Conformational Analysis of Dual-Scale Simulations of Ubiquitin Chains

A. Berg, C. Peter

published in

NIC Symposium 2020
M. Müller, K. Binder, A. Trautmann (Editors)

Forschungszentrum Jülich GmbH,
John von Neumann Institute for Computing (NIC),
Schriften des Forschungszentrums Jülich, NIC Series, Vol. 50,
ISBN 978-3-95806-443-0, pp. 137.
http://hdl.handle.net/2128/24435

Conformational Analysis of Dual-Scale Simulations of Ubiquitin Chains

Andrej Berg and Christine Peter

Department of Chemistry, Universität Konstanz, 78464 Konstanz, Germany
E-mail: christine.peter@uni-konstanz.de

The analysis of large-scale simulations of (bio)molecular systems generated on high performance computer (HPC) clusters poses a challenge on its own due to the sheer amount of high-dimensional data. To make sense of these data and extract relevant information, techniques such as dimensionality reduction and clustering are used. They can be applied to characterise the sampling of conformational phase space as well as to bridge between simulations on different levels of resolution in multiscale setups. Here, we present an approach to analyse long-timescale simulations and to characterise conformational ensembles of flexibly-linked multidomain proteins using the example of differently covalently conjugated ubiquitin chains. We have analysed exhaustive coarse grained (CG) and atomistic simulations with the help of collective variables (CVs) that are particularly suitable to describe the mutual orientation of different subunits and the protein-protein interfaces between them. These data have been further processed through different dimensionality reduction techniques (relying on multidimensional-scaling like approaches as well as neural network autoencoders). The resulting low-dimensional maps have been used for the characterisation of conformational states and the quantitative comparison of conformational free energy landscapes (from simulations at different levels of resolution as well as of different chain types). With this multiscale simulation and analysis approach it is possible to identify characteristic properties of ubiquitin chains in solution which can be subsequently correlated with experimentally observed linkage- and chain length-specific behaviour.

1 Introduction

A special form of post translational modification of proteins is ubiquitylation, which describes the covalent attachment of the C-terminus of the Ubiquitin (Ub) protein to a lysine side chain in a substrate protein by an isopeptide bond. Ub itself possesses seven lysine (K) residues and additionally its N-terminal methionine (M) for further ubiquitylation which allows for the formation of different types of polyubiquitin (polyUb) chains. These chains can be selectively recognised by ubiquitin binding domains (UBDs). For some UBDs, relative or absolute selectivity for distinct linkage types and chain lengths was shown.[1, 2] Since all possible homotypic chain types, and even mixed and branched polyUb, were found in cells so far, it is supposed that the Ub signalling system serves as a tool to store and transmit information inside the eukaryotic cellular system.[3, 4]

The specific Ub recognition mechanisms are largely unknown, although some first insights can be obtained from experimentally determined (static) structures. Recent studies, however, are indicating that Ub dimers (diUb), which are the shortest possible Ub chains, can assume linkage specific, dynamic ensembles of various conformations in solution.[5–7] In the following we present our previous and ongoing work on the conformational characterisation of Ub conjugates from molecular dynamics (MD) simulations. We are using a dual-scale framework linking protein simulations on the atomistic and coarse grained (CG) level to obtain free-energy landscapes and to identify low free-energy configurations, which are representatives for the thermodynamically weighted conformational ensemble of

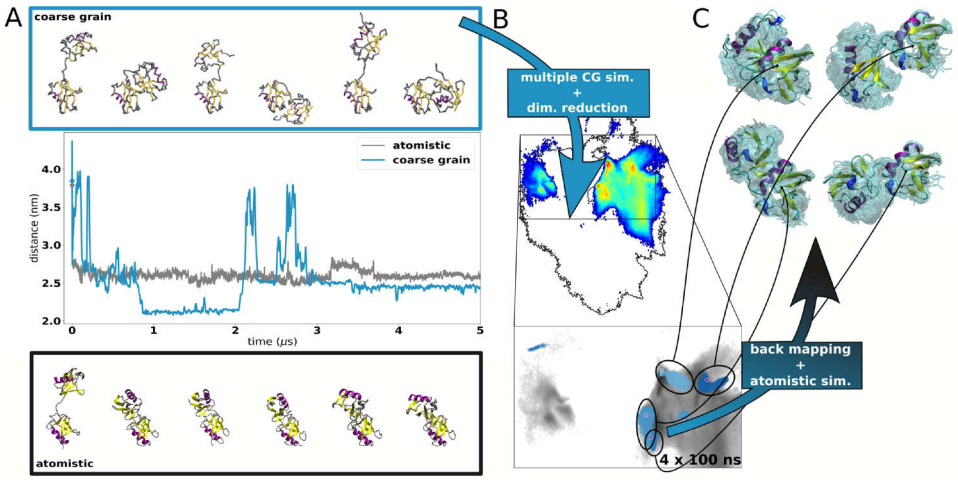

Figure 1. Schematic representation of the multi dimensional procedure. A: Centre of geometry distance between the two Ub moieties inside of diUb as function of time. Atomistic simulation (grey) exhibits a single aggregation event. Corresponding snapshots are shown below in cartoon representation. CG simulation (blue) exhibits repeated transitions between different conformations. Snapshots are shown above as sphere and stick model of back bone beads. B: Two-dimensional representation of a diUb ensemble for a certain linkage type (K27) obtained from multiple independent CG simulations (histogram coloured blue to red for low to high densities of conformations projected to the map). This landscape is shown in zoom inset in grey scale. Blue scatters show atomistic simulations which were started from back-mapped CG low energy simulations. C: Cartoon representation of diUb conformations from atomistic simulations shown in B (blue scatters) aligned on proximal chain.

Ub chains.[8,9] Atomistic simulations have a high level of accuracy but are not able to yield a proper sampling of the conformational equilibrium for diUb, with transitions between different collapsed structures that exhibit contact interfaces between the Ub subunits. For example, Fig. 1A (gray graph) illustrates how one observes a single aggregation event in a 5 μs simulation of K48 linked diUb where a stable contact interface forms that does not open up again. It is therefore not feasible to obtain an equilibrated atomistic ensemble for diUb with a correctly weighted representation of different collapsed structures and protein-protein interfaces. In contrast, on the CG level we can observe repeated transitions between open and aggregated configurations on the μs time scale (Fig. 1A, blue graph).

We used large conformational ensembles from CG MD simulations, a suitable set of collective variables (CVs) and dimensionality reduction to obtain a two-dimensional map of the conformational space (Fig. 1B) that can be interpreted as a free-energy landscape if one assumes that the sampling is converged. This representation can be easily used for structural clustering and intuitive interpretation of complex data sets. From the landscapes we identified low free-energy conformations for back-mapping to the atomistic level for visualisation and as input for MD simulations to obtain high resolution ensembles of these low free-energy states (Fig. 1C). We used the same dimensionality reduction technique to project the conformations from atomistic MD simulations on top of the landscape from the CG ones to validate, that atomistic simulations are staying in the low free-energy area from where they were started on ns time scales.[9] Additionally, such two-dimensional representations can be analysed in many different ways. Certain conformational properties

and features can be assigned to certain landscapes areas.[9] In contrast to high dimensional representations, two-dimensional distributions can be compared easily qualitatively but also quantitatively, for example with the Earth-Mover-Distance[10] algorithm. Clustering using k-means or other density based algorithms can be applied to such low dimensional representations to identify conformational states.

For dimensionality reduction we have successfully used the Sketch-map algorithm in the past, which is a multi-dimensional scaling like method which generates a map by iteratively minimising a nonlinear distance metric for a subset of representative data points (landmarks), focusing on an intermediate range of distances between data points with the help of a sigmoid function. All other points are subsequently projected into the map based on their relative positions to those landmarks.[11, 12] We found that Sketch-map gives good results in combination with a suitable set of input CVs to characterise the interaction of two Ub proteins/subunits to each other.[8, 9] However, this algorithm can be computationally demanding for very large data sets. Therefore, we extended the analysis and applied a new algorithm called EncoderMap, which was developed and improved recently in our group.[13, 14] This method combines the advantage of the pairwise distance based cost function of Sketch-map with a neuronal network autoencoder for dimensionality reduction. Here, we show that with this machine learning technique we were able to reproduce our results with higher efficiency. Thus, the results are robust with respect to the exact dimensionality reduction method used. Furthermore the use of EncoderMap allows to extend the analysis to substantially larger data sets such as simulations of ubiquitin trimers (triUb) and longer chains.

2 Methods

MD simulations

All simulations were performed with the GROMACS simulation package v5.[15] A modified version of the MARTINI force field v2.2 was used for CG simulations.[8, 16] For further details on the simulation setup and force field parameters please see Refs. 8 and 9. For each diUb linkage type (8) 20 simulations (of 10 μs each) were performed (200 μs simulation time for each linker). Two unlinked Ub were simulated in 80 simulations (10 μs each). Ub trimers (triUb) were simulated for 100 μs in 20 simulations which sums up to 2000 μs for each linker type. Structures were used every 100 ps for analysis.

EncoderMap

A detailed description on the functionality of EncoderMap can be found in Ref. 13 which is available as a python package on *github.com*.[a] The network which was used to obtain projections presented in Fig. 2 was trained on diUb conformations only. Based on the high-dimensional distance distribution of CG data, the sigmoid function parameters σ = 5.9, A = 12, B = 4, a = 2, b = 4 were chosen. This network was build up from 144 input nodes followed by three hidden layers of 150 neurons for each layer and two bottle neck nodes for the encoder part. The decoder part of the network was mirrored at the bottle neck layer.

[a] `https://github.com/AG-Peter/encodermap`

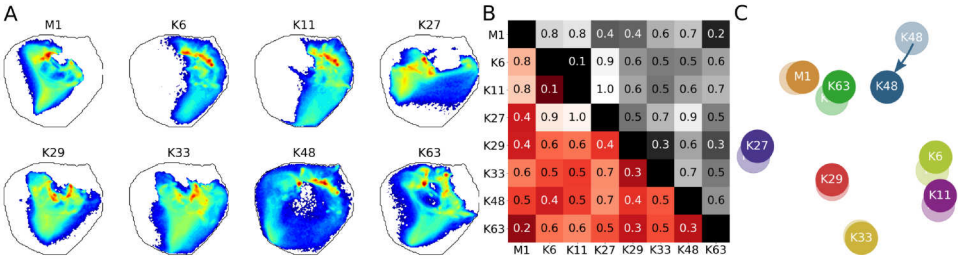

Figure 2. EncoderMap projections of diUb and similarity map. A: Histograms of two-dimensional representations of diUb conformations obtained by EncoderMap dimensionality reduction separated by linkage type (coloured blue to red for low to high densities of conformations projected to the map). Black circular line represents the outer rim of the sum over all linkage types. B: Relative pair-wise EMDs calculated between histograms shown in A (red lower half matrix) and histograms obtained by Sketch-map in Ref. 8 (gray upper half matrix). C: Similarity between two-dimensional conformational landscapes in A is represented by their distance to each other (rich coloured markers). Positioning was obtained by MDS of pair-wise EMD shown in lower half matrix in B. Initial positions for MDS optimisation were taken from Ref. 8 (pale markers).

The network was trained on $1.6 \cdot 10^6$ randomly selected conformations from diUb CG simulations over 10^4 training steps. The network which was used to obtain projections presented in Fig. 3 was trained on randomly selected configurations ($n = 4.6 \cdot 10^6$) from CG simulations of two unlinked Ub proteins. The network topology was the same as for the first neuronal network, but sigmoid function parameters were adjusted to σ = 10, A = 6, B = 10, a = 2, b = 6.

Miscellaneous

To obtain atomistic representations of simulated CG conformations BACKWARD was used for back-mapping.[17] All figures were created using Python v.3.5 and Matplotlib v.2.2.2. Two-dimensional distributions were quantitatively compared using the EMD algorithm as it is implemented in Pyemd v.0.5.1[10] to give relative pair-wise EMDs between two-dimensional distributions of diUb simulations (Fig. 2B). These distances were then used together with metric multi-dimensional scaling (implemented in *sklearn.manifold.MDS*) to image the similarity between two-dimensional representations (Fig. 2C).

3 Results

Dual Scale Simulations of Ub Dimers

We have used CG simulations, which we validated by comparison with atomistic and experimental data, to characterise the conformation of diUb in solution.[8] To this end, we performed extensive MD simulations on the CG level which by now consist of 200 μs simulation time for each of the 8 possible linkage positions for diUb. Residue-wise minimum distances (RMD), calculated for each dimer conformation every 100 ps, have been found to give a good set of CVs for dimensionality reduction. The RMD vector consists of 144 minimum distances between the two subunits, this means that for each of the 72 backbone beads (the last four residues of the highly flexible C-terminus are excluded) in the distal

moiety the minimum distance to any backbone bead in the proximal moiety is calculated and *vice versa*. We used Sketch-map to obtain a low dimensional (2D) representation of this 144 dimensional RMD space. Conformations which have similar RMD values are positioned nearby in 2D while dissimilar conformations are separated. We calculated two dimensional free energy landscapes for each linkage type which we used for qualitative and quantitative interpretation of how the linkage positions influence the conformation of diUb. Quantitative analysis was done by calculation of Earth-Mover-Distances (EMD), which is a metric developed for image recognition but can also be used to compare two-dimensional distributions. This combination of state of the art methods yields a versatile tool for the investigation of protein-protein interaction.

While Sketch-map is rather good for dimensionality reduction of non linear high dimensional data it has also some drawbacks.[18] The most striking for us was, that for training of the model only a limited amount of representative structures, so called landmarks can be used (a few hundreds in the present case). They have to be selected carefully since they should span the whole phase space of the data set. Once the model is trained, each sample from the data set is projected into the map based on optimised positions relative to the landmark points. Although, this method is much faster than conventional MDS with all samples at the same time (which is only feasible with some thousand data points), it can take several days to project a microsecond CG simulation data set. EncoderMap solves these two problems by replacing the MDS algorithm with a neural network autoencoder.[13] A neural network can be trained with a much larger data set (up to several million samples), *i. e.* no landmarks are needed, and such a network, once trained, can be used to project a huge amount of samples with an efficiency, that read and write operations (IO) become the limiting factor. By adding a MDS like pairwise-distance cost contribution to the normal autoencoder cost function upon training, EncoderMap aims (similarly to Sketch-map) at reproducing relative distances between data points in the low-dimensional projection (latent space).

Here, we have applied EncoderMap to the CG data set from Ref. 8 to see if this alters the results of our analysis of diUb conformational landscapes (Fig. 2A). Using very similar parameters for the sigmoid function in the Sketch-map cost contribution of EncoderMap, which determine the outcome of the projection to a large extent, we were able to obtain conformational landscapes which are strikingly similar compared to the previously obtained Sketch-map projections (see Fig. 2 in Ref. 8). A significant difference is observed only for the K48 linked dimer which now spans almost the whole conformational space accessible to diUb. However, this is not due to the dimensionality reduction technique but due to the fact that in the meantime we have also expanded the conformational ensembles by 66 % for each linkage type by addition of multiple new independent runs. It turns out that the K48 ensemble had not been fully converged. As in Ref. 8 the eight two-dimensional distributions are compared to each other by calculating pair-wise EMD which are normalised to a range between zero (for identical distributions) and one, for two very dissimilar maps, K11 and K27 in this case. In Fig. 2B the EMDs for the EncoderMap projections are compared to the data obtained with Sketch-map. Although, some values are different for EncoderMap projections (red bottom half matrix) compared to Sketch-map projections (upper gray half matrix), the overall similarity map for diUb is only slightly affected (Fig. 2C). This similarity map is obtained by again using MDS with the pair-wise EMDs as input for optimisation (the positioning from Ref. 8 has been used

for initialisation). The good agreement between the two different dimensionality reduction methods validates our assumption that RMDs are a suitable set of CVs for a structure based analysis of protein-protein interactions. Furthermore, EncoderMap enables us to analyse much larger data sets, which is required for our ongoing work on the characterisation of simulation data obtained from simulation of Ub trimers and tetramers.

Impact of Linkage on Protein-Protein Interface Formation

In the previous section we demonstrated how a suitable set of CVs together with dimensionality reduction can be used to obtain a (conformational) free energy landscape for differently linked diUb. Since this set of CVs is independent of the linker positions on the Ub surface, this method can be also applied to unlinked Ub, or other proteins in general. Therefore, to characterise the various binding modes Ub can exhibit to a second Ub, we performed extensive CG simulations ($80 \times 10\ \mu s$) of two unlinked Ub. A Sketch-map model was trained on conformations from these simulations only.[9] We obtained a landscape which was able to depict a larger interaction phase space than if diUb data was used

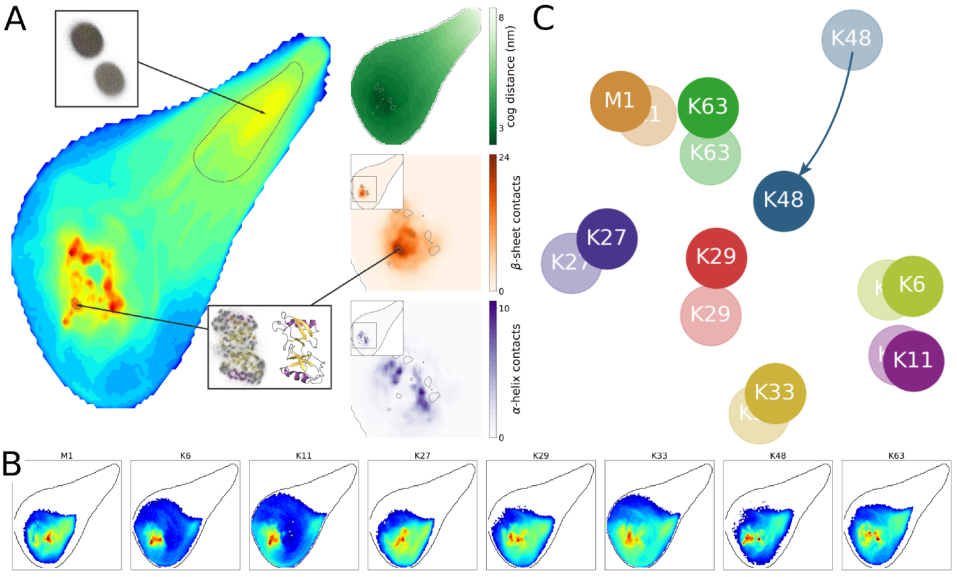

Figure 3. EncoderMap projections of two unlinked Ub and diUb and similarity map. A: Histogram of two-dimensional representations of two unlinked Ub configurations obtained by EncoderMap dimensionality reduction (coloured blue to red for low to high densities of conformations projected to the map). Panels on the right show certain properties of configurations dependent on their position on the landscape: Centre of geometry distance between the two proteins (green); Number of β-sheet contacts (orange) and α-helix contacts between the two proteins. Structure bundles from CG simulations extracted from certain areas on the map demonstrate conformational clustering. Cartoon representation shows back-mapped atomistic structure. B: Histograms, separated by linkage type, of two-dimensional representations of diUb conformations obtained by an EncoderMap network which was trained on unlinked configurations. Black circular line represents the outer rim of landscape in A. C: Similarity between two-dimensional conformational landscapes in B is represented by their distance to each other (rich coloured markers). Positioning was obtained by MDS of pair-wise EMDs. Initial positions for MDS optimisation were taken from Ref. 8 (pale markers).

for training. We found, that the set of CVs which we used as input for dimensionality reduction introduces an arbitrary ordering of chain A and B, which had no evident impact while used for diUb. In a dimer topology chain A is called distal moiety and is linked via its C-terminus to the lysine side chain of moiety B, also called proximal. For two unlinked Ub, which are obviously identical, this is not true any more and the high dimensional data set contains samples which are identical if one would change the chain order from AB to BA. Interestingly, this feature manifests as a two-fold symmetry in the low dimensional projection (as discussed and demonstrated in detail in Ref. 9).

With this model we were able to obtain proper projections for diUb and investigate the immediate (linkage specific) effect of covalent linkage formation on the protein-protein interface between two Ub proteins. Since conformations of diUb are projected into the two-dimensional representation of two unlinked Ub, one can compare these two landscapes directly and observe that some linkage types (*e. g.* K63) select already present conformations while others (*e. g.* K6) obtain conformations which are not stable without any linker. This might be critical for the investigation of the Ub ligase mechanism which happens by a cascade of reactions involving a linkage and substrate specific combination of three enzymes.[3] Additionally, by training the model on a set of configurations which can be formed by two unlinked proteins and should therefore contain all possible configurations at least to some extent, the chance is reduced to obtain unreasonable positions in 2D if new diUb conformations are added to the projection.

In our present work, thanks to the ability of EncoderMap to process a much larger data set for training, we were able to improve the projection of Ub conformations even further – and consequently enhance the characterisation of the protein-protein interface formed by two Ub proteins. Exploiting the inherent symmetry of the projection problem, due to the equivalence of the Ub proteins, we can further improve the map by doubling the high dimensional RMD data set for two unlinked Ub by just changing the chain order. With EncoderMap we are able to use this large data set ($4.6 \cdot 10^6$ samples) for training to obtain an almost perfectly symmetric landscape (Fig. 3A). This projection divides clearly configurations with no contacts between the two chains (high centre of geometry distance, see small green histogram) from aggregated structures (the red highly populated spots in the main map). One can find configurations with a symmetric interaction interface along the symmetry axis and – as it was shown in Ref. 9 already – configurations from two minima which are mirrored at this symmetry axis can be aligned almost perfectly if the chain order is interchanged. Certain interaction interface features, like a high amount of β-sheet contacts (see small orange histogram) can be assigned to certain areas on the map and enhances the interpretation of the configurational space.

The EncoderMap neuronal network, once trained, can be now used to project diUb configurations in a very effective way (shown in Fig. 3B for all linkages). In these landscapes, diUb conformations occupy only the area with lower centre of geometry distance which is obviously due the linkage constraint. We can observe similar overlaps for certain linkage types and we can repeat the quantitative comparison by calculating pair-wise EMD for these distributions, which are then used for MDS optimisation to obtain a similarity map (Fig. 3C). Initial positions were again taken from Ref. 8. Although all markers (saturated colours) are shifted slightly compared to the reference data (pale colours), the overall picture persists, except for K48 which we already identified to be different due to additional sampling compared to the previously published data.

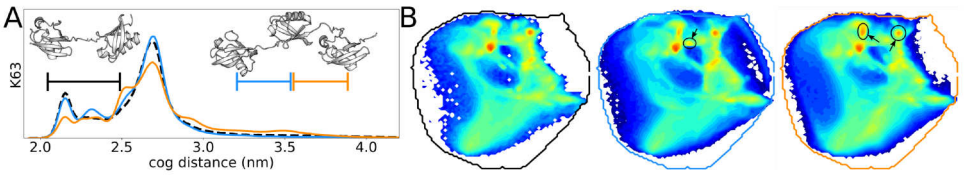

Figure 4. Conformation of K63 linked polyUb. A: Centre of geometry distance distributions between distal and proximal chain in diUb (black) and adjacent chains in triUb (blue, orange) from CG simulations. B: Two-dimensional representations obtained by EncoderMap network which was trained on RMD data from diUb. Conformational landscape of K63 linked diUb (left panel). Interaction landscapes of adjacent Ub moieties in K63 linked triUb (middle and right panel).

Protein-Protein Interface Characterisation in Ub Trimers

To investigate the linkage specific conformational space of longer Ub chains we performed CG simulations of all eight homotypic triUb. In this section we will present preliminary results from these simulations, where we obtained 2000 μs of simulation time for each linkage type (topology). Fig. 4A illustrates how centre of geometry distance distributions between two adjacent chains are altered in triUb compared to a diUb with the same linkage type (K63). Although, this gives already some insight how polyUb chain elongation affects the interaction between Ub moieties, we can use the EncoderMap neuronal network, which was trained to characterise diUb conformations, to obtain interaction maps for triUb.

As discussed above, the two Ub proteins in diUb are not equivalent, which means that the distal moiety ($diUb_d$) is bound by its C-terminus to the proximal moieties ($diUb_p$) lysine side-chain. TriUb has an additional Ub moiety in the middle ($triUb_m$) which is constrained by a covalent bond in both directions, by its lysine side-chain to $triUb_d$ and by its C-terminus to $triUb_p$. Consequently, while diUb has a single interaction interface between $diUb_d$ and $diUb_p$, triUb has two interfaces (for the moment neglecting, that $triUb_d$ and $triUb_p$ can in principle make contact, too). For diUb we had used the RMD vector which describes distances between the distal and proximal moiety to train an EncoderMap neuronal network to obtain a conformational landscape, which can be considered at the same time as an interaction map (Fig. 2A). We applied this trained network to triUb to obtain two interaction maps for each triUb linkage type by providing two different RMD vectors for each triUb conformation, *i. e.* one RMD vector between $triUb_d$ and $triUb_m$ and a second one between $triUb_m$ and $triUb_p$. These maps are shown in Fig. 4B for K63 linked diUb and triUb, where one can see that the overall appearance of the two triUb interaction landscapes is comparable with the conformational landscape from diUb. A closer view reveals that differences in distance distributions are also imaged as altered intensities of certain states in these maps. Since we are able to assign representative structures to each state (red areas in 2D distribution), we can obtain a very detailed picture about the interaction pattern of polyUb chains. By this, we will observe for each linkage type which interaction modes are altered in a triUb compared to diUb and we will conclude on reasons for chain length specificity of certain UBDs.

4 Concluding Remarks

In the present work we present the application of EncoderMap to efficiently obtain low-dimensional representations of complex conformational ensembles of protein conjugates. These representations allow for conformational clustering but also for the visualisation of certain properties and quantitative comparison between histograms of different data sets. With this, intuitive examination of large simulated data sets is possible. Furthermore, selected CVs (RMD) as input for dimensionality reduction are highly transferable and can be used to characterise protein-protein interaction of unlinked proteins as well as certain parts of larger conjugates. In our ongoing work we use the combination of tools presented here to characterise polyUb (up to tetramers) conformations for all homotypic chain types. By this we hope to extend our knowledge about the linkage type and chain length specific conformational properties of polyUb and with this contribute to the understanding of the Ub recognition mechanism.

Acknowledgements

We would like to thank Tobias Lemke for support and helpful suggestions regarding the use of EncoderMap and for critically reading this manuscript. AB is funded by the DFG through SFB969 (Project B09). We are grateful for computational resources of the bwHPC project (DFG grant INST 35/1134-1 FUGG and the state of Baden-Württemberg) for the initial simulations of diUb. We also gratefully acknowledge computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JUWELS at Jülich Supercomputing Centre (JSC).

References

1. F. Ikeda, N. Crosetto, and I. Dikic, *What Determines the Specificity and Outcomes of Ubiquitin Signaling?*, Cell **143**, 677–681, 2010.
2. X. Zhang, A. H. Smits, G. B. A. van Tilburg, P. W. T. C. Jansen, M. M. Makowski, H. Ovaa, and M. Vermeulen, *An Interaction Landscape of Ubiquitin Signaling*, Molecular Cell **65**, 941–955.e8, 2017.
3. D. Komander and M. Rape, *The Ubiquitin Code*, Annual Review of Biochemistry **81**, 203–229, 2012.
4. K. N. Swatek and D. Komander, *Ubiquitin Modifications*, Cell Research **26**, 399–422, 2016.
5. C. Castañeda, A. Chaturvedi, C. M.Camara, J. E.Curtis, S. Krueger, and D. Fushman, *Linkage-specific Conformational Ensembles of Non-canonical Polyubiquitin Chains*, Physical Chemistry Chemical Physics **18**, 5771–5788, 2016.
6. C. A. Castañeda, E. K. Dixon, O. Walker, A. Chaturvedi, M. A. Nakasone, J. E. Curtis, M. R. Reed, S. Krueger, T. A. Cropp, and D. Fushman, *Linkage via K27 bestows ubiquitin chains with unique properties among polyubiquitins*, Structure **24**, 423–436, 2016.
7. A. Kniss, D. Schuetz, S. Kazemi, L. Pluska, P. E. Spindler, V. V. Rogov, K. Husnjak, I. Dikic, P. Güntert, T. Sommer, T. F. Prisner, and V. Dötsch, *Chain Assembly*

and Disassembly Processes Differently Affect the Conformational Space of Ubiquitin Chains, Structure **26**, 249–258.e4, 2018.
8. A. Berg, O. Kukharenko, M. Scheffner, and C. Peter, *Towards a molecular basis of ubiquitin signaling: A dual-scale simulation study of ubiquitin dimers*, PLoS computational biology **14**, e1006589, 2018.
9. A. Berg and C. Peter, *Simulating and analysing configurational landscapes of protein-protein contact formation*, Interface Focus **9**, 20180062, 2019.
10. O. Pele and M. Werman, *A linear time histogram metric for improved sift matching*, in Computer Vision–ECCV 2008, Springer, 495–508, 2008.
11. M. Ceriotti, G. A. Tribello, and M. Parrinello, *Simplifying the Representation of Complex Free-energy Landscapes Using Sketch-map*, Proceedings of the National Academy of Sciences **108**, 13023–13028, 2011.
12. O. Kukharenko, K. Sawade, J. Steuer, and C. Peter, *Using Dimensionality Reduction to Systematically Expand Conformational Sampling of Intrinsically Disordered Peptides*, Journal of Chemical Theory and Computation **12**, 4726–4734, 2016.
13. T. Lemke and C. Peter, *EncoderMap: Dimensionality Reduction and Generation of Molecule Conformations*, Journal of Chemical Theory and Computation **15**, 1209–1215, 2019, PMID: 30632745.
14. T. Lemke and C. Peter, *EncoderMap(II): Visualizing important molecular motions with improved generation of protein conformations*, Journal of Chemical Information and Modeling, 2019, doi:10.1021/acs.jcim.9b00675.
15. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*, SoftwareX **1**, 19–25, 2015.
16. L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S.-J. Marrink, *The MARTINI Coarse-Grained Force Field: Extension to Proteins*, J. Chem. Theory and Comput. **4**, 819–834, 2008.
17. T. A. Wassenaar, K. Pluhackova, R. A. Böckmann, S. J. Marrink, and D. P. Tieleman, *Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models*, Journal of Chemical Theory and Computation **10**, 676–690, 2014.
18. G. A. Tribello and P. Gasparotto, *Using Dimensionality Reduction to Analyze Protein Trajectories*, Frontiers in Molecular Biosciences **6**, 46, 2019.