

Jozef Adam Liwo

Statement of Research

Overview

The goal of my research is to understand how intra- and inter-molecular interactions make a given biomolecular system attain its unique structure and perform a given biological action. I use the coarse-graining approach for this purpose, in which interactions are treated at the site-site rather than detailed atomistic level. Because of a tremendous reduction of the number of degrees of freedom and, thereby, the CPU and memory requirements of computations, a practical application of this approach, which is pursued in my lab, is the development of methods with which to simulate the structure and dynamics of biomolecules and biomolecular systems at a large time and size scales, which are not accessible to all-atom simulations.

Together with my coworkers, I developed a systematic physics-based approach to coarse-graining, which is based on the decomposition of the potential of mean force (PMF) of a given system [1] into contributions due to its parts, thus enabling to produce a transferable force field. Based on this approach, we developed the UNRES model for the simulation of protein structure and dynamics [2, 3], the NARES-2P model for nucleic acids [4] and the SUGRES-1P model for polysaccharides [5], which are now being merged into the Unified Coarse-Grained Model [5]. We are currently working on the extension of the model to incorporate multi-scale approach, where some parts of a system are treated at the all-atom level.

UNRES, NARES-2P, and SUGRES-1P are minimalistic models of the respective biological macromolecules, with one or two sites per residue. Such a representation has resulted in an over 1000-fold speed-up of computations with respect to the all-atom approach [6]. Yet, owing to the physics-based principles of force-field construction, the model is able to predict protein structure [4], reproduce protein-folding kinetics [7] and free-energy landscapes [8], simulate DNA and RNA structure, dynamics, and thermodynamics [9], and simulate biologically-important processes [10, 11]. It can also handle D-amino-acid residues [12] and dynamic formation of disulfide bridges [13]. The UNRES component of the model is available from the www.unres.pl web page and has recently also been implemented on the Polish Computer Grid Network (PL-GRID) as the **UNRES-NG** utility. The current version of NARES-2P can fold small RNA molecules and hybridize small DNA molecules, as well as reproduce well the thermodynamical parameters of folding of small DNA molecules and reproduce “bubble” formation (or pre-melting transition) in AT-rich sequences [9].

Background

Factor expansion approach to coarse-graining

In our approach [1], the prototype of the effective energy function of a system is the potential of mean force (PMF), also termed restricted energy function (RFE) of the system of interest, $F(\mathbf{X})$, where \mathbf{X} is a shorthand for the coarse-grained degrees of freedom. All degrees of freedom that are lost when passing from the all-atom to the coarse-grained model (\mathbf{Y}) are averaged over, as

expressed by Eq. (1).

$$F(\mathbf{X}) = -RT \ln \left\{ \frac{1}{V_{\mathbf{Y}}} \int_{\Omega_{\mathbf{Y}}} \exp[-E(\mathbf{X}; \mathbf{Y})/RT] dV_{\mathbf{Y}} \right\} \quad (1)$$

where $V_{\mathbf{Y}} = \int_{\Omega_{\mathbf{Y}}} dV_{\mathbf{Y}}$, $E(\mathbf{X}; \mathbf{Y})$ is the original (all-atom) energy function, R is the universal gas constant, T is the absolute temperature, $\Omega_{\mathbf{Y}}$ is the region of the \mathbf{Y} subspace of variables over which the integration is carried out, and $V_{\mathbf{Y}}$ is the volume of this region. The PMF is decomposed into factors of increasing order [Eq. (2)]; factors of order 1 ($f^{(1)}$) pertain to interactions with single coarse-grained sites and pairs of sites, while factors of higher order correspond to clusters of sites and give rise to multibody terms.

$$F(\mathbf{X}) = \sum_i f_i^{(1)}(\mathbf{X}) + \sum_{i<j} f_{ij}^{(2)}(\mathbf{X}) + \sum_{i<j<k} f_{ijk}^{(3)}(\mathbf{X}) + \dots + \sum_{i_1<i_2<\dots<i_n} f_{i_1 i_2 \dots i_n}^{(n)}(\mathbf{X}) \quad (2)$$

A tremendous advantage of this approach compared to statistical and other physics-based potentials [14, 15] is that the total potential of mean force of the system is split, in a rigorous way, into a sum of contributions from its smaller parts, which enables us to construct a transferable force field, and that analytical expressions for the terms in the simplified energy function, including the multibody terms essential to maintain regular structure [16, 1], can be obtained [1] by the expansions of the factors into Kubo’s cluster cumulants [17].

Use of the approach outlined above for a specific system/class of systems involves the following steps [2, 5]: (i) determination of coarse-grained sites and determination of the maximum order of the factor expansion; (ii) derivation of approximate analytical expressions for the factors by Kubo’s cumulant expansion [17] or other means; (iii) computing the surfaces of the factors based on the PMF surfaces of model systems calculated by direct integration or all-atom MD simulations, and (iv) calibration of the force field to reproduce the structure and thermodynamical properties of training molecules. This approach was followed when developing the Unified Coarse Grained Model.

The Unified Coarse Grained Model of biological macromolecules

The components of the Unified Coarse Grained Model of biological macromolecules are schematically depicted in Figure 1 [2, 4, 5]. Their common feature is the presence of an interaction site capable of hydrogen bonding, namely the peptide group (blue spheroids in the left panel of the Figure) for UNRES, united sugar-base groups (purple spheroids in the middle panel of the Figure) for NARES-2P, and sugar rings, which are the only interaction sites in the SUGRES-1P model (blue spheroids in the right panel of the Figure).

To use the model in molecular simulations, the Langevin molecular dynamics approach [6] and its multiplexed replica-exchange (MREMD) and multicanonical molecular dynamics (MUCAMD) were implemented [18, 19].

UCGM

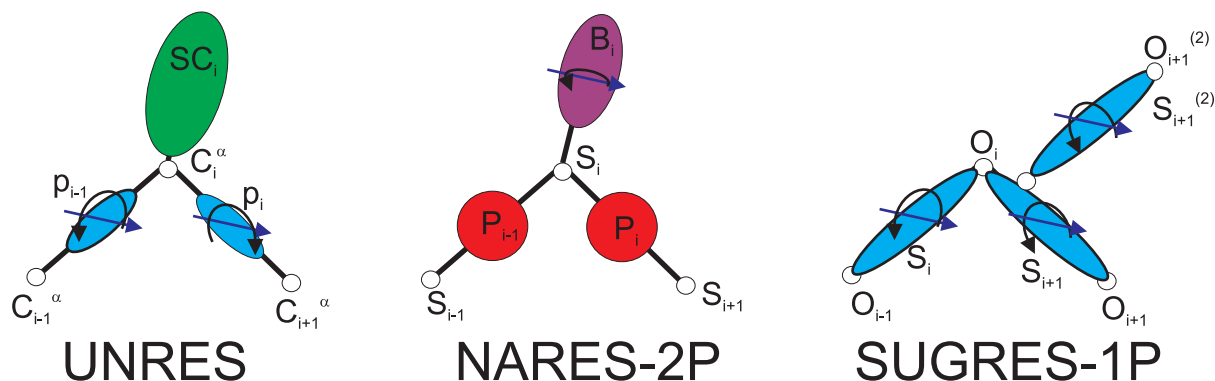


Figure 1: Schematic representation of the components of the Unified Coarse Grained Model of biomolecules. Little arrows indicate the representation of each polar group by a point dipole, while curved arrows indicate that the rotation of a dipole about the respective virtual-bond axis is the degree of freedom that is integrated out.

Current Research

Force-field development. The current focus in the development of NARES-2P and SUGRES-1P is directed at the derivation of local potentials from the PMF surfaces of model systems obtained by all-atom MD simulations. Analytical expressions are fitted to these surfaces as in previous work [20, 12]. UNRES is now being enhanced with a mean-field model of a lipid membrane.

My coworkers and I have recently developed a novel approach to force-field calibration, which is based on maximum-likelihood fitting of the conformational ensembles generated by simulations to the ensembles determined experimentally at various temperatures by means of the maximum-likelihood principle [21]. The method was tested with tryptophan cage as the training protein for which NMR structures were determined at three temperatures, including the folding-transition temperature [22] and found to be very stable and to produce a force field transferable with which to predict the structures of α -helical proteins [21]. Calibration of UNRES with α -, $\alpha + \beta$ -, and β -proteins is in progress.

The new method was also applied to calibrate NARES-2P [23]. The target function was supplemented with the χ^2 terms corresponding to the difference between the calculated and the experimental heat-capacity curves. The resulting force field reproduced the thermodynamics of DNA hybridization better than the original NARES-2P force field.

Use of UNRES in protein-structure prediction. UNRES is being tested in every Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP). To enhance its predictive power, my coworkers and I have recently designed and tested an approach in which restraints from templates [24] or other knowledge-based sources such as, e.g., contact prediction [25] are used as restraints in UNRES. This approach is particularly successful when individual domains are well predicted by template-based modeling while their packing is not because of poor similarity scores in the linker region [24].

Protein folding theory based on dark-soliton concept. With Dr. Antti Niemi, we are investigating protein folding pathways in terms of the dark-soliton solutions of Discrete Nonlinear Schrödinger equation. Dark solitons are identified in the folding trajectories obtained by coarse-grained simulations with UNRES and soliton evolution is studied. With this approach, we are studying the folding of staphylococcal protein A [26].

Kinetics and free-energy landscapes of protein folding. My coworkers and I are currently studying the free-energy landscapes [8] and kinetics of the folding the Fbp28 WW domain and its mutants [7]. The rate constants determined by fitting kinetic equations to the simulated profiles of the folded and intermediate structures turned out to correlated well with the experimental constants. The treatment is now being extended to include hydrodynamic interactions in the Langevin dynamics scheme of UNRES simulations.

Mechanism of action of molecular chaperones. My coworkers and I are working on the investigation of the mechanism of conformational transition of Hsp70 chaperones; we simulated the structure of bacterial Hsp70 [10] before the experimental structure was determined [27]. We are also studying the conformational changes of Hsp70 at the atomic level [28]. Our current research is focused on the influence of chaperone conformation on peptide-substrate binding. We are also working iron-sulfur cluster biogenesis [11]. We proposed the most probable binding modes of the iron-sulfur 1 (Isu1) protein to the Jac1 protein and are now extending the approach to the ternary Ssq1/Isu1/Jac1 complex, where Ssq1 is the Hsp70 chaperone from yeast.

Future research plan

The immediate objective of my research is introduction multiscale treatment of the systems under investigation to the Unified Coarse Grained Model, in which the most of the system is treated at the coarse-grained level and parts of interest (e.g., binding sites and ligands) at the detailed atomic level. Because the prototype of the energy function of the model is the potential of mean force, the potentials of mean force due to the parts treated at atomic level will be calculated on the fly during simulation and used to compute their interactions with the coarse-grained part. This approach will enable to consider metal ions, small ligands and the docking of which cannot be handled at the coarse-grained level because of too low resolution of this treatment. I also plan to combine this approach with the QM/MM treatment to enable the consideration of allostery in studying enzymatic reactions, proton and electron transfer, etc.

Collaborators

In the development of UNRES and NARES-2P, studying protein-folding kinetics and free-energy surfaces, and applying the coarse-grained model to solve biological problems, I have an established long-term collaboration with Dr. Harold A. Scheraga, Cornell University, U.S.A., Dr. Stan Oldziej, University of Gdańsk and Medical University of Gdańsk, Poland, and Dr. Maciej Kozak, Adam Mickiewicz University, Poznań, Poland. In using knowledge-based information in UNRES, I am collaborating with Dr. Jooyoung Lee, Korea Institute for Advanced Study, Seoul, Republic of Korea, and Dr. Silvia Crivelli, National Instituted of Health, U.S.A., and Dr. Michela Taufer,

Delaware University, U.S.A. I am also collaborating with Dr. Antti Niemi, Uppsala University, Sweden, University of Tours, France and Beijing Institute of Technology (BIT), P.R. China, and with Dr. Jianfeng He, BIT, and Dr. Yi Xiao, Huazhong University of Science and Technology, Wuhan, P.R. China, in using UNRES and the Landau Hamiltonian to study protein folding.

Funding and access to supercomputer facilities

My current research is supported by two grants from the Polish National Science Center and Foundation for Polish Science. I have access to the supercomputers of the Academic Computer Center in Gdańsk TASK (www.task.pl), Interdisciplinary Center of Mathematical and Computer Modeling, University of Warsaw (www.icm.edu.pl) and, through collaboration with H.A. Scheraga and S. Crivelli, to XSEDE and NERSC supercomputing facilities in the U.S.A.

References

- [1] Liwo, A., Czaplewski, C., Pillardy, J. and Scheraga, H. A., *J. Chem. Phys.*, **115**, 2323–2347 (2001).
- [2] Liwo, A., Czaplewski, C., Ołdziej, S., Rojas, A. V., Kaźmierkiewicz, R., Makowski, M., Murarka, R. K. and Scheraga, H. A. “Simulation of protein structure and dynamics with the coarse-grained UNRES force field”. In Voth, G., editor, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, chapter 8, 1391–1411, 2008, (2008).
- [3] Liwo, A., He, Y. and Scheraga, H. A., *Phys. Chem. Chem. Phys.*, **13**, 16890–16901 (2011).
- [4] He, Y., Mozolewska, M. A., Krupa, P., Sieradzan, A. K., Wirecki, T. K., Liwo, A., Kachlishvili, K., Rackovsky, S., Jagieła, D., Ślusarz, R., Czaplewski, C. R., Ołdziej, S. and Scheraga, H. A., *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 14936–14941 (2013).
- [5] Liwo, A., Baranowski, M., Czaplewski, C., Gołaś, E., He, Y., Jagieła, D., Krupa, P., Maciejczyk, M., Makowski, M., Mozolewska, M. A., Niadzedtski, A., Ołdziej, S., Scheraga, H. A., Sieradzan, A. K., Ślusarz, R., Wirecki, T., Yin, Y. and Zaborowski, B., *J. Molec. Modeling*, **20**, 2306 (2014).
- [6] Liwo, A., Khalili, M. and Scheraga, H. A., *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 2362–2367 (2005).
- [7] Zhou, R., Maisuradze, G. G., Sunol, D., Todorovski, T., Macias, M. J., Xiao, Y., Scheraga, H. A., Czaplewski, C. and Liwo, A., *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 18243–18248 (2014).
- [8] Maisuradze, G. G., Zhou, R., Liwo, A., Xiao, Y. and Scheraga, H. A., *J. Mol. Biol.*, **420**, 350–365 (2012).
- [9] He, Y., Maciejczyk, M., Ołdziej, S., Scheraga, H. A. and Liwo, A., *Phys. Rev. Lett.*, **110**, 098101 (2013).
- [10] Golas, E., Maisuradze, G. G., Senet, P., Ołdziej, S., Czaplewski, C., Scheraga, H. A. and Liwo, A., *J. Chem. Theor. Comput.*, **8**, 1750–1764 (2012).

- [11] Mozolewska, M. A., Krupa, P., Scheraga, H. A. and Liwo, A., *Proteins: Struct. Funct. Bioinfo.*, **83**, 1414–1426 (2015).
- [12] Sieradzan, Adam K., Niadzvedtski, Andrei, Scheraga, Harold A. and Liwo, Adam, *J. Chem. Theory Comput.*, **10**, 2194–2203 (2014).
- [13] Chinchio, M., Czaplewski, C., Liwo, A., Ołdziej, S. and Scheraga, H.A., *J. Chem. Theory and Comput.*, **3**, 1236–1248 (2007).
- [14] Ayton, G. S., Noid, W. G. and Voth, G. A., *Curr. Opinion Struct. Biol.*, **17**, 192–198 (2007).
- [15] Marrink, S. J. and Tieleman, D. P., *Chem. Soc. Rev.*, **42**, 6801–6822 (2013).
- [16] Kolinski, A., Godzik, A. and Skolnick, J., *J. Chem. Phys.*, **98**, 7420–7433 (1993).
- [17] Kubo, R., *J. Phys. Soc. Japan*, **17**, 1100–1120 (1962).
- [18] Nancias, M., Czaplewski, C. and Scheraga, H. A., *J. Chem. Theor. Comput.*, 513–528 (2006).
- [19] Czaplewski, C., Kalinowski, S., Liwo, A. and Scheraga, H. A., *J. Chem. Theor. Comput.*, **5**, 627–640 (2009).
- [20] Liwo, A., Ołdziej, S., Czaplewski, C., Kozłowska, U. and Scheraga, H. A., *J. Phys. Chem. B*, **108**, 9421–9438 (2004).
- [21] Zaborowski, B., Jagieła, D., Czaplewski, C., Hałabis, A., Lewandowska, A., Żmudzińska, W., Ołdziej, S., Karczyńska, A., Omieczynski, C., Wirecki, T. and Liwo, A., *J. Chem. Inf. Model.*, **55**, 2050–2070 (2015).
- [22] Hałabis, A., Żmudzińska, W., Liwo, A. and Ołdziej, S., *J. Phys. Chem. B*, **116**, 6898–6907 (2012).
- [23] He, Y., Liwo, A. and Scheraga, H. A., *J. Chem. Phys.*, **143**, 243111 (2015).
- [24] Krupa, P., Mozolewska, M. A., Joo, K., Lee, J., Czaplewski, C. and Liwo, A., *J. Chem. Inf. Model.*, **55**, 1271–1281 (2015).
- [25] Khoury, G.A., Liwo, A., Khatib, F., Zhou, H., Chopra, G., Bacardit, J., Bortot, L. O., Faccioli, R. A., Deng, X., He, Y., Krupa, P., Li, J., Mozolewska, M. A., Sieradzan, A. K., Smadbeck, J., Wirecki, T., Cooper, S., Flatten, J., Xu, K., Baker, D., Cheng, J., Delbem, A. C. B., Floudas, C. A., Keasar, C., Levitt, M., Popovič, Z., Scheraga, H. A., Skolnick, J., Crivelli, S. N. and Foldit Players, *Proteins: Struct. Funct. Bioinfo.*, **82**, 1850–1868 (2014).
- [26] Krokhotin, A., Liwo, A., Maisuradze, G. G., Niemi, A. J. and Scheraga, H. A., *J. Chem. Phys.*, **140**, 4855735 (2014).
- [27] Kityk, R., Kopp, J., Sinning, I. and Mayer, M. P., *Mol. Cell*, **48**, 863–874 (2012).
- [28] Gołaś, E. I., Czaplewski, C., Scheraga, H.A. and Liwo, A., *Proteins: Struct. Func. Bioinf.*, **83**, 282–299 (2015).