ELSEVIER

Coarse-grained models for proteins

Valentina Tozzini

Coarse-grained models for proteins and biomolecular aggregates have recently enjoyed renewed interest. Coarse-grained representations combined with enhanced computer power currently allow the simulation of systems of biologically relevant size (submicrometric) and timescale (microsecond or millisecond). Although these techniques still cannot be considered as predictive as all-atom simulations, noticeable advances have recently been achieved, mainly concerning the use of more rigorous parameterization techniques and novel algorithms for sampling configurational space. Moreover, the simulation size scales and timescales coincide with those that can be reached with the most advanced spectroscopic techniques, making it possible to directly compare simulation and experiment.

Addresses
NEST-INFM Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy

Corresponding author: Tozzini, Valentina (tozzini@nest.sns.it)

Current Opinion in Structural Biology 2005, **15**:144–150

This review comes from a themed issue on
Theory and simulation
Edited by J Andrew McCammon and Rebecca C Wade

Available online 9th March 2005

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2005.02.005

Introduction

During the past two decades, the size of system addressable with computer simulations has gradually increased, allowing the inclusion of biochemistry as an area of research that can benefit from this powerful investigation tool. Considering specifically force-field-based all-atom approaches, this has been possible thanks to the following achievements: enhanced computer power has now reached memory and speed requirements sufficient to treat explicitly solvated proteins; new and more efficient techniques to sample configurational space have been proposed [1,2]; and force-field parameterization — optimized through a continuous and ongoing validation process — has become accurate enough to explain and predict experimental results [3,4].

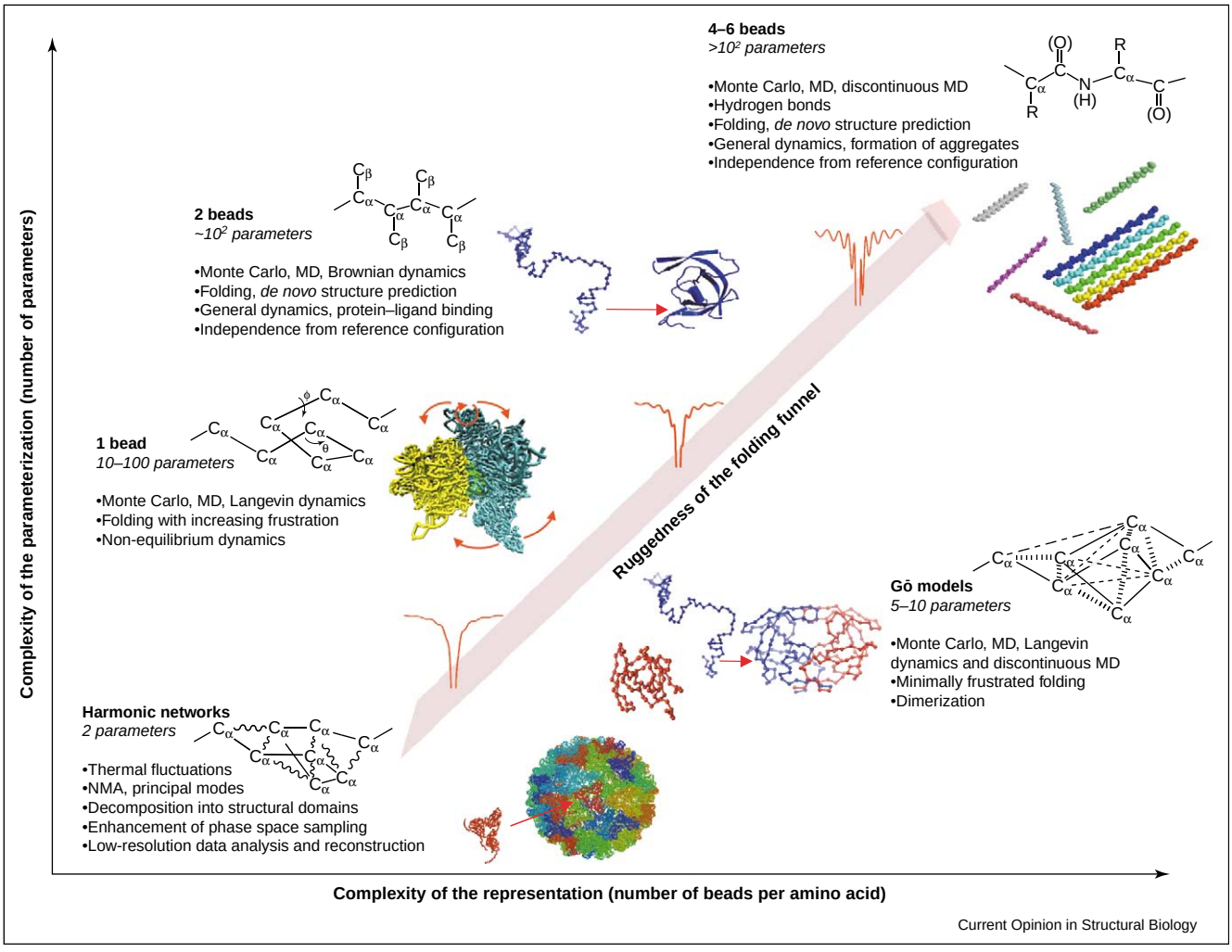
Ten or hundred nanosecond simulations of proteins in an explicit aqueous environment are currently feasible. However, most of the relevant dynamics and interactions within cells (typically, protein–protein docking, rearrangement upon ligand binding or after biochemical reactions, folding) occur on the timescale of microseconds or milliseconds, and involve large macromolecular aggregates. In these processes, the number of degrees of freedom is at least one order of magnitude and the timescale four to six orders of magnitude larger than what is currently feasible with all-atom simulations. Furthermore, in some cases, atomically detailed simulations might not be the most appropriate, as an increasing number of biomolecular aggregates are being studied experimentally using low to medium resolution techniques (such as cryo-EM and small-angle X-ray scattering [5,6]). Thus, the idea of using simplified descriptions through the 'integration' of a large number of degrees of freedom into a few — 'coarse graining' — arises spontaneously. Coarse-grained approaches, which have been around for years [7], have recently enjoyed renewed interest. With respect to earlier studies, a larger variety of different simplified descriptions and more rigorous methodologies for the parameterization are currently being proposed.

This review provides a classification of currently used coarse-grained models for macro-biomolecular systems, focusing on the most recent applications and pointing out the innovative aspects. The bead models, that is, models based on a united-atom representation of the amino acid (involving one to six interacting centers), are presented. In general, as the number of beads decreases, the simulation is less expensive and the system that can be simulated is larger. However, parameterizing force-fields that are both accurate and transferable — that is, capable of describing the general dynamics of systems with different compositions and different configurations — becomes increasingly difficult as the graining becomes 'coarser', because more specific interactions must effectively be included in fewer parameters and functional forms. For this reason, most currently popular coarse-grained models are parameterized based on a single reference configuration and the dynamics they reproduce are strongly biased towards it [8,9]. Different models represent different compromises between accuracy and transferability, with different degrees of independence from the reference configuration. Recent methodologies for 'extreme' coarse graining are also reported, based on different mathematical approaches for the integration of the degrees of freedom.

Elastic network models

In elastic network models (ENMs), the system is represented by a network of beads connected by elastic springs (see Figure 1), usually one bead per amino acid (although

Figure 1



Pictorial representation of the features of bead models. For each class of model, the following aspects are reported: schematic representation of the model, indicative number of parameters, methods of solution, main characteristics and applications. Sample applications are also illustrated with representative pictures (prepared using crystallographic coordinates from the PDB [codes 1hhp, 1cwp, 1mwr, 486d]) intended to show the size of system that can be studied and the kind of study that can be done. The location of the models in the *x*-*y* plane is intended to qualitatively illustrate their complexity, which increases following the direction of the arrows.

elastic networks have also been used together with all-atom descriptions [8,10]). The extreme simplicity of the parameterization is balanced by the need to know the equilibrium reference configuration, from which only harmonic fluctuations are possible. However, an ENM correctly includes the topology of the system and is able reproduce the correct pattern of the principal modes (i.e. the modes with the largest amplitude). These are usually the most relevant to the protein function. ENMs have also proven to be suitable tools for analyzing certain general aspects of protein behavior. Methods to automatically decompose proteins into structural domains [11•] or to identify the signature of secondary structures [12] were developed based on ENMs. ENMs were also recently used to show a general relationship between the size (length) of a protein and its topological connectivity [13•]. In some new methods to enhance the sampling of the phase space, the principal modes evaluated with ENM are used to 'guide' the system towards less populated zones of the phase space during a molecular dynamics (MD) simulation [14•] or to interpolate between two given conformations separated by potential barriers [15•]. A recent development of the model is βNM, a 'two-bead' network model, which includes the centroids of the sidechains as interacting centers [16•].

ENMs apply naturally to the analysis and refinement of low-resolution data [17]. Recent applications include fitting atomic structure into electron density maps [18•], the analysis of the fundamental motions of

mega-Dalton structures based on low-resolution cryo-EM maps [19••] and the reconstruction of molecular aggregates by flexibly docking high-resolution structures into low-resolution data [20••]. Beside these special applications, ENMs are currently used with normal mode analysis (NMA) for the analysis of the principal modes of a large variety of different systems. A (non-exhaustive) list of recent examples includes diphtheria toxin [21], the 70S ribosome from *Thermus thermophilus* [22•], influenza virus hemagglutinin [23], the bacterial chaperonin GroEL–GroES [24] and motor proteins [25•,26]. ENMs may also be applied to relatively small systems (e.g. HIV-1 protease and rhodopsin [27,28]), when efficient NMA of large sets of structures is required.

Gō-like models

The Gō model was proposed long ago specifically for the simulation of protein folding; it was recently discovered to be more general and capable of innovative applications and developments. In the original version, the protein is represented as a chain of one-bead amino acids whose structure is biased toward the native configuration by means of simple attractive or repulsive non-bonded interactions between beads [9]. The success of this extremely simplified representation in reproducing several aspects of the thermodynamics and kinetics of folding is due to the fact that the folding process has generally evolved to satisfy the principle of minimal frustration [29]. Consequently, the energy landscape of the protein can be described by a weakly rugged funnel pointing towards the native structure (Figure 1). The folding rate is then primarily correlated with the topological complexity of the native state, from which the folding pathways and the thermodynamics (and kinetics) of folding can be reasonably inferred [30]. According to this point of view, the Gō model works because it represents a perfectly funneled model for folding that correctly includes the topology of the native state [29–31].

However, completely unfrustrated Gō models fail in describing intermediate metastable folding states. Thus, the model has been recently reconsidered and made more sophisticated. When new energy terms are added to the model, frustration is increased and new intermediate states appear, making the folding funnel more rugged. For instance, the inclusion of a solvation-desolvation barrier to the non-bonded interactions leads to the appearance of a partially desolvated-partially folded intermediate [32,33]. An interesting recent development of the model concerning the integration algorithm is an adaptation for use with discontinuous MD [34]. This version of the Gō model was recently used to study the folding and stability of β-sheet complexes [35,36•]. With the aim of introducing amino acid specificity, a Gō-like parameterization was used in conjunction with an all-(heavy) atom description [37,38] and shown to catch important sequence-dependent features. The Gō model was also recently used to study the binding of monomers in a variety of homodimeric proteins [39••,40] and to investigate the confining effects of the chaperonin cage on the folding kinetics [41•].

Towards independence from the reference configuration

The bias of ENM and Gō models towards a reference configuration makes them only weakly transferable to general dynamics studies. Early attempts to build transferable coarse-grained models date back to the 1970s, with the work of Levitt, who reported a knowledge-based parameterization [7] that inspired many subsequent studies. As will emerge in the following, the smaller the number of beads representing an amino acid, the harder it is to build a transferable parameterization. In this review, the models are grouped into three main classes according to the number of interacting centers for each amino acid. Passing from the coarsest (one bead) to the finest (four–six beads) class, the explicit representation of the sidechain and peptide bond atoms is progressively included (Figure 1).

One-bead models

After very early work using knowledge-based parameterization [42] or parameterizations designed to describe specific structural transitions [43], most of the currently available one-bead models are an evolution of Gō models that includes more sophisticated potentials, but still retains a partial bias towards a reference configuration. This is due to the difficulty of including, in only a few parameters, generic effects of the amino acid size, geometry and conformation; this would result in inter-bead potentials that are highly anisotropic, multimodal and specifically dependent on the local biochemical environment [44,45]. The dependence on a reference configuration is maintained by different means, depending on the model. In the model developed by Head-Gordon and colleagues [46••], *a priori* knowledge of a reference secondary structure is required for the parameterization of the angle and dihedral terms, whereas the sequence specificity is included in non-bonded terms describing hydrophobicity/hydrophilicity. The model is able to discriminate the different folding behavior of proteins with the same native topology and was also recently used to analyze the effects of crowding in the cell environment on folding kinetics [47]. By introducing a parameter that adjusts the strength of the non-native hydrophobic interactions, a continuous set of models interpolating between unfrustrated Gō-like models and Sorenson-like models were built to show how frustration influences the folding kinetics [48•].

In a very recent study, a model for general dynamics was presented with a sequence-specific and structure-independent parameterization, except for the local non-bonded interactions, which are biased towards a reference

structure to maintain local order. In this way, very good structural accuracy is achieved, but fluctuations as large as complete flap opening in HIV-1 protease, which occurs on the microsecond timescale, can be simulated with a modest computational cost [49]. The model has also been recently applied to the 70S bacterial ribosome, for which several hundred-nanosecond MD runs have been performed, revealing slow highly anharmonic motions [50].

Two-bead models

Adding a second bead on the centroid of the sidechain improves the specificity of local interactions. Bahar and Jernigan [44] developed a two-bead model with a numerical force-field for general dynamics parameterized on the statistical analysis of a set of experimental structures [51]. The model is independent of a reference configuration, but the energy terms are very complex and also include angle-dihedral correlations. This model was very recently applied to the study of peptide binding to HIV-1 protease [52•]. Recent versions of this approach use a simpler analytical force-field, such as that of Mukherjee and Bagchi, whose force-field contains an additional term for helix propensity [53••]. A similar model with a parameterization optimized by fitting free energy functions from all-atom simulations of oligopeptides was developed by Scheraga and co-workers [54,55•]. One additional bead for the more extended sidechains (such as lysine) was used by Zacharias [56•] in a model of protein–protein docking that allows sidechain flexibility.

Four–six-bead models

In four-bead models, the sidechain is represented by a single bead, whereas the coordinates of the three heavy atoms of the backbone are represented explicitly, allowing an explicit description of the hydrogen bonds. The model by Smith and Hall [57,58] is one of the simplest, with constrained amino acid geometry, and is specifically designed to be used with discontinuous MD. It has been applied to the study of the formation of β-amyloid dimers [59], and the structural transitions and aggregation of fibrils of a polyalanine peptide [60,61,62•]. In the model by Fernandez and Colubri [63], the non-bonded energy potential is more complex, and includes explicit ionic, dipolar and disulfide bond terms. This model has been used with Monte Carlo sampling and is proven to be very efficient in *de novo* predicting the structure of small proteins [64].

The model by Takada and co-workers — characterized by particularly sophisticated hydrogen bond and dihedral terms — recently reached an optimized version [65,66,67•] and was applied to the simulation of the folding of protein G and α-spectrin SH3 [68]. A six-bead model with simplified hydrogen bond terms and sequence-specific hydrophobic interactions was developed by Irbäck and co-workers [69–71], and was recently applied to the study of the formation of secondary structures [72]. The model by Forcellino and Derremaux [73] uses harmonic terms to restrain the local geometry, and van der Waals like and hydrogen bond terms optimized for the folding of small peptide sequences. It was used in conjunction with Monte Carlo sampling for *de novo* structure prediction [74].

'Coarser' graining

Coarser descriptions are achieved using very different methodologies and approaches. One possibility is to use ENM in conjunction with rigid-block decomposition approaches [75,76•,77], which represent the system as a collection of rigid bodies (structural units) connected by springs. Electrostatics-driven biomolecular diffusional association studied using Brownian dynamics can be considered another example of 'coarser' graining [78•,79•]. Proteins undergo diffusional dynamics with translational and rotational diffusional coefficients evaluated under spherical assumptions, although the excluded volume and electrostatic interactions are evaluated by taking into account their shape in atomic detail. Although based on previously developed methodologies, interesting studies of protein–protein docking or the packing of secondary structure blocks to build a tertiary structure based on shape and electrostatic complementarity have recently appeared [80,81]. Finally, the representation of proteins as a continuum through their density fields via the density functional formalism — recently used by Kinjo and Takada [82] to study the competition between protein folding and protein aggregation with molecular chaperones in crowded solutions — can be considered the extreme level of coarse graining.

Conclusions and perspectives

The development of more rigorous parameterization methodologies, more accurate representations of the system and better validation procedures, together with the use of more efficient algorithms for sampling the phase space, has made coarse-grained approaches an extremely powerful tool for analysis and comparison with experimental results on mesoscopic size scales and timescales. This is a great achievement, allowing a more direct comparison between simulation and a vast variety of experimental techniques, including low-resolution ones. Furthermore, simulating events on 'biological' timescales and size scales is becoming feasible, and the possibility of simulating the machinery of a whole cell tangible, leading to a range of possible advances in biophysics, biochemistry and (molecular) biology.

Current coarse-grained methodologies are still not as predictive as all-atom simulations. This is partly because these methodologies have been less used; thus, the validation of associated force-fields is still not very advanced. Furthermore, there is intrinsic difficulty in the parameterization of coarse-grained force-fields, related to the fact that complex and diverse interactions

must be described by a small number of parameters. However, many significant steps have been made in the past few years. Improving the predictivity and transferability of coarse-grained approaches involves basic research in statistical mechanics and mathematics, beside biochemistry and biology. Thus, it is a stimulating challenge that will hopefully involve an increasing number of researchers.

Acknowledgements

I wish to thank J Andrew McCammon, fruitful collaborations with whom have greatly influenced the content of this paper. I also thank Joanna Trylska, for many useful discussions and stimulating interactions, and for helping me with the figure. I acknowledge the hospitality of the Department of Chemistry and Biochemistry, and the Center for Theoretical Biological Physics at the University of California, San Diego, which hosted me for a research visit.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Tai K: **Conformational sampling for the impatient**. *Biophys Chem* 2004, **107**:213-220.

2. Norberg J, Nilsson L: **Advances in biomolecular simulations: methodology and recent applications**. *Q Rev Biophys* 2003, **36**:257-306.

3. Karplus M: **Molecular dynamics of biological macromolecules: a brief history and perspective**. *Biopolymers* 2003, **68**:350-358.

4. Schleif R: **Modeling and studying proteins with molecular dynamics**. *Methods Enzymol* 2004, **383**:28-47.

5. Bhella D, Ralph A, Yeo RP: **Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction**. *J Mol Biol* 2004, **340**:319-331.

6. White SH, von Heijne G: **The machinery of membrane protein assembly**. *Curr Opin Struct Biol* 2004, **14**:397-404.

7. Levitt M: **A simplified representation of protein conformations for rapid simulation of protein folding**. *J Mol Biol* 1976, **104**:59-107.

8. Tirion MM: **Large amplitude elastic motions in proteins from a single-parameter, atomic analysis**. *Phys Rev Lett* 1996, **77**:1905-1908.

9. Ueeda Y, Taketomi H, Gō N: **Studies on protein folding, unfolding and fluctuations by computer simulation. A three-dimensional lattice model of lysozyme**. *Biopolymers* 1978, **17**:1531-1548.

10. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I: **Anisotropy of fluctuations dynamics of proteins with an elastic network model**. *Biophys J* 2001, **80**:505-515.

11. • Kundu S, Sorensen DC, Phillips GN Jr: **Automatic domain decomposition of proteins by Gaussian network model**. *Proteins* 2004, **57**:725-733.
A method is presented that uses ENM to automatically decompose a protein into structural domains, defined as portions of the protein that move independently. The accuracy of the method is verified for a large set of proteins and connections with graph theory are indicated.

12. Wu Y, Yuan X, Gao X, Fang H, Zi J: **Universal behavior of localization of residue fluctuations in globular proteins**. *Phys Rev E* 2003, **67**:041909.

13. • Burioni F, Cassi D, Cecconi F, Vulpiani A: **Topological thermal instability and length of proteins**. *Proteins* 2004, **55**:529-535.
ENM is used to analyze the relationship between topological thermal instability and the size of proteins. A general relationship is demonstrated between the size of the protein and the 'degree' of topological connectivity.

14. • Zhang Z, Shi Y, Liu H: **Molecular dynamics simulations of peptides and proteins with amplified collective motions**. *Biophys J* 2003, **84**:3583-3593.
ENM is used to compute the principal modes along a full-atom MD trajectory and amplify them to better explore the phase space. The method is applied to T4 lysozyme and to the refolding of an S-peptide analog.

15. • Kim MK, Li W, Shapiro BA, Chirikjian GS: **A comparison between elastic network interpolation and MD simulation of 16S ribosomal RNA**. *J Biomol Struct Dyn* 2003, **21**:395-405.
An efficient method based on ENM for generating conformations that interpolate between two different conformations is demonstrated and applied to 16S rRNA. The resulting conformations are compared with an all-atom MD simulation.

16. • Micheletti C, Carloni P, Maritan A: **Accurate description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models**. *Proteins* 2004, **55**:635-645.
An ENM based on Cα and Cβ positions is proposed; it has the same computational complexity of a one-bead ENM and is proven to give a better accuracy.

17. Delarue M, Dumas P: **On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models**. *Proc Natl Acad Sci USA* 2004, **101**:6957-6962.

18. • Hinsen K, Reuter N, Navaza J, Stokes DL, Lacapere J-J: **Normal mode based fitting of atomic structure into electron density maps: application to SR Ca-ATPase**. *Biophys J* 2005, **88**:818-827.
An ENM with distance-dependent elastic constants was used together with NMA to fit sarcoplasmic reticulum Ca-ATPase into electron density maps. Atomically detailed structures are obtained and conclusions are drawn about phosphorylation-induced conformational changes of the protein.

19. •• Chacon P, Tama F, Wriggers W: **Mega-dalton biomolecular motion captured from electron microscopy reconstructions**. *J Mol Biol* 2003, **326**:485-492.
ENM and NMA are used to reconstruct the fundamental motions of bacterial RNA polymerase, the 30S and 50S subunits of the ribosome, and the chaperonin CCT.

20. •• Tama F, Miyashita O, Brooks CL III: **Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM**. *J Struct Biol* 2004, **147**:315-326.
A method for fitting high-resolution structures into low-resolution experimental data is shown. The method is based on ENM and NMA, and allows protein flexibility by taking into account conformational changes induced by interactions between structures. The method is applied to elongation factor G bound to the ribosome, *Escherichia coli* RNA polymerase and cowpea chlorotic mottle virus.

21. Kundy S, Jernigan RL: **Molecular mechanism of domain swapping in proteins: an analysis of slower motions**. *Biophys J* 2004, **86**:3846-3854.

22. • Tama F, Valle M, Frank J, Brooks CL III: **Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy**. *Proc Natl Acad Sci USA* 2003, **100**:9319-9323.
The slower motions of the 70S ribosome from *T. thermophilus* are analyzed with ENM and NMA. Mechanical rearrangements that appear to be related to function are highlighted and are shown to agree with cryo-EM data.

23. Isin B, Doruker P, Bahar I: **Functional motions of influenza virus hemagglutinin: a structure-based analytical approach**. *Biophys J* 2002, **82**:569-581.

24. Keskin O, Bahar I, Flatow D, Covell DG, Jernigan RL: **Molecular mechanisms of chaperonin GroEL-GroES function**. *Biochemistry* 2002, **41**:491-501.

25. • Navizet I, Lavery R, Jernigan RL: **Myosin flexibility: structural domains and collective vibrations**. *Proteins* 2004, **54**:384-393.
An ENM of scallop myosin is reported and a method for the analysis of the rigid blocks of the protein is presented. Regions that have a key role in the change of rigidity of the protein are detected and analyzed.

26. Zheng W, Doniach S: **A comparative study of motor-protein motions by using a simple elastic-network model**. *Proc Natl Acad Sci USA* 2003, **100**:13253-13258.

27. Kurt N, Scott WRP, Schiffer CA, Haliloglu T: **Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: a structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations**. *Proteins* 2003, **51**:409-422.

28. Rader AJ, Andersona G, Isin B, Dhorana HG, Bahar I, Klein-Seetharaman J: **Identification of core amino acids stabilizing rhodopsin**. *Proc Natl Acad Sci USA* 2004, **101**:7246-7251.

29. Baker D: **A surprising simplicity to protein folding**. *Nature* 2000, **405**:39-42.

30. Koga N, Takada S: **Roles of native topology and chain-length scaling in protein folding: a simulation study with a Gō-like model**. *J Mol Biol* 2001, **313**:171-180.

31. Clementi C, Nymeyer H, Onuchic JN: **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**. *J Mol Biol* 2000, **298**:937-953.

32. Cheung MS, Garcia AE, Onuchic J: **Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after structural collapse**. *Proc Natl Acad Sci USA* 2000, **99**:685-690.

33. Kaya H, Chan HS: **Solvation effects and driving forces for protein thermodynamics and kinetic cooperativity: how adequate is native-centric topological modeling?** *J Mol Biol* 2003, **326**:911-931.

34. Jang H, Hall CK, Zhou Y: **Folding thermodynamics of model four-strand antiparallel β-sheet protein**. *Biophys J* 2002, **82**:646-659.

35. Jang H, Hall CK, Zhou Y: **Thermodynamics and stability of a β-sheet complex: molecular dynamics simulations on simplified off-lattice protein models**. *Protein Sci* 2004, **13**:40-53.

36. • Jang H, Hall CK, Zhou Y: **Assembly and kinetic folding pathways of a tetrameric β sheet complex: molecular dynamics simulations on simplified off-lattice protein model**. *Biophys J* 2004, **86**:31-49.
In this paper, a Gō-like model is used with discontinuous MD to investigate the folding dynamics and thermodynamic properties of the different kinds of intermediates that form during the folding process.

37. Clementi C, Garcia AE, Onuchic JN: **Interplay among tertiary contacts, secondary structure formation and side chain packing in the protein folding mechanism: all-atom representation study of protein L**. *J Mol Biol* 2003, **326**:933-954.

38. Hubner IA, Oliveberg M, Shakhnovich EI: **Simulation experiment and evolution: understanding nucleation in protein S6 folding**. *Proc Natl Acad Sci USA* 2004, **101**:8354-8359.

39. •• Levy Y, Wolynes PG, Onuchic J: **Protein topology determines binding mechanism**. *Proc Natl Acad Sci USA* 2004, **101**:511-516.
A Gō model is used to investigate the binding of a variety of homodimeric proteins. Analysis of the free energy as a function of the intramonomer and intermonomer contacts allows discrimination between two-state obligate dimers (binding occurs together with folding) and three-state dimers (folding precedes binding and monomers are stable). The model reproduces the experimentally known association mechanisms.

40. Levy Y, Caflisch A, Onuchic JN, Wolynes PG: **The folding and dimerization of HIV-1 protease: evidence for a stable monomer from simulations**. *J Mol Biol* 2004, **340**:67-79.

41. • Takagi F, Koga N, Takada S: **How protein thermodynamics and folding mechanisms are altered by the chaperoning cage: molecular simulations**. *Proc Natl Acad Sci USA* 2003, **100**:11367-11372.
The folding of different substrate proteins confined within a model chaperonin cage is studied with Langevin dynamics and a Gō model of the proteins. It is shown that the folding kinetics are enhanced and that the native state is stabilized due to a reduction of the conformational entropy.

42. McCammon JA, Northrup SH, Karplus M, Levy RM: **Helix-coil transitions in a simple polypeptide model**. *Biopolymers* 1984, **19**:2033-2045.

43. Wade RC, Luty BA, Demchuk E, Madura JD, Davis ME, Briggs JM, McCammon JA: **Simulation of enzyme-substrate encounter with gated active sites**. *Nat Struct Biol* 1994, **1**:65-69.

44. Bahar I, Jernigan: **Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation**. *J Mol Biol* 1997, **266**:195-214.

45. Buchete N-V, Straub JE, Thirumalai D: **Continuous anisotropic representation of coarse-grained potentials for proteins by spherical harmonics synthesis**. *J Mol Graph Model* 2004, **22**:441-450.

46. •• Brown S, Fawzi NJ, Head-Gordon T: **Coarse grained sequences for protein folding and design**. *Proc Natl Acad Sci USA* 2003, **100**:10712-10717.
This is the last in a set of papers in which a one-bead minimalist model of protein folding is presented and optimized. The energy function includes bonded, dihedral and non-bonded terms, whose parameterization is dependent on the native topology and on the sequence via a 'three-letter' amino acid code (hydrophilic, hydrophobic and neutral). Angular parameters are designed to reproduce secondary structure, whereas non-bonded interactions are van der Waals like, and are set as attractive for hydrophobic–hydrophobic interactions, repulsive for hydrophobic–hydrophilic and strongly repulsive for hydrophilic–hydrophilic.

47. Friedel M, Scheeler DJ, Shea J-E: **Effects of confinement and crowding on the thermodynamics and kinetics of folding of a minimalist β-barrel protein**. *J Chem Phys* 2003, **118**:8106-8113.

48. • Friedel M, Shea J-E: **Self-assembly of peptides into a beta-barrel motif**. *J Chem Phys* 2004, **120**:5809-5823.
By using a non-bonded interaction term with a parameter that adjusts the strength of non-native hydrophobic interactions, a continuous set of models with increasing frustration was used to study the aggregation of a β-barrel minimal protein. The frustration is shown to influence the aggregation process, stabilizing, in certain situations, amyloid-like aggregates.

49. Tozzini V, McCammon A: **The dynamics of flap opening in HIV-1 protease: a coarse grained approach**. *Protein Sci* 2004, **13(suppl 1)**:194.

50. Trylska J, Tozzini V, McCammon JA: **A coarse grained model for the ribosome: molecular dynamics simulations**. *Protein Sci* 2004, **13**:121.

51. Reith D, Pütz M, Müller-Plathe F: **Deriving effective mesoscale potentials from atomistic simulations**. *J Comput Chem* 2003, **24**:1624-1636.

52. • Kurt N, Haliloglu T, Schiffer CA: **Structure -based prediction of potential binding and non binding peptides to HIV-1 protease**. *Biophys J* 2003, **85**:853-863.
The two-bead numerical force-field described in [39••] is used to evaluate the affinity of a test set of naturally occurring substrates of HIV-1 protease and to predict the binding affinity of new substrates. It is shown that the molecular flexibility and the specificity of sidechain short-range interactions are crucial in predicting the binding affinity.

53. •• Mukherjee A, Bagchi B: **Contact pair dynamics during folding of two small proteins: chicken villin head piece and the Alzheimer protein β-amyloid**. *J Chem Phys* 2004, **120**:1602-1612.
The authors present an analytical two-bead model with an interacting potential containing simplified bonded interactions between the backbone atoms, van der Waals terms including hydrophobicity, modeled following Levitt and based on the Boltzmann inversion of probability distributions of experimental structures. A term for helix propensity is added. The model was used with Brownian dynamics and applied to the folding of chicken villin head piece and the Alzheimer protein β-amyloid; it was shown to reproduce the different stages of folding.

54. Saunders JA, Scheraga HA: ***Ab initio* structure prediction of two alpha-helical oligomers with a multiple-chain united-residue force field and global search**. *Biopolymers* 2003, **68**:300-317.

55. • Khalili M, Saunders JA, Liwo A, Oldziej S, Scheraga HA: **A united residue force-field for calcium-protein interactions**. *Protein Sci* 2004, **13**:2725-2735.
The last of a series of papers describing the development and optimization of the UNRES (united-residue) model, which was derived by

Scheraga and co-workers over the past few years. The beads are located on the Cα atom, on the centroid of the sidechain and on a 'peptide' (p) center located midway between two Cα atoms. However, Cα is only used to calculate the position of p; thus, the degrees of freedom correspond to a two-bead model. The parameterization is based on the statistical analysis of structures and on fitting the free energy functions obtained from full-atom simulations of oligopeptides. Additional terms account for correlation between the degrees of freedom. The model has been recently optimized to include ellipsoid sidechains and interaction parameters with calcium, and used in protein structure and interaction prediction.

56. Zacharias M: **Protein docking with a reduced protein model**
• **accounting for side-chain flexibility**. *Protein Sci* 2003, **12**:1271-1282.

A model with one bead for the Cα atom and up to two beads for the sidechains was used in a protein–protein docking approach. Size and hydrophilic/hydrophobic interactions are accounted for by van der Waals like amino-acid-specific terms. Sidechain conformational changes are taken into account by introducing several different rotamers of the sidechain during docking. This procedure is shown to generate more accurate complex structures than rigid docking.

57. Smith AV, Hall CK: **α-helix formation: discontinuous molecular dynamics on an intermediate-resolution protein model**. *Proteins* 2001, **44**:344-360.

58. Smith AV, Hall CK: **Assembly of a tetrameric α-helical bundle: computer simulations on an intermediate-resolution protein model**. *Proteins* 2001, **44**:376-391.

59. Urbanc B, Cruz L, Ding F, Sammond D, Khare S, Buldyrev SV, Stanley HE, Dokholyan NV: **Molecular dynamics simulation of amyloid β dimer formation**. *Biophys J* 2004, **87**:2310-2321.

60. Nguyen HD, Marchut AJ, Hall CK: **Solvent effects on the conformational transition of a polyalanine peptide**. *Protein Sci* 2004, **13**:2909-2924.

61. Nguyen HD, Hall CK: **Phase diagrams describing fibrillization by polyalanine peptides**. *Biophys J* 2004, **87**:4122-4134.

62. Nguyen HD, Hall CK: **Molecular dynamics simulations of**
• **spontaneous fibril formation by random-coil peptides**. *Proc Natl Acad Sci USA* 2004, **101**:16180-16185.

The four-bead model described in [60] is used to study the aggregation of polyalanine in fibrils. Thermodynamic and environmental conditions (dis)favorable to fibril formation are indicated. The results are in qualitative agreement with experiments.

63. Fernandez A, Colubri A: **Pathway heterogeneity in protein folding**. *Proteins* 2002, **48**:293-310.

64. Colubri A: **Prediction of protein structure by simulating coarse-grained folding pathways: a preliminary report**. *J Biomol Struct Dyn* 2004, **21**:625-637.

65. Takada S, Luthey-Schulten Z, Wolynes PG: **Folding dynamics with nonadditive forces: a simulation study of a designed helical protein and random heteropolymer**. *J Chem Phys* 1999, **110**:11616-11628.

66. Takada S: **Protein folding simulation with solvent-induced force field: folding pathway ensemble of three-helix-bundle proteins**. *Proteins* 2001, **42**:85-98.

67. Fujisuka Y, Takada S, Luthey-Schulten ZA, Wolynes PG:
• **Optimizing physical energy functions for protein folding**. *Proteins* 2004, **54**:88-103.

The latest version of the Takada model is presented. It uses constraints for bonds and bond angles, and sophisticated terms for dihedrals, van der Waals interactions and hydrogen bonds. Sequence-specific terms account for hydrophobicity, electrostatics, aromaticity and sidechain entropy loss due to secondary structure formation, allowing the accurate reproduction of Ramachandran plots. The parameters are optimized to minimize the folding frustration of a set of training structures.

68. Lee SY, Fujitsuka Y, Kim do H, Takada S: **Roles of physical interactions in determining protein-folding mechanisms: molecular simulation of protein G and alpha spectrin SH3**. *Proteins* 2004, **55**:128-138.

69. Irbäck A, Sjunesson F, Wallin S: **Three-helix-bundle protein in a Ramachandran model**. *Proc Natl Acad Sci USA* 2000, **97**:13614-13618.

70. Irbäck A, Sjunnesson F, Wallin S: **Hydrogen bonds, hydrophobicity forces and the character of the collapse transition**. *J Biol Phys* 2001, **27**:169-179.

71. Favrin G, Irbäck A, Wallin S: **Folding of a small helical protein using hydrogen bonds and hydrophobicity forces**. *Proteins* 2002, **47**:99-105.

72. Irbäck A, Sjunnesson F, Sjunnesson F, Wallin S: **Thermodynamics of β and α structure formation in proteins**. *Biophys J* 2003, **85**:1466-1473.

73. Forcellino F, Derremaux P: **Computer simulations aimed at structure prediction of supersecondary motifs in proteins**. *Proteins* 2001, **45**:159-166.

74. Wei G, Mosseau N, Derremaux P: **Complex folding pathways in a simple b-hairpin**. *Proteins* 2003, **56**:464-474.

75. Dorucker P, Jernigan RL, Bahar I: **Dynamics of large proteins through hierarchical levels of coarse-grained structures**. *J Comput Chem* 2002, **23**:119-127.

76. Schuyler AD, Chirikjan GS: **Normal mode analysis of proteins: a**
• **comparison of rigid cluster modes with C_α coarse graining**. *J Mol Graph Model* 2004, **22**:183-193.

A new method for modeling a system as a set of rigid bodies connected by an elastic network is presented that reduces considerably the number of degrees of freedom with respect to conventional elastic network approaches. It is shown that the lowest modes of the system are reproduced with a good accuracy.

77. Li G, Cui Q: **A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca^{2+}-ATPase**. *Biophys J* 2002, **83**:2457-2474.

78. Flock D, Helms V: **A Brownian dynamics study: the effect of a**
• **membrane environment on an electron transfer system**. *Biophys J* 2004, **87**:65-74.

A Brownian dynamics study of the association of cytochrome *c* with cytochrome *c* oxidase is presented, both *in vacuo* and as an all-atom model of the membrane. The membrane is shown to dramatically influence the association rates.

79. Elcock AH: **Atomic-level observation of macromolecular**
• **crowding effects: escape of a protein from the GroEL cage**. *Proc Natl Acad Sci USA* 2003, **100**:2340-2344.

A Brownian dynamics simulation of the interaction of a protein with the chaperonin GroEL cage in the presence of crowding agents is presented. A crowding agent is shown to be more effective if it forms favorable interactions with itself.

80. Jiang S, Tovchicrechko A, Vakser IA: **The role of geometric complementarity in secondary structure packing: a systematic docking study**. *Protein Sci* 2003, **12**:1646-1651.

81. Espinoza-Fonseca LM: **Base docking model of the homomeric α7 nicotinic receptor-β-amyloid$_{1-42}$ complex**. *Biochem Biophys Res Comm* 2004, **320**:587-591.

82. Kinjo AR, Takada S: **Competition between protein folding and aggregation with molecular chaperones in crowded solutions: Insight from mesoscopic simulations**. *Biophys J* 2003, **85**:3521-3531.