**Molecular dynamics-assisted interpretation of experimentally determined IDP conformational components: the case of human alpha synuclein.**

Oscar Palomino-Hernandez 1,2,3,4, Carlo Santambrogio 5, Giulia Rossetti 1,6,7, Claudio O. Fernandez 8,9, Rita Grandori 5, Paolo Carloni 1,2,10

1 Computational Biomedicine, Institute for Neuroscience and Medicine (INM-9) and Institute for Advanced Simulations (IAS-5), Forschungszentrum Jülich, 52425 Jülich, Germany

2 Faculty of Mathematics, Computer Science and Natural Sciences, RWTH Aachen, 52425 Aachen, Germany

3 Computation-Based Science and Technology Research Center, The Cyprus Institute, 2121 Nicosia, Cyprus

4 Institute of Life Science, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel

5 Department of Biotechnology and Biosciences, University of Milan-Bicocca, 20126 Milan, Italy

6 Department of Oncology, Hematology, Oncology, Hemostaseology, and Stem Cell Transplantation University Hospital Aachen, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany

7 Jülich Supercomputing Center (JSC), Forschungszentrum Jülich, 52425 Jülich, Germany

8 Max Planck Laboratory for Structural Biology, Chemistry and Molecular Biophysics of Rosario (MPLbioR, UNR-MPIbpC) and Instituto de Investigaciones para el Descubrimiento de Fármacos de Rosario (IIDEFAR, UNR-CONICET), Universidad Nacional de Rosario, S2002LRK Rosario, Argentina

9 Department of NMR-Based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany

10 Institute for Neuroscience and Medicine (INM-11) Forschungszentrum Jülich, 52425 Jülich, Germany

**ABSTRACT**

Mass spectrometry and single molecule force microscopy are two experimental approaches able to provide structural information of intrinsically disordered proteins (IDPs). These techniques allow the dissection of conformational ensembles in their main components, although at a low-resolution level.

In this work, we interpret the results emerging from these experimental approaches on human alpha synuclein (AS) through a molecular perspective, analyzing a 73 microsecond-long simulation of the protein in explicit solvent.

The combined theoretical and experimental data provide a description of AS main conformers, shedding light into intramolecular interactions and overall structural compactness. This approach can be easily transferred to other IDPs.

**INTRODUCTION**

Intrinsically disordered proteins (IDPs) lack a defined and ordered three-dimensional structure [1,2](https://sciwheel.com/work/citation?ids=345483,533528&pre=&pre=&suf=&suf=&sa=0,0). They can exhibit rather diverse conformational ensembles in solution, retaining a high degree of structural disorder. The experimental technique most suited to determine IDP structural ensembles at atomic resolution is NMR [3,4](https://sciwheel.com/work/citation?ids=1631275,3810130&pre=&pre=&suf=&suf=&sa=0,0), at times aided by computational methods including molecular dynamics simulations [5](https://sciwheel.com/work/citation?ids=11971732&pre=&suf=&sa=0). In addition, several other techniques can provide complementary, low-resolution information. These include native mass spectrometry (MS) and single-molecule force spectroscopy (SMFS). The first technique provides information related to protein structural compactness: its charge-state distribution (CSD) analysis allows the estimation of the solvent-accessible surface area (SASA)in solution [6–9](https://sciwheel.com/work/citation?ids=4518628,12065792,12065796,12065795&pre=&pre=&pre=&pre=&suf=&suf=&suf=&suf=&sa=0,0,0,0), while ion mobility (IM) yields values for the rotationally-averaged collisional cross-section of the ions in the gas phase [10–12](https://sciwheel.com/work/citation?ids=1019109,3049046,12065799&pre=&pre=&pre=&suf=&suf=&suf=&sa=0,0,0). The second technique gives information on the strength of the intramolecular interactions, which aids in the characterization of the distinct conformers in solution.

In this paper we explore the use of molecular dynamics (MD) simulations to provide insight on the structural determinants of IDP conformers as emerging from MS and SMFS. In this work, we focus on human alpha-synuclein (AS). This IDP plays a key role for the progression of Parkinson’s disease [13–15](https://sciwheel.com/work/citation?ids=924800,24270,3301222&pre=&pre=&pre=&suf=&suf=&suf=&sa=0,0,0). The protein consists of 140 amino acids and is highly disordered, both in solution and in cell cultures [16](https://sciwheel.com/work/citation?ids=7946698&pre=&suf=&sa=0). Nonetheless, it can adopt well-defined structures when interacting with membrane or cellular partners [17](https://sciwheel.com/work/citation?ids=8184087&pre=&suf=&sa=0). Our work is motivated by a recently published (73 microsecond-long**[[1]](#footnote-1)**) MD simulation of AS in explicit solvent [18](https://sciwheel.com/work/citation?ids=5231125&pre=&suf=&sa=0). This provides an unprecedented molecular view on the AS conformational ensemble and a rather insightful interpretation of the experimental data. This information is used to provide insight on the major conformational components (CCs) emerging from MS and SMFS experiments performed by some of us [19,20](https://sciwheel.com/work/citation?ids=4039011,11971913&pre=&pre=&suf=&suf=&sa=0,0). In the first case, the deconvolution of MS spectra yields CCs that range from compact to expanded. In the second, the CCs identified describe strong-interacting, weak-interacting and random coil conformations, according to the pulling force required to achieve mechanical unfolding [19,20](https://sciwheel.com/work/citation?ids=4039011,11971913&pre=&pre=&suf=&suf=&sa=0,0).

**METHODS**

An structural ensemble of alpha synuclein (residues 1-140), simulated for roughly 73 microseconds was retrieved from D. E. Shaw research [18](https://sciwheel.com/work/citation?ids=5231125&pre=&suf=&sa=0), and used with structures sampled evenly every 10 ns, accumulating a total of 7320 conformations. This ensemble has been validated experimentally against SAXS measurements [18](https://sciwheel.com/work/citation?ids=5231125&pre=&suf=&sa=0).

Several structural descriptors were computed. Hydrogen-bonds and charge contacts were computed with the hydrogen bond and Salt Bridges modules of VMD [21](https://sciwheel.com/work/citation?ids=357281&pre=&suf=&sa=0), respectively. Solvent-accessible surface area was computed with the FreeSASA module [22](https://sciwheel.com/work/citation?ids=2143536&pre=&suf=&sa=0). The Accessible Solvent Area was computed from the mean values of each residue over the Maximum Accessible Solvent Area values obtained in ref. [23](https://sciwheel.com/work/citation?ids=1044563&pre=&suf=&sa=0). Radius of gyration was computed with the *gmx gyrate* module [24](https://sciwheel.com/work/citation?ids=1599369&pre=&suf=&sa=0). Hydrophobic contacts were computed as the distance between the centers of mass of the side chains of hydrophobic residues. A contact was considered when the distance was lower than a certain cutoff. This was computed with MDAnalysis [25](https://sciwheel.com/work/citation?ids=7099724&pre=&suf=&sa=0).

Feature clustering was performed using the AgglomerativeClustering module of scikit-learn [26](https://sciwheel.com/work/citation?ids=12065880&pre=&suf=&sa=0). A cluster cut-off of 3 was used for **M0-M3,** while a cluster cut-off of 4 was used for **C0-C2.** The quality of the clustering parameters was established by the *elbow* method [27](https://sciwheel.com/work/citation?ids=8310249&pre=&suf=&sa=0). The Silhouette and Calinski-Harabasz scores were employed [28,29](https://sciwheel.com/work/citation?ids=2681601,1210252&pre=&pre=&suf=&suf=&sa=0,0).

The population standard deviation from the MD simulation was obtained through a jackknife approach, based on 50 samples covering 90% of the original trajectory. To quantify the overlap between the two sets of CCs, a dimensionality reduction was performed using the implementation of t-Stocastic Neighbor Embedding in Python (t-SNE) [30](https://sciwheel.com/work/citation?ids=12065836&pre=&suf=&sa=0). It was observed that multiple runs of t-SNE converged into the same representation at a value of perplexity of 40.

Two distinct estimations of SASA from the average charge state of conformers in Native-MS data (Zav) were employed. Model **a** was obtained by the empirical relationship of Zav with the SASA of globular proteins, obtained from folded PDB structure [7](https://sciwheel.com/work/citation?ids=12065792&pre=&suf=&sa=0). Analogously, model **b** exploits SASA values deriving from Monte Carlo simulations of the unfolded ensemble generated by ProtSA [31](https://sciwheel.com/work/citation?ids=12065886&pre=&suf=&sa=0). These two models provide similar results for small proteins, while this is not the case for large proteins [32](https://sciwheel.com/work/citation?ids=12097249&pre=&suf=&sa=0).

**RESULTS AND DISCUSSION**

In a first approach, we attempt to identify as many CCs from the MD trajectory as many from MS and SMFS experiments [19,20](https://sciwheel.com/work/citation?ids=4039011,11971913&pre=&pre=&suf=&suf=&sa=0,0) (4 and 3, respectively). To this aim, we perform a clustering procedure using different descriptors, including solvent-accessible surface area (SASA), radius of gyration (RG), end-to-end distance (EE), number of intramolecular hydrogen bonds (HB), intramolecular hydrophobic contacts (HC), and salt-bridges (SB, see Methods section for details). Of course, the analysis provides different CCs depending on the numbers and type of descriptor. It turns out that the clustering provides 4 CCs if one uses SASA, RG and EE as descriptors and 3 CCs using HB and HC. The CCs obtained are thus compared with those by MS and SMFS, respectively. The comparison allows us to uncover some of the salient structural determinants of the CCs based on the simulation data.

**1. Comparison with MS**

**1.1. General features**

The CCs obtained for the comparison between the MS and MD are listed in order of decreasing compactness (**C0-C3**),in Table 1.

The populations of the CCs obtained by MD and MS are qualitatively similar. However, they show some difference in relative amounts: the compact species are more represented and the extended speciesare less represented in MD than in MS. Similarly, the collisional cross-section mean values display a good general agreement between simulations and experiments. They increase on passing from **C0** to **C3**, with MS values being smaller for compact species and larger for extended states compared to MD. We notice that the calculated CCS are in the same range as the experimental ones and not statistically different from each other.

For the SASA values, two slightly different linear regressions models for calculating it from MS data have been generated, for either folded or unfolded proteins (Table X, see Methods section for details). The calculated SASA values agree fairly with both the experimentally-derived ones.

This discrepancies in the measurements above may be caused by several reasons, including: **(i)** the AS ensemble was simulated in solution by MD while it is studied in vanishing solvent, under electrospray conditions, by MS; **(ii)** limitations of the time-scale and/or the force fields for the MD [33](https://sciwheel.com/work/citation?ids=5616637&pre=&suf=&sa=0); **(iii)** uncertainties as regarding the clustering in the MD. Moreover, comparisons between CCS of proteins simulated in water solution and those investigated in the gas phase by MS experiments should obviously be made with great caution.

**1.2. Structural features**

Structural features were analyzed in terms of the four CCs. The calculated radius of gyration (RG) increases on passing from **C0** to **C3 (**Fig. 1, **A**). Instead, the intramolecular hydrophobic contacts (HC), intramolecular hydrogen bonds (HB) and salt bridges (SB) decrease (Fig 1, **A**). The helical content is similar across all the conformations. The beta-strand content and the end-to-end distance (EE) increase from **C0** to **C1,** and from **C1** to **C2/C3**. The latter two CCs are not statistically different in this regard.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **C0** | **C1** | **C2** | **C3** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Population (%) | MS | 5-10 | 20-30 | 40-50 | 20-30 |
| MD | 16.37 ± 6.94 | 31.94 ± 7.19 | 32.94 ± 8.18 | 18.74 ± 5.51 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Collisional cross-section (nm2) | IM-MS | 21.88 | 23.88 | 27.22 | 29.44 |
| MD | 23± 2.84 | 25± 2.24 | 27± 2.54 | 28± 2.25 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SASA  (nm2) | MSa | 70-75 | 90-120 | 125-160 | 165-195 |
| MSb | 72-80 | 105-150 | 175-235 | 265-310 |
| MD | 124.29 ± 2.54 | 135.91 ± 2.48 | 146.09 ± 2.46 | 158.17 ± 2.3 |

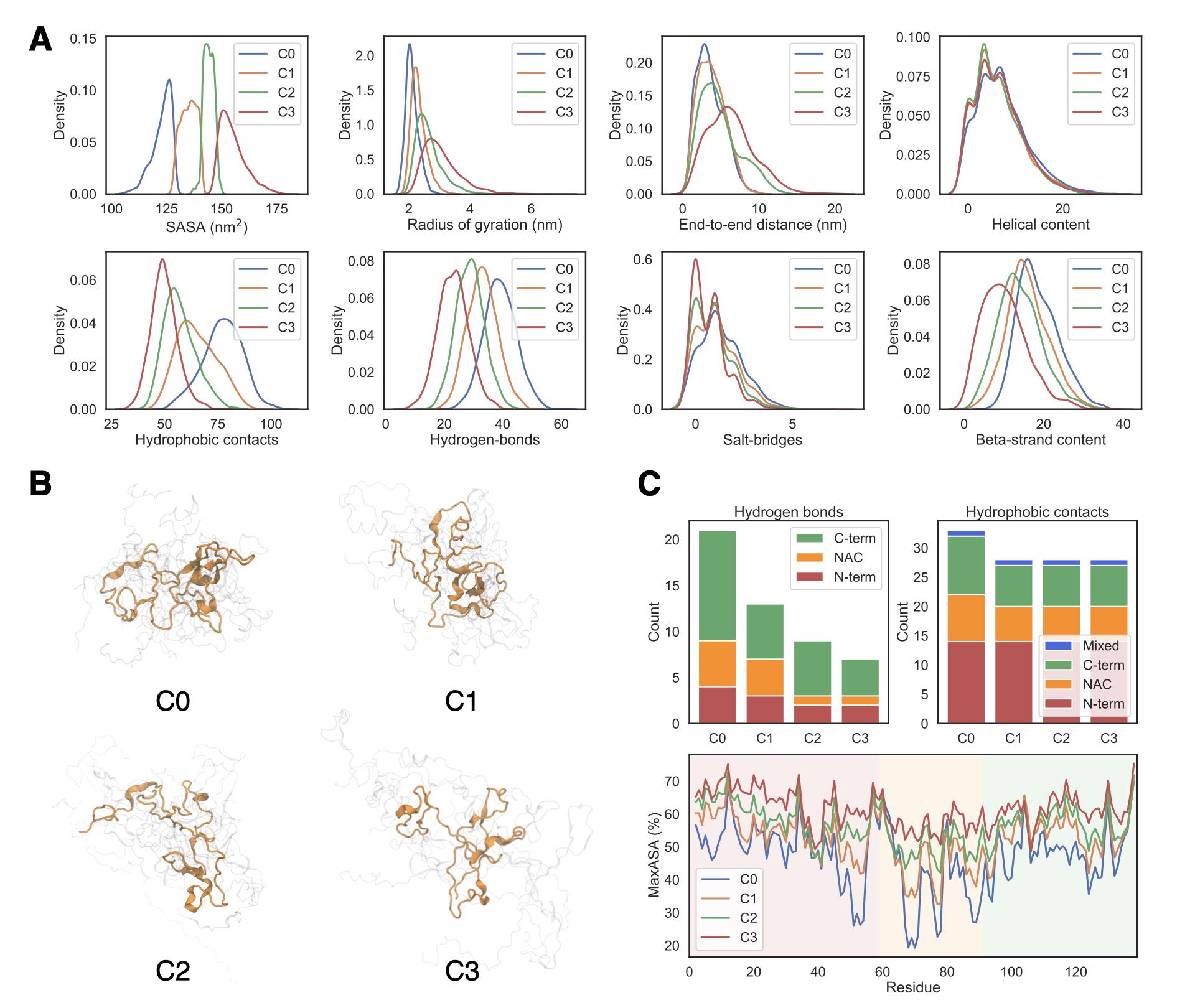
***Table 1.******CCs emerging from MS.*** *Experimental and calculated values of populations, SASA and collisional cross-section of the four CCs. The MS-derived SASA values were derived using models for (a) folded structures and (b) unfolded structures (see Methods).*

Delving further, the features were analyzed in terms of residues and regions in AS primary structure. **C0** is stabilized through intra-region HB in each of the three regions, particularly in the C-terminal one.In **C1**, the number of HB content decreases, notably more in the C-terminal one. For **C2** and **C3**, the HB content reduction is more noticeable in the NAC region. Some relevant interactions conserved across all four clusters include residues located mainly within the N-terminal and C-terminal regions.

The hydrophobic contact content (HC, defined in Methods) decreases on passing from **C0** to **C1**-**C3**, especially because of the loss of hydrophobic contacts inside the NAC and C-terminal region. Remarkably, clusters **C1-C3** share the same amount of HC.

At the single-residue level, the change in SASA was registered as the percentage of maximal accessible surface area (‘MaxASA’, see Methods). The largest increaseacross **C0-C3** is found in residues V52, T54, V70, V71, V74, V82, F94, L100, V118, Y125. Overall, this indicates that, going from **C0-C3**, the residues in the N-terminal and C-terminal regions seem to have similar SASA, while the residues in the NAC region significantly increase their contact with the solvent (Fig. 1, **C**).

On the whole, it can be seen that the four clusters conserve HB interactions within the N-terminal and C-terminal regions (E13, V15, E35, V37, D98, E110, I112, E114, M116, D134, Y136), and HC interactions within the N-terminal and NAC regions (F4, V16, A17, A18, L38, Y39, V70, V77, A89, A90, P117).



***Figure 1. Selected results from the molecular dynamics simulation. (A)*** *Selected properties of* ***C0-C3****;**(****B****) Top representative structures of the CCs from a structure-based clustering.* ***(C)*** *Number of hydrogen bonds, number of hydrophobic contacts across* ***C0-C3****. Maximal accessible solvent area (MaxASA) for all residues of* ***C0-C3****.*

**2. Comparison with SMFS data**

**2.1. General features**

The CCs obtained by SMFS are listed in Table 2, in order of decreasing intramolecular strength (from strong-interacting **(M0**), to weak-interacting (**M1**) and random coil (**M2**) [19,20](https://sciwheel.com/work/citation?ids=4039011,11971913&pre=&pre=&suf=&suf=&sa=0,0)). At the very qualitative level, we associate here the number of intramolecular interactions (HB, HC) to these classes. In other words, we make the plausible assumption that the content of intramolecular interactions distinguishes among the strong-interacting, weak-interacting and random coil conformations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **M0** | **M1** | **M2** |
| Population (%) | SMFS | ~10 | ~30 | ~60 |
| MD | 15.86 ± 2.25 | 29.06 ± 2.03 | 55.09 ± 2.96 |
| Intermolecular strength | SMFS | Strong interacting | Weak interacting | Random coil |
| MD | Multiple HB and HC contacts | Reduced HB and HC contacts | Mostly HC contacts |

***Table 2. CCs emerging from SMFS experiments.*** *Experimental and calculated values of the populations of the three CCs.*

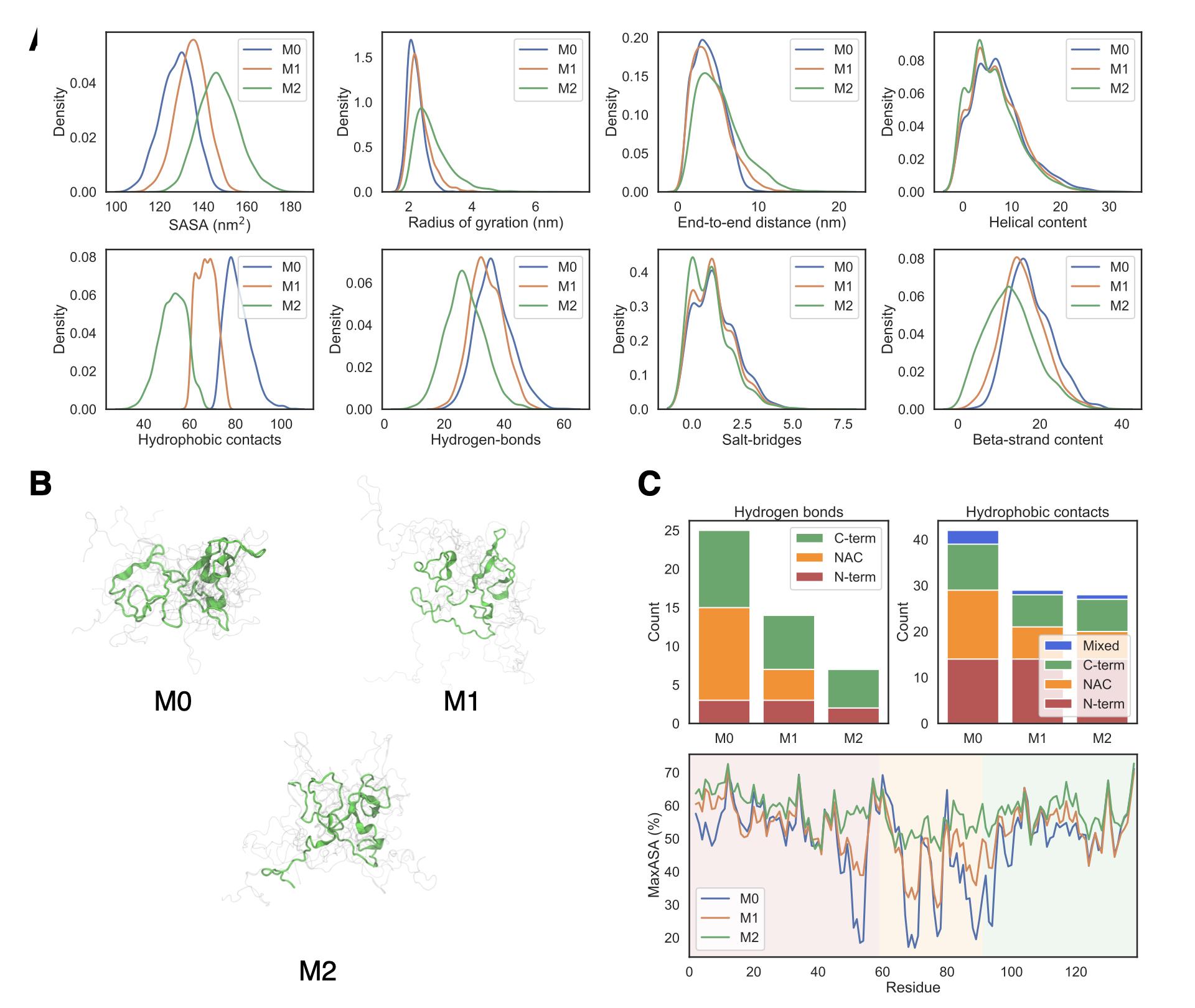
The populations of the CCs obtained by MD and SMFS are also qualitatively similar. When discussing intramolecular interactions it is observed that, on passing from **M0** to **M2,** the amounts of HB and HC get strongly reduced (*vide infra*), in line with the experimental observations.

**2.2. Structural features**

With respect to global structural features, RG and the EE are very similar in **M0** and **M1**, and increase, instead, for **M2**. The coil content increases, while the beta strand content decreases on passing from **M0** to **M2**. (Figure 2, **A**). The helical content does not change significantly across the conformers.

The SASA increases on passing from **M0** to **M2 (**126.4 ± 1.33, 136.13 ± 1.89,147.38 ± 2.24 nm2, respectively). It is also shown that the residues in the N-terminal and C-terminal regions do not modify significantly their SASA, as the NAC region does. The largest increase is found for residues V52, T54, V55, A69, V70, V71, V74, V78, V82, I88, A89, A91, F94, L100.

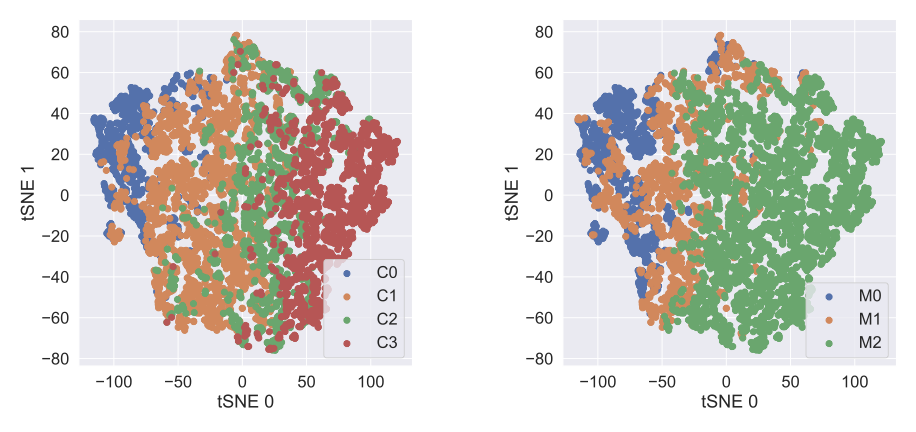
The largest contributions of HB and HC come from the NAC and the C-terminal regions. The latter contributions decrease progressively in M1, and further, in **M2**. Interestingly, the HB formed by residues E13, V15, D98, E110, I112, E114, M116, Q134, Y136, along with HC interactions formed by F4, V16, A17, A18, L38, Y39, V70, V77, A89, A90, P117, are conserved across all the conformers.

****

***Figure 2. Selected results from the molecular dynamics simulations. (A)***  *Properties of* ***M0-M2****, emerging from the MD simulations;**(****B****) Top representative structures of the CCs from a structure-based clustering. (C) Number of hydrogen bonds, number of hydrophobic contacts across* ***M0-M2****. Maximal accessible solvent area (MaxASA) for all residues of* ***C0-C3****.*

3. **Overlap between M0-M2 and C0-C3**

The overlap between the two sets of CCs was calculated by using a dimensionality reduction projection by t-SNE, and the cluster members were projected on it. Fig. 3 shows that **M2** has a large overlap with **C2**+**C3**, sharing *ca.* 80% of the conformations, and covering *ca.* 50% of the total conformational space. The comparison between **M1** vs. **C1** and **M0** vs. **C0** shows a significantly reduced overlap, sharing *ca.* 50% of the conformations for both comparisons. Both clustering schemes define in similar ways the expanded/random coil structures, but the compact clusters appear to be defined in different ways. This finding is in line with the fact that MS and SMFS agree at best in assessing changes in the random coil CC while capturing different aspects of the compact or partially collapsed CCs.



**Figure 3.** Clusters C and M in the common reduced feature space of tSNE.

**Conclusions**

This work has provided details on the structural determinants of CCs observed experimentally by MS and SMFS for free AS in solution. Remarkably, computational simulations were able to merge the partially contrasting results from the experiments. By associating the latter with CCs obtained from a very long simulation of the protein, we were able to describe the intermolecular interactions and the SASA in the different regions of the protein.

The **C0**-**C3** CCs show structures with compact cores, which involve intra N-terminal and intra NAC interactions. On the other hand, the **M0**-**M2** CCs highlights the role of the NAC region, and its ability to form HB and HC when found in compacted states.

This study points out that different features for clustering focus on specific traits of the CCs. Although this is not an issue when trying to describe the unfolded states, it leads to partially different results when describing compact states, covering somewhat different populations.

The methodology could be used for other IDPs in a rather straightforward way. Thus, the combined use of extensive MD with MS and/or SMFS may be an additional, useful tool for the investigations of this fascinating class of proteins.

[**Bibliography**](https://sciwheel.com/work/bibliography)

[(1) Tompa, P. Intrinsically Unstructured Proteins. *Trends Biochem. Sci.* **2002**, *27*, 527–533.](https://sciwheel.com/work/bibliography/345483)

[(2) Frimpong, A. K.; Abzalimov, R. R.; Uversky, V. N.; Kaltashov, I. A. Characterization of Intrinsically Disordered Proteins with Electrospray Ionization Mass Spectrometry: Conformational Heterogeneity of Alpha-Synuclein. *Proteins* **2010**, *78*, 714–722.](https://sciwheel.com/work/bibliography/533528)

[(3) Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. Mapping Long-Range Interactions in Alpha-Synuclein Using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2005**, *127*, 476–477.](https://sciwheel.com/work/bibliography/1631275)

[(4) Gibbs, E. B.; Cook, E. C.; Showalter, S. A. Application of NMR to Studies of Intrinsically Disordered Proteins. *Arch. Biochem. Biophys.* **2017**, *628*, 57–70.](https://sciwheel.com/work/bibliography/3810130)

[(5) Peris-Díaz, M. D.; Guran, R.; Domene, C.; de Los Rios, V.; Zitka, O.; Adam, V.; Krężel, A. An Integrated Mass Spectrometry and Molecular Dynamics Simulations Approach Reveals the Spatial Organization Impact of Metal-Binding Sites on the Stability of Metal-Depleted Metallothionein-2 Species. *J. Am. Chem. Soc.* **2021**, *143*, 16486–16501.](https://sciwheel.com/work/bibliography/11971732)

[(6) Kaltashov, I. A.; Mohimen, A. Estimates of Protein Surface Areas in Solution by Electrospray Ionization Mass Spectrometry. *Anal. Chem.* **2005**, *77*, 5370–5379.](https://sciwheel.com/work/bibliography/4518628)

[(7) Testa, L.; Brocca, S.; Grandori, R. Charge-Surface Correlation in Electrospray Ionization of Folded and Unfolded Proteins. *Anal. Chem.*](https://sciwheel.com/work/bibliography/12065792) **[2011](https://sciwheel.com/work/bibliography/12065792)**[,](https://sciwheel.com/work/bibliography/12065792) *[83](https://sciwheel.com/work/bibliography/12065792)*[, 6459–6463.](https://sciwheel.com/work/bibliography/12065792)

[(8) Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. Conformational Effects in Protein Electrospray-Ionization Mass Spectrometry.](https://sciwheel.com/work/bibliography/12065796) *[Mass Spectrom. Rev.](https://sciwheel.com/work/bibliography/12065796)* [**2016**, *35*, 111–122.](https://sciwheel.com/work/bibliography/12065796)

[(9) Santambrogio, C.; Natalello, A.; Brocca, S.; Ponzini, E.; Grandori, R. Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. *Proteomics* **2019**, *19*, e1800060.](https://sciwheel.com/work/bibliography/12065795)

[(10) Konijnenberg, A.; Butterer, A.; Sobott, F. Native Ion Mobility-Mass Spectrometry and Related Methods in Structural Biology. *Biochim. Biophys. Acta* **2013**, *1834*, 1239–1256.](https://sciwheel.com/work/bibliography/1019109)

[(11) May, J. C.; Morris, C. B.; McLean, J. A. Ion Mobility Collision Cross Section Compendium. *Anal. Chem.* **2017**, *89*, 1032–1044.](https://sciwheel.com/work/bibliography/3049046)

[(12) Gabelica, V.; Shvartsburg, A. A.; Afonso, C.; Barran, P.; Benesch, J. L. P.; Bleiholder, C.; Bowers, M. T.; Bilbao, A.; Bush, M. F.; Campbell, J. L.; et al. Recommendations for Reporting Ion Mobility Mass Spectrometry Measurements. *Mass Spectrom. Rev.* **2019**, *38*, 291–320.](https://sciwheel.com/work/bibliography/12065799)

[(13) Fernández, C. O.; Hoyer, W.; Zweckstetter, M.; Jares-Erijman, E. A.; Subramaniam, V.; Griesinger, C.; Jovin, T. M. NMR of Alpha-Synuclein-Polyamine Complexes Elucidates the Mechanism and Kinetics of Induced Aggregation. *EMBO J.* **2004**, *23*, 2039–2046.](https://sciwheel.com/work/bibliography/924800)

[(14) Stefanis, L. α-Synuclein in Parkinson’s Disease. *Cold Spring Harb. Perspect. Med.* **2012**, *2*, a009399.](https://sciwheel.com/work/bibliography/24270)

[(15) Flagmeier, P.; Meisl, G.; Vendruscolo, M.; Knowles, T. P. J.; Dobson, C. M.; Buell, A. K.; Galvagnion, C. Mutations Associated with Familial Parkinson’s Disease Alter the Initiation and Amplification Steps of α-Synuclein Aggregation. *Proc Natl Acad Sci USA*](https://sciwheel.com/work/bibliography/3301222) **[2016](https://sciwheel.com/work/bibliography/3301222)**[,](https://sciwheel.com/work/bibliography/3301222) *[113](https://sciwheel.com/work/bibliography/3301222)*[, 10328–10333.](https://sciwheel.com/work/bibliography/3301222)

[(16) Strohäker, T.; Jung, B. C.; Liou, S.-H.; Fernandez, C. O.; Riedel, D.; Becker, S.; Halliday, G. M.; Bennati, M.; Kim, W. S.; Lee, S.-J.; et al. Structural Heterogeneity of α-Synuclein Fibrils Amplified from Patient Brain Extracts. *Nat. Commun.* **2019**, *10*, 5535.](https://sciwheel.com/work/bibliography/7946698)

[(17) Fernández, R. D.; Lucas, H. R. Mass Spectrometry Data Confirming Tetrameric α-Synuclein N-Terminal Acetylation. *Data Brief* **2018**, *20*, 1686–1691.](https://sciwheel.com/work/bibliography/8184087)

[(18) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc Natl Acad Sci USA*](https://sciwheel.com/work/bibliography/5231125) **[2018](https://sciwheel.com/work/bibliography/5231125)**[,](https://sciwheel.com/work/bibliography/5231125) *[115](https://sciwheel.com/work/bibliography/5231125)*[, E4758–E4766.](https://sciwheel.com/work/bibliography/5231125)

[(19) Sandal, M.; Valle, F.; Tessari, I.; Mammi, S.; Bergantino, E.; Musiani, F.; Brucale, M.; Bubacco, L.; Samorì, B. Conformational Equilibria in Monomeric Alpha-Synuclein at the Single-Molecule Level.](https://sciwheel.com/work/bibliography/4039011) *[PLoS Biol.](https://sciwheel.com/work/bibliography/4039011)* [**2008**, *6*, e6.](https://sciwheel.com/work/bibliography/4039011)

[(20) Corti, R.; Marrano, C. A.; Salerno, D.; Brocca, S.; Natalello, A.; Santambrogio, C.; Legname, G.; Mantegazza, F.; Grandori, R.; Cassina, V. Depicting Conformational Ensembles of α-Synuclein by Single Molecule Force Spectroscopy and Native Mass Spectroscopy. *Int. J. Mol. Sci.* **2019**, *20*.](https://sciwheel.com/work/bibliography/11971913)

[(21) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38, 27.](https://sciwheel.com/work/bibliography/357281)

[(22) Mitternacht, S. FreeSASA: An Open Source C Library for Solvent Accessible Surface Area Calculations. [Version 1; Peer Review: 2 Approved]. *F1000Res.* **2016**, *5*, 189.](https://sciwheel.com/work/bibliography/2143536)

[(23) Tien, M. Z.; Meyer, A. G.; Sydykova, D. K.; Spielman, S. J.; Wilke, C. O. Maximum Allowed Solvent Accessibilites of Residues in Proteins. *PLoS ONE* **2013**, *8*, e80635.](https://sciwheel.com/work/bibliography/1044563)

[(24) 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* **2015**, *1–2*, 19–25.](https://sciwheel.com/work/bibliography/1599369)

[(25) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; et al. Mdanalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*; Proceedings of the python in science conference; SciPy, 2016; pp. 98–105.](https://sciwheel.com/work/bibliography/7099724)

[(26) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**.](https://sciwheel.com/work/bibliography/12065880)

[(27) Thorndike, R. L. Who Belongs in the Family? *Psychometrika* **1953**, *18*, 267–276.](https://sciwheel.com/work/bibliography/8310249)

[(28) Rousseeuw, P. J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *Journal of Computational and Applied Mathematics* **1987**, *20*, 53–65.](https://sciwheel.com/work/bibliography/2681601)

[(29) Calinski, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Communications in Statistics - Theory and Methods* **1974**, *3*, 1–27.](https://sciwheel.com/work/bibliography/1210252)

[(30) van der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *Journal of Machine Learning Research* **2008**.](https://sciwheel.com/work/bibliography/12065836)

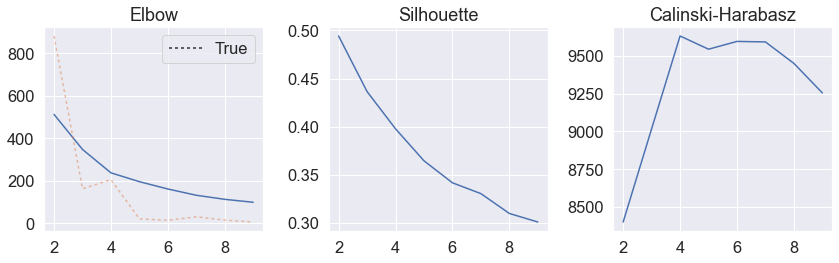
[(31) Estrada, J.; Bernadó, P.; Blackledge, M.; Sancho, J. ProtSA: A Web Application for Calculating Sequence Specific Protein Solvent Accessibilities in the Unfolded Ensemble. *BMC Bioinformatics* **2009**, *10*, 104.](https://sciwheel.com/work/bibliography/12065886)

[(32) Allison, T. M.; Barran, P.; Cianférani, S.; Degiacomi, M. T.; Gabelica, V.; Grandori, R.; Marklund, E. G.; Menneteau, T.; Migas, L. G.; Politis, A.; et al. Computational Strategies and Challenges for Using Native Ion Mobility Mass Spectrometry in Biophysics and Structural Biology. *Anal. Chem.* **2020**, *92*, 10872–10880.](https://sciwheel.com/work/bibliography/12097249)

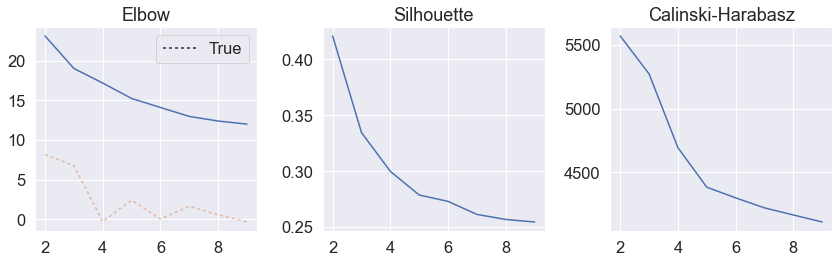
[(33) Bottaro, S.; Lindorff-Larsen, K. Biophysical Experiments and Biomolecular Simulations: A Perfect Match? *Science* **2018**, *361*, 355–360.](https://sciwheel.com/work/bibliography/5616637)

**Supplementary information:**

Clustering metrics for the **compactness features** (Clusters **C0**-**C3**).



Clustering metrics for the **mechanostability features** (Clusters **M0**-**M2**).



The elbow method [1] at the total within-cluster sum of square (WSS) as a function of the number of clusters. The location of a knee in the plot is usually considered as an indicator of the appropriate number of clusters because it means that adding another cluster does not improve the partition. This method seems to suggest 4 clusters for the compactness space, and 3 clusters for the mechanostable space.

The Silhouette score [2] is calculated using the mean intra-cluster distance, as well as the mean distance to the nearest cluster for each sample in the dataset, while the Calinski-Harabasz score [3] is a variance ratio measurement which measures the ratio between within-cluster dispersion and between-cluster dispersion. For both scores, the higher the value, the better the clustering.

Most frequent residues involving hydrophobic contacts through the compactness descriptors

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Region** | **C0** | **C1** | **C2** | **C3** |
| N-terminal | F4 | F4 | F4 | F4 |
| N-terminal | V16 | V16 | V16 | V16 |
| N-terminal | A17 | A17 | A17 | A17 |
| N-terminal | A18 | A18 | A18 | A18 |
| N-terminal | L38 | L38 | L38 | L38 |
| N-terminal | Y39 | Y39 | Y39 | Y39 |
| NAC | V70 | V70 | V70 | V70 |
| NAC | V71 |  |  |  |
| NAC | V74 |  |  |  |
| NAC | A76 |  |  |  |
| NAC | V77 | V77 | V77 | V77 |
| NAC | A89 | A89 | A89 | A89 |
| NAC | A90 | A90 | A90 | A90 |
| C-terminal | A91 |  |  |  |
| C-terminal | F94 |  |  |  |
| C-terminal | P117 | P117 | P117 | P117 |
| C-terminal | Y125 |  |  |  |
| C-terminal | M127 |  |  |  |

Most frequent residues involving hydrophobic contacts through the mechanical stability descriptors

|  |  |  |  |
| --- | --- | --- | --- |
| **Region** | **M0** | **M1** | **M2** |
| N-terminal | F4 | F4 | F4 |
| N-terminal | V16 | V16 | V16 |
| N-terminal | A17 | A17 | A17 |
| N-terminal | A18 | A18 | A18 |
| N-terminal | L38 | L38 | L38 |
| N-terminal | Y39 | Y39 | Y39 |
| N-terminal | V52 |  |  |
| NAC | A69 |  |  |
| NAC | V70 | V70 | V70 |
| NAC | V71 |  |  |
| NAC | V74 |  |  |
| NAC | A76 | A76 |  |
| NAC | V77 | V77 | V77 |
| NAC | A78 |  |  |
| NAC | V82 |  |  |
| NAC | A89 | A89 | A89 |
| NAC | A90 | A90 | A90 |
| C-terminal | A91 |  |  |
| C-terminal | F94 |  |  |
| C-terminal | P117 | P117 | P117 |
| C-terminal | Y125 |  |  |
| C-terminal | M127 |  |  |

[1] Robert L. Thorndike (December 1953). "Who Belongs in the Family?". Psychometrika. 18 (4): 267–276. doi:10.1007/BF02289263.

[2] Peter J. Rousseeuw (1987). “Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis”. Computational and Applied Mathematics 20: 53-65.

[3] T. Calinski and J. Harabasz, 1974. “A dendrite method for cluster analysis”. Communications in Statistics

1. The original 60 microsecond-long trajectory reported in this paper had issues with periodic image interactions, thus, a second, 73 microsecond-long simulation produced by the same group was used here. [↑](#footnote-ref-1)