

Biophysical Journal, Volume 114

Supplemental Information

PyContact: Rapid, Customizable, and Visual Analysis of Noncovalent Interactions in MD Simulations

Maximilian Scheurer, Peter Rodenkirch, Marc Siggel, Rafael C. Bernardi, Klaus Schulten, Emad Tajkhorshid, and Till Rudack

Table S1: Benchmark of parallelized code ¹

# CPU cores	trajectory read-in and hydrogen bonds [seconds]	SASA calculation [seconds]
1	35	25
2	18.5	13
4	11	8
6	10.5	6
8	10	5.5

¹ A cohesin-dockerin SMD trajectory with 612 frames was used for the benchmark. The data set is extensively discussed in the main text. Atom selections were made as presented in Fig. S1 and Fig. S7. Hardware: Intel Core i7 6700K (8 virtual cores). Memory usage: approx. 300 MB per core upon trajectory read-in and hydrogen bond analysis.

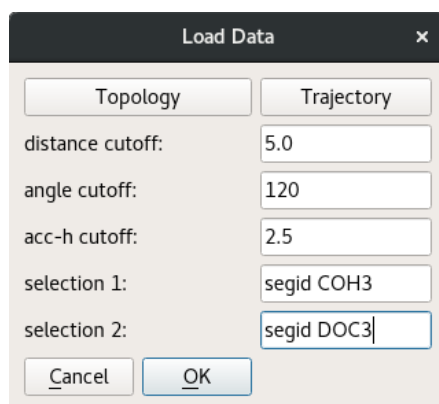


Figure S1: Trajectory loading dialog. The user can choose the input topology and trajectory file for contact analysis. Furthermore, customization of the distance cutoff, the hydrogen bond angle cutoff and the acceptor-hydrogen bond distance cutoff is provided. The two selection text fields let the user specify the atom selections for contact analysis.

Algorithm

PyContact non-covalent interaction analysis consists of sequential tasks:

1. distance-based interatomic contact scoring
2. hydrogen bond analysis
3. contact score accumulation

4. contact type determination based upon residue properties

In the following, the individual steps are explained in detail.

Distance-based interatomic contact scoring

The user-defined selections $S1$ and $S2$ contain N and M atoms, respectively. First, the program calculates the distance matrix d_{ij}^f for all atoms i in $S1$ and all atoms j in $S2$ for every frame f , given by

$$d_{ij}^f = \|\vec{r}_i^f - \vec{r}_j^f\|, \quad (1)$$

where \vec{r}_i denotes the position vector of atom i .

Next, for each matrix element d_{ij}^f , its corresponding score s_{ij} is examined according to

$$s_{ij}^f = \begin{cases} \frac{1}{1+\exp(5.0 \cdot (d_{ij}^f - 4.0))} & , d_{ij}^f \leq \text{cutoff} \\ 0 & , \text{else} \end{cases} \quad (2)$$

This sigmoid function was used in previously published work to calculate contact scores (1–3). Atom tuples (i, j) with a score s_{ij}^f greater than 0 are stored in a list.

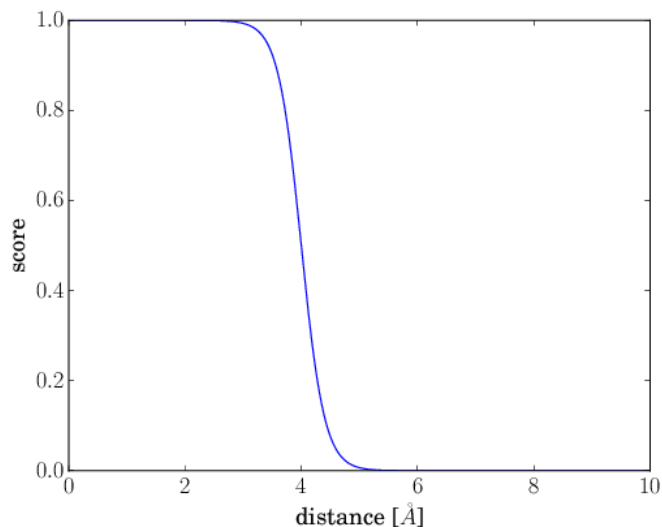


Figure S2: Scoring function plot. Contact scores are calculated based on the atom-atom distance by evaluating the given sigmoid function (Eq. 2).

Hydrogen bond analysis

Hydrogen bonds require a precise definition concerning geometry (Fig. S3): The acceptor to hydrogen distance is usually in the interval from 1.5 to 2.5 Å (4). Thus, a hydrogen bond between the acceptor and the donor's hydrogen atom establishes when aforementioned distance constraint is fulfilled and the angle α_{DHA} spanned by the three atoms D, H and A is smaller or equal to 120° (4), defined by the following function:

$$\alpha_{\text{DHA}} = \begin{cases} \arccos\left(\frac{\langle \vec{r}_{\text{H}} - \vec{r}_{\text{I}}, \vec{r}_{\text{H}} - \vec{r}_{\text{J}} \rangle}{\|\vec{r}_{\text{H}} - \vec{r}_{\text{I}}\| \cdot \|\vec{r}_{\text{H}} - \vec{r}_{\text{J}}\|}\right) & , d_{\text{HA}} \leq \text{hbondcutoff} \\ \text{not defined} & , \text{else} \end{cases} \quad (3)$$

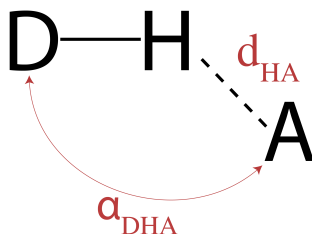


Figure S3: Hydrogen bond geometry is defined by the hydrogen-acceptor distance d_{HA} and the angle α_{DHA} .

Contact score accumulation

After storing the hydrogen bonds, the data is further processed: To represent contact data clearly, we sum up the individual scores s_{ij}^f according to certain atom properties. For example, one is able to sum all individual atom contacts of two amino acids interacting with each other. To provide a generic choice of attributes, the user can choose from a list of atom attributes.

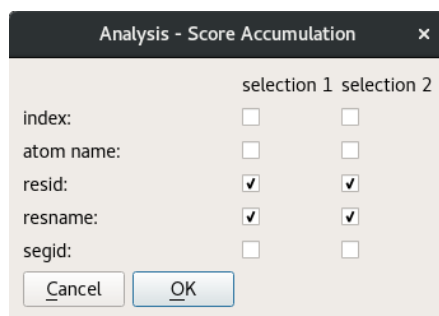


Figure S4: Score accumulation dialog. To interactively change between different levels of detail, the individual scores s_{ij}^f can be accumulated to different properties of the molecules i and j . For example, if the user wants to inspect residue-wise interactions between the two selections, the checkboxes for ‘resid’ and ‘resname’ ought to be checked. The score accumulation dialog is easily accessible, and the accumulations can be calculated on the fly during an analysis session.

Contact type determination

Hydrogen bonds are found by the aforementioned geometric criteria. For other typical interaction types, PyContact uses pre-implemented information about the set of amino acids. An interaction between negatively and positively charged side chains (e.g., LYS and ASP) will be marked as a salt bridge (Fig. S5).

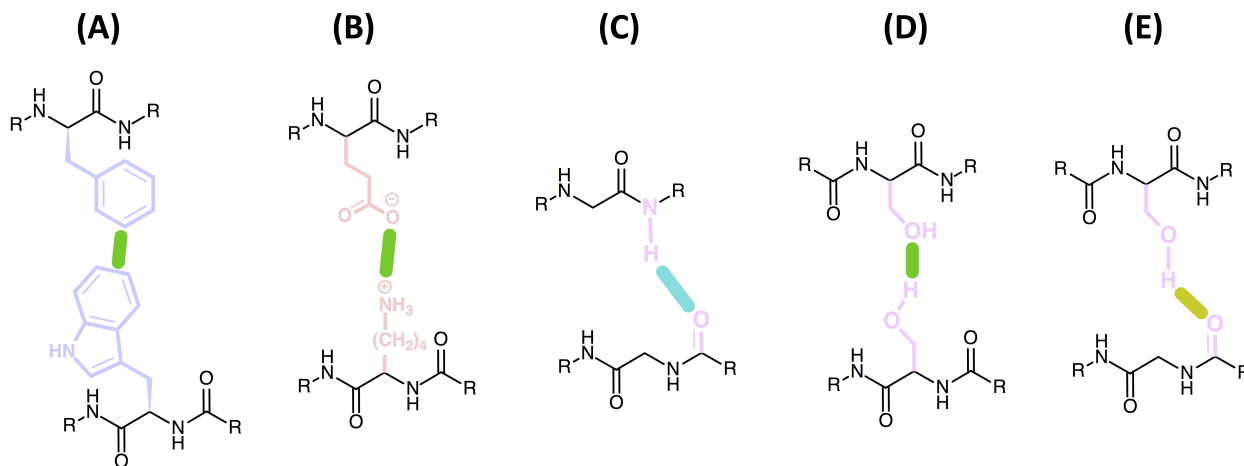


Figure S5: Contact type example overview. **(A)** Hydrophobic contact, colored in blue, between Trp and Phe, established by their side chains (green). **(B)** Salt bridge contact between Glu and Lys (red), established by side chains (green). **(C)** Hydrogen bond between two peptide backbone atoms. **(D)** Hydrogen bond between two Ser residues, established by side chain atoms. **(E)** Hydrogen bond between a Ser side chain and a peptide backbone.

The presented color code is used to color contacts in the PyContact timeline, and the color intensity in the timeline shows the respective strength of the contact in every frame.

Contact lifetime determination

To calculate a contact's lifetime, the user can define a *score threshold* in the PyContact Preferences panel. If the accumulated score of a contact in a specific frame is larger than the specified threshold, the contact is counted as active, otherwise, it is counted as inactive. Then, the mean/median lifetime of a contact is calculated from this binary active/inactive representation.

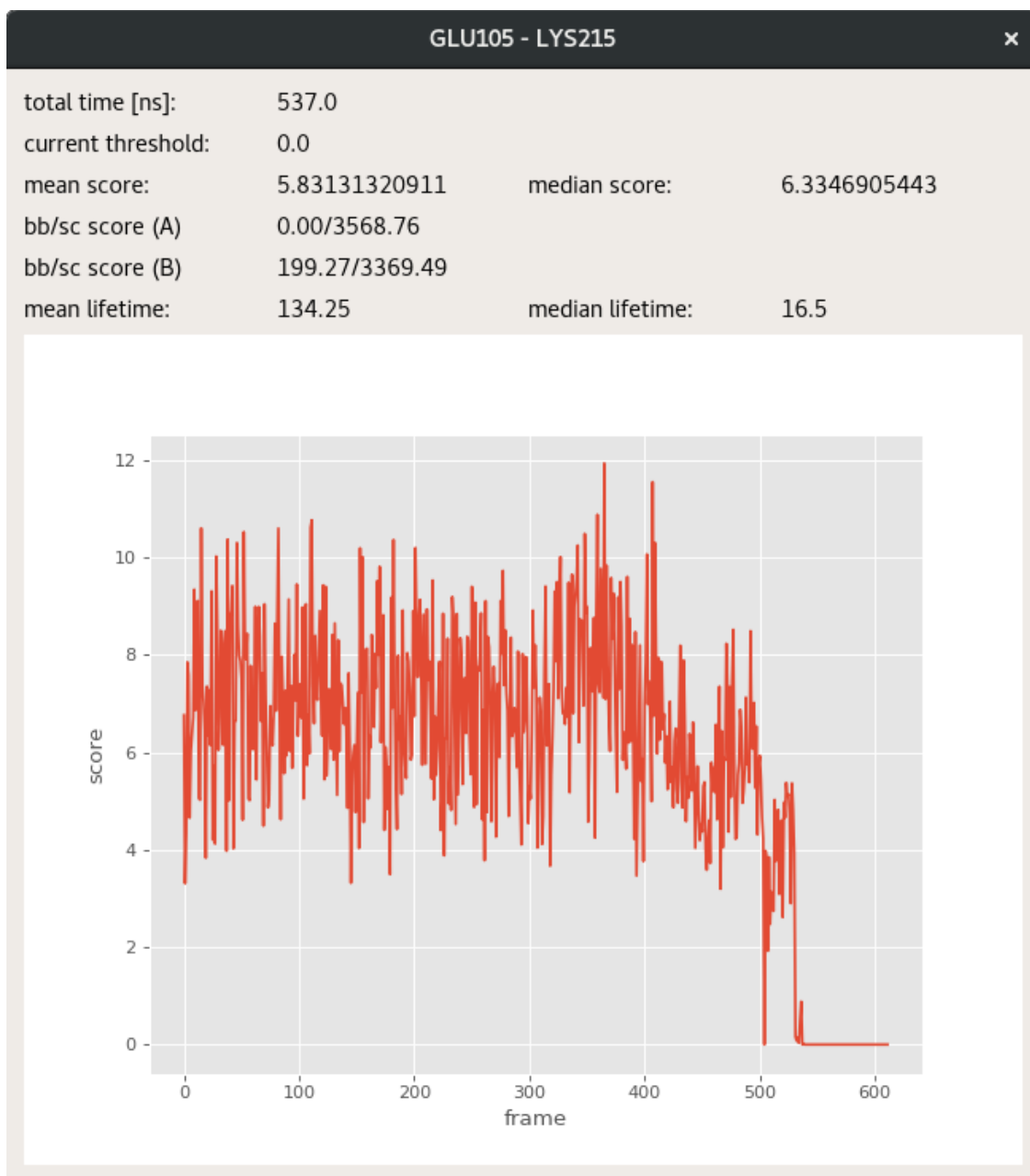


Figure S6: Panel with detailed contact information. Clicking on a contact label in the timeline invokes the displayed panel. A 2D plot of the contact score over time is displayed. In addition, statistical properties, such as total time, mean and median scores, or life times of the selected contact are displayed.

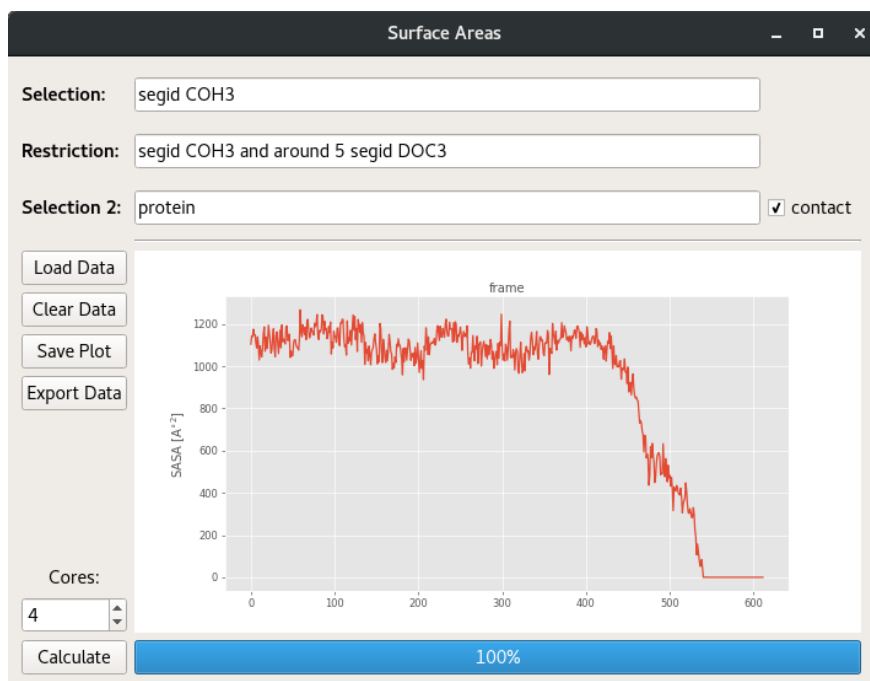


Figure S7: SASA widget. The SASA widget provides easy access to surface and contact area calculations. Atom selections are specified in the top text fields, the number of used CPU cores can be readily changed and the progress of ongoing calculations is displayed by a progress bar on the bottom. The produced data can be exported as an image or a plain text file.

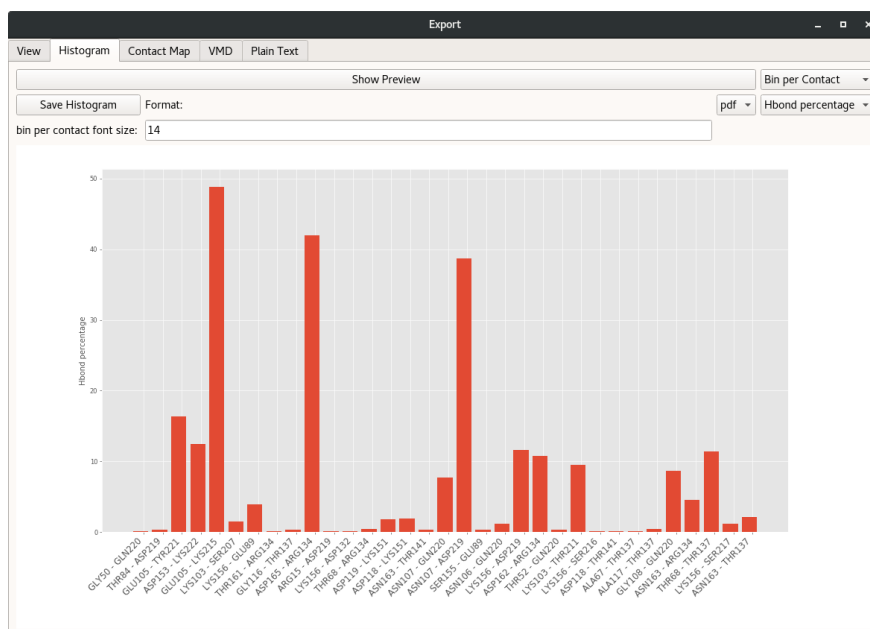


Figure S8: Export Data widget. To share data and create high-quality figures, the Export Data widget provides several plotting and export options, accessible from the top tab bar. Shown is a histogram of hydrogen bond percentages.


```

1  from PyContact.core.Scripting import PyContactJob, JobConfig
2
3  # define input files and parameters
4  job = PyContactJob("/path/to/topology", "/path/to/trajectory", "title",
   ↪ JobConfig(5.0, 2.5, 120, [0,0,1,1,0], [0,0,1,1,0], "segid A",
   ↪ "segid B"))
5  # running the job on 4 cores
6  job.runJob(4)
7  # writing the session to file, the file can be loaded into the user
   ↪ interface
8  job.writeSessionToFile("title.session")
9  # get contact list
10 contacts = job.analyzer.finalAccumulatedContacts
11 # create a Score filter for mean score > 1
12 scoreFilter = ScoreFilter("score", "greater", 1.0, "Mean")
13 # execute the filtering
14 filteredContacts = scoreFilter.filterContacts(contacts)

```

Figure S9: Job Automation example. A simple Python script using the PyContact framework can automatically run a PyContact job with given parameters and finally save the session for subsequent contact inspection in the GUI. Here, the interactions between "segid A" and "segid B" are analyzed and then accumulated to residue names and IDs.

Table S2: P2X₃ protein-lipid contacts

interface ¹	F7, W21, T22, R28, Y37, W41		
headgroups ²	K251, K315, N317, N345, K348		
	chain A	chain B	chain C
step 1 ³	-	S331, Y37, T330, V334, I34	T330, V326
step 2	F329, V326, I322	E11	V337, I431
step 3	E11	-	F329, V332, R28, N31

¹ Interface residues at the border between polar headgroups and hydrophobic membrane core.

² Residues in contact with polar headgroups.

³ Protein contacts to the Na⁺-coordinating lipid residue. Steps of ion permeation pathway as defined in the main text.

Supporting References

1. Sheinerman, F. B., and C. L. Brooks, 1998. Calculations on folding of segment B1 of streptococcal protein G. *Journal of Molecular Biology* 278:439–456.
2. Vermaas, J. V., and E. Tajkhorshid, 2014. Conformational heterogeneity of α -synuclein in membrane. *Biochimica et Biophysica Acta - Biomembranes* 1838:3107–3117.
3. Vermaas, J. V., L. Petridis, X. Qi, R. Schulz, B. Lindner, and J. C. Smith, 2015. Mechanism of lignin inhibition of enzymatic biomass deconstruction. *Biotechnology for Biofuels* 8:217.
4. Torshin, I. Y., I. T. Weber, and R. W. Harrison, 2002. Geometric criteria of hydrogen bonds in proteins and identification of "bifurcated" hydrogen bonds. *Protein Engineering* 15:359–63.