John von Neumann Institute for Computing

Coarse-Grained Molecular Models for High-Throughput and Multi-Scale Functional Investigations

K. Hamacher

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 97-100, 2008.

© 2008 by John von Neumann Institute for Computing
Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

http://www.fz-juelich.de/nic-series/volume40

Coarse-Grained Molecular Models for High-Throughput and Multi-Scale Functional Investigations

Kay Hamacher

Technische Universität Darmstadt, Fachbereich Biologie, Institut für Mikrobiologie und Genetik,
Schnittspahnstr. 10, 64287 Darmstadt, Germany
`http://www.kay-hamacher.de`
E-mail: hamacher@bio.tu-darmstadt.de

We discuss the development of a physics-based, coarse-grained molecular model and its subsequent application to in silico functional investigation. Such models overcome shortcomings of both sequence-based bioinformatics (no physics) and molecular dynamics simulations (large CPU requirements, therefore sampling issues) as an integrative approach. We demonstrate this claim by showing: capabilities to investigate macromolecular thermodynamics [here: assembly of the bacterial ribosome[1]] integrating information into the physical realm [here HIV protease[2] and drug resistance[3]] We discuss also further (technical) advantages: the computations are highly parallelizable, allow for high-throughput screening, and can be integrated with an information driven analysis technique[4].

1 Introduction

Understanding protein-protein interactions, signaling pathways, protein dynamics, and biomolecular mechanics in general is of uttermost importance in systems biology and in all areas of life science.. Proof of the great use of mathematical modeling of signaling-networks are well known, e.g. the work by Lee et.al. on the Wnt signaling for oncogenesis[5].

To this end established protocols rely on off-the-shelf bioinformatics algorithms such as Hidden Markov Models (HMM) and other machine learning approaches to investigate e.g. the vertices in interaction networks. It was, however, noted for example by Albert[6] that most computationally derived protein-protein-interactions networks show too many false positives. This can be attributed to several drawbacks of statistical modelling as the underlying theory of sequence-based bioinformatic procedures such as HMMs: these bioinformatics models are ’local’ in the sense that they not necessarily take into account multi-amino-acids correlations - a property that contributes to recognition and structural stability as recently re-confirmed by the Plotkin lab[7].

Molecular dynamics[8] on the other hand allows for the detailed investigation of physical and chemical properties. The high computational costs are the major drawback of this approach: it is – despite exponential progress in hardware and software – still impossible to simulate large number of protein mutants, although such databases are routinely investigated by sequence-based algorithms.

Now reduced molecular models allow to bridge the gap between pure sequence-focused bioinformatics and the detailed molecular biophysics – allowing for real *in silicio* structural & functional proteomics and derivation of the above mentioned interaction networks due to the fast computation of reduced (thermo-)dynamics in those models. These models take into account physical interactions and their effects such as binding stability or responses to perturbations of the protein's structure. Both the elimination of false positives and the inclusion of potential, not yet detected interactions will help to derive more precise protein interaction networks, perform myriads of thought experiments, and allow therefore the functional characterisation of the underlying genomes.

2 Reduced Molecular Models

Interactions within biomolecules are usually modelled with sohphisticated force-fields and contain various terms such as Lennard-Jones-potentials, electrostatics, hydrogen-bonding etc. For small perturbations such a full potential can be expanded to second order in a Taylor expansion, which then poses the problem of diagonalizing a quadratic form[9]. Eigenmovements and eigenfrequencies are then easily determined. Tirion[10] was the first to investigate the resulting dynamics. Extensions with respect to anisotropy[11], extended thermodynamics[12], or non-linear conformational transitions[13] were introduced in recent years.

3 Applications

3.1 Protein Aggregation Network in the Ribosome

Using Micheletti's Self-Consistent-Pair-Contact-Potential iteration scheme[12] we investigated the influence of the presence or absence of ribosomal proteins and pairs of them on the binding affinities of the others[1]. This is a protocol suitable for general biomolecular formation processes. For the organism *T. thermophilus* we found that the assembly map is in very good agreement with the experimental known one for *E. coli*. This study examplifies: reduced molecular models can be used in multiscale-investigations and enable one to derive computationally protein-protein-interaction networks.

3.2 Amino-Acid-Mutation Network in the HIV1-Protease

We were furthermore able to parameterize reduced models to incorporate effects of mutated amino acids[2]. Using this system we investigated the origin of the drug resistance mechanism of the V82F-I84V-mutant of HIV1-protease (induced increase of flexibility in protein-substructures). By computing ”hybrid” mutants with systemtically changed wild-type-like and mutant-like interactions one can reveal the mechanism: this particular mutation weakens the interaction at the ”joints” of the protease-dimer. With this we have shown for the first time that a *functional* annotation of mutation events is possible. Continuing this work we applied the method to some 40,000 mutants[3] and correlated the changes in the protein mechanics to the stability in sequence space. The dimerization interface in the protease turned out to be an interesting target for new protease inhibitors: it has a large influence on the efficient molecular dynamics, while at the same time there are evolutionary barriers so that drug resistance evolution is repressed.

4 Concluding Remarks

We have motivated a new approach to cope with shortcomings of well-established and widely known algorithms in computational biology: sequence-based bioinformatics (no insight into the molecular biophysics) and molecular dynamics (prohibitive high computational costs). With such reduced molecular models it is possible to merge the information space of genomes with the physical realm of proteomes - thus allowing for physical underpinning of molecular systems biology.

Two examples illustrate the usefullness of this approach: the assembly network of proteins in the ribosome and the amino-acid network during evolution of drug resistance.

Acknowledgments

We are indebted to the organizers of the CBSB08 for the opportunity to present our work. The underlying studies were financially supported by the Fonds der chemischen Industrie through a Liebig-Fellowship and through ongoing financial support.

References

1. K. Hamacher, J. Trylska, and J.A. McCammon, *Dependency Map of Proteins in the Small Ribosomal Subunit*, PLoS Computational Biology, **2**, e10, 2006.
2. K. Hamacher and J. A. McCammon, *Computing the Amino Acid Specificity of Fluctuations in Biomolecular Systems*, J. Chem. Theory Comput., **2**, 873–878, 2006.
3. K. Hamacher, Relating sequence evolution of hiv1-protease to its underlying molecular mechanics, 2008, Gene, accepted.
4. K. Hamacher, *Information Theoretical Measures to Analyze Trajectories in Rational Molecular Design*, J. Comp. Chem., **28**, 2576–2580, 2007.
5. E. Lee, A. Salic, et al. *The Roles of APC and Axin Derived from Experimental and Theoretical Analysis of the Wnt Pathway*, PLoS Biology, **1**, e10, 2003.
6. R. Albert, *Scale-free networks in cell biology*, J Cell Sci, **118**, 4947–4957, 2005.
7. M. R. Ejtehadi, S. P. Avall, and S. S. Plotkin, *Three-body interactions improve the prediction of rate and mechanism in protein folding models*, Proc. Nat. Acad. Sci., **101**, 15088–15093, 2004.
8. T. Schlick, *Molecular Modeling and Simulation*, Interdisciplinary Applied Mathematics - Mathematical Biology. Springer, New York, 2002.
9. W.H. Press et al, *Numerical Recipies in C*, Cambridge University Press, Cambridge, 1995.
10. M.M. Tirion, *Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis*, Phys. Rev. Lett., **77**, 1905–1908, 1996.
11. A.R. Atilgan, S.R. Durrell, R.L. Jernigan, et al. *Anisotropy of fluctuation dynamics of proteins with an elastic network model*, Biophys. J., **80**, 505–515, 2001.
12. C. Micheletti, J.R. Banavar, and A. Maritan, *Conformations of Proteins in Equilibrium*, Physical Review Letters, **87**, 088102–1, 2001.
13. O. Miyashita, J. N. Onuchic, and P. G. Wolynes, *Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins*, PNAS, **100**, no. 22, 12570–12575, 2003.