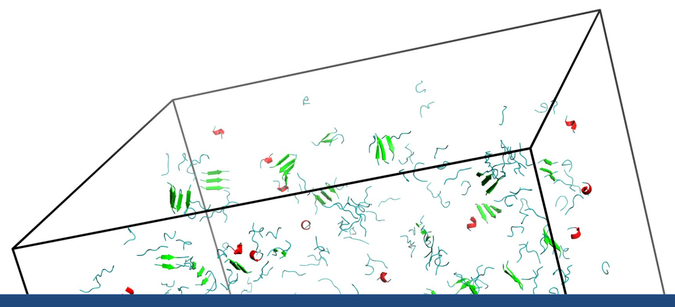


All-Atom MCMC-Simulation of Protein Folding with ProFASi

Olav Zimmermann, Jan Meinke, Sandipan Mohanty
SDL Biology, Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

CECAM Tutorial on MCMC Simulation using ProFASi Sep 22-26, 2025 #1401

The ‘second half’ of the Protein Folding Problem

Novel transformer-based models like those implemented in Alphafold can learn the sequence-structure relationships found in nature from large sequence and structure training data and have largely solved 3d-structure prediction.
The second half of the Protein Folding Problem, i.e. understanding ***how*** protein molecules autonomously fold into their specific native 3d structures, still remains. The scarcity of experimental data on folding processes of proteins and the detailed insights requested suggest that physics based simulation methods are the best option to approach this problem.

Why MCMC?

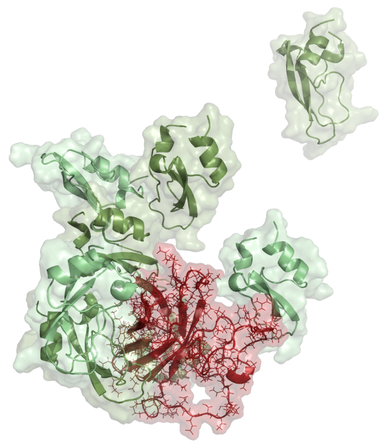
Even on special purpose supercomputers Molecular Dynamics trajectories have not reached more than a few milliseconds. Folding of a protein can take seconds or even minutes and it’s unbiased simulation requires full atomic detail. For this regime (long time scales, all atom) we have been developing ProFASi [1], an extremely fast and versatile software for all-atom Markov Chain Monte Carlo (MCMC) simulations, as an efficient alternative.

Example use cases

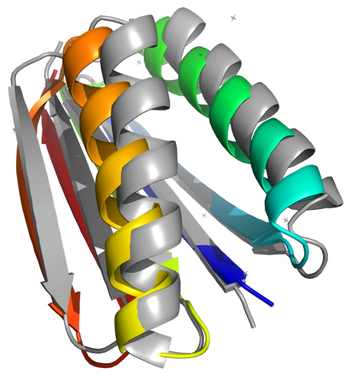
Large scale MCMC simulations on the JSC Supercomputers have provided unprecedented insight into long timescale processes.

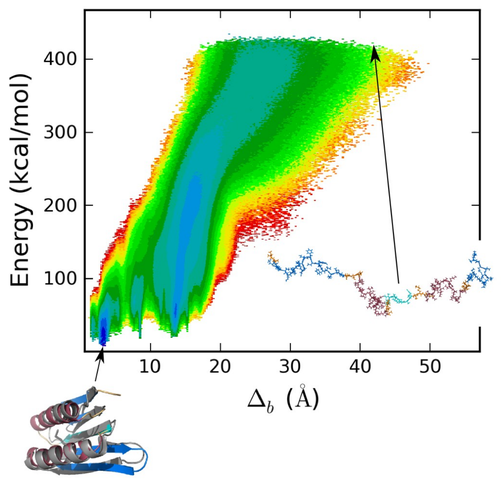
ProFASi has been used to study:

- Protein folding
- Peptide aggregation
- Protein folding with protein crowders
- Unraveling conformational ensembles of cancer related unstructured proteins
- Thermostability of enzymes



Folding of TOP7 [3]:



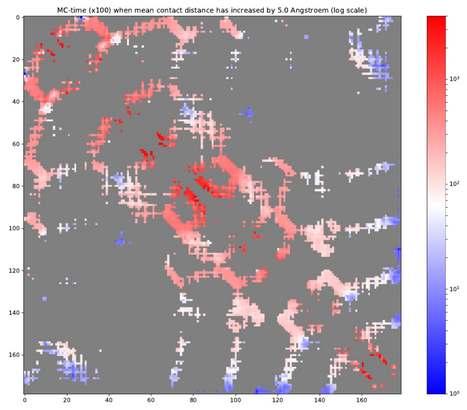


- Experimental folding time of TOP7 (92 aa, designed): ~one second!
- 19k core-hrs per folding event (25 foldings in total)
- Free energy minimum at native folded structure (rmsd: 3.5 Å)

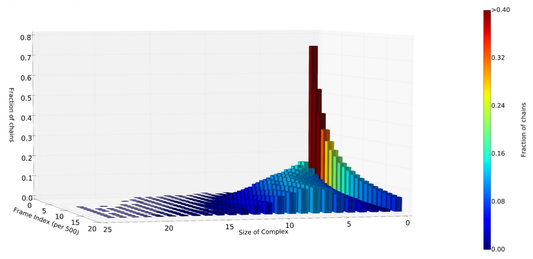
Thermal unfolding of BsLipA mutants [4]:

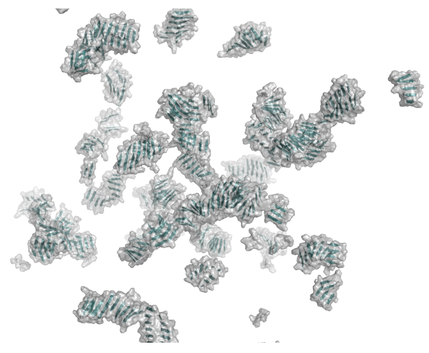
Systematic study of mutation effects on the thermostability of LipaseA (*B.subtilis)*

- 96 MC-unfolding simulations per mutant
- Unfolding starts at C-terminus
- Predicted unfolding pathways match the ones predicted by a rigidity theory based approach using Molecular Dynamics.



Aggregation of PHF6





- Main promoter of tau protein aggregation
- Up to 512 peptides (6 aa), full aggregation observed (s.a. top of poster)
- 32 temperatures x 32 replicas, 10^7 samples, 8TB data
- High propensity of single-layer parallel beta-sheets

ProFASi: high perfomance MCMC simulation software

The key advantage of MD simulations is that they generate physically realistic trajectories of classical many body systems.
The key advantage of Monte Carlo simulations is that they do not.
(Frenkel & Smit, Understanding Molecular Simulation, 3rd ed., chapter 13)

Simulation: MC updates of single or multiple torsion angles as well as global rotation and translation of individual chains or clusters accepted/rejected using Metropolis-Hastings criterion. Fixed bond lengths and bond angles.

Force fields: physics based force fields (Lund FF) [2], implicit solvent model.

Algorithms: Basic MCMC, Simulated Tempering, Parallel Tempering (Replica Exchange), Wang-Landau, Multicanonical

Features: observables (e.g. RMSD, Rg, Q), and constraints can be used as pseudo energy terms. Several accessory tools for setup and analysis.

Availability: https://gitlab.jsc.fz-juelich.de/slbio/profasi

Contact: olav.zimmermann@fz-juelich.de

ProFASi: recent development and outlook

Code features: modern code architecture (core: C++23, ~75k LoC), extremely fast: >5000 moves per core-second, multithreaded energy evaluation, speculative moves, MPI parallelized replica exchange simulations. Open source (GPLv3).

User extensible: Creation of custom MC updates, observables, abd algorithms due to plugin architecture. ProFASi can be used as a library. Connection to trajectory analysis tools via Python bindings.

Upcoming version (v3.0, Q3/25): new force field, MD-plugin, new Jupyter-based tutorials (see also CECAM course above), Web GUI (exper.), …

References

[1] Irbäck, A.; Mohanty, S. PROFASI: a Monte Carlo simulation package for protein folding and aggregation. J. Comput. Chem. 2006, 27, 1548– 1555, DOI: 10.1002/jcc.20452

[2] Irbäck, A.; Mitternacht, S.; Mohanty, S. An effective all-atom potential for proteins. PMC Biophys. 2009, 2, 2, DOI: 10.1186/1757-5036-2-2

[3] Mohanty, S.; Meinke, J. H.; Zimmermann, O. Folding of Top7 in unbiased all-atom Monte Carlo simulations. Proteins 2013, 81, 1446– 1456, DOI: 10.1002/prot.24295

[4] Nutschel, C.; Fulton, A.; Zimmermann, O.; Schwaneberg, U.; Jaeger, K.E.; Gohlke, H. Systematically Scrutinizing the Impact of Substitution Sites on Thermostability and Detergent Tolerance for *Bacillus subtilis* Lipase A. J. Chem. Inf. Model. 2020 60 (3), 1568-1584, DOI: 10.1021/acs.jcim.9b00954