$_E$SBMTOOLS: PYTHON TOOLS FOR ENRICHED STRUCTURE BASED MODELING

Benjamin Lutz[a,b] , Claude Sinner[a,b], Geertje Heuermann[a,b], Abhinav Verma[b], Alexander Schug[b,*]

[a] Department of Physics, Karlsruhe Institute of Technology, Karlsruhe, Germany

[b]Steinbuch Centre for Computing, Karlsruhe Institute of Technology, Karlsruhe, Germany

[*] alexander.schug@kit.edu

ABSTRACT

Biomolecular simulations provide a computational microscope to dynamically visualize biomolecular systems with atomic resolution. Given the advances in speed of computational resources, simulations can complement experiments and help understand the relationship of folding, structure, and function of proteins, structured RNA, or DNA. Structure-based models (SBM) provide a computationally inexpensive tool to study the folding and structural assembly of (macro)biomolecules. Their theoretical foundations are energy landscape theory and the principle of minimal frustration. Here, we present $_E$SBMTools: python tools that assist to setup and analyze structure-based simulations of proteins and nucleic acids, both at the C_α and all-atom level. The tools interface with GROMACS and support its standard output formats. Information from other sources like bioinformatics or experimental data can be added as enrichments. One example would be docking protein complexes out of the composing individual known proteins plus bioinformatically derived information of the inter-protein interface contacts.

Keywords: protein folding, protein structure prediction, structure-based model, setup and analysis

1. INTRODUCTION

In the last few decades, important progress has been achieved in the field of biomolecular sciences. Experiments can explore these systems in high detail (Sadqi, Fushman et al. 2006) even on the single molecule level (Mickler, Dima et al. 2007; Gambin, VanDelinder et al. 2011). The resolution of experiments, however, is still often limited by technical constraints. Simulation techniques based on Monte Carlo (Schug and Wenzel 2004; Schug, Herges et al. 2005; Verma, Schug et al. 2006; Verma and Wenzel 2009; Perez-Sanchez and Wenzel 2011) or Molecular Dynamics (MD) (Adcock and McCammon 2006; Lee, Hsin et al. 2009) have similarly advanced and can complement experiments (Lange, Lakomek et al. 2008; Gambin, Schug et al. 2009). The advances in MD result from improved hardware performance, utilizing new architectures like Cell (Olivier, Prins et al. 2007), GPU or specialized supercomputers (Shaw, Maragakis et al. 2010), better force fields (Lindorff-Larsen, Piana et al. 2010) and more effective simulation algorithms (Hess, Kutzner et al. 2008). Taken together, this enables the detailed exploration of dynamical properties for a biomolecular system. This step is crucial for understanding its function at an atomic resolution. Deeper insights in the dynamics of proteins and structured RNA result from, for example, the investigation of conformational changes in the systems of interest (Okazaki, Koga et al. 2006; Schug, Whitford et al. 2007) and challenges as complex as protein folding become accessible (Kussell, Shimada et al. 2002; Thirumalai, Klimov et al. 2002; Onuchic and Wolynes 2004; Thirumalai and Hyeon 2005; Dill, Ozkan et al. 2008; Schug and Onuchic 2010). As many dynamic processes occur on slow μs to ms timescales, however, MD simulations struggle to reach these scales given a time step of 1-2fs.

In order to reach the desired timescales, coarse-graining has proven to be a reliable approach by reducing the complexity of the simulated system (Klimov and Thirumalai 1998). This can occur on multiple scales. For example, coarse-graining the biomolecular system to a C_α level reduces each amino acid to a single bead (Clementi, Nymeyer et al. 2000). This approach reduces the number of particles in the biophysical system by around two orders of magnitude compared to an all-atom system with explicit water molecules. The price paid is reduced insight into, for example, details of side-chain interactions or the influence of base-pairing and stacking interactions in RNA or DNA.

The structure-based model (SBM) approach is based on energy landscape theory and the principle of minimal frustration for protein folding and structured RNA (Onuchic and Wolynes 2004; Schug and Onuchic 2010). Accordingly, biomolecular folding occurs on a funneled energy landscape with its free-energy minimum in the native fold (Bryngelson, Onuchic et al. 1995). The SBM-Hamiltonian (Clementi, Nymeyer et al. 2000; Lammert, Schug et al. 2009; Whitford, Noel et al. 2009; Whitford, Schug et al. 2009) is directly based on this native fold. A crucial part of these force fields are interactions of amino acid pair contacts described as a contact map, i.e., of a matrix of spatially close interacting amino acids (Noel, Whitford et al. 2012).

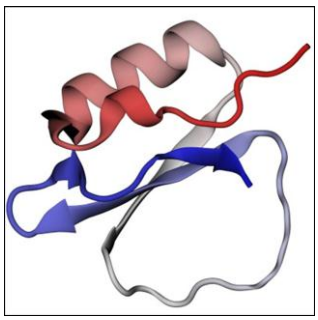

Figure 2: Cartoon representation of CI-2 (PDB-code 2CI2)

Typically, the physicochemical details of each individual interaction are condensed into simple terms like Lennard-Jones or Gaussian interactions(Lammert, Schug et al. 2009) further reducing computational complexity. SBM have been parameterized for different resolutions from the C_α level (Clementi, Nymeyer et al. 2000), over the $C_\alpha C_\beta$ (Oliveira, Schug et al. 2008)to the all-atom level (Whitford, Noel et al. 2009). In spite of their simplified energetics, they have shown good agreement with experimental measurements (Clementi, Jennings et al. 2001; Schug, Whitford et al. 2007; Gambin, Schug et al. 2009; Schug, Weigt et al. 2009; Baxter, Jennings et al. 2012).

When enriched with bioinformatic information (Weigt, White et al. 2009) as additional distance constraints and more detailed biophysical force fields, even accurate predictions of three-dimensional structures of protein complexes (Schug, Weigt et al. 2009), globular proteins (Sulkowska, Morcos et al. 2012), trans-membrane proteins (Hopf, Colwell et al. 2012) or active conformations (Dago, Schug et al. 2012) have been made. Similarly, integrating distance constraints from experimental measurements like FRET or EM density maps could be included.

The main purpose of the present tool collection is to facilitate the scriptable setup of huge systems in SBM simulations for the GROMACS (Hess, Kutzner et al. 2008) software package and enhancing these simulation with information from other sources. This reduces the effort for a single simulation run. The tools can be included in automated workflows for a wide range of biophysical investigations. The tools include routines that run post processing protocols of standard analysis procedures, like contact map analysis, Q value generation, Phi value or RMSD evaluation. In this paper we discuss the methods that SBM are based on, the implementation of pre and post processing functionality of ESBMTools, and give an overview over several exemplary scenarios that apply their functionality.

2. THEORETICAL BACKGROUND

Molecular dynamics simulation technique solves Newtonian equations of motion via numerically integrating over time. The central characterization of a system of interest is introduced by a potential for the equations of motion. We describe the underlying potentials for the all-atom and the C_α case. Furthermore, we discuss the C_α method as a coarse-grained approach in the context of protein simulation. The contact map as the substantial ingredient to SBM is described afterwards.

2.1. Structure-based Potential

The most basic information that characterizes a molecular dynamics simulation is aggregated in its potential from which the force field for the Newtonian equations of motion is derived. The all-atom formulation of the structure-based potential (Whitford, Noel et al. 2009) reads as

$$\begin{aligned} V_{AA} = & \sum_{\text{bonds}} K_b(r - r_0)^2 + \sum_{\text{angles}} K_a(\theta - \theta_0)^2 \\ & + \sum_{\text{impropers}} K_i(\chi - \chi_0)^2 + \sum_{\text{dihedrals}} K_d f_d(\phi) \\ & + \sum_{\text{contacts}} K_c\left[\left(\frac{\sigma_{ij}}{r}\right)^{12} - 2\left(\frac{\sigma_{ij}}{r}\right)^{6}\right] + \sum_{\text{non contacts}} K_{nc}\left(\frac{\sigma_{nc}}{r}\right)^{12}, \end{aligned} \tag{1}$$

where the dihedral or torsional angle potential is given by

$$f_d(\phi) = [1 - \cos(\phi - \phi_0)] + \frac{1}{2}[1 - \cos(3(\phi - \phi_0))], \tag{2}$$

and K_b, K_a, K_i, K_d, K_c and K_{nc} are the corresponding force constants. $r_0, \theta_0, \chi_0, \phi_0$ and σ_{ij} are taken from the native structure. Accordingly, the potential has its minimum at the native conformation.

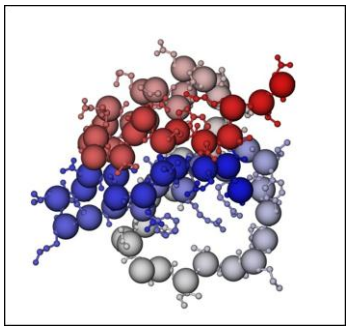

Figure 1 All-atom and C_α representation of CI-2

2.2. Coarse-graining

We present an approach of coarse-graining for protein systems that reduces each amino acid of a protein to a single bead at the position of the protein's C_α atom. This reduction decreases the number of particles in the computational system. Removing an explicit water representation reduces the number of atoms by one order of magnitude and reducing an all-atom (non-Hydrogen) representation to the C_α level results in another order of magnitude less atoms. Therefore, the approach results in computationally far less demanding simulations. The structure-based potential for proteins in a coarse-grained C_α formulation (Clementi, Nymeyer et al. 2000) reads as

$$V_{C\alpha} = \sum_{\text{bonds}} K_b(r - r_0)^2 + \sum_{\text{angles}} K_a(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_d f_d(\phi) + \sum_{\text{contacts}} K_c\left[5\left(\frac{\sigma_{ij}}{r}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r}\right)^{10}\right] + \sum_{\text{non contacts}} K_{nc}\left(\frac{\sigma_{nc}}{r}\right)^{12}, \quad (3)$$

where $f_d(\phi)$ is the dihedral potential as defined in Eq. 2 and structural information is incorporated as in Eq. 1, correspondingly. Relative to Eq. 1, the potential in coarse-grained formulation exhibits two changes: The potential lacks terms for improper dihedral angles since the C_α formulation loses the possibility to model planarity within the sidechain. Secondly, the Lennard-Jones potential is changed from its standard 6-12 formulation to a 10-12 formulation. Equivalent Hamiltonians can be found using Gaussian potentials (Lammert, Schug et al. 2009).

It should be emphasized that the negligence of structural detail in the C_α method, makes it insufficient to describe nucleic acid chains. Their base-base interactions (pairing and stacking) play a crucial role in nucleic acid strands, both for structural and dynamic investigations, and cannot be neglected (Thirumalai and Hyeon 2005; Whitford, Schug et al. 2009).

2.3. Contact Map

The information of bonded interactions (bonds, angles and dihedrals) is complemented by contact information (Noel, Whitford et al. 2012). This information is aggregated in the contact map of a biomolecular structure. In its simplest form, a contact between two atoms is formed if the distance σ_{ij} between the two of them is below a certain threshold (typically 4-5Å). Typically, a minimal distance in sequence can be required in case of proteins, while nucleic acids need to be able to form contacts between neighboring residues as stacking interactions. Contact information is represented by repulsive and attractive terms of a Lennard-Jones potential, as denoted in Eq. 1 and 3. All other possible pairings of atoms are assigned to a repulsive Lennard-Jones term that is characterized by the exclusion radius σ_{nc}.

3. SIMULATION SETUP

The simulation setup consists of the standardized generation and, as the case may be, customizable manipulation of the coordinate and topology file of a biomolecular system. The generation is based on a PDB conform structure file and several user defined parameters. The native conformation is taken from the structure file and combined with an XML based topology that defines bonded interactions to create the following input files for a GROMACS(Hess, Kutzner et al. 2008) (4.5.4) simulation. The $_E$SBMTools package is written in python 2.7 and requires biophython (Cock, Antao et al. 2009) and scientific python. It can be downloaded from sourceforge. Examples are included.

3.1. Coordinate File

The coordinate file for a GROMACS simulation contains the atoms represented by their names, residue names and atom types in combination with their individual Cartesian coordinates. The coordinates represent the native conformation which introduces the conformation of minimal potential energy in the SBM. These coordinates are generated automatically from a PDB conform data file of the biomolecular system of interest. In case of coarse-graining, the only atom type that is present in the coordinate file is the C_α type. The corresponding coordinates are again taken from the PDB structure file.

3.2. Topology File

The topology file introduces the biomolecular system by a list of its atoms. Referring to the list of atoms, the topology file also contains the force constants and geometrical quantities of equilibrium of the SBM potential. The geometrical quantities are calculated from the system's coordinates. The particular geometrical associations of bonded interactions are defined in an XML file which can be adapted to user defined scenarios. Topologies for amino acids and nucleic acids are provided with $_E$SBMTools, but the topologies are easily expandable by the user's own defined topological rules, if, e.g., ligands are needed. The contact map as part of the potential is also included in the topology file.

3.3. Look-up Table

In case of a coarse-grained simulation, the Lennard-Jones potential terms is formulated with powers of 10-12 instead of 6-12. GROMACS offers the introduction of such modifications on the standard potential expressions. To this end, the user has to provide a look-up table for GROMACS to specify the desired modifications. $_E$SBMTools generates a file (table.xvg) that contains such a look-up table in the format required by GROMACS.

3.4. Configuration File

A standard configuration file (md.mdp) for a GROMACS simulation can be created by $_E$SBMTools in order to be equipped with a complete set of required files for a molecular dynamics simulation. The creation of this configuration file is also customizable by, e.g., setting the number of integration steps, the temperature or generation seeds for random events.

3.5. Input Modification

Existing input files can be modified with several functions in $_E$SBMTools. The manipulation of existing contacts and the introduction of new contacts, atoms, bonds, angles or dihedral angles is a desirable feature in the course of SBM simulation. This approach enables the user to set up a heterogeneous potential or a combination of two separately generated SBM systems.

3.5.1. Contact Map Modification

The contact map of a SBM plays a crucial role in introducing tertiary structural elements in the biomolecular system. Contact map extension to an existing protein-protein interface can model, for example, structural transitions correctly by adding a second local minimum. To this end, ESBMTools offers a variety of functions that facilitate the access to a given contact map and the scriptable generation of new entries and modification of existing entries. Externally generated contact maps can be read in, random subsets of contact maps can be generated, contact strengths be modified and the force constant in the Lennard-Jones terms for user-defined ranges of contact partners can be modified.

3.5.2. Merging & Enrichment

In order to combine two existing SBM systems, it is necessary to merge their coordinate and topology files. ESBMTools provides functionality to generate a single SBM for the combination of two existing systems to allow, for example, simulating protein complex formation. The two systems stay isolated from each other in terms of interactions by a plain merging procedure. This functionality can be complemented by introducing additional inter-molecular contacts from non-structural sources. Examples are distance contraints derived from bioinformatic information(Weigt, White et al. 2009) like in MAGMA (Schug, Weigt et al. 2010) to predict a coarse-grained protein complex model, which can be afterwards relaxed in more detailed biophysical force fields to accurately predict three-dimensional structures of protein complexes(Schug, Weigt et al. 2009), active conformations(Dago, Schug et al. 2012) or globular proteins(Sulkowska, Morcos et al. 2012). Other possibilities would be including distance constraints from experimental measurements like FRET, small angle X-ray scattering (Jamros, Oliveira et al. 2010) or cryo EM (Whitford, Ahmed et al. 2011).

4. SIMULATION ANALYSIS

Simulation analysis is mainly based on the trajectory as outcome of a molecular dynamics simulation. ESBMTools provides interfaces to several GROMACS evaluation extensions that process simulation data. The output of these extensions can be read in for further processing. We present a variety of possible post processing scenarios that can be conducted in the following.

4.1. RMSD

The root mean square deviation (RMSD) of a trajectory in each dumped frame describes the average deviation of the current structure in comparison to a reference structure, often the native structure. These values are a measure for fluctuations in the course of a simulation. This analysis can serve as a rough estimate for a folded or unfolded state of a biomolecular state.

4.2. Q Value Evaluation

The Q value evaluation is a standard analysis for SBM simulations(Cho, Levy et al. 2006) where the Q value is defined for each simulation frame as the fraction of formed native contacts. By this means, the Q value itself can serve as a reaction coordinate that represents a mapping on the folding progress in time if the Q value is monotonically increasing over time. ESBMTools provides several functions that filter and bin Q value trajectories for requested ranges of involved atoms or residues in order to analyze, e.g., folding characteristics in biomolecular systems.

4.2.1. Contact Maps

Contact maps are a standard visualization technique of the present contacts not just in a native structure but also along a simulated trajectory. To this end, ESBMTools offers the functionality to plot contact maps from a given topology as well as a user-defined set of contact maps from a trajectory. These contact maps along a trajectory can be also arranged to a movie by standard movie encoding software.

4.2.2. State Population

States are visualized as a histogram of Q (or other reaction coordinates) (Cho, Levy et al. 2006) timedirectly from a trajectory.

4.2.3. Φ Value

The **Φ-**value analysis investigates the stabilizing influence of amino acids on the transition state in a two-state folding protein(Fersht 1995). This analysis originates from an experimental approach. The experimental phi value analysis can be translated in a computational analogue that is based on the probabilities of occupancy of the three possible states folded, unfolded and transition state.

5. EXAMPLES

We illustrate the variety of preparation and investigation related procedures provided by ESBMTools.

ESBMTools generates all necessary files for a molecular dynamics simulation with GROMACS. The contact map is calculated from the native state(see Fig. 3).

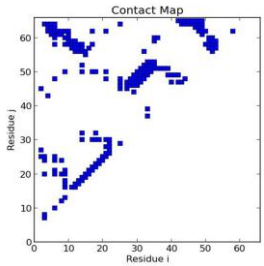

Figure 3: C_α based Contact map of CI-2

After running a simulation, user-defined frames from the trajectory of a denaturing protein (Fig. 4) can be

plotted in order to investigate the dynamics of folding paths by the order of opening structural elements.

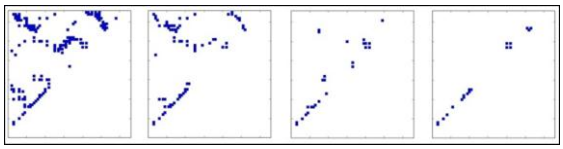

Figure 5: Contact maps of denaturing CI-2. The leftmost contact map is in the folded state, while the contacts maps to the right are found at lower Q-values (i.e. more unfolded).

A statistically relevant analysis is the investigation of temporal progress of the Q value as the number of formed native contacts (Fig. 5, Top). This value identifies the nativeness of the trajectory at every timestep. A histogram of the Q value distribution (Fig. 5, Bottom) embodies the frequency distribution of the system's state variable. Therefore, it serves as a starting point from which the free energy can be calculated.

As part of a more specific analysis, $_{E}$SBMTools filters the trajectory's Q values for user-defined ranges of involved contact partners. The investigation of melting characteristics in the context of RNA hairpins (Fig. 6) can be based on a Q value filtering for base pair contacts in combination with averaging over several hundreds of trajectories.

Our last example illustrates the usage of $_{E}$SBMTools for a **Φ**-value analysis. Fig. 7 shows the **Φ** values calculated from a simulation based on a SBM simulation run and evaluated by our toolset.

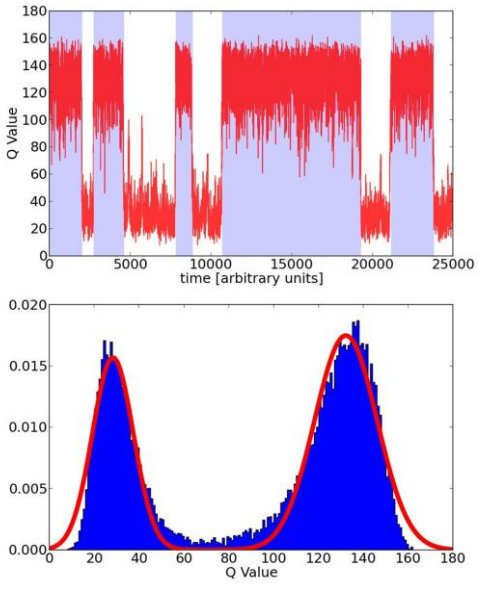

Figure 4: Top: Temporal progress of the Q value for CI-2 close to the folding temperature that indicates folded (light blue) and unfolded (white) domains. Bottom: Histogram of the population of Q values over a trajectory.

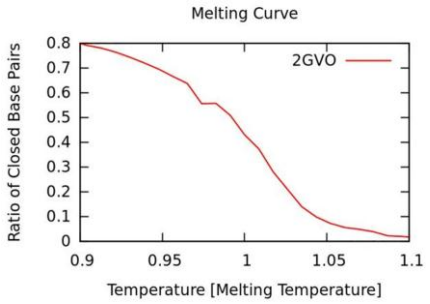

Figure 7: Melting curve of RNA hairpin 2GVO

6. CONCLUSION

We present $_{E}$SBMTools as a customizable, extendable and flexibly scriptable python package that aims at setting up and evaluating SBM molecular dynamics simulations with GROMACS. The toolbox provides a diverse collection of functions, that can be integrated in existing projects or build the foundation for new projects. It can be downloaded and it is open source, which gives the user complete control over all of its functionality. It is compatible with standard installations of clusters that provide GROMACS, biopython and scientific python.

ACKNOWLEDGMENTS

We thank the Helmholtz Association for funding our work within the "Impuls- und Vernetzungsfond" and the bwGRiD as part of the National Grid Initiative (NGI) for providing cluster resources for our simulations.

REFERENCES

Adcock, S. A. and J. A. McCammon (2006). "Molecular dynamics: Survey of methods for simulating the activity of proteins." Chem Rev **106**(5): 1589-1615.

Baxter, E. L., P. A. Jennings, et al. (2012). "Strand swapping regulates the iron-sulfur cluster in the diabetes drug target mitoNEET." Proceedings of the National Academy of Sciences of the United States of America **109**(6): 1955-1960.

Bryngelson, J. D., J. N. Onuchic, et al. (1995). "Funnels, Pathways, and the Energy Landscape of Protein-Folding - a Synthesis." Proteins-Structure Function and Genetics **21**(3): 167-195.

Cho, S. S., Y. Levy, et al. (2006). "P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes." Proceedings of the National Academy of Sciences of the United States of America **103**(3): 586-591.

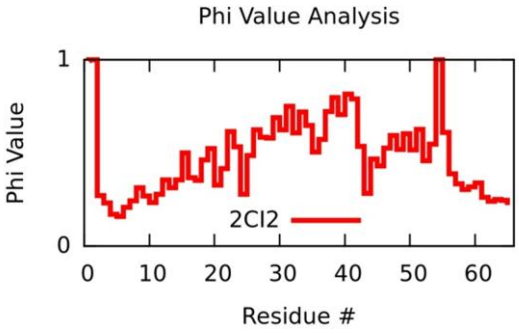

Figure 6: Phi value analysis for CI-2

Clementi, C., P. A. Jennings, et al. (2001). "Prediction of folding mechanism for circular-permuted proteins." Journal of Molecular Biology **311**(4): 879-890.

Clementi, C., H. Nymeyer, et al. (2000). "Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins." Journal of Molecular Biology **298**(5): 937-953.

Cock, P. J., T. Antao, et al. (2009). "Biopython: freely available Python tools for computational molecular biology and bioinformatics." Bioinformatics **25**(11): 1422-1423.

Dago, A. E., A. Schug, et al. (2012). "Structural basis of histidine kinase autophosphorylation deduced by integrating genomics, molecular dynamics, and mutagenesis." Proceedings of the National Academy of Sciences of the United States of America **109**(26): E1733-1742.

Dill, K. A., S. B. Ozkan, et al. (2008). "The protein folding problem." Annu Rev Biophys **37**: 289-316.

Fersht, A. R. (1995). "Characterizing Transition-States in Protein-Folding - an Essential Step in the Puzzle." Current Opinion in Structural Biology **5**(1): 79-84.

Gambin, Y., A. Schug, et al. (2009). "Direct single-molecule observation of a protein living in two opposed native structures." Proc Natl Acad Sci U S A **106**(25): 10153-10158.

Gambin, Y., V. VanDelinder, et al. (2011). "Visualizing a one-way protein encounter complex by ultrafast single-molecule mixing." Nature methods **8**(3): 239-241.

Hess, B., D. Kutzner, et al. (2008). "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Dynamics." J. Chem. Theory Comp. **4**: 435-447.

Hopf, T. A., L. J. Colwell, et al. (2012). "Three-dimensional structures of membrane proteins from genomic sequencing." Cell **149**(7): 1607-1621.

Jamros, M. A., L. C. Oliveira, et al. (2010). "Proteins at work: a combined small angle X-RAY scattering and theoretical determination of the multiple structures involved on the protein kinase functional landscape." The Journal of biological chemistry **285**(46): 36121-36128.

Klimov, D. K. and D. Thirumalai (1998). "Cooperativity in protein folding: from lattice models with sidechains to real proteins." Folding & Design **3**(2): 127-139.

Kussell, E., J. Shimada, et al. (2002). "A structure-based method for derivation of all-atom potentials for protein folding." Proceedings of the National Academy of Sciences of the United States of America **99**(8): 5343-5348.

Lammert, H., A. Schug, et al. (2009). "Robustness and generalization of structure-based models for protein folding and function." Prot: Struct, Funct and Bioinf **77**(4): 881-891.

Lange, O. F., N. A. Lakomek, et al. (2008). "Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution." Science **320**(5882): 1471-1475.

Lee, E. H., J. Hsin, et al. (2009). "Discovery through the computational microscope." Structure **17**(10): 1295-1306.

Lindorff-Larsen, K., S. Piana, et al. (2010). "Improved side-chain torsion potentials for the Amber ff99SB protein force field." Proteins **78**(8): 1950-1958.

Mickler, M., R. I. Dima, et al. (2007). "Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations." Proc Natl Acad Sci U S A **104**(51): 20268-20273.

Noel, J. K., P. C. Whitford, et al. (2012). "The Shadow Map: A General Contact Definition for Capturing the Dynamics of Biomolecular Folding and Function." The journal of physical chemistry. B.

Okazaki, K., N. Koga, et al. (2006). "Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations." Proc Natl Acad Sci U S A **103**(32): 11844-11849.

Oliveira, L. C., A. Schug, et al. (2008). "Geometrical features of the protein folding mechanism are a robust property of the energy landscape: a detailed investigation of several reduced models." J Phys Chem B **112**(19): 6131-6136.

Olivier, S., J. Prins, et al. (2007). Porting the GROMACS Molecular Dynamics Code to the Cell Processor. IEEE International Parallel and Distributed Processing Symposium, IPDPS.

Onuchic, J. N. and P. G. Wolynes (2004). "Theory of protein folding." Curr Opin Struct Bio **14**(1): 70-75.

Perez-Sanchez, H. and W. Wenzel (2011). "Optimization methods for virtual screening on novel computational architectures." Current computer-aided drug design **7**(1): 44-52.

Sadqi, M., D. Fushman, et al. (2006). "Atom-by-atom analysis of global downhill protein folding." Nature **442**(7100): 317-321.

Schug, A., T. Herges, et al. (2005). "Comparison of Stochastic optimization methods for all-atom folding of the Trp-cage protein." Chemphyschem **6**(12): 2640-2646.

Schug, A. and J. N. Onuchic (2010). "From protein folding to protein function and biomolecular binding by energy landscape theory." Curr Opin Pharmacol **10**(6): 709-714.

Schug, A., M. Weigt, et al. (2010). "Computational modeling of phosphotransfer complexes in two-component signaling." Methods in enzymology **471**: 43-58.

Schug, A., M. Weigt, et al. (2009). "High resolution complexes from integrating genomic information with molecular simulation." Proc Natl Acad Sci U S A **106**(52): 22124-22129.

Schug, A. and W. Wenzel (2004). "All-atom folding of the trp-cage protein with an adpative parallel tempering method." Europhysics Letters **67**(2): 307-313.

Schug, A., P. C. Whitford, et al. (2007). "Mutations as trapdoors to two competing native conformations of the Rop-dimer." Proc Natl Acad Sci USA **104**: 17674-17679.

Shaw, D. E., P. Maragakis, et al. (2010). "Atomic-level characterization of the structural dynamics of proteins." Science **330**(6002): 341-346.

Sulkowska, J. I., F. Morcos, et al. (2012). "Genomics-aided structure prediction." Proceedings of the National Academy of Sciences of the United States of America **109**(26): 10340-10345.

Thirumalai, D. and C. Hyeon (2005). "RNA and protein folding: common themes and variations." Biochemistry **44**(13): 4957-4970.

Thirumalai, D., D. K. Klimov, et al. (2002). Insights into specific problems in protein folding using simple concepts. Computational Methods for Protein Folding. **120:** 35-76.

Verma, A., A. Schug, et al. (2006). "Basin hopping simulations for all-atom protein folding." Journal of Chemical Physics **124**(4).

Verma, A. and W. Wenzel (2009). "A free-energy approach for all-atom protein simulation." Biophys J **96**(9): 3483-3494.

Weigt, M., R. A. White, et al. (2009). "Identification of direct residue contacts in protein-protein interaction by message passing." Proc Natl Acad Sci U S A **106**(1): 67-72.

Whitford, P. C., A. Ahmed, et al. (2011). "Excited states of ribosome translocation revealed through integrative molecular modeling." Proceedings of the National Academy of Sciences of the United States of America **108**(47): 18943-18948.

Whitford, P. C., J. K. Noel, et al. (2009). "An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields." Proteins-Structure Function and Bioinformatics **75**(2): 430-441.

Whitford, P. C., A. Schug, et al. (2009). "Nonlocal helix formation is key to understanding S-adenosylmethionine-1 riboswitch function." Biophys J **96**(2): L7-9.