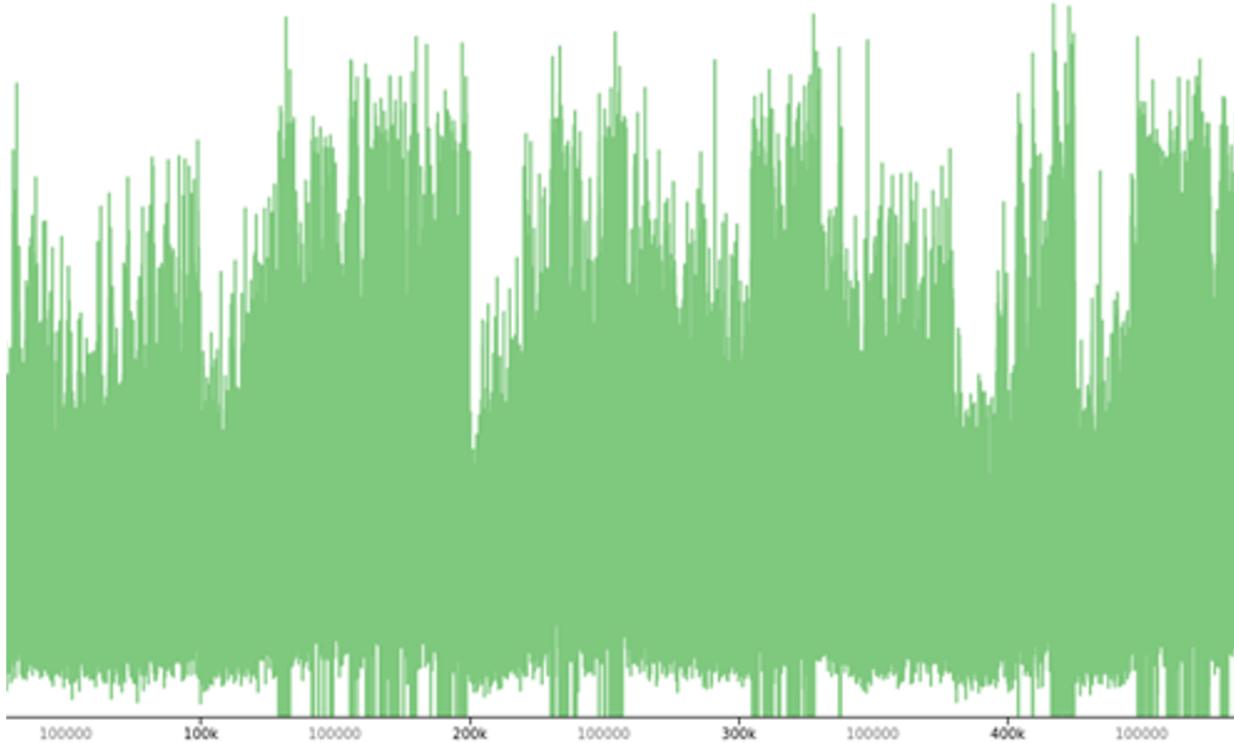
# system overview



# hierarchical trajectory exploration

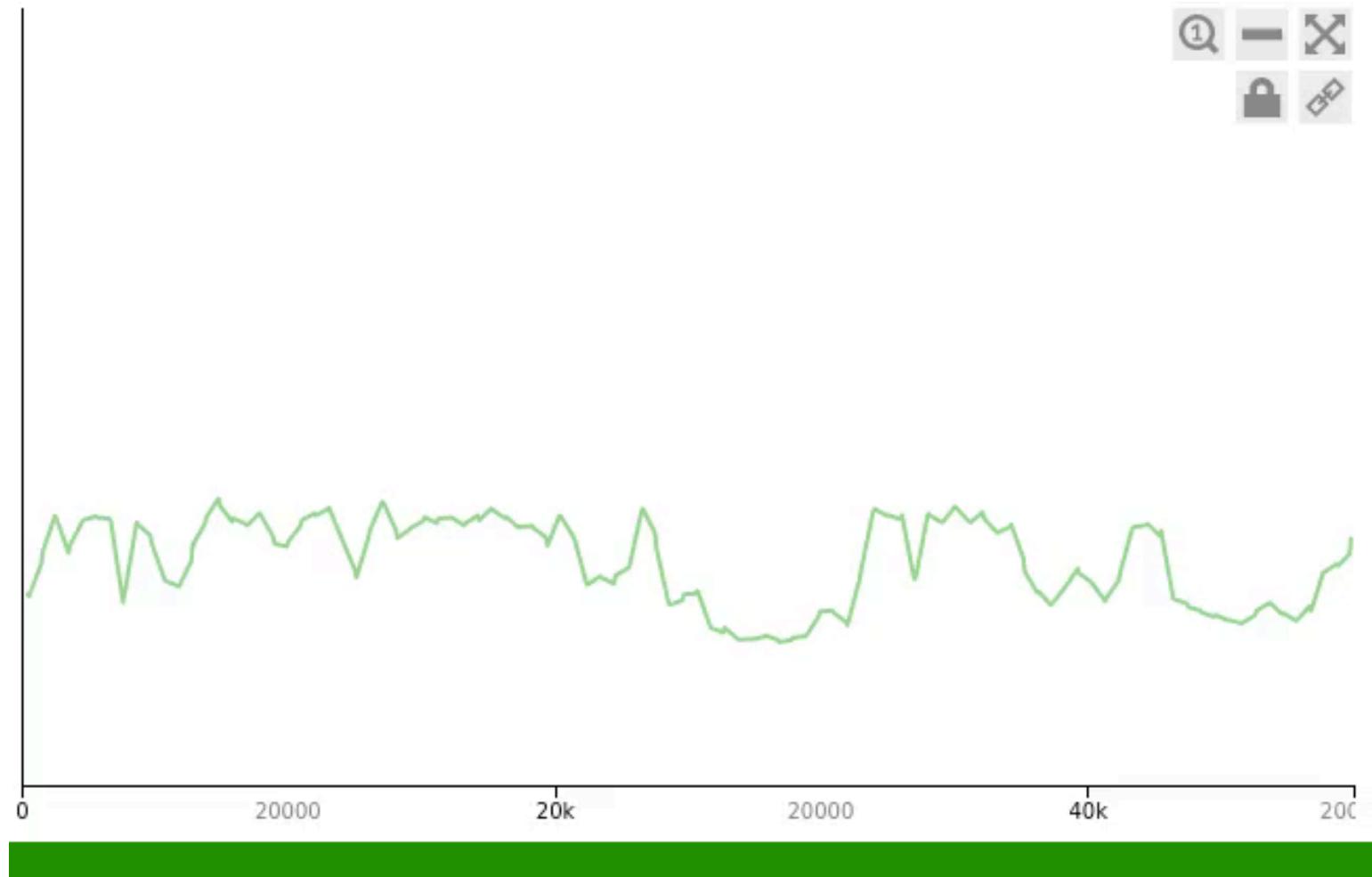
Too much data!

- 50-800 K values in ~1K screen pixels
- Up to 4 magnitudes: vdW energy, electrostatic energy + velocity, relative distance



# hierarchical trajectory exploration

Progressive exploration



# hierarchical trajectory exploration

Progressive exploration not enough

- Minimum and maximum values may be very far from the average
- The whole trajectory still too long to explore
  - Guidance needed

Enhancing energy charts

- Better communicate value variations (beyond the average)
- Analyze input to generate guidance
- Provide interaction tools to facilitate exploration

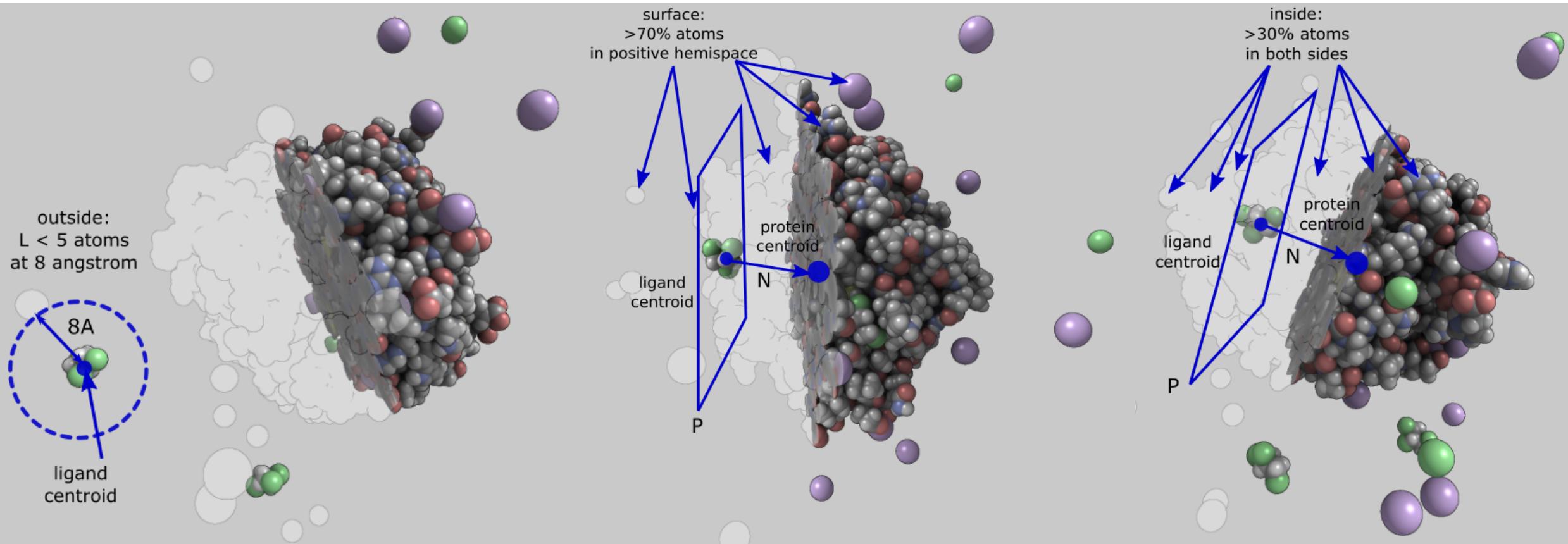
# hierarchical trajectory exploration

## Data derivation

- Avg and Q<sub>1</sub>, Q<sub>3</sub> from vdW and electrostatic values
  - Improves the communication of non-visible values
- Velocity & position
  - From the atoms' positions
  - Low speed may indicate interaction
  - Ligands' position inside the protein are important
- Potentially important regions
  - Classification of ligands' relative position from the protein: **outside, surface, and inside**
  - Similarity function: Maximum temporal distance between frames
- Interactions list
  - For each ligand, for each atom, a list of first and last frame pairs

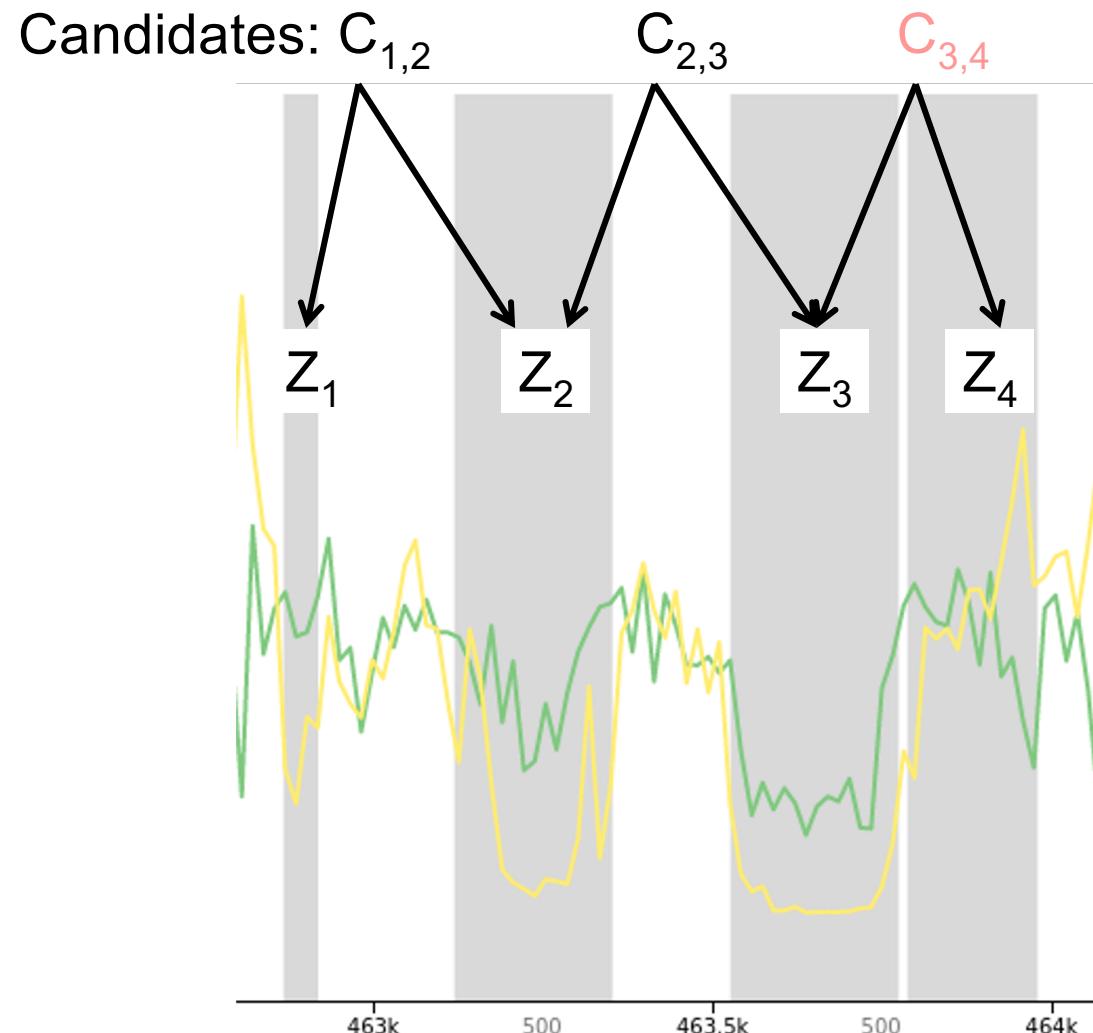
# hierarchical trajectory exploration

Data derivation: outside, surface, inside



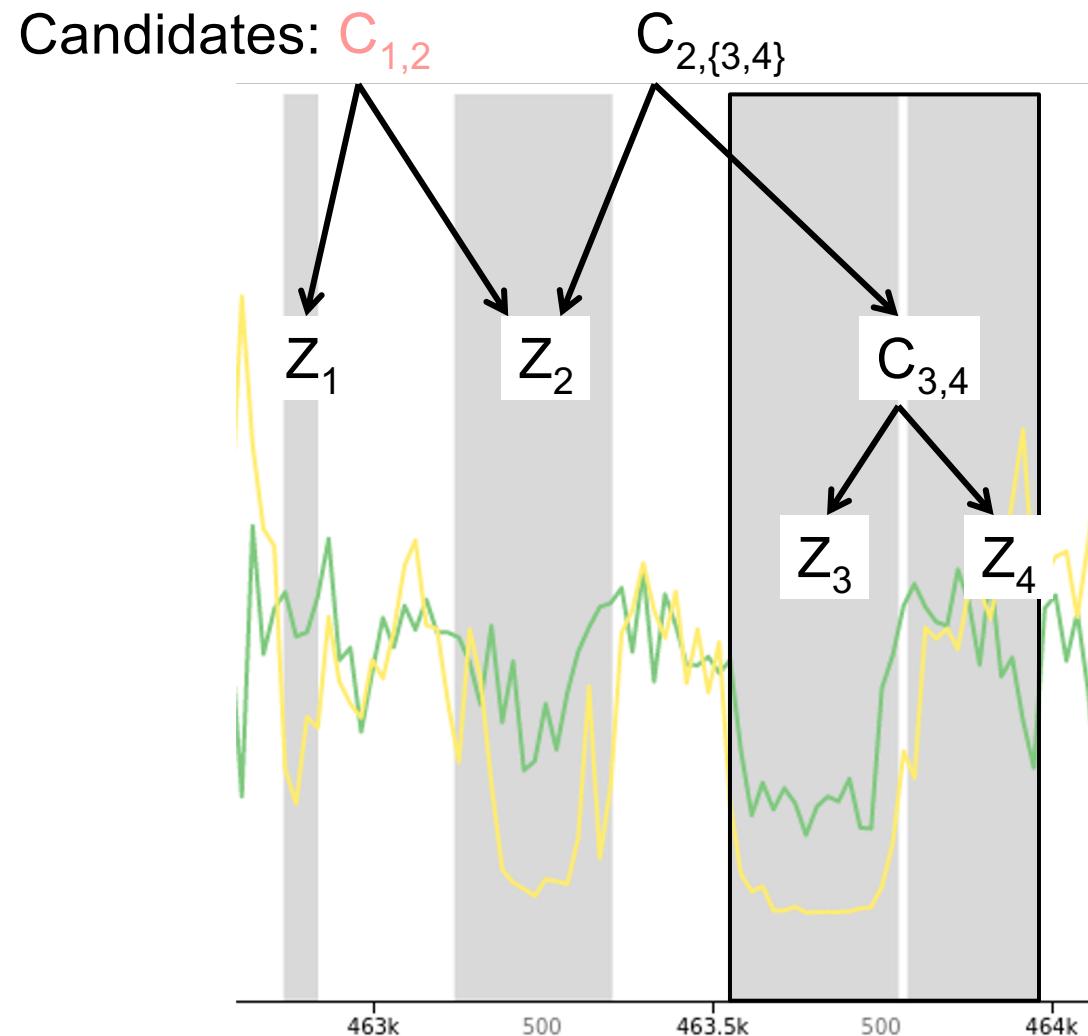
# hierarchical trajectory exploration

Hierarchical agglomerative clustering



# hierarchical trajectory exploration

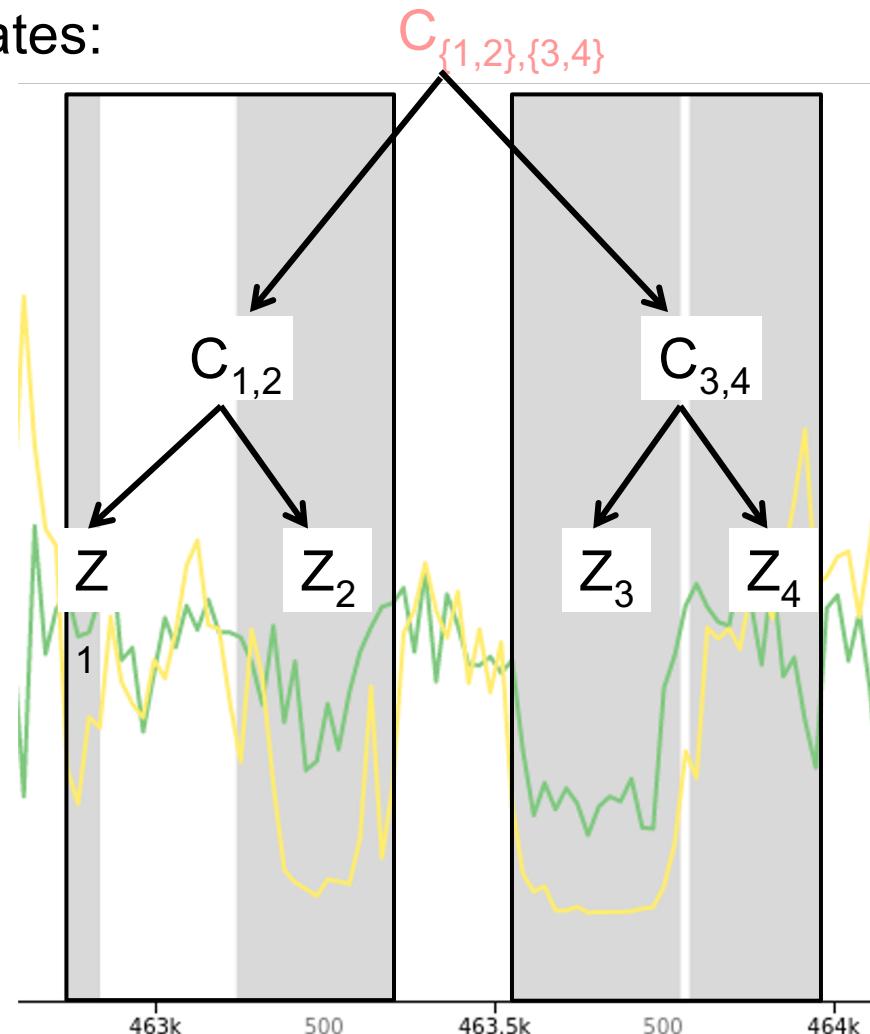
Hierarchical agglomerative clustering



# hierarchical trajectory exploration

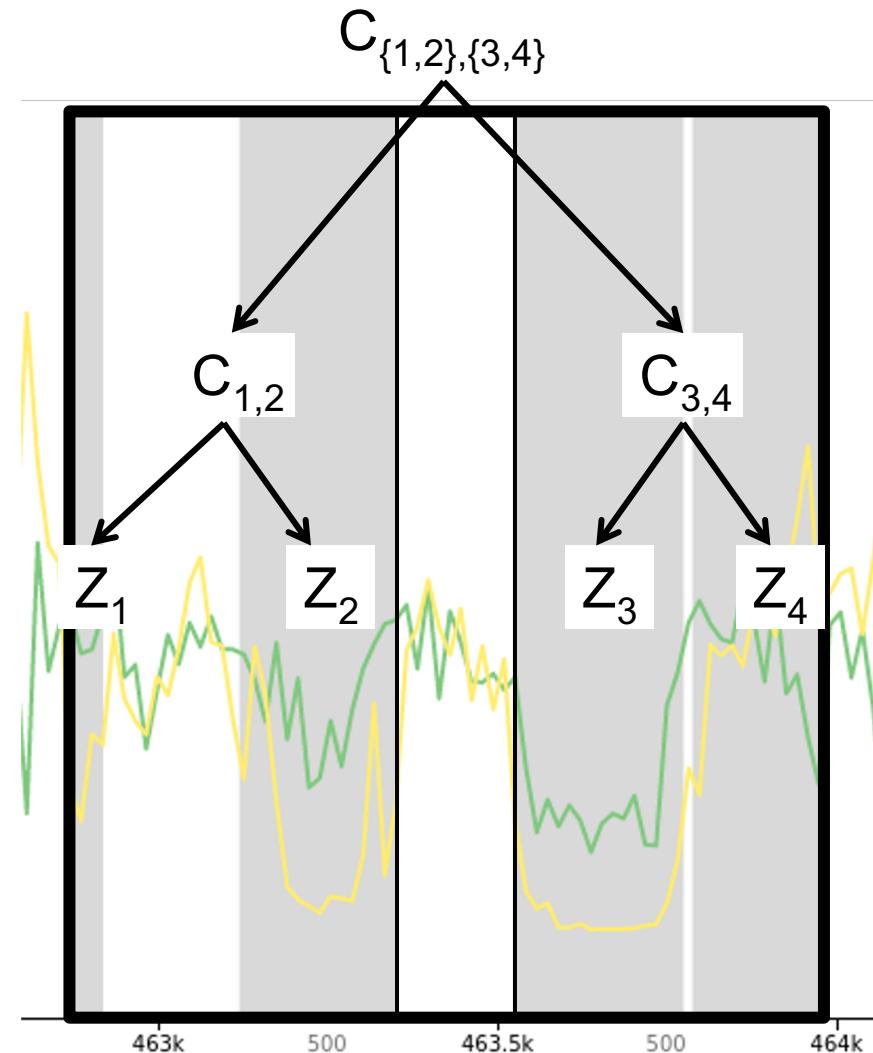
Hierarchical agglomerative clustering

Candidates:



# hierarchical trajectory exploration

Hierarchical agglomerative clustering



# hierarchical trajectory exploration

Visually encoding interactions

Class	Color single class	Color most prominent class
INTERACTION		
<i>Interior</i>		
<i>Surface</i>		

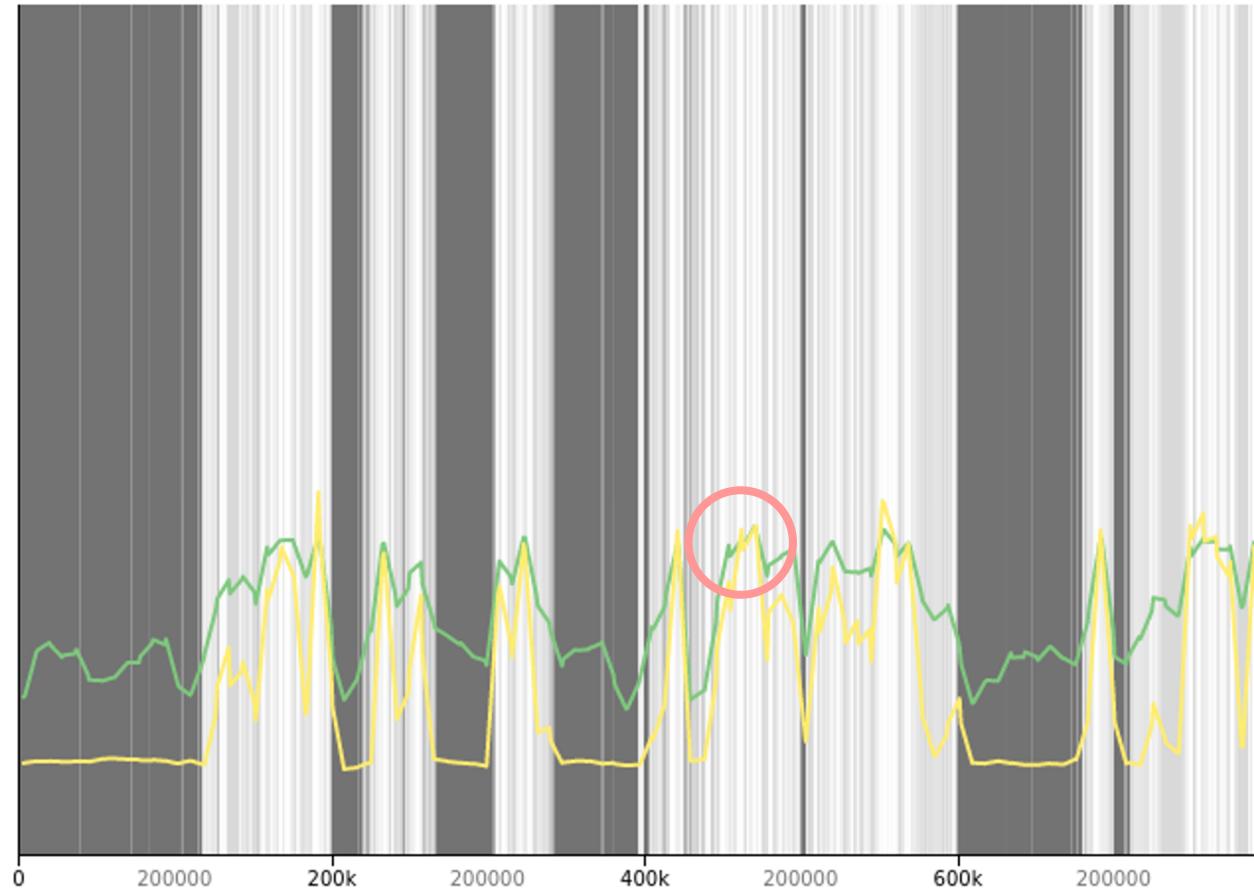
# hierarchical trajectory exploration

Hierarchical agglomerative clustering. Cost

- $\theta(n \log n)$  in time
- $\theta(n)$  in memory
- Since we are restricted to join neighboring clusters

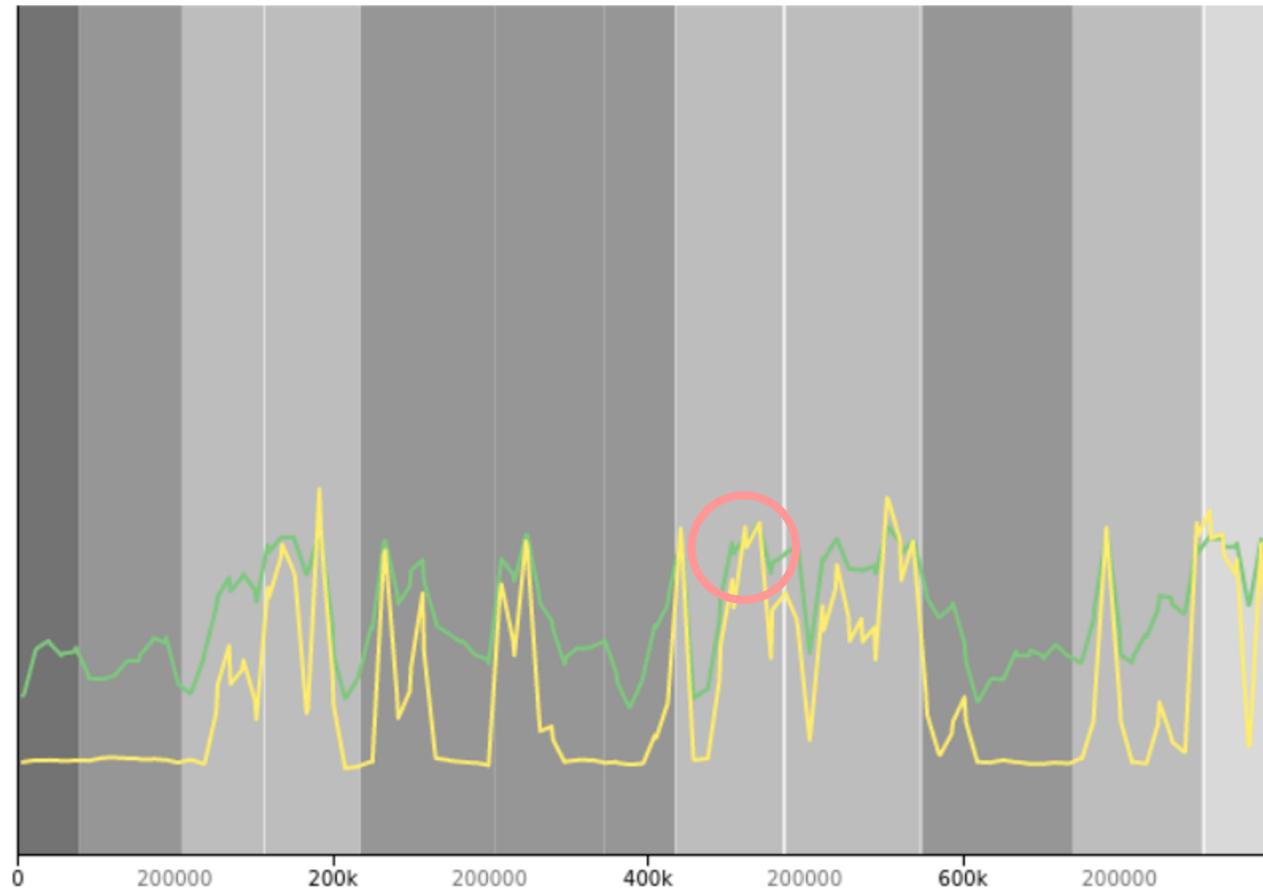
# hierarchical trajectory exploration

Without clustering



# hierarchical trajectory exploration

With clustering



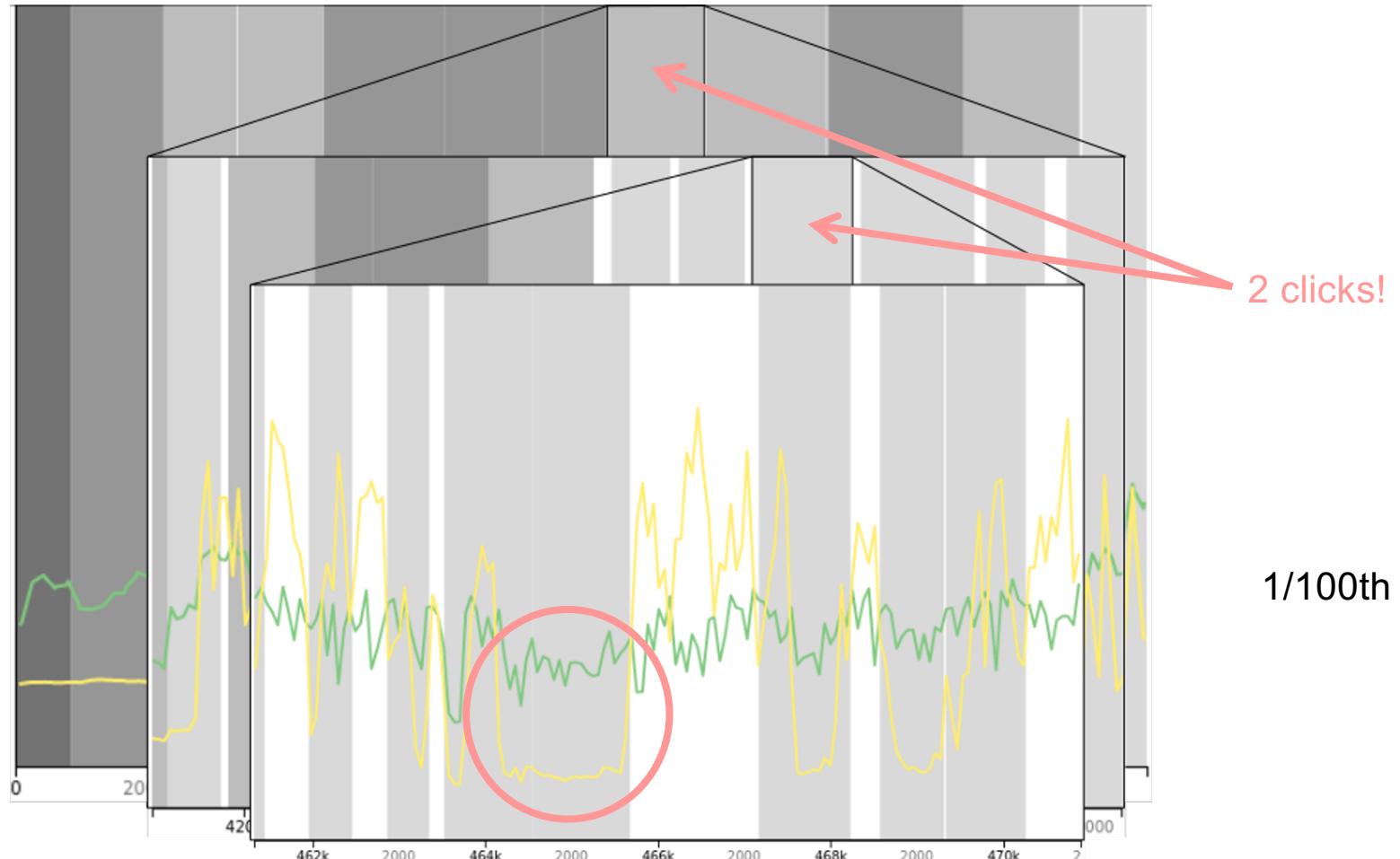
# hierarchical trajectory exploration

Interactive exploration



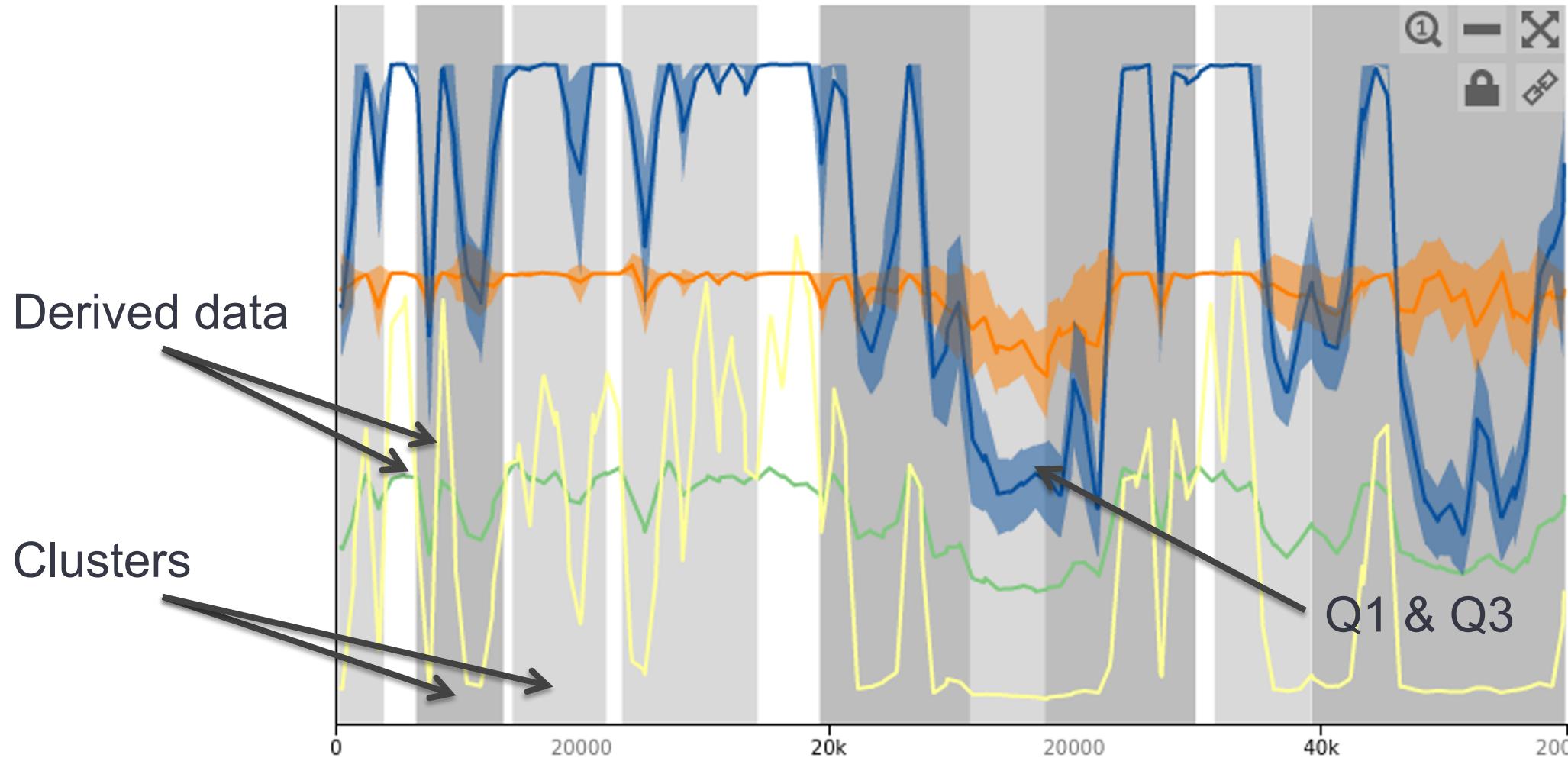
# hierarchical trajectory exploration

Interactive exploration



# hierarchical trajectory exploration

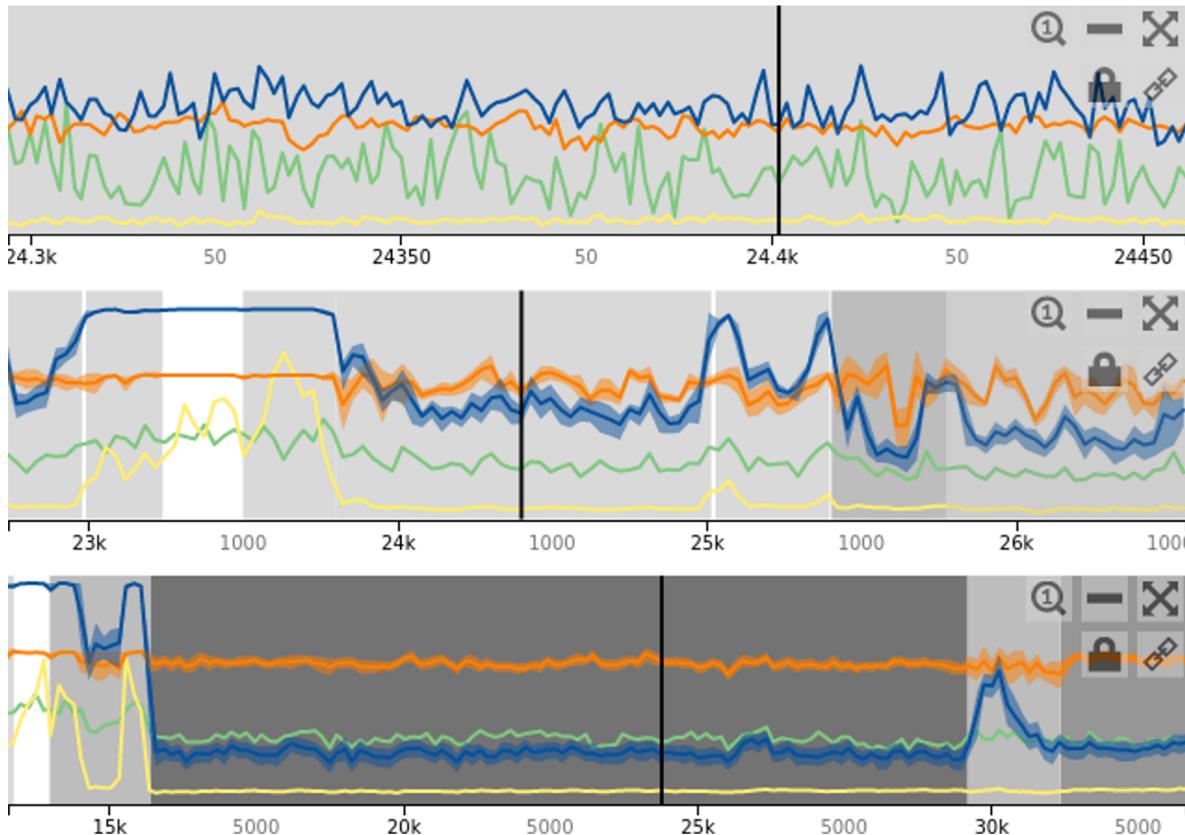
Enriched charts





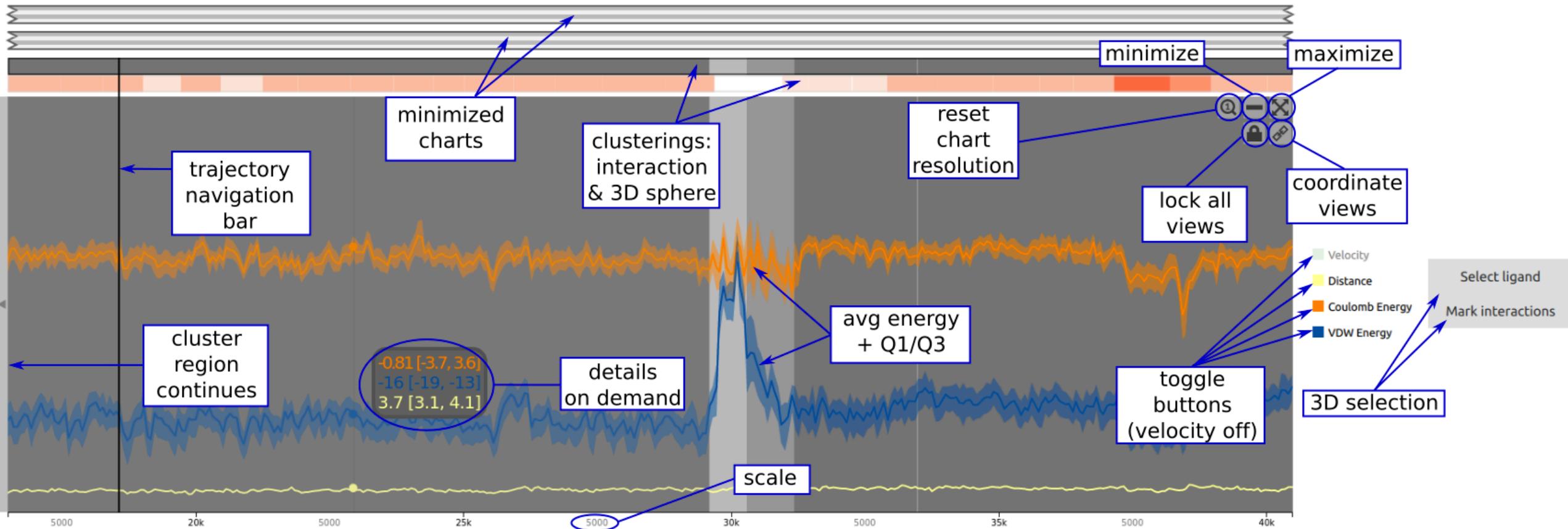
# hierarchical trajectory exploration

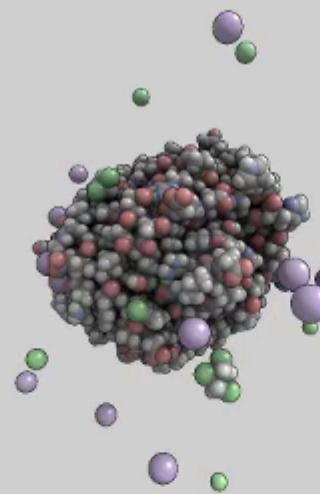
Multiple ligands



# hierarchical trajectory exploration

## Enriched charts interaction overview





**View mode**

Molecule VDW

Ligand(s) VDW

Animation Play Stop

Culling plane

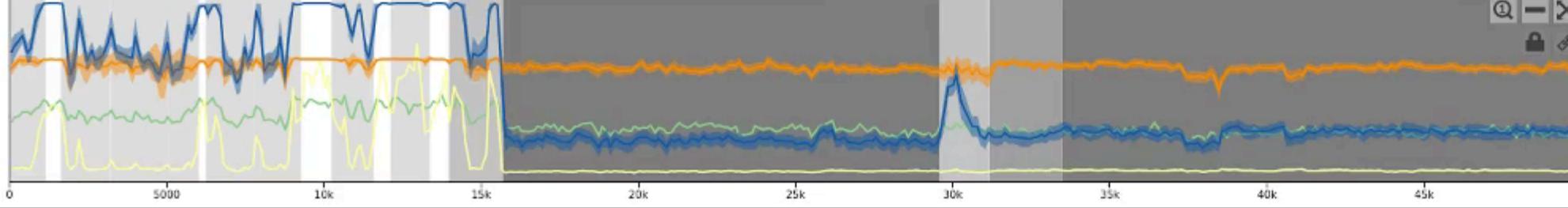
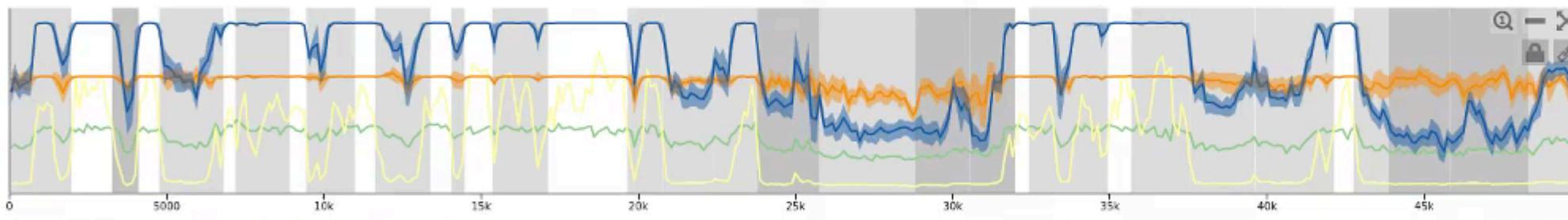
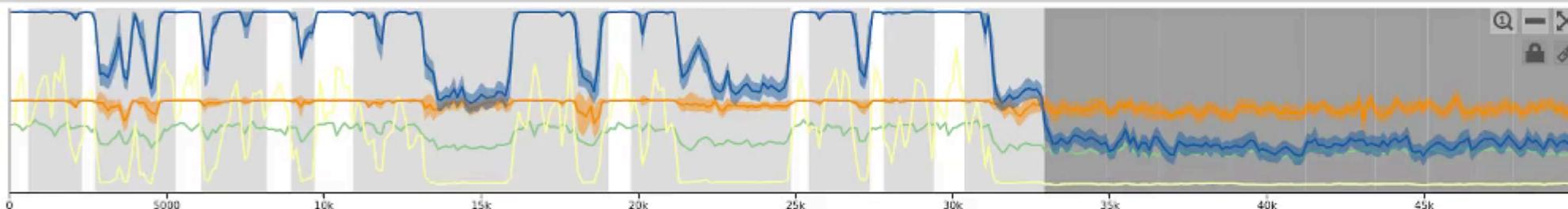
- Active
- Static
- Cull Ligand
- Silhouette

Radius (4 Å)

Compute (0)

Clear

Dist.

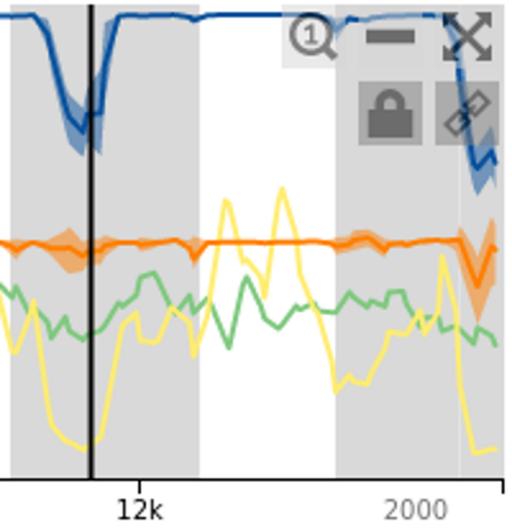


# coordinated views

2D to 3D interaction

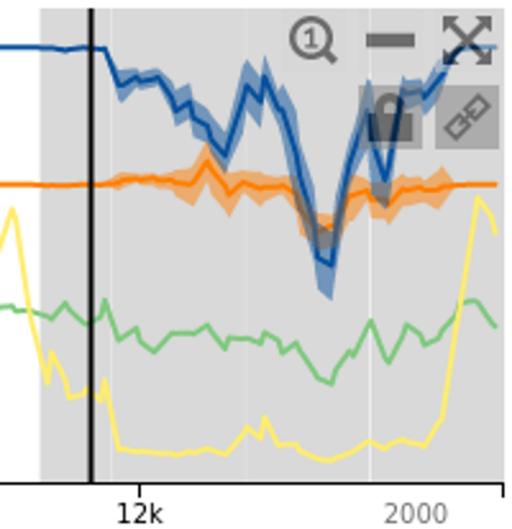
- Highlight ligand
- Highlight area of influence
  - Dynamic updates
- Control animation

# coordinated views



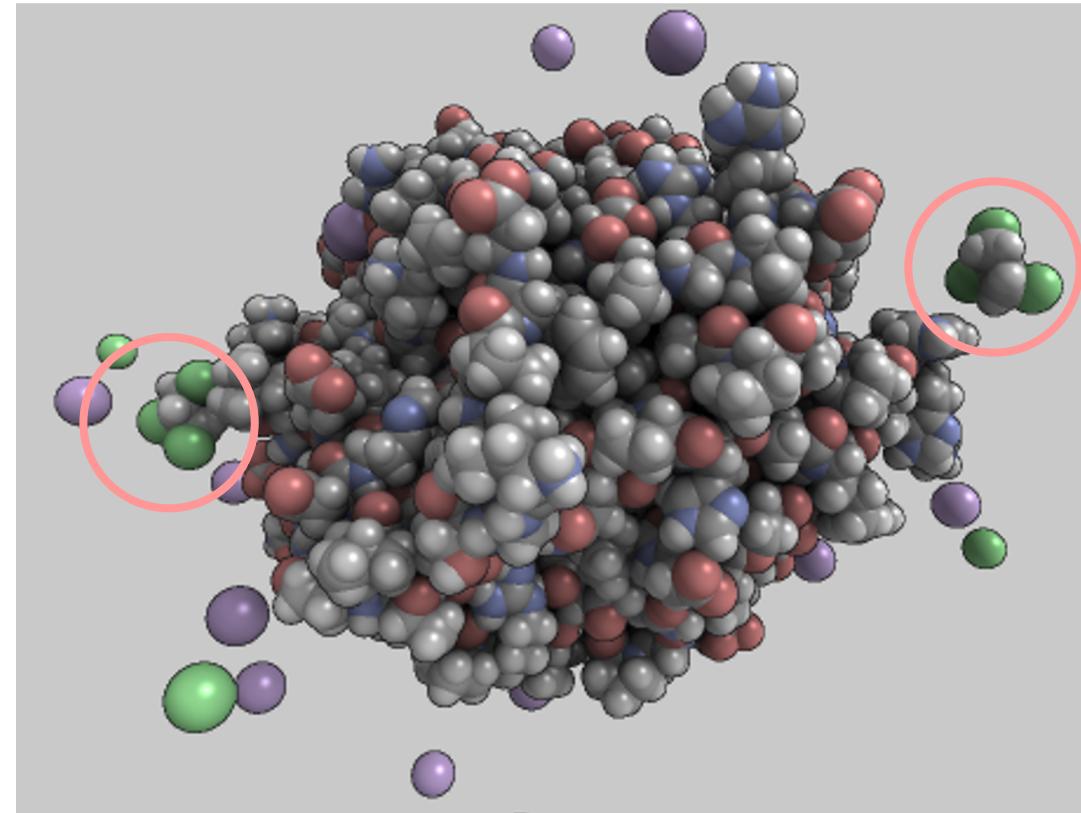
Velocity  
Distance  
Coulomb Energy  
VDW Energy

Select ligand  
Mark interactions

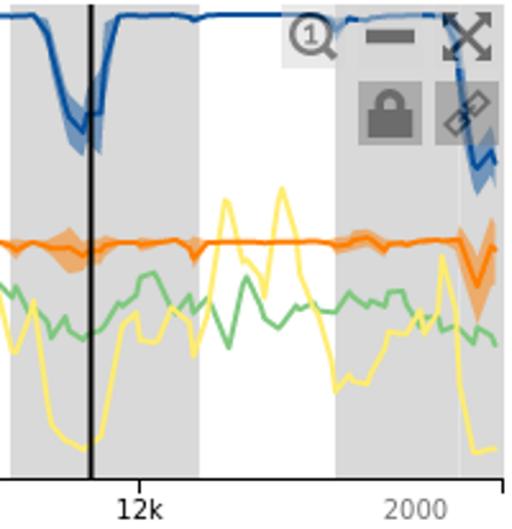


Velocity  
Distance  
Coulomb Energy  
VDW Energy

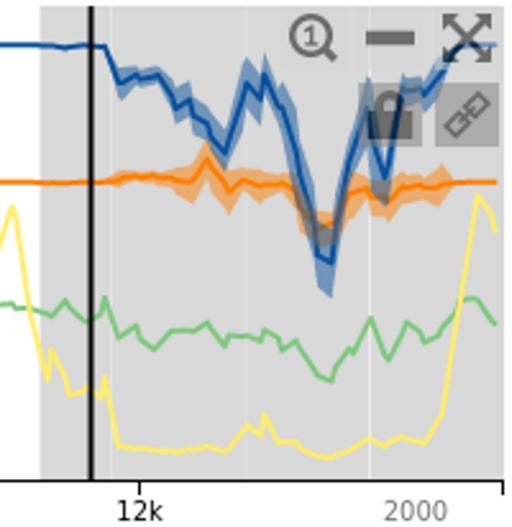
Select ligand  
Mark interactions



# coordinated views

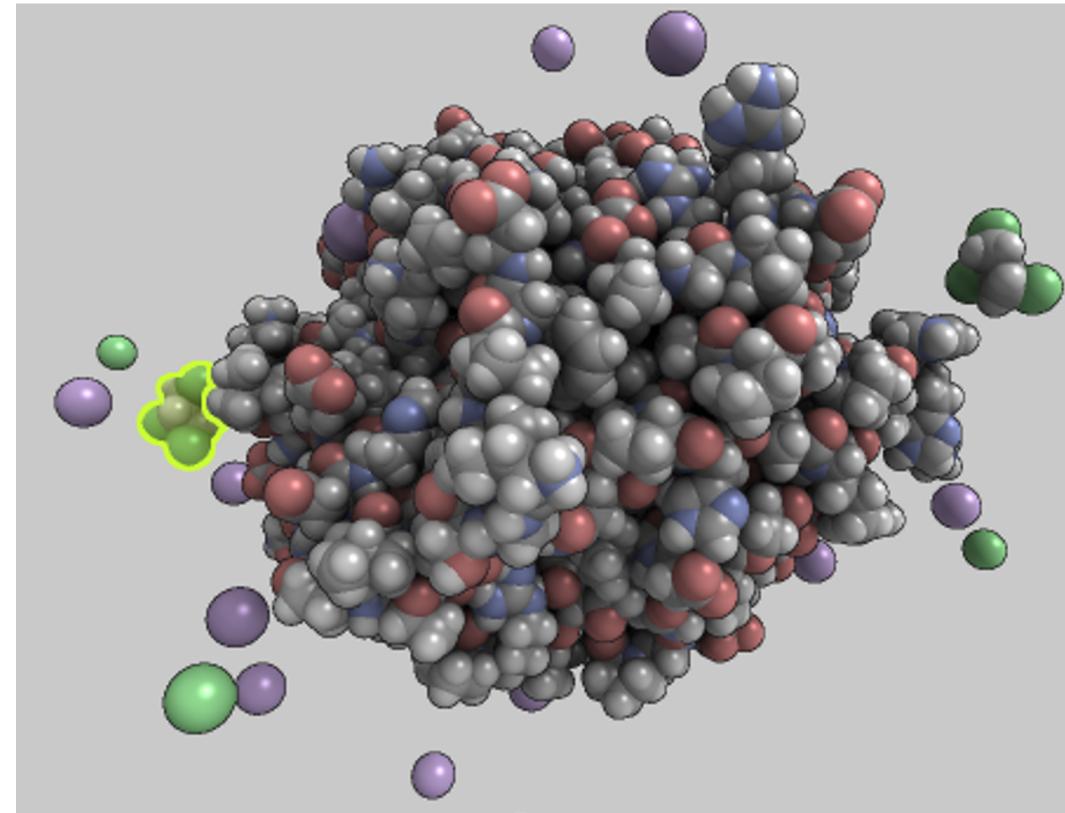


Velocity  
Distance  
Coulomb Energy  
VDW Energy

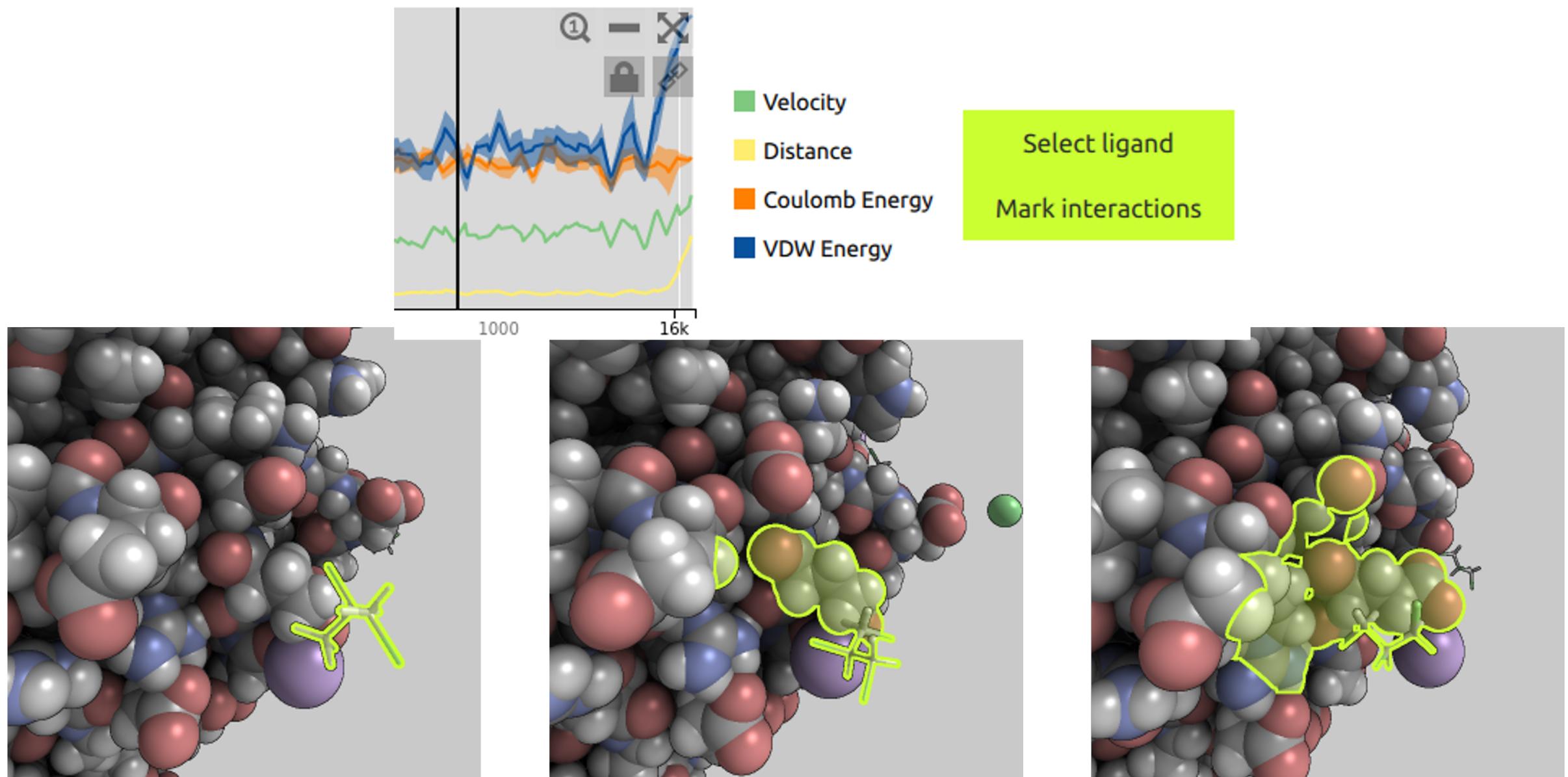


Velocity  
Distance  
Coulomb Energy  
VDW Energy

Select ligand  
Mark interactions



# coordinated views





# coordinated views

3D to 2D interaction

- Selection of Region of Interest
- Visual analysis of interacting ligands

# coordinated views

## Selection of ROI

- Spherical 3D region analysis
  - Using the clipping plane to facilitate ROI placement
- Radius can be modified
- Typically for cavity exploration
- Solvent Excluded Surfaces mode facilitate cavity detection

# coordinated views

Visual encoding of interacting ligands

- Calculate/detect ligands entering the region (for the whole path)
- Highlight current step interactions
- Navigation & interaction

# coordinated views

Calculating ligand's interaction

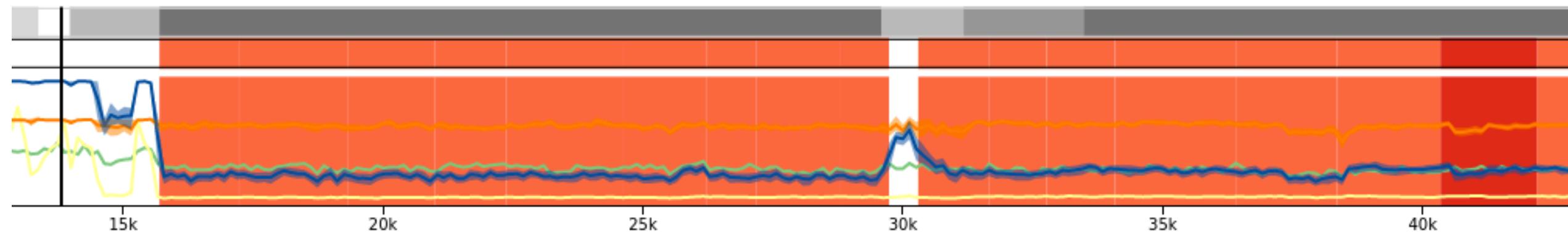
- For each time snapshot
  - Count number of selected atoms interacting with ligand
  - Normalize with maximum count
  - Classify interaction in 4 groups (no, low, medium, high)
  - Cluster hierarchically interaction values
  - Colorize charts accordingly

Class	Color single class	Color most prominent class
SELECTION		
<i>Low interaction</i>		
<i>Medium interaction</i>		
<i>High interaction</i>		

# coordinated views

Highlight step interactions

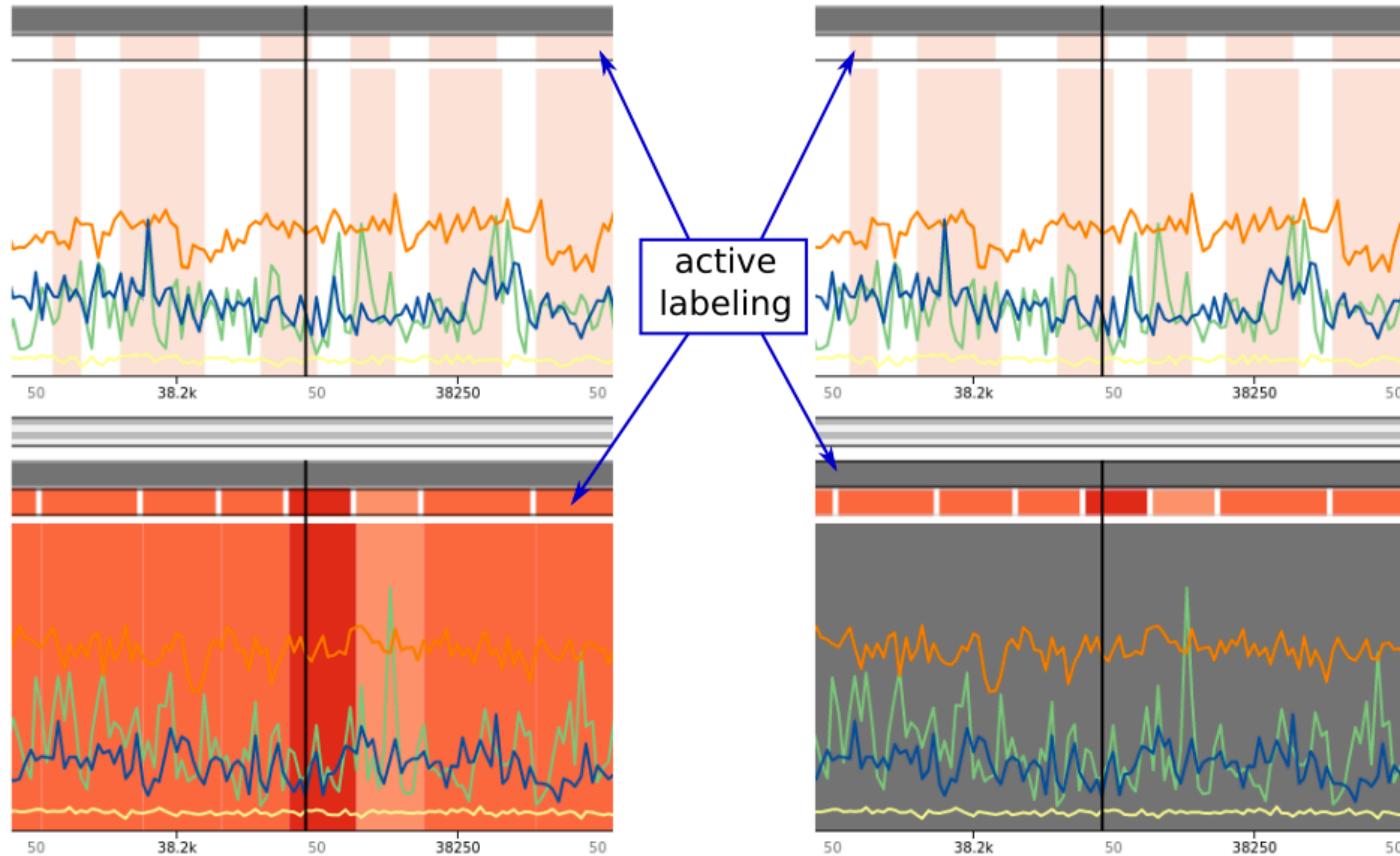
- Cluster hierarchically interaction values
- Colorize charts accordingly
  - Using levels of orange
  - Hide previous clustering information



# coordinated views

## Navigation & interaction

- Current coloring can be clicked
- Top navigation bar & color encoding change





# results & evaluation

## Performance

- 3D model renders immediately (less than 1s)
- Data processed and charts rendering in < 3s
  - Appear progressively
- Label recalculation interactive
- Precomputation time in the order of one or two dozens of seconds
  - Grows linearly with number of snapshots

# results & evaluation

## User study

- 6 domain experts (Loschmidt Laboratories at Masaryk University)
- Demo session + Q&A: around one hour
- Internal discussion
- Questionnaire

# results & evaluation

## User study. Conclusions

- Tool could be useful for their work (avg. 4 out of 5)
- Easy to learn (avg. 3.8 out of 5)
- Utility of multiple ligands visualization (4.5 out of 5)
- Chart visual labelling (4.8 out of 5)
- Utility of nearby atoms' highlighting (4.6 out of 5)
- 3D cavities selection (3.6 out of 5)

# conclusions & future work

## Limitations

- Same color for ligands
- Up to 3 ligands
- Fixed size of 3D window

# conclusions & future work

Potential future extensions

- Would add direct 3D coordinates input
- Configure the distance to infer interactions
- Compare multiple trajectories of the same ligand
- Deal with multiple simulations
- Multiple 3D views

# conclusions & future work

Advantages over other software

- Ligplot+
  - No integrated 3D view
  - Not progressive exploration
- Schrödinger's SID
  - No interactive chart explorations
  - No 3D exploration of multiple snapshots
- VMD, TAMD
  - No integrated layout with 2D ↔ 3D bindings
- PLIP, PyMol
  - Trajectories' analysis not provided

# acknowledgments

- GEN3DLIVE project (TIN2017-88515-C2-1-R) from the Spanish Ministerio de Economía y Competitividad
- Deutsche Forschungsgemeinschaft (DFG) under grant RO 3408/2-1 (ProLint)
- Czech Science Foundation (GACR) under grant GC18-18647J
- Researchers at the Loschmidt Laboratories at Masaryk University
- Anonymous reviewers

# visualization of large molecular trajectories

---

D. Duran-Rossich, P. Hermosilla, T. Ropinski, B.  
Kozlikova, À. Vinacua, P. Vázquez

*VIS 2018*

[pere.pau@cs.upc.edu](mailto:pere.pau@cs.upc.edu)

*IEEE*

