



Contribution ID: 39

Type: Talk

## Integrative ensemble modeling that utilizes distance distribution restraints

Tuesday, 13 February 2024 13:45 (25 minutes)

AlphaFold2 predictions for the human proteome reveal that the great majority of all human proteins contain intrinsically disordered regions (IDRs) or do not contain a folded domain at all. The structure of such proteins must be represented by an ensemble of conformers. However, most experimental techniques provide ensemble-average restraints, which do not contain direct information on width of the ensemble. Even single-molecule techniques, such as Förster Resonance Energy Transfer (smFRET) average over many conformers that are visited by molecular dynamics at time scale of the measurement. Using restraints from such techniques one cannot evaluate the probability for an individual conformer to belong to the ensemble.

In contrast, distance distributions between two spin-labelled sites in a protein, as they can be obtained by pulse electron paramagnetic resonance experiments, such as DEER, directly inform on ensemble width and can be used to compute such probabilities. Therefore, distance distribution restraints can already be used at the stage of sampling conformer space and protect against unrealistic narrowing of the ensemble during ensemble reweighting. To make use of these features in ensemble modelling, we have developed the toolbox MMMx.

MMMx implements the RigiFlex approach that can model flexible multi-domain proteins by first computing a distributed rigid-body arrangement of the folded domains in the Rigi step. The IDRs that link the folded domains or are terminal are then constructed in the Flex step. The raw ensemble generated in this way is then reweighted by the EnsembleFit step, which can additionally take into account small-angle scattering curves and NMR paramagnetic relaxation enhancements (PREs). Use of smFRET mean-distance restraints will be implemented soon.

We illustrate MMMx ensemble modelling on the examples of (1) the intrinsically disordered N-terminal domain of FUS in the dispersed and condensed state formed by liquid-liquid phase separation, (2) full-length hnRNP A1 that consist of a folded domain and a long C-terminal IDR, and (3) SRSF1 $_{\Delta RS}$  in its free form and in complexes with two short single-stranded RNAs. More information on MMMx can be found in the online documentation.

### Submitting to:

Integrative Computational Biology workshop

**Primary author:** JESCHKE, Gunnar (ETH Zürich)

**Presenter:** JESCHKE, Gunnar (ETH Zürich)

**Session Classification:** Protein conformational flexibility and solution experiments