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].

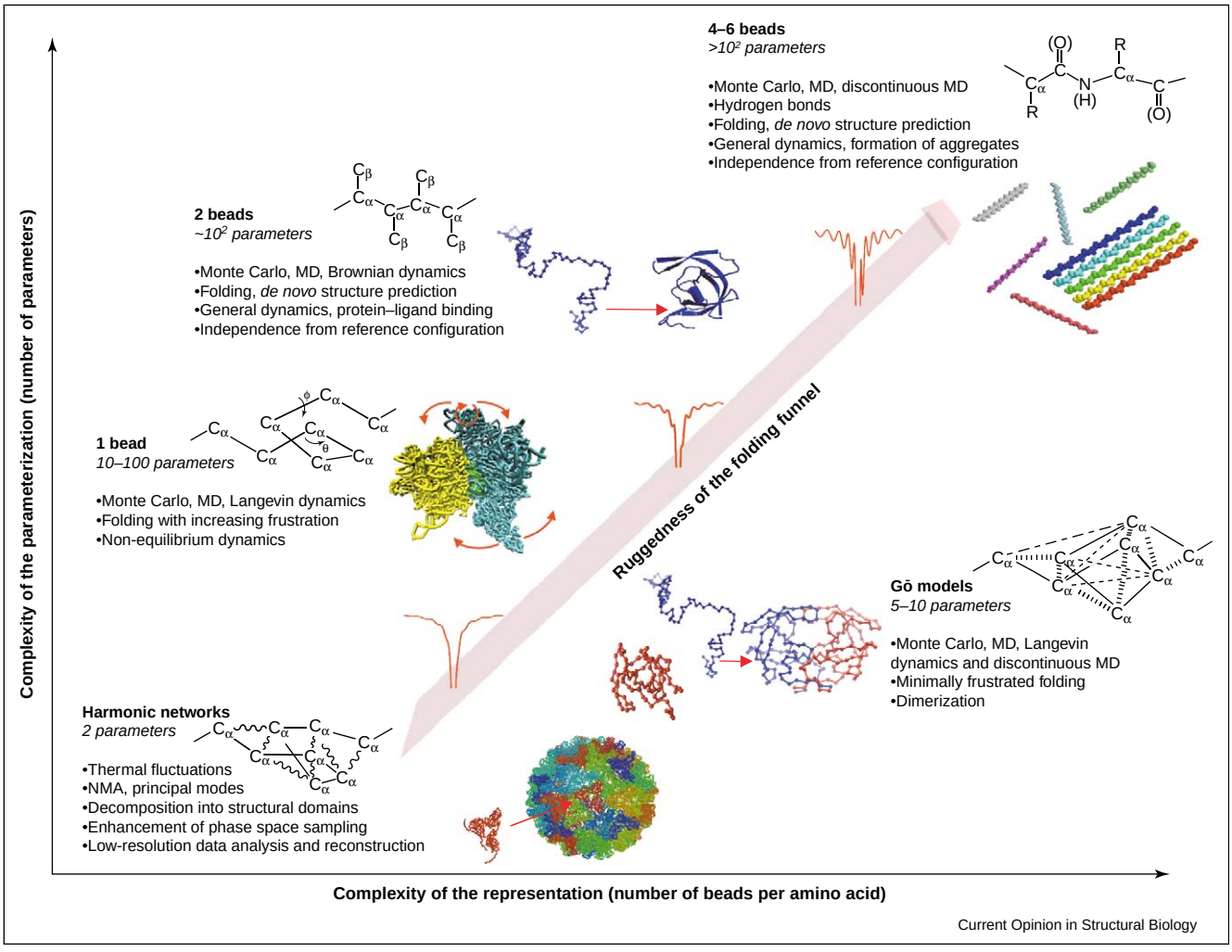
Ten or hundred nanosecond simulations of proteins in an explicit aqueous environment are currently feasible. However, most of the relevant dynamics and interactions within cells (typically, protein–protein docking, rearrangement upon ligand binding or after biochemical reactions, folding) occur on the 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