Analysis of conformation manifolds for intrinsically disordered proteins

Marina Botnari[1], Nathalie Sibille[2], Jung-Hsin Lin[3], Thérèse E Malliavin[4]

Université de Lorraine, Laboratoire de Physique et Chimie Théoriques, École Doctorale Chimie-Mécanique-Matériaux-Physique, Metz, France[1]
Centre de Biologie Structurale, Centre National de la Recherche Scientifique, Montpellier, France[2]
Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan[3]
Université de Lorraine, Laboratoire de Physique et Chimie Théoriques, Centre National de la Recherche Scientifique, Metz, France[4]

Determining a protein's functional conformation is one of the greatest challenges in structural biology. Although folded proteins easily interact with numerous partners, roughly 35–50% of the human proteome does not adopt stable three-dimensional structures[1]. Instead, many proteins are partially (IDRs) or entirely (IDPs) intrinsically disordered. Their inherent flexibility facilitates interactions with multiple partners and participation in various physiological and pathological processes, such as and neurodegenerative diseases and cancers[2]. However, standard structural techniques fail to capture these proteins' plasticity.

To address this limitation, we propose a computational strategy to explore the local folding of IDPs. First, secondary structure elements are assigned from experimentally derived backbone chemical shifts using δ2d[3]. These observations define probabilistic "boxes" on the Ramachandran map that reflect the likelihood of specific backbone torsion angles (ϕ and ψ). Due to the dynamic nature of IDPs, the protein is divided into intervals of residues. To manage the conformational complexity, we employ a threading-augmented interval branch and prune (TAiBP)[4] method to generate an ensemble of candidate conformations, which are clustered using a self-organizing map (SOM) to select representative structures. After fragment assembly, side chains are added via molecular dynamics simulations using NAMD[5].

For validation, we apply our method to a set of PED proteins, including the wild-type[6] C-terminal domains of the vasopressin receptor, growth hormone secretagogue receptor type 1a, and B2-adrenergic receptor, along with their phosphomimetic analogues[7]. Several analyses compare the conformations obtained from the PED proteins with those generated by TAiBP.

References:

(1) Förster, D.; Idier, J.; Liberti, L.; Mucherino, A.; Lin, J.-H.; Malliavin, T. E. Low-Resolution Description of the Conformational Space for Intrinsically Disordered Proteins. *Sci. Rep.* **2022**, *12* (1), 19057. https://doi.org/10.1038/s41598-022-21648-9.

(2) Malliavin, T. E.; Mucherino, A.; Lavor, C.; Liberti, L. Systematic Exploration of Protein Conformational Space Using a Distance Geometry Approach. *J. Chem. Inf. Model.* **2019**, *59* (10), 4486–4503. https://doi.org/10.1021/acs.jcim.9b00215.

(3) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **2012**, *51* (11), 2224–2231. https://doi.org/10.1021/bi3001825.

(4) Worley, B.; Delhommel, F.; Cordier, F.; Malliavin, T. E.; Bardiaux, B.; Wolff, N.; Nilges, M.; Lavor, C.; Liberti, L. Tuning Interval Branch-and-Prune for Protein Structure Determination. *J. Glob. Optim.* **2018**, *72* (1), 109–127. https://doi.org/10.1007/s10898-018-0635-0.

(5) Phillips, J. C.; Hardy, D. J.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem. Phys.* **2020**, *153* (4), 044130. https://doi.org/10.1063/5.0014475.

(6) Guillien, M.; Mouhand, A.; Fournet, A.; Gontier, A.; Martí Navia, A.; Cordeiro, T. N.; Allemand, F.; Thureau, A.; Banères, J.-L.; Bernadó, P.; Sibille, N. Structural Insights into the Intrinsically Disordered GPCR C-Terminal Region, Major Actor in Arrestin-GPCR Interaction. *Biomolecules* **2022**, *12* (5), 617. https://doi.org/10.3390/biom12050617.

(7) Guillien, M.; Mouhand, A.; Sagar, A.; Fournet, A.; Allemand, F.; Pereira, G. A. N.; Thureau, A.; Bernadó, P.; Banères, J.-L.; Sibille, N. Phosphorylation Motif Dictates GPCR C-Terminal Domain Conformation and Arrestin Interaction. *Structure* **2023**, *31* (11), 1394-1406.e7. https://doi.org/10.1016/j.str.2023.08.011.