1st International Conference on Mathematical and Computational Biomedical Engineering – CMBE2009
June 29 – July 1, 2009, Swansea, UK
P. Nithiarasu and R. Löhner (eds)

MULTI-SCALE MODELING OF MACROMOLECULAR CONFORMATIONAL CHANGES

Aqeel Ahmed
Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany, aqeel@bioinformatik.uni-frankfurt.de
Holger Gohlke
Institut für Pharmazeutische und Medizinische Chemie, Heinrich-Heine-Universität, Düsseldorf, Germany, gohlke@uni-duesseldorf.de

ABSTRACT

Key Words: *Constrained geometric simulation, rigid cluster normal mode analysis.*

1. INTRODUCTION

Specific functions of biological systems often require conformational transitions of macromolecules. Such changes range from movements of single side-chains and loop rearrangements to large scale domain motions. In binding events involving macromolecules, molecular motions provide the origin of plasticity of the binding partners, enabling them to conformational adapt to each other.[1, 2] Thus, being able to describe and predict conformational changes of biological macromolecules is not only important for understanding their impact on biological function, but will also have implications for the modeling of (macro)molecular complex formation [3] and in structure-based drug design approaches.[4] Modelling conformational transitions of macromolecules is computationally challenging. Hence, coarse-grained approaches have emerged as efficient alternatives for investigating large-scale conformational changes.[5] Here, we introduce a three-step approach for multi-scale modeling of macromolecular conformational changes (Figure 1).

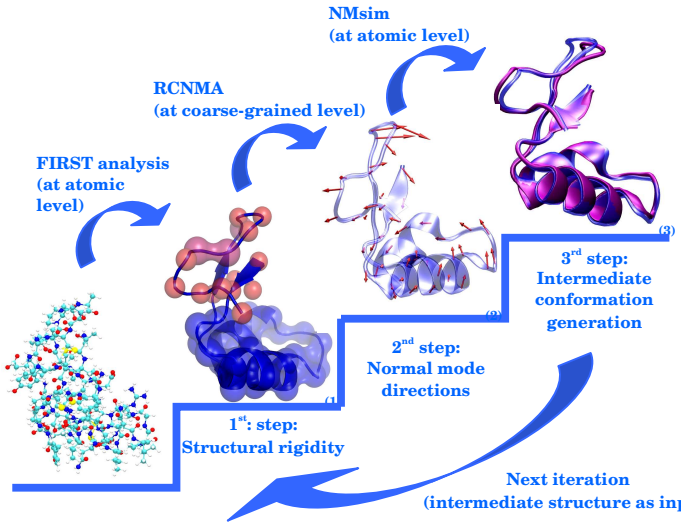


Figure 1: Three-step approach for multi-scale modeling of macromolecular conformational changes

2. RESULTS AND DISCUSSION

The first two steps are based on recent developments in rigidity and elastic network theory.[6] Initially, *static* properties of the macromolecule are determined by decomposing the macromolecule into rigid clusters using the graph-theoretical approach FIRST [7] and an all-atom representation of the protein. That way, the rigid cluster decomposition is not limited to consist of residues adjacent in sequence or secondary structure elements as in previous studies.[8] Furthermore, flexible links between rigid clusters are identified and can be modeled as such subsequently. In a second step, *dynamical* properties of the molecule are revealed by the rotations-translations of blocks approach (RTB) [9] using an elastic network model representation of the coarse-grained protein (termed **R**igid **C**luster **N**ormal **M**ode **A**nalysis). I.e., in this step, only rigid body motions are allowed for rigid clusters while links between them are treated as fully flexible.
In the final step, the recently introduced idea of constrained geometric simulations of diffusive motions in proteins [10] is extended. New macromolecule conformers are generated by deforming the structure along low-energy normal mode directions predicted by RCNMA plus random direction components. The generated structures are then iteratively corrected regarding steric clashes or constraint violations. This module is termed NMsim. Constraints to be satisfied include torsions of the main-chain and side-chains, distances and angles due to noncovalent interactions such as hydrogen bonds or hydrophobic interactions, and bond, angle, and planarity constraints. In total, when applied repetitively over all three steps, the procedure generates efficiently series of conformations that lie preferentially in the low energy subspace of normal modes.
The RCNMA approach was initially tested on a data set of 10 proteins that show conformational changes upon ligand binding.[6] In terms of efficiency, coarse-graining the protein results in a remarkable reduction of memory requirements and computational times by factors of 9 and 27 on average and up to 25 and 125, respectively. In terms of accuracy, directions and magnitudes of motions predicted by our approach agree well with experimentally determined ones (Figure 2), despite embracing in extreme cases more than 50 % of the protein into one rigid cluster. In fact, the results of our method are in general comparable to if no or a uniform coarse-graining is applied and become superior if the movement is dominated by loop or fragment motions (Figure 2b). This indicates that explicitly distinguishing between flexible and rigid regions is advantageous when using a simplified protein representation in the second step. Finally, motions of atoms in rigid clusters are also well predicted by our approach, which points to the need to consider *mobile* protein regions in addition to *flexible* ones when modeling correlated motions.

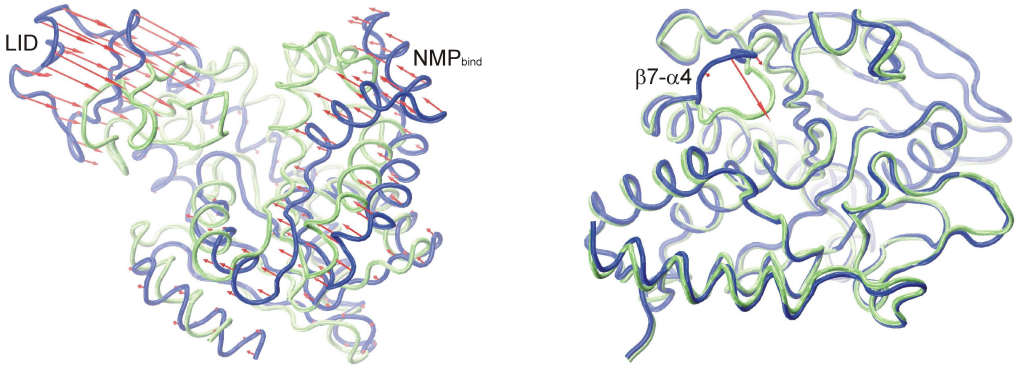


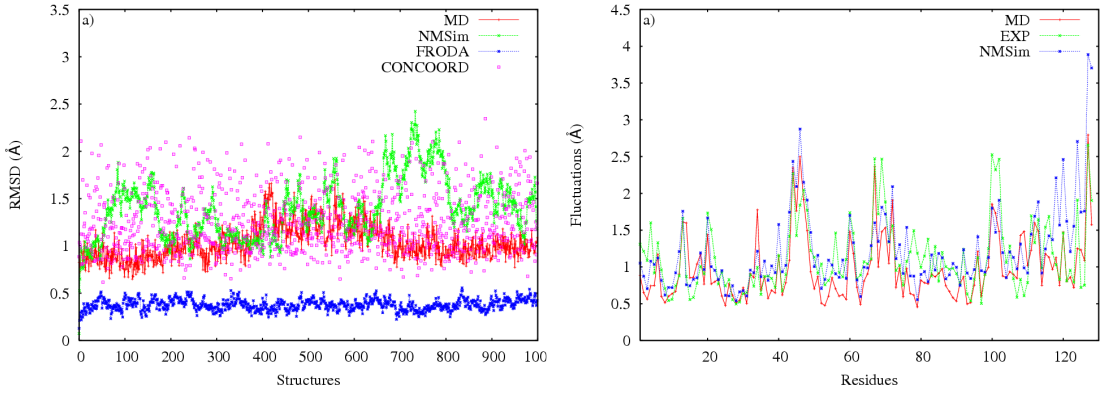
Figure 2: Superimposition of open (blue) and closed (green) conformations of adenylate kinase (panel a) and tyrosine phosphatase (panel b). In addition, the amplitudes and directions of motions as predicted by RCNMA modes most involved in the conformational changes, respectively, are depicted as red arrows. In both cases, the amplitudes of the motions were scaled for best graphical representation.

The NMsim approach was tested on hen egg white lysozyme (HEWL), a well-investigated protein consisting of 129 residues, in terms of exploration of the conformational space, residue fluctuations, and quality of the generated structures. Experimental structures and conformations from state of the art molecular dynamics simulations [11] were compared with conformations obtained from the constrained

geometric simulation-based approach FRODA,[10] the distance geometry-based approach CONCOORD,[12] and NMsim.

In the case of constrained geometric simulations, the progression of a series of conformers is usually measured in terms of the root means square deviation (RMSD) from a reference conformer. Figure 3a) shows RMSD values of backbone atoms with respect to the HEWL starting structure for the different methods. Except FRODA, all other methods show considerable RMSD predominantly in the range of 1 to 2 Å. While FRODA apparently underestimates the conformational variability of HEWL, CONCOORD provides the broadest sampling. NMSim explores conformations in a similar range as CONCOORD, with higher RMSD values and more frequent peaks compared to MD. An analysis in terms of the "essential dynamics" [13] of HEWL confirms this finding: CONCOORD and NMsim broadly sample the space spanned by the first two principal components obtained from the MD trajectory (data not shown).

Regarding residue fluctuations, i.e., mass-weighted averages of heavy-atom fluctuations for each residue, NMsim and CONCOORD results are strongly correlated with MD results with correlation coefficients of 0.79, respectively. Encouragingly, for the residue fluctuations obtained from NMsim and CONCOORD a good correlation with fluctuations obtained from 130 experimental HEWL structures is found, too, with correlation coefficients of ~0.70. Residue fluctuations from NMsim, MD, and experiment are given in Figure 3b). Qualitatively and quantitatively, the NMsim results are in good agreement with those from MD and experiment, with the C-terminal region being the exception.



a) b)

Figure 3: Panel a: Backbone RMSD between the starting structure and the conformations generated by MD (red), NMSim (green), FRODA (blue), and CONCOORD (magenta). Panel b: Residue fluctuations obtained from MD (red), 130 experimental structures (green), and NMSim (blue).

Finally, the stereochemical quality of structures obtained from NMsim was analyzed using Procheck.[14] For comparison, 100 high resolution crystal structures from Richardson's lab [15] were analysed, as were 130 experimental HEWL structures. Table 1 summarizes the results. Procheck divides the Ramachandran map into four types: most favored or core, additionally allowed, generously allowed, and disallowed. NMsim shows a good Ramachandran plot distribution with almost no structures in generously allowed and disallowed regions and with a highly populated core region (92 %) due to the specific modeling of phi/psi constraints. These results are in remarkable agreement with the characteristics of 100 high resolution experimental structures. The *G*-factor provides a measure of how likely a stereochemical property is. In Procheck it is computed for torsions and covalent geometry. A low *G*-factor indicates that the property corresponds to a low-probability conformation. Ideally, the G-factor should be above -0.5. Values below -1.0 indicate that a re-investigation of the structure is necessary. Table 1 shows that for NMsim the overall *G*-factor is -0.3. This value is comparable to the other geometry-based simulation methods CONCOORD and FRODA (data not shown). Regarding the planarity of, e.g., aromatic rings, NMsim succeeds in all cases.

3. CONCLUSIONS

In summary, the multi-scale approach introduced here provides an efficient means for sampling macromolecular conformational changes. In this respect, a computational time of ~30 h in the HEWL case compares favorably to days or weeks of computational time required by FRODA or MD, respectively. At the same time, the sampling is exhaustive and provides protein conformations of good stereochemical quality. At present, the scope and limitations of the approach are further tested regarding a hierarchy of movements, including domain, loop, and side-chain motions.

Table 1: Stereochemical quality of NMsim generated and experimental structures

Dataset	**Ramachandran plot** [a]				***G*-factor** [b]			**Planar** [c]
	Core	**Allow**	**Gen.**	**Disal.**	**Dihe.**	**Cova.**	**Overall**	
NMSim [d]	92.55	7.42	0.01	0.00	-0.26	-0.36	-0.30	100.0
EXP [e]	81.31	17.62	0.89	0.18	-0.07	0.24	0.06	98.34
EXPTOP [f]	91.26	8.30	0.28	0.14	0.06	-0.28	-0.05	92.82

a) Percentages of phi/psi values found in different regions (i.e. core, allowed, generously allowed, and disallowed) of the Ramachandran map. b) Procheck-derived *G*-factor. Procheck computes *G*-factors for torsions, covalent geometry and overall. c) Percentages of side-chain planarity. d) NMsim generated conformations. e) 130 experimental HEWL structures. f) 100 high resolution structures taken from ref. [15].

REFERENCES

[1] C.-S. Goh, D. Milburn and M. Gerstein, M., Conformational changes associated with protein-protein interactions. *Curr Opin Struct Biol* 14, 104-109, 2004.

[2] S. J. Teague, Implications of Protein Flexibility for Drug Discovery. *Nat Rev Drug Discov* 2, 527-541, 2003.

[3] H. A. Carlson, Protein flexibility and drug design: How to hit a moving target. *Curr Opin Chem Biol* 6, 447-452, 2002.

[4] A. Ahmed, S. Kazemi, H. Gohlke, Protein flexibility and mobility in structure-based drug design. *Frontiers Drug Des Discov* 3, 455-476, 2007.

[5] V. Tozzini, Coarse-grained models for proteins. *Curr Opin Struct Biol* 15, 144-150, 2005.

[6] A. Ahmed, H. Gohlke, Multi-scale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins* 63, 1038-1051, 2006.

[7] D. J. Jacobs, A. J. Rader, L. A. Kuhn, M. F. Thorpe, Protein flexibility predictions using graph theory. *Proteins* 44, 150-65, 2001.

[8] F. Tama, F. X. Gadea, O. Marques, Y.-H. Sanejouand, Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins* 41, 1-7, 2000.

[9] P. Durand, G. Trinquier, Y.-H. Sanejouand, A New Approach for Determining Low-Frequency Normal Modes in Macromolecules. *Biopolymers* 34, 759-771, 1994.

[10] S. Wells, S. Menor, B. M. Hespenheide, M. F. Thorpe, Constrained geometric simulation of diffusive motions in proteins. *Phys Biol* 2, 1-10, 2005.

[11] A. N. Koller, H. Schwalbe, H. Gohlke, Starting structure dependence of NMR order parameters derived from MD simulations. *Biophys J* 95, L04-6, 2008.

[12] B. L. de Groot, G. Vriend, H. J. Berendsen, Conformational changes in the chaperonin GroEL: new insights into the allosteric mechanism. *J Mol Biol* 286, 1241-9, 1999.

[13] A. Amadei, A. B. Linssen, H. J. Berendsen, Essential dynamics of proteins. *Proteins* 17, 412-25, 1993.

[14] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst* 26, 283-291, 1993.

[15] J. M. Word, S. C. Lovell, T. H. LaBean, H. C. Taylor, M. E. Zalis, B. K. Presley, J. S. Richardson, D. C. Richardson, Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J Mol Biol* 285, 1711-33, 1999.