STATEMENT OF RESEARCH INTERESTS

Ulrich H.E. Hansmann

A) Overview

Proteins are crucial components in the molecular machinery of cells, responsible for transporting molecules or catalyzing biochemical reactions. Despite remarkable progress in experimental techniques for producing and characterizing proteins, a detailed understanding of folding and interaction of proteins is still lacking. While simulations can complement experiments in probing folding, aggregation, binding, and other fundamental processes in cells, they are computationally challenging for realistic protein models. This is because these processes happens on time scales (seconds to days) that exceed the ones available in atomistic molecular dynamics simulations. As the computational requirements for sampling the energy landscape increase exponentially with the size of a sytem, It follows that exhaustive explorations of the landscape of proteins and protein aggregates, required to determine the relative weight of the various states, and the pathways connecting them, are rarely possible.

A significant part of my research is concerned with overcoming this sampling problem in simulations of biological macromolecules. It's center piece is the development and advancement of numerical techniques such as the *generalized-ensemble* approach [1]. Related is the development and dissemination of new software. Our programs are collected in the free program package SMMP (**S**imple **M**olecular **M**echanics for **P**roteins) [2]; and in recent years increasingly as add-ons to the often-used GROMACS package [3].

Current applications of our techniques focus on fold-switching proteins such as the transcription factor RfaH or the 93-residue lymphotactin; and the assembly, conversion and propagation of amyloids. Protein-ligand binding and protein interaction networks belong to the same research direction and provide an interface for collaborations with experimental groups.

Our research is supported by the National Institutes of Health (NIH) under contract R01-GM120634, and under sub-contracts from NIH grants GM120578 and AG062292.

B) Background: Generalized-Ensemble Sampling and Related Techniques

The key idea behind generalized-ensemble based techniques is to replace the canonical simulations, where the crossing of an energy barrier of height ΔE is suppressed by a factor $\propto \exp(-\Delta E/k_B T)$ (k_B is the Boltzmann constant and T the temperature of the system), with schemes that both ensure sampling of low-energy configurations *and* avoid trapping in local minima. For instance, in multicanonical sampling [4] the weight $w(E)$ leads to a distribution

$$P(E) \propto n(E)\; w_{mu}(E) = \text{const}, \tag{1}$$

with $n(E)$ the density of states. A free random walk in the energy space is performed that allows the simulation to escape from any local minimum. From this simulation one can calculate the thermodynamic average of any physical quantity A by re-weighting: [5]

$$< \mathcal{A} >_T \;\; = \;\; \frac{\int dx\; \mathcal{A}(x)\; w^{-1}(E(x))\; e^{-E(x)/k_B T}}{\int dx\; w^{-1}(E(x))\; e^{-E(x)/k_B T}}\;, \tag{2}$$

where x labels the configurations. Note that the weight $w(E)$ is not *a priori* known and estimators have to be determined iteratively.

We have introduced Energy Landscape Paving (ELP) [6] as an optimization technique that relies on a modified energy steering the search away from regions already explored:

$$w(\tilde{E}) = e^{-\tilde{E}/k_B T} \quad \text{with} \quad \tilde{E} = E + f(H(q,t)) \; . \tag{3}$$

Here, $\tilde{E}$ is an "effective" energy, and $f(H(q,t))$ is a function of the histogram $H(q,t)$ in a pre-chosen "order parameter" q. The weight of a local minimum decreases with the time the system stays in that minimum till it is no longer favored, and the system continues it search. For $f(H(q,t)) = f(H(q))$ the method reduces to the various generalized-ensemble methods [1] (for instance for $f(H(q,t)) = \ln H(E)$ to multicanonical sampling).

In parallel tempering (also known as replica exchange sampling) [7], first introduced to protein folding by me in Ref. [8], standard Monte Carlo or molecular dynamics moves are performed in parallel at different values of a control parameter, most often the temperature. At certain times the current conformations of replicas at neighboring temperatures T_i and $T_{j=i+1}$ are exchanged with a probability

$$w(\ \mathbf{C}^{old} \rightarrow\ \mathbf{C}^{new}) = \min(1, \exp(-\beta_i E(C_j) - \beta_j E(C_i) + \beta_i E(C_i) + \beta_j E(C_j))) \; . \tag{4}$$

For a given replica the swap moves induce a random walk from low temperatures, where barriers lead to long relaxation times, to high temperatures, where equilibration is rapid, *and back*. This results in a faster convergence at low temperatures.

For both Monte Carlo [9] and molecular dynamics [10] we have demonstrated that generalized-ensemble based techniques are superior in locating low-energy conformers.For a critical evaluation of these now widely used methods, see Ref. [11].

C) Current work

In the past, my coworkers and I have demonstrated that the generalized-ensemble approach allows *de novo* folding simulations of peptides and proteins with up to $30 - 50$ residues [12, 13]. However, despite their successful use in many folding studies, the efficiency of these techniques is still not sufficient for transforming simulations into virtual microscopes that can routinely complement experiments in exploring the molecular machinery of cell. For this reason, my group continues with improving the computational efficiency of these techniques [14]; for instance, by identifying "order parameters" or "reaction coordinates" [15] that allow tailoring of generalized ensembles specifically designed for the simulation of proteins, or classes of proteins. Many of our algorithms [6, 8, 13] are implemented in the freeware program package SMMP [2] available from either the program library of *Computer Physics Communications* or directly from the authors (*www.hansmann-lab.com/cbpc/smmp/smmp.php*); and increasingly also as add-ons to the popular molecular dynamics program GROMACS [3]. This software is made available on Github (*github.com/hansmann-lab*).

Our algorithms are tested on carefully chosen proteins ranging from the 28-residue miniprotein Fsd-Ey to the 93-residue lymphotactin (which under physiological conditions interconverts between two folds). Our aim is to probe whether the sequence of amino acids in a protein

contains not only information on the native structure but encodes a more complex multi-funnel landscape where mutations, changes in environment or interaction with other molecules switch between members of an ensemble of availble folds.

For instance, the transcription factor RfaH changes biological function by switching between distinct three-dimensional folds. RfaH regulates transcription if the C-terminal domain folds into a double helix bundle, and promotes translation when this domain assumes a β-barrel form. We have studied this fold-switching in RfAH [17] using an enhanced sampling method introduced by us, replica-exchange-with-tunneling (RET) [18], to calculate the free-energy landscape of the protein. Analyzing this landscape, we could already propose a mechanism for the conversion process that is consistent with experimental measurements [17]. As extension of this research we are looking now into the early steps of the conversion between native PrP^P and the toxic scrapie PrP^{Sc} form [19] of prion proteins, and how this process can be triggered by interaction with RNA [19, 20].

In recent years, our interest has shifted toward aggregation and amyloid formation as examples for self-assembly of proteins. For instance, we have studied self-assembly of phenylalanine into nanotubes that are build out of rings of four molecules. The size of these aggregates is consistent with experimental measurements of fibrils obtained from mice with phenylketonuria [21]. More widely studied is the role of amyloids in Alzheimer's disease. Interestingly, the β-amyloid aggregates exhibit *in vitro* a remarkable polymorphism that is missing in patient-derived Aβ aggregates. We could show that the lack of polymorphism in the brain-derived fibrils is likely due to the different environmental condition by which the fibrils grow *in vitro* and in the brain. [22] The different characteristics of *in vitro* and patient-derived Aβ-fibril needs to be considered in the design of drug candidates that can bind to amyloids. We have demonstrated this in Refs. [23] pointing out that the binding affinities of angiotensin-converting enzyme (ACE) inhibitors (which inhibit *in vitro* amyloid formation) are different for *in vitro* and patient-derived fibrils. Similarly, we have also evaluated the stabilities of self-replicating Osaka-mutant fibrils [24]; and comparing the recently resolved S-shaped fibril motif in Aβ_{42} fibrils with the U-shaped forms commonly seen in Aβ_{40} fibrils [25], we could propose possible new fibril organizations that are consistent with experimental measurements [26].

We are especially interested in the structural transitions between amyloids. For instance, we have studied extensively n eleven-residue segment of αB - crystalline that forms in solution either a hexamer with out-of-register antiparallel β-strands arranged as a cylindrical barrel, but can also assemble into amyloid fibrils with their typical in-register β - strands. The presumed universality of the barrel-shaped motif in toxic oligomers and it's small size makes cylindrin an interesting system for studying the association of mis-folded proteins into small soluble oligomers, and their conversion into insoluble fibrils. Our simulations showed that the transition between the two polymorphs does not involve unfolding of the chains but only their dissociation and re-association. Crucial for formation of the barrel-like oligomer is the salt-bridge between K3-D7 which guides the association of the peptides into this form instead of the energetically more favorable fibril [27].

Given the complexity of amyloid formation, conversion and propagation, we expect that our above studies will inspire further development of methods and algorithms, directed toward the long-term goal of simulating cells on all time and length scales relevant for medical and technological applications.

References

[1] O. Zimmermann and U.H.E. Hansmann, *BBA - Proteins and Proteomics*, **1784** (2008) 252.

[2] F. Eisenmenger, U.H.E. Hansmann, Sh. Hayryan, C.-K. Hu, *Comp. Phys. Comm.* **138** (2001) 192.

[3] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, *J. Chem. Theory Comput.* **4** (2008) 435-447.

[4] B. Berg and T. Neuhaus, *Phys. Lett.* **B267** (1991) 249.

[5] A.M. Ferrenberg and R.H. Swendsen, *Phys. Rev. Lett.* **61** (1988) 2635.

[6] U.H.E. Hansmann and L.T. Wille, *Phys. Rev. Let.*, **88** (2002) 068105.

[7] K. Hukushima and K. Nemoto, *J. Phys. Soc. (Japan)*, **65** (1996) 1604; G.J. Geyer, *Stat. Sci.* **7** (1992) 437.

[8] U.H.E. Hansmann, *Chem. Phys. Lett.* **281** (1997) 140.

[9] U.H.E. Hansmann and Y. Okamoto, *J. Comp. Chem* **14** (1993) 1333.

[10] U.H.E. Hansmann, Y. Okamoto, F. Eisenmenger, *Chem. Phys. Lett.* **259** (1996) 321.

[11] D. Gront, A. Kolinski, J. Skolnick, *J. Chem. Phys.* **113** (2000) 5065.

[12] C.-Y. Lin, C.-K. Hu and U.H.E. Hansmann, *Proteins*, **52** (2003) 436.

[13] W. Kwak and U.H.E. Hansmann, *Phys. Rev. Lett.* **95** (2005) 138102.

[14] W. Nadler and U.H.E. Hansmann, *Phys. Rev. E* **75** (2007) 026109; **76** (2007) 057102.

[15] U.H.E. Hansmann, *J. Chem. Phys.* **120** (2004) 417.

[16] N.A. Alves and U.H.E. Hansmann, *Phys. Rev. Lett.* **84** (2000) 1836.

[17] N.A. Bernhardt and U.H.E. Hansmann, J. Phys. Chem. B, **122** (2018) 1600.

[18] N. A. Bernhardt, W. Xi, W. Wang, and U.H.E. Hansmann, J. Chem. Theor. Comp. **12** (2016) 5656; **13** (2017) 393.

[19] E.J. Alred, M.V. Nguyen, M. Martin and U.H.E. Hansmann, Protein Science, **26** (2017) 1524.

[20] E.J. Alred, I. Lodangco, J. Gallaher, and U.H.E. Hansmann, ACS Omega, **3** (2018) 3936.

[21] H.W. German, S. Uyaver and U.H.E. Hansmann, J. Phys. Chem. A, **119** (2015) 1609.

[22] E.J. Alred, M. Phillips, W.M. Berhanu and U.H.E. Hansmann, Protein Science, **24** (2015) 923.

[23] M. Bhavaraju, M. Philips, D. Bowman, J.M. Aceves-Herndez & U.H.E. Hansmann, J. Chem. Phys. **144** (2016) 015101.

[24] W.M. Berhanu, E.J. Alred and U.H.E. Hansmann, J. Phys. Chem. B, **119** (2015) 13063.

[25] W. Xi, W. Wang, G. Abbott and U. H.E. Hansmann, J. Phys. Chem. B, **120** (2016) 4548.

[26] W. Xi and U.H.E. Hansmann, Scientific Reports, **7** (2017) 6588.

[27] H. Zhang, W. Xi, U.H.E. Hansmann and Y. Wei, J. Chem. Theor. Comp. **13** (2017) 3936.