*Review*

# Modeling of Protein Structural Flexibility and Large-Scale Dynamics: Coarse-grained Simulations and Elastic Network Models

**Sebastian Kmiecik[1], Maksim Kouza[1,2], Aleksandra E. Badaczewska-Dawid[1], Andrzej Kloczkowski[2]\*, Andrzej Kolinski[1]\***

[1]  Faculty of Chemistry, Biological and Chemical Research Center, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland; sebastian.kmiecik@gmail.com

[2]  Nationwide Children's Hospital, Columbus, Ohio; Andrzej.Kloczkowski@nationwidechildrens.org

\*  Correspondence: kolinski@chem.uw.edu.pl; Tel.: +48-22-55-26-365; Andrzej.Kloczkowski@nationwidechildrens.org

**Abstract:** Fluctuations of protein three-dimensional structures and large-scale conformational transitions are crucial for the biological function of proteins and their complexes. Experimental studies of such phenomena remain very challenging and therefore molecular modeling can be a good alternative or a valuable supporting tool for the investigation of large molecular systems and long-time events. In this mini-review, we present two alternative approaches to the coarse-grained (CG) modeling of dynamic properties of protein systems. We discuss two CG representations of polypeptide chains used for Monte Carlo dynamics simulations of protein local dynamics and conformational transitions and, on other hand, highly simplified structure-based Elastic Network Models of protein flexibility. In contrast to classical Molecular Dynamics the modeling strategies discussed here allow quite accurate modeling of much larger systems and longer time dynamic phenomena. We briefly describe the main features of these models and outline some of their applications, including modeling of near-native structure fluctuations, sampling of large regions of the protein conformational space, or possible support for the structure prediction of large proteins and their complexes.
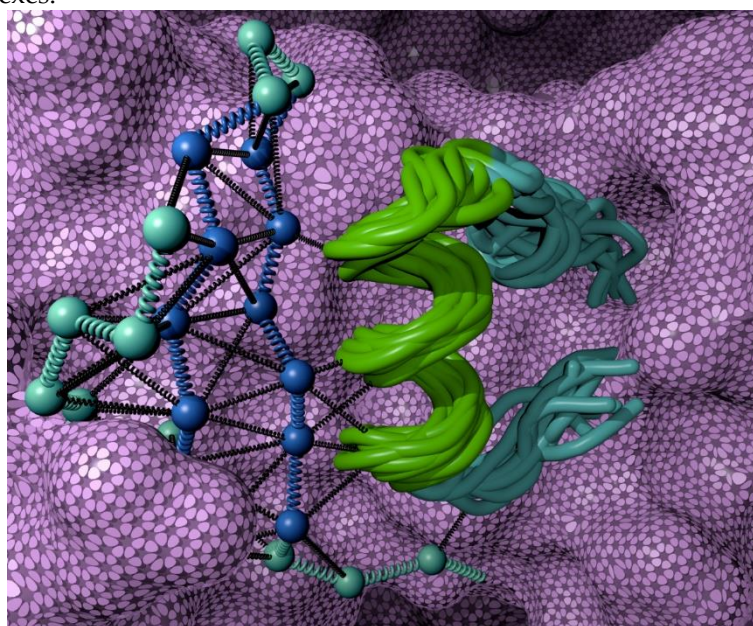
Graphical Abstract

**Keywords:** protein dynamics; coarse-grained simulation; Monte-Carlo dynamics; structural flexibility; large-scale dynamics; elastic network model;

## 1. Introduction

The biological activity of proteins involves adopting a specific conformation, their local fluctuations and, in many cases, structural transitions between different conformations. Structural flexibility can range from small side-chain fluctuations to large rearrangements of the entire protein backbone or its fragments, for example disordered regions [1–3]. Most of the knowledge about protein structures comes from X-ray crystallography that in most cases offers only a static view of well-ordered protein regions. Determination of protein structure dynamics remains difficult using either experiment or computer simulations, despite significant effort in the field of drug design [4–6]. All-atom Molecular Dynamics (MD), the classical simulation technique, enables relatively inexpensive studies of small protein fluctuations (e.g. side-chain or very local backbone moves) and rather small protein systems [7]. In practice, many of the most interesting protein systems (especially from the medical point of view) are either too large to be effectively simulated using MD or require very large supercomputer resources. Some acceleration of MD dynamics can be achieved using so-called Go models (or structure-based models) that provide a quite valuable but limited insight into large-scale protein dynamics [8–10]. Significant computational speed-up is possible by simplifying the protein description to a less complex level than that of MD. This is offered by coarse-grained (CG) protein models [8,11] and Elastic Network Models (ENM) used in conjunction with normal mode analyses (NMA) [11–14]. Both approaches can be used as key components of multiscale modeling methods [15,16] merging computational tools of various resolutions, from the low-resolution to atomistic level [17]).

In this review, we briefly discuss two representative CG simulation models of protein dynamics (**Section 2.2 and 2.3**) and various ENM-based modeling techniques (**Section 2.4**). Next, we review their applications to modeling protein local dynamics: structural flexibility of folded proteins (**Section 3.1**) and large-scale structural transitions (**Section 3.2**). The discussed approaches (see **Figure 1**) differ in their levels of accuracy of protein structure representation, employed models of the interaction scheme and their sampling of system dynamics. Interestingly, for many problems all these methods lead to surprisingly similar results, with one difference: the ENM approach is the simplest and therefore can be effectively used for studies of very large systems, obviously with all the limitations of structure-based models. Finally, we discuss the advantages and limitations of the presented approaches.
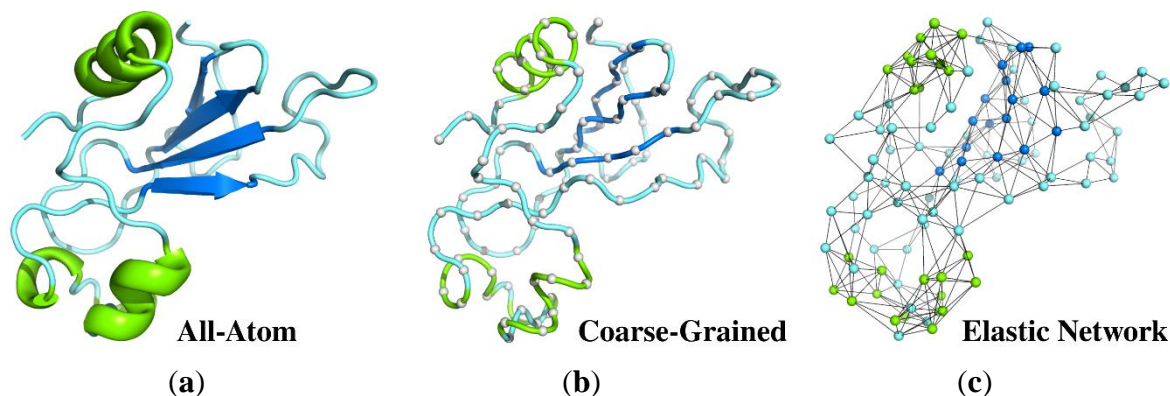


**Figure 1.** Different representations of protein structure (PDB code: 1a2p) used as a reference state for structure dynamics studies. (a) all-atom structure used in MD simulations showed as a ribbon diagram; (b) deeply coarse-grained model (gray balls represent a single center of interaction per residue connected by a tube) for Monte Carlo (MC) dynamics simulations; (c) coarse-grained elastic network model with spring-type constraints and nodes in the positions of Cα atoms. The colors refer to the secondary structure assignment of protein fragments: helices are shown in green and β-strands are in dark blue.

## 2. Coarse-Grained modeling of proteins

*2.1 From all-atom to coarse-grained modeling*

Molecular modeling of protein dynamics plays important roles in many branches of biophysics, molecular biology and related sciences. The most popular strategies for modeling proteins and other biomolecular systems are classical Molecular Dynamics (MD) simulation methods. MD has however several limitations [18]. Important biomacromolecular systems are very large (consisting of many atoms) and the timescale of interesting processes is usually beyond the MD simulation range accessible to contemporary computers. Using the most powerful and dedicated computing devices it is now possible to simulate structure assembly dynamics of small, fast-folding, proteins [19]. An alternative is to use coarse-grained (CG) modeling tools [8,20]. The CG modeling of proteins, initiated half a century ago by Levitt (and others) [15,16], has become recently a rapidly growing branch of molecular modeling. The general idea is to reduce the number of explicitly treated degrees of freedom in the modeled systems (see **Figure 2**) and simplify energy calculations to speed up computations and thereby allow simulations of larger systems and/or longer processes.
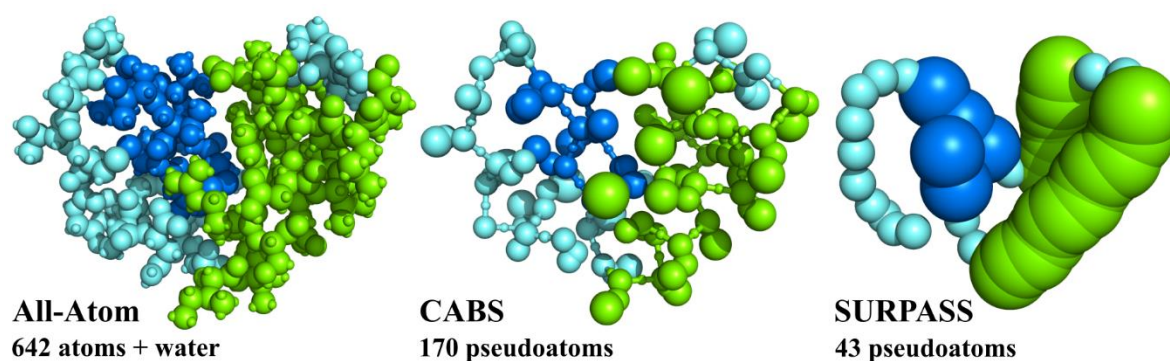


**All-Atom**
**642 atoms + water**

**CABS**
**170 pseudoatoms**

**SURPASS**
**43 pseudoatoms**

**Figure 2.** All-atom (AA) and two CG representations of protein structures. For a small protein (PDB code: 1CRN), consisting of 46 residues the AA representation needs explicit treatment of 642 atoms (+ a large number of solvent atoms when an explicit solvent model is analyzed). A lower resolution CG CABS model requires explicit treatment of 170 pseudoatoms and the deeply CG SURPASS model reduces the number of simulated pseudoatoms (or rather pseudoresidues in this case) to just 43, reducing simulation cost by orders of magnitude (see **Figure 3**). In both cases of CG representation, the atomistic structure can be approximately reconstructed thanks to the properly defined geometry and interaction schemes of these models.

Many CG approaches have been proposed, using different levels of reduced representations of biomacromolecules, different force-field schemes and different strategies for sampling the conformational space [8,20–23]. We discuss the models that in spite of their reduced numbers of explicitly treated atoms, or "pseudoatoms", allow realistic reconstruction of atomic details, and thereby open a possibility for the multiscale, integrative modeling of protein systems. In addition, instead of the "physics-based" force-field we opt for statistical, "knowledge-based" models of interactions. Both approaches have their pros and cons, but since proteins (and generally biomacromolecules) can be characterized by many specific features and owing to the rapidly growing databases of structural biology, statistical force fields can be a good choice. Consequently, instead of using various ersatz MD schemes we prefer properly designed Monte Carlo dynamics sampling. Such approaches prove to be very productive in the computational prediction of protein structures, modeling of protein dynamics and molecular docking mechanisms [8].

Below, we discuss in more detail two representative models of various resolutions, based on the above outlined strategy of CG modeling: a relatively high-resolution CABS [24] CG model (assuming C-Alpha, C-Beta and Side chains representation of polypeptide chains) and a low-resolution SURPASS [25] CG model (Single United Residue per Pre-averaged Secondary Structure fragment). As illustrated in **Figure 3** CG simplifications of the models can significantly increase (by orders of magnitude) both the size and time scale of tractable systems. Obviously, the deeper the coarse-graining of the modeled structures the larger systems can be effectively simulated. On the other hand, coarse-graining of the studied structures always results in neglecting some subtle features. How

costly the CG simplification is depends not only on the assumed level of coarse-graining, but also on the force field model and sampling schemes used in simulations. CABS, SURPASS or other related CG models can address a broad range of structure dynamics problems, including protein folding-unfolding or associations with other biomacromolecules, obviously with the above mentioned computational limitations.
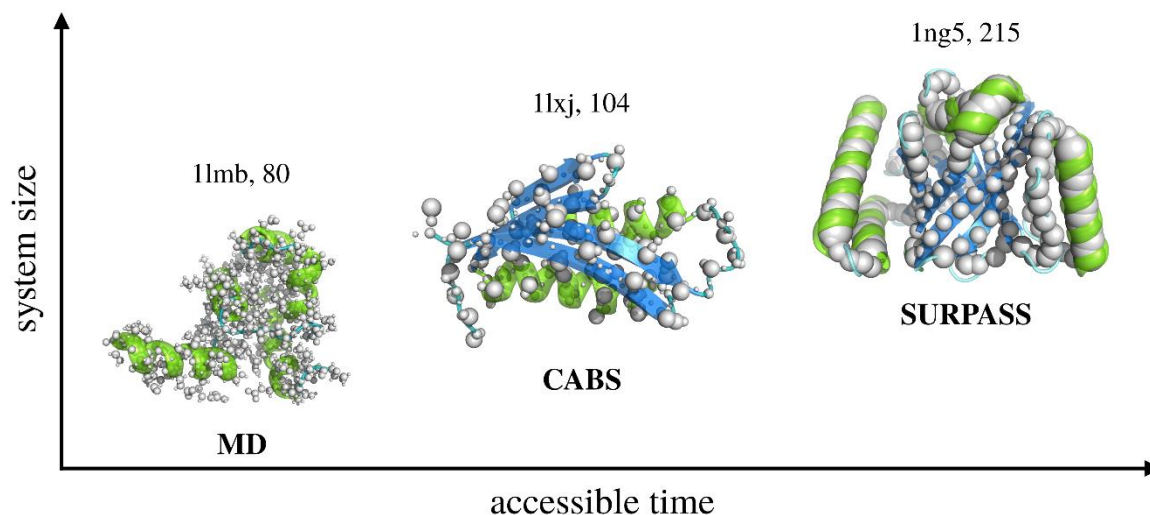


**Figure 3.** Increasing range of applicability of dynamics modeling, from MD to deeply coarse-grained simulations. CABS (assuming C-Alpha, C-Beta and Side chains representation of polypeptide chains) is a moderate resolution model of protein structure and dynamics. SURPASS (Single United Residue per Pre-averaged Secondary Structure fragment) is a deeply CG model, where single united residues represent entire amino acids. Both CG structure models use Monte Carlo Dynamics sampling for fast simulations of long-time protein dynamics.

*2.2 Coarse-grained CABS model*

CABS [24] is a coarse-grained model for simulations of protein structures and dynamics (see **Figure 4** and recent review on its design and applications [8]). CG representation of amino acid residues is reduced to up to four pseudoatoms or united atoms. These pseudoatoms of CABS models represent main-chain alpha carbons (CA), beta carbons (B) and the center of mass of side chains (S). The fourth pseudoatom is placed in the center of the C$\alpha$-C$\alpha$ pseudobond and is used for the CG definition of the main-chain hydrogen bonds.
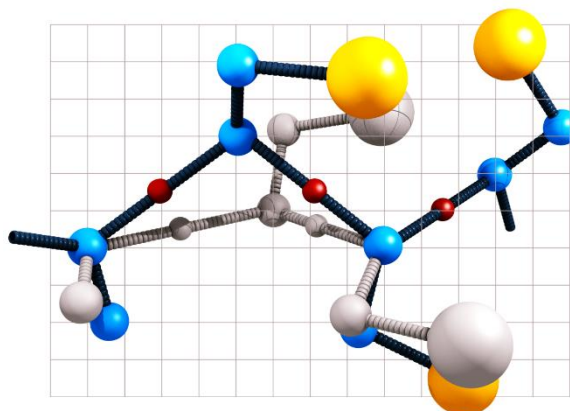


**Figure 4.** CABS representation of a small protein chain fragment. Positions of C$\alpha$ pseudoatoms are restricted to grid nodes of the underlying simple cubic lattice, with spacing of 0.61 Å. Allowing small fluctuation of the expected C$\alpha$-C$\alpha$ distance enables hundreds of possible orientations of these pseudobonds. Side chain pseudoatoms (C$\beta$ in blue and an additional center of the side chain, where applicable, in yellow) are not restricted to the lattice and their positions are defined by the geometry of the main chain and statistical properties of specific amino acids. A move of a single C$\alpha$ leads to

specific displacements of three side chains (new positions of the side chains are shown in gray). Thanks to the high-coordination lattice representation of the main chain geometry all possible local moves are stored in large precalculated data tables.

The force field of CABS is knowledge-based, derived from the statistical analysis of a representative set of known protein structures. The force field contains local conformational biases dependent on amino acid identities and excludes volumes of pseudoatoms, a model of strongly directional main-chain hydrogen bond attractions and contact potentials for side chains [26]. The latter have a specific feature, making the CABS force field very sensitive to the amino acid content in the studied proteins. Namely, this potential (for single domain proteins) strongly depends on the mutual orientations of the contacting side chains [24,26]. For instance, when modeling a single-domain globular protein, oppositely charged side chains are strongly attractive if their contacts are parallel, and anti-parallel contacts are repulsive. Indeed, when looking into globular proteins the oppositely charged residues are usually contacting on the protein surface, and consequently the corresponding side chains are almost parallel. On the contrary, charged residue contacts in the central regions of a globule, where parallel and anti-parallel orientations of the contacting side chains are possible, are extremely rare, and therefore the anti-parallel contacts of charged residues are treated as repulsive (unlike). This orientation-dependent definition of side chain contact interactions is a powerful way to introduce a crude account for implicit solvent effects in CG models.

The sampling scheme of the CABS model (constant temperature simulations, cooling/heating simulations, replica exchange simulations, and other sampling strategies) is based on the MC dynamics method. The dynamics of the entire protein structure is modeled as a very long random sequence of small local moves, including single amino acid moves to a small fragment (composed of two, three, and rarely a slightly larger number of amino acids) local moves. MC-dynamics provides very reasonable dynamics trajectories for long-time processes, although the time-unit of such dynamics is not a priori defined. MC dynamics can be properly scaled by fitting (using mean-square displacements, or related parameters, as global motion criteria) fragments of the MC trajectories to corresponding MD trajectories and/or to appropriate experimental data. CABS MC dynamics simulations are very fast, not only due to the reduced representation of protein structures but also additionally thanks to the high-coordination lattice representation of the model structures. The underlying lattice enables use of precalculated data tables of local geometry and interactions. This makes computing new, randomly selected structures and interactions extremely fast. On the other hand the very high coordination number of interacting pseudoatoms (due to the high resolution of the underlying lattice) does not impose any noticeable directional biases, so painful for lower resolution lattice models of protein structures [27].

The CABS model has been successfully used in numerous studies, including folding of globular proteins [28–33], dynamics of protein-peptide binding [34,35], flexibility of globular proteins [36–40], chaperonin mode of action [41], protein insertion into the biological membrane [42], structure prediction of proteins [43–45] and protein-peptide complexes [46–50]. The CABS has been implemented as a simulation component of several multiscale modeling tools, publicly available as web servers and standalone packages (see http://biocomp.chem.uw.edu.pl and http://lcbio.pl). The CABS-based tools are targeted onto specific simulation tasks, including simulations of near-native (or in the vicinity of another provided structure) protein flexibility [37–40], modeling of protein folding (including de novo or template based structure prediction) [43,51], and unrestrained molecular docking of flexible peptides to flexible protein receptors [46–48,50]. Very recently, the CABS model has been made available as the CABS-flex standalone package [40], which is a flexible simulation environment for modeling protein dynamics including small-scale fluctuations (based on the input structure) or large-scale moves. The CABS-flex package offers easy modification of the CABS simulation parameters (and other modeling stages like scoring, structural clustering or all-atom reconstruction) and using various input data (in the form of user-defined distance restraints). Therefore, the CABS-flex package can be easily incorporated into other multiscale methodologies of structural biology.

*2.3 Coarse-grained SURPASS model*

By implementing specific structural regularities characteristic for protein globules, it is possible to design much simpler than CABS (and other models of similar resolution), deeply coarse-grained representations of protein chains that regardless of their simplicity maintain some of the most important features of real systems. SURPASS (Single United Residue per Pre-Averaged Secondary Structure element) is an example of such approach. SURPASS [25,52] is based on the fundamental idea that three-dimensional structures of globular proteins (and also the majority of membrane proteins) are to a large extent defined by the local ordering of polypeptide chains, defined as the secondary structure. A united residue of the SURPASS chain represents a single amino acid unit, located in the center of mass of four consecutive alpha carbons. The choice of the four-residue averaging of united residue positions is not incidental. For regular elements of secondary structure, such as helices or β-strands (but also, to a lesser extent, other secondary structure fragments), such averaging leads to a very simple, almost linear, representation of polypeptide chains (**Figure 5**). On this level of protein drawing it is possible to derive several simple statistical rules controlling three-dimensional structures [25,52]. For instance, helical fragments can be treated as linear chains composed of partially overlapping thick spheres while β-strand fragments are thinner, with the asymmetric shape of their excluded volume profiles, representing differences in distances between neighboring strands belonging to the same β-sheets and between strands of different sheets. Characteristic distances between helices and β-sheets, and other structural regularities, can be also easily incorporated in the statistical force field of the model. In its simplest version the SURPASS force field is based on secondary structure assignment (or predictions) as the only sequence-defined variable defining interaction patterns. The solvent effects, similarly as it was assumed in the CABS model, are treated in an implicit fashion. Reconstruction of more detailed structure representations from SURPASS models is not trivial, although, as tested on a large set of globular structures, it is possible and does not create drastic inaccuracy of the resulting atomistic models. In this context, the structural accuracy (resolution) of the model is in a range of 2-3 Angstroms. This is an acceptable range for all-atom structure refinement protocols [53]. The dynamics of SURPASS proteins is simulated by Monte Carlo dynamics schemes, within single trace or replica-exchange (REMC, Replica Exchange Monte Carlo dynamics) runs. Simplicity of the model allows very fast simulations of large systems. As shown for representative sets of single-domain globular proteins, even for relatively large proteins, a single REMC run provides trajectories covering the entire conformational space of the model, with multiple visits of native-like structure regions. Due to its computational speed and the fact that its low-resolution results can be used as the starting frames for higher resolution simulations SURPASS modeling opens up a possibility for efficient multi-scale modeling of the long-time dynamics of large proteins and protein systems.
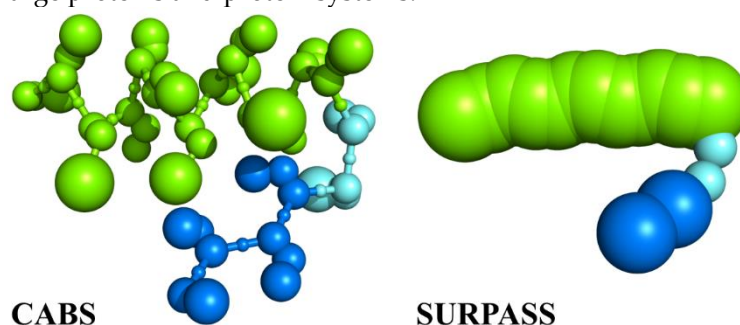


**Figure 5.** Comparison of CABS and SURPASS CG representations of proteins. The picture presents a short structural element containing helical (green) and β-strand (blue) fragments, connected by a loop.

*2.4 Elastic Network Models*

In the last decades, a large number of studies have shown the usefulness of Normal Mode Analysis (NMA) and Principal Component Analysis (PCA) for the prediction and analysis of protein cooperative motions [13,14,54–63]. Many of these studies were conducted using computationally

simple Elastic Network Models (ENM) which use uniform harmonic potentials for interacting atom (or residue) pairs instead of more complicated potentials.

In the early 1980s the NMA method was adapted by Go [64] and Karplus [65] and others for studies of thermal fluctuations of proteins around their native structures. PCA and NMA of proteins and other biomolecules applied to their equilibrium dynamics around native states has shown that low-frequency normal modes correspond to the cooperative motions of large parts of the protein structure that are essential for protein function, while high-frequency modes corresponding to the isolated, non-cooperative vibrations of small parts of the structure are functionally meaningless. The long-range character of low-frequency modes enables allosteric interaction of various parts of proteins that are spatially distant from one another. The old paradigm that the protein sequence encodes protein structure and the structure determines protein function has been in the last 20 years modified by a new paradigm assuming that sequence encodes the structure, the structure determines protein dynamics and the dynamics encodes protein function. It has been also shown that for globular proteins their dynamics is strongly determined by protein shape [66]. Besides describing the near-native motion of globular proteins PCA can also be applied to structure snapshots from Molecular Dynamics or Monte Carlo simulations, to structures from NMR ensembles [67], or to the analysis of large sets (clusters) of protein or RNA structures determined by X-ray crystallography, such as for example a set of over 350 HIV-1 protease structures [68], see **Figure 6**.
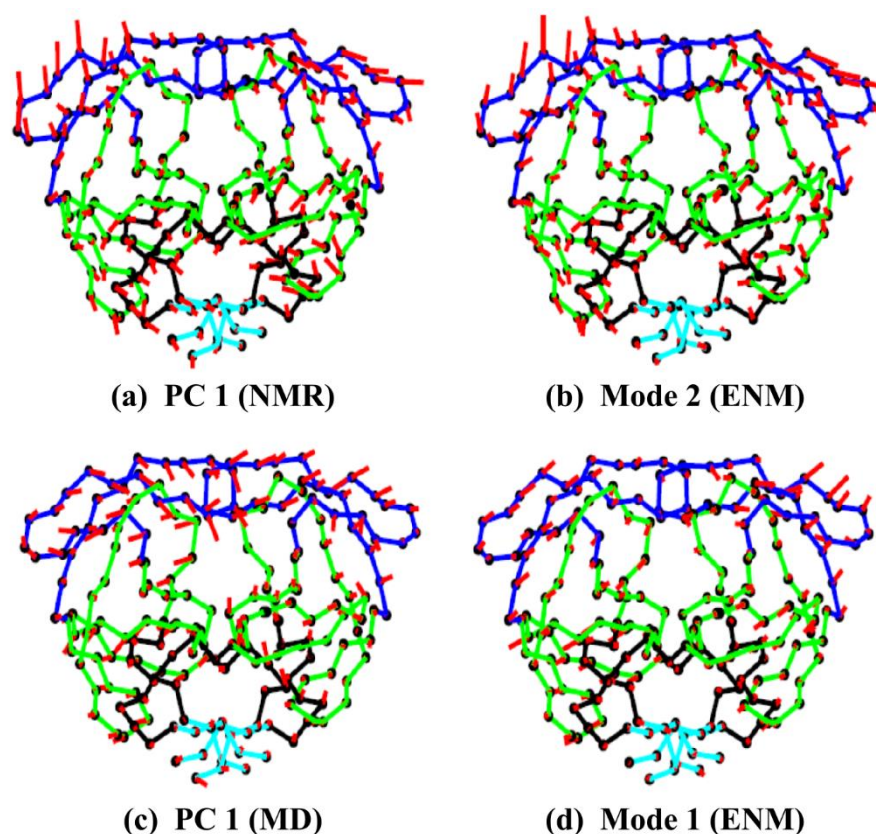


**(a) PC 1 (NMR)**     **(b) Mode 2 (ENM)**

**(c) PC 1 (MD)**      **(d) Mode 1 (ENM)**

**Figure 6.** Visualization of motions of the dominant PCs (a,c) and the most similar corresponding modes predicted by ENM (b,d). In the NMR dataset, the overlap between (a) PC 1 and (b) Mode 2 is 0.91. In the MD dataset, the overlap between (c) PC 1 and (d) Mode 1 is 0.74. Blue: flap domain; green: core domain; cyan: terminal domain; yellow: other residues. The motions of PCs and modes are shown as red sticks to indicate the directions of motion of the attached atom. Stick lengths represent the relative amplitudes of fluctuations of the corresponding residue. This figure is adopted from ref. [68].

ENMs have become extremely successful in biology after their reformulation as the Gaussian Network Model (GNM) of proteins [55,69]. GNM assumes that each residue of a protein is represented by a single node with its coordinates usually given by coordinates of the C$\alpha$ atom, although different definitions of nodes (e.g. the center of mass of a side chain) are also frequently used. Nodes that are separated by less than a cutoff distance (usually assumed to be around 7 Å) are connected by identical harmonic springs. Nodes that are further than a cutoff distance are not connected by springs. The model is based on the assumption of Tirion [70,71] that bonded and nonbonded contact interactions in biological structures can be described by a single uniform spring constant parameter. Although such approximation seems at first rather unrealistic the results of normal mode analysis are almost undistinguishable from the results derived using sophisticated energy function potentials, such as the CHARMM force-field [72]. The excellent performance of spring-like potentials for normal mode analysis is due to the fact that any complicated potential function is very well approximated around its minimum by a harmonic function. The basic difference between elastomeric polymer networks and proteins is that a protein is a densely packed collapsed heteropolymer with unique packing of amino acids and many nonbonded interactions, whereas for polymers these nonbonded interactions are not important. The fluctuations of residues in proteins computed with GNM agree surprisingly well with experimental temperature factors in the Protein Data Bank (PDB). Excellent agreement is typically achieved by this approach when crystallographic B-factors [69] or H/D exchange data [73] are plotted against mean-square fluctuations computed with GNM as a function of the residue index. Additionally, the covariance of fluctuations of two residues *i* and *j* that provides information about correlations of equilibrium fluctuations of residues in protein structure can be simply computed with elastic network models. Pairs of residues that have large values of such correlations may communicate with each other because of functional reasons, for example by participating in common allosteric communication paths. In the GNM it is assumed that fluctuations of residues around their mean positions are spherically symmetric.

The Anisotropic Network Model (ANM) assumes that fluctuations of residue positions around their starting points are, in contrast to the spherical approximation of GNM, anisotropic and are represented by ellipsoids. The model can be used to calculate normal modes from a single structure [74,75]. The cutoff distance for defining the springs is usually assumed to be 13 Å that is significantly larger than the cutoff distance of around 7 Å that is usually used in GNM. For most of the proteins in PDB crystallographic B-factors are scalars and GNM predicts very well the magnitude of fluctuations of residues. Protein structures solved with high or ultra-high resolution (less than 0.8 Å) usually have anisotropic B-factor tensors deposited in PDB, and ANM that predicts the tensor of anisotropic B-factors is a very useful tool for modeling proteins with very high resolution reference structures.

Most cooperative (lowest frequency) modes of motion are almost insensitive to the details of the model and parameters, and can be satisfactorily captured by adopting low-resolution ENMs. The idea has been applied by Doruker and Jernigan [76] to low-resolution EN representation, where each node stands for a group of m residues condensed into a single interaction site. Calculations performed for different levels of coarse-graining (m = 1, 2, 10, 40) have shown that the slowest modes of motion are preserved regardless of the level of coarse-graining. Mixed Coarse-Graining has also been introduced for EN models, where a protein's native conformation is represented by different regions having high and low resolution [77]. The aim here was to capture the dynamics of the interesting parts in structures at higher resolution and retain the remainder of the structure at lower resolution, keeping the total number of modes sufficiently low for computational tractability.

Except as mentioned above, ENMs have been successfully used in many combinations with different input information and methods, such as for example data from electron microscopy [14], atomistic MD simulations [17,78], Brownian simulations [79], structure-based models [80,81] and many other combinations that have been thoroughly reviewed [13,14,54–63].

## 3. Applications of coarse-grained modeling: ENM and CG Monte Carlo simulations

*3.1 Modeling of the structural flexibility of folded proteins*

When the range of protein structural dynamics studies can be limited to a roughly defined vicinity of a specific folded structure (determined by experiment), ENM, CG models or their modifications can be very effective in predicting protein flexibility [8,36,82–84].

In the last decades, ENM-based modeling has been definitely the most common approach to modeling fluctuations near the input folded structures. There are many ENM-based methods available, some of them as easy to use web servers [68,85–96]. CG methods using more sophisticated potentials that are not derived from the input structure are much less common in predictions of protein fluctuations [8,36,37,39,97,98]. These include the CABS-flex method using the CABS CG model for the MC simulations of protein flexibility [37,39] near a predefined reference structure that can be taken from experiment or predicted theoretically. The CABS-flex method has been shown to provide a consistent view of protein flexibility with short timescale MD simulations [36] and NMR ensembles [38]. While qualitatively similar to ENM predictions, CABS-flex generated better correlated pictures of protein fluctuations (on average for the studied set) in comparison to all-atom MD simulations [36]. The CABS-flex method is currently available as a web server (the improved version 2.0 has been published recently [39]) and a standalone package [40].

Below, in **Figure 7**, we present the comparison of fluctuation profiles (RMSF, root mean square fluctuation) for protein 1hpw (using the first model from the 1hpw PDB ensemble as the input) obtained from:

- **NMR ensemble:** data calculated using 10 models deposited in the PDB code: 1hpw
- **MD all-atom simulation**: data obtained using a 10 nanosecond trajectory taken from the MoDEL database of MD trajectories [38]; RMSF was calculated for the entire trajectory consisting of 10,000 models
- **CG simulation using the CABS model**: data obtained using the CABS-flex 2.0 web server [39]; results calculated using the default server settings; real time calculation: 22 minutes; RMSF was calculated for the set of 10 representative models from the simulation trajectory
- **CG simulation using the SURPASS model**: data obtained with the following SURPASS [25] settings: isothermal MC simulation in low reduced temperature (T=0.2), 10000 MC steps, 1 and 3-bead motions; real time calculation: 15 seconds; RMSF was calculated for the entire trajectory of 100 models
- **ENM modeling**: data obtained using the DynOmics web server [86,96] that integrates two ENM methods: the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM); calculated using the default server settings; real time calculation: <1 minute; RMSF was calculated for the set of 20 models based on the 10 slowest modes (2 models per mode for extreme positions during movement); the amplitude of motion along each mode was chosen so that the RMSD between the models was less than 2Å; all models were generated using the 'Molecular Motions - Animation' option available on the results page of the DynOmics server.
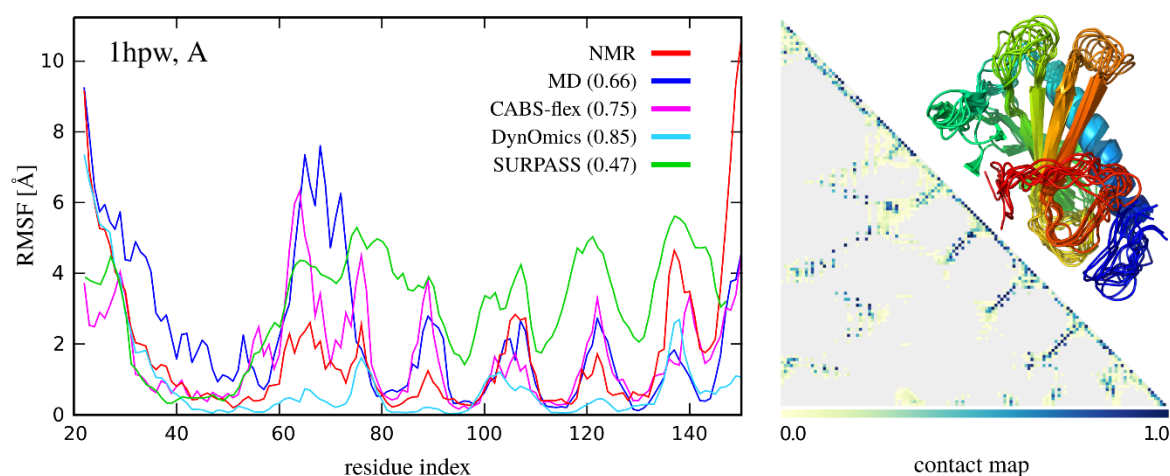
**Figure 7.** Comparison of residue fluctuation (RMSF, root mean square fluctuation) profiles for an example protein (PDB code: 1hpw, chain A). Chart on the left presents RMSF values (in Å) derived from NMR ensembles (red) and simulation trajectories: all-atom MD (blue), CABS-flex (pink), DynOmics (light blue) and SURPASS (green). In the legend section the values in brackets show Spearman's correlation coefficients for residue fluctuations between NMR and each method. The chart on the right presents a contact map (frequency of contacts is showed in different colors, see the color scale) and example fluctuation profiles visualized on the structures given by the CABS-flex 2.0 server [39].

**Figure 7** shows that ENM and CABS MC simulations provide dynamic profiles closest to experimental characteristics (such experiments have been made for many protein systems). The deeply CG SURPASS method is less accurate, probably mostly due to its extremely simplified interaction model, although the dynamic profiles are quite realistic. Obviously the CG methods are computationally much less expensive than atomistic simulations by MD.

The ENM-based exploration of protein conformational space near the given folded structure can also be practically used in the refinement of models derived from protein structure prediction methods [99–105]. For example, Feig proposed and successfully tested a modeling scheme based on NMA and MD simulations with constraints and an efficient sampling scheme [106,107]. The procedure used by Feig's group was an iterative one. First, they performed molecular mechanics energy minimization and then employed NMA computations around the local energy minimum. Subsequently, they generated and evaluated an ensemble of possible new conformations along the lowest frequency normal modes. The normal modes provide guidance toward the effective refining of protein structures. A similar approach was used by Tama and Brooks to refine cryo-Electron Microscopy structures [108]. Similarly to ENM and related models also CG Monte Carlo simulations can be easily combined with MD simulations [109–112].

*3.2 Modeling large-scale structural transitions*

ENMs have been shown to be useful to study large-scale protein conformational transitions in cases when starting and ending conformational states of the transition are available [14,54,97,113–119]. Protein conformational transitions frequently occur in biology, and usually relate to protein function. The most frequently occurring large-scale conformational transition involves two protein conformations: the "open" and the "closed" form, that are the starting and the ending states of the transition. Assuming that the physical mechanism of normal modes is used in nature to drive large-scale conformational transition, ENM can be used to model such transitions using both the "open" and the "closed" structures [113]. Since such large-scale conformational transitions are frequently slow and difficult for detailed modeling with MD simulations the above mentioned ENM approach represents a very useful and computationally inexpensive alternative method to study them. Both approaches usually give similar results [78]. ENM-based methods have been successfully applied to structural transitions of very large molecular structures [14], such as a ribosome [120–122].

CG protein simulation models, such as CABS or SURPASS, offer an alternative approach to predicting large-scale transitions that do not require starting and ending conformational states. The CABS model has been used for simulations of folding mechanisms from the fully denatured to the folded state [28–31,41,44]. This way, with some successes, the CABS model can be used for de novo structure prediction. Obviously, with the growing structural databases de novo modeling of protein structures becomes rather a theoretical challenge with decreasing practical importance. For the majority of new proteins it is now possible to identify close or distant homologs with known, experimentally determined structures. CG models, including CABS, have become powerful tools for comparative modeling, especially for the cases of distant homology, where differences between the template and target structures are significant. The CABS-fold server can be easily used for de novo structure prediction (purely de novo, without any knowledge about the plausible starting structure, is reasonable only for rather small proteins) and for template/templates based comparative modeling. In a different context, using weak or limited structural restraints, and/or known starting structures, models such as CABS, UNRES, or related models [45,123–127] can be powerful tools for simulations

of large-scale structural rearrangements. Folding/unfolding pathways, chaperone effects, large-scale structural rearrangements during molecular docking and other related processes can be efficiently simulated. The medium-resolution CG models, while opening the time-scale about three to four orders of magnitude wider than that accessible for classical MD simulations, still remain computationally too expensive for modeling multiple large-scale structural transitions of large proteins. The SURPASS model opens up such possibility. Example results of SURPASS simulations of folding-unfolding processes in a globular protein are illustrated in **Figure 8**. The entire conformational space of large proteins can be effectively sampled. This allows very detailed studies of the very long-time dynamics of proteins and protein complexes, although biased by the necessary simplifications of structure representations and the model (knowledge-based, statistical) force field. Although maintaining fundamental features of protein structure, such crude models can be combined with more accurate modeling tools, for example providing representative starting conformations for de novo simulations with higher resolution tools, or a plethora of random, partially folded and more compact folded/misfolded structures for derivation of realistic statistical potentials for other CG models.
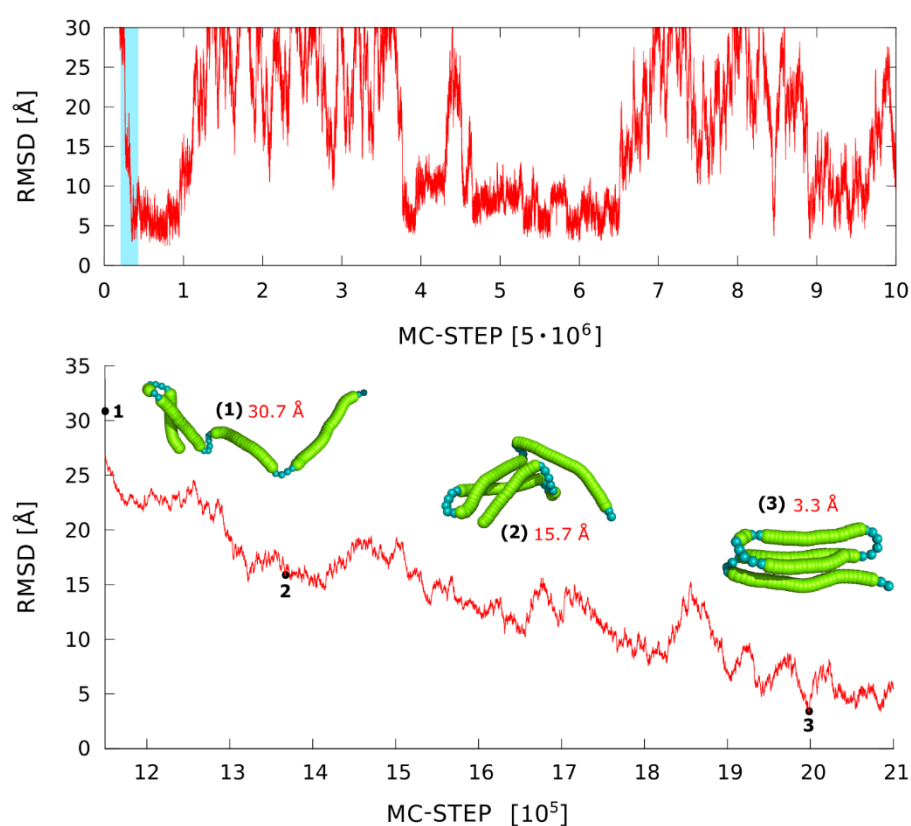


**Figure 8.** MC dynamics trace for a single replica for alpha protein (PDB code: 1k40) in SURPASS representation. The lowest panel shows the enlargement of the coil-to-globule transition taken from a small fragment of the RMSD trajectory marked in blue in the upper plot. The selected snapshots of protein structures illustrate the observed mechanism of fold assembly.

## 4. Concluding remarks and perspectives

Computational exploration of protein conformational transitions has already been quite successful in many tasks of structural biology, including rational drug design [4,5,128,129] but it is often very challenging. The major difficulties arise from inexact sets of parameters (e.g. force-field parameters [4,130]) and inability to access long simulation timescales of many important biological processes. In this review, we present two approaches to extend accessible timescales and sizes of modeled systems: ENMs and CG models using MC dynamics. These can be combined with atomistic level MD providing efficient strategies of multiscale simulations of proteins and protein complexes.

While the ENM approach is well suited for speeding up collective motions, CG MC can be applied to a broader range of functional moves.

ENM (or NMA based methods) are well suited only for certain problems [55]. They can describe well protein dynamics around the known experimentally solved native state, or conformational transitions between two states, such as "open" – "closed" or "bound" – "unbound", for which both structures are known. The applicability of ENM methods depends strongly on how collective the motion is [54,131]. In other words, ENM-based methods do not provide real dynamics and real transition pathways but only an approximation based on the combination and interpolation of normal modes. The high accuracy of such a simplified approach for the description of system dynamics near the given reference structure clearly shows that the local dynamics is to a large extent controlled by the general structural features of proteins and is less dependent on specific interaction patterns between amino acid residues. For example, the use of ENM in protein-ligand docking is restricted to selected cases in a few only collective motions responsible for binding a ligand [131]. In this context, CG MC simulations appear to be more realistic. In principle, simulation methods, such as MD and CG MC (or CG MD), can be used for simulations of entire folding/unfolding processes, although the cost (and limitations of practical applicability) of such simulations rapidly grows with increasing resolution of the models. Deeply CG models, such as SURPASS discussed here, allow simulations of the largest systems, although such CG levels should be used in connection with more accurate modeling tools.

In the above context, the future will see combinations of ENM and other CG methods of various resolution with atomistic MD [14,132–134]. Such hierarchical integrative approaches should be properly tailored for specific problems and utilize structural data from available experimental sources such as X-ray crystallography, NMR, EM, SAXS and the like [135–137]. The challenges of integrative modeling call for computational tools available as easy to use software packages based on different kinds of data and merging with other tools, just as the CABS model described here [40].

## References

1.  Habchi, J.; Tompa, P.; Longhi, S.; Uversky, V. N. Introducing protein intrinsic disorder. *Chem. Rev.* **2014**, *114*, 6561–88, doi:10.1021/cr400514h.
2.  Papaleo, E.; Saladino, G.; Lambrughi, M.; Lindorff-Larsen, K.; Gervasio, F. L.; Nussinov, R. The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **2016**, *116*, 6391–423, doi:10.1021/acs.chemrev.5b00623.
3.  Guo, J.; Zhou, H. X. Protein Allostery and Conformational Dynamics. *Chem. Rev.* **2016**, *116*, 6503–15, doi:10.1021/acs.chemrev.5b00590.
4.  Cossins, B. P.; Lawson, A. D. G.; Shi, J. Computational exploration of conformational transitions in protein drug targets. *Methods Mol. Biol.* **2018**, *1762*, 339–365, doi:10.1007/978-1-4939-7756-7_17.
5.  Antunes, D. A.; Devaurs, D.; Kavraki, L. E. Understanding the challenges of protein flexibility in drug design. *Expert Opin. Drug Discov.* **2015**, *10*, 1301–13, doi:10.1517/17460441.2015.1094458.
6.  Ciemny, M.; Kurcinski, M.; Kamel, K.; Kolinski, A.; Alam, N.; Schueler-Furman, O.; Kmiecik, S. Protein–peptide docking: opportunities and challenges. *Drug Discov. Today* **2018**, *23*, 1530–1537, doi:10.1016/j.drudis.2018.05.006.
7.  Pan, A. C.; Weinreich, T. M.; Piana, S.; Shaw, D. E. Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J. Chem. Theory Comput.* **2016**, *12*, 1360–1367, doi:10.1021/acs.jctc.5b00913.
8.  Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898–7936, doi:10.1021/acs.chemrev.6b00163.
9.  Kouza, M.; Hu, C. K.; Li, M. S.; Kolinski, A. A structure-based model fails to probe

the mechanical unfolding pathways of the titin I27 domain. *J. Chem. Phys.* **2013**, *139*, 065103, doi:10.1063/1.4817773.

10. Hills, R. D.; Brooks, C. L. Insights from coarse-grained Gō models for protein folding and dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889–905, doi:10.3390/ijms10030889.

11. Orozco, M. A theoretical view of protein dynamics. *Chem. Soc. Rev.* **2014**, *43*, 5051–5066, doi:10.1039/c3cs60474h.

12. Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. Global Dynamics of Proteins: Bridging Between Structure and Function. *Annu. Rev. Biophys.* **2010**, *39*, 23–42, doi:10.1146/annurev.biophys.093008.131258.

13. Bahar, I.; Rader, A. J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–92, doi:10.1016/j.sbi.2005.08.007.

14. López-Blanco, J. R.; Chacón, P. New generation of elastic network models. *Curr. Opin. Struct. Biol.* **2016**, *37*, 46–53, doi:10.1016/j.sbi.2015.11.013.

15. Levitt, M. Birth and future of multiscale modeling for macromolecular systems (Nobel Lecture). *Angew. Chemie - Int. Ed.* **2014**, *53*, 10006–10018, doi:10.1002/anie.201403691.

16. Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 10020–10031, doi:10.1002/anie.201403689.

17. Stumpff-Kane, A. W.; Maksimiak, K.; Lee, M. S.; Feig, M. Sampling of near-native protein conformations during protein structure refinement using a coarse-grained model, normal modes, and molecular dynamics simulations. *Proteins Struct. Funct. Bioinf.* **2008**, *70*, 1345–1356, doi:10.1002/prot.21674.

18. Kolinski, A. Toward more efficient simulations of slow processes in large biomolecular systems: Comment on "Ligand diffusion in proteins via enhanced sampling in molecular dynamics" by Jakub Rydzewski and Wieslaw Nowak. *Phys. Life Rev.* **2017**, *22–23*, 75–76, doi:10.1016/j.plrev.2017.07.003.

19. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (80-. ).* **2010**, *330*, 341–346, doi:10.1126/science.1187409.

20. Tozzini, V. Minimalist models for proteins: A comparative analysis. *Q. Rev. Biophys.* **2010**, *43*, 333–371, doi:10.1017/S0033583510000132.

21. Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150, doi:10.1016/j.sbi.2005.02.005.

22. Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; Niadzvedtski, A.; Ołdziej, S.; Scheraga, H. A.; Sieradzan, A. K.; Ślusarz, R.; Wirecki, T.; Yin, Y.; Zaborowski, B. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306–2321, doi:10.1007/s00894-014-2306-5.

23. Kar, P.; Gopal, S. M.; Cheng, Y. M.; Predeus, A.; Feig, M. PRIMO: A transferable coarse-grained force field for proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3769–3788, doi:10.1021/ct400230y.

24. Kolinski, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* **2004**, *51*, 349–371, doi:035001349.

25. Dawid, A. E.; Gront, D.; Kolinski, A. SURPASS Low-Resolution Coarse-Grained Protein Modeling. *J. Chem. Theory Comput.* **2017**, *13*, 5766–5779, doi:10.1021/acs.jctc.7b00642.

26. Kmiecik, S.; Kolinski, A. One-dimensional structural properties of proteins in the coarse-grained cabs model. *Methods Mol. Biol.* **2017**, *1484*, 83–113, doi:10.1007/978-

1-4939-6406-2_8.

27.   Godzik, A.; Kolinski, A.; Skolnick, J. Lattice representations of globular proteins: How good are they? *J. Comput. Chem.* **1993**, *14*, 1194–1202, doi:10.1002/jcc.540141009.

28.   Kmiecik, S.; Kolinski, A. Characterization of protein-folding pathways by reduced-space modeling. *Proc. Natl. Acad. Sci.* **2007**, *104*, 12330–12335, doi:10.1073/pnas.0702265104.

29.   Kmiecik, S.; Kurcinski, M.; Rutkowska, A.; Gront, D.; Kolinski, A. Denatured proteins and early folding intermediates simulated in a reduced conformational space. *Acta Biochim. Pol.* **2006**, *53*, 131–144.

30.   Kmiecik, S.; Kolinski, A. Folding pathway of the B1 domain of protein G explored by multiscale modeling. *Biophys. J.* **2008**, *94*, 726–736, doi:10.1529/biophysj.107.116095.

31.   Kmiecik, S.; Gront, D.; Kouza, M.; Kolinski, A. From coarse-grained to atomic-level characterization of protein dynamics: Transition state for the folding of B domain of protein A. *J. Phys. Chem. B* **2012**, *116*, 7026–7032, doi:10.1021/jp301720w.

32.   Kmiecik, S.; Kolinski, A. Simulation of Chaperonin Effect on Protein Folding: a Shift from Nucleation-Condensation to Framework Mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 10283–10289, doi:10.1021/ja203275f.

33.   Kmiecik, S.; Wabit, J.; Kolinski, M.; Kouza, M.; Kolinski, A. Coarse-Grained Modeling of Protein Dynamics. *Comput. Methods to Study Struct. Dyn. Biomol.* **2014**, *1*, 55–79, doi:10.1007/978-3-642-28554-7.

34.   Kurcinski, M.; Kolinski, A.; Kmiecik, S. Mechanism of folding and binding of an intrinsically disordered protein as revealed by ab initio simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2224–2231, doi:10.1021/ct500287c.

35.   Ciemny, M. P.; Debinski, A.; Paczkowska, M.; Kolinski, A.; Kurcinski, M.; Kmiecik, S. Protein-peptide molecular docking with large-scale conformational changes: The p53-MDM2 interaction. *Sci. Rep.* **2016**, *6*, 37532, doi:10.1038/srep37532.

36.   Jamroz, M.; Orozco, M.; Kolinski, A.; Kmiecik, S. Consistent view of protein fluctuations from all-atom molecular dynamics and coarse-grained dynamics with knowledge-based force-field. *J. Chem. Theory Comput.* **2013**, *9*, 119–25, doi:10.1021/ct300854w.

37.   Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex: Server for fast simulation of protein structure fluctuations. *Nucleic Acids Res.* **2013**, *41*, W427-431, doi:10.1093/nar/gkt332.

38.   Jamroz, M.; Kolinski, A.; Kmiecik, S. CABS-flex predictions of protein flexibility compared with NMR ensembles. *Bioinformatics* **2014**, *30*, 2150–2154, doi:10.1093/bioinformatics/btu184.

39.   Kuriata, A.; Gierut, A. M.; Oleniecki, T.; Ciemny, M. P.; Kolinski, A.; Kurcinski, M.; Kmiecik, S. CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res.* **2018**, *46*, W338-343, doi:10.1093/nar/gky356.

40.   Kurcinski, M.; Oleniecki, T.; Ciemny, P. M.; Kuriata, A.; Kolinski, A.; Kmiecik, S. CABS-flex standalone: a simulation environment for fast modeling of protein flexibility. *Bioinformatics* **2018**, *bty685*, doi:10.1093/bioinformatics/bty685.

41.   Kmiecik, S.; Kolinski, A. Simulation of chaperonin effect on protein folding: A shift from nucleation - Condensation to framework mechanism. *J. Am. Chem. Soc.* **2011**, *133*, 10283–10289, doi:10.1021/ja203275f.

42.   Pulawski, W.; Jamroz, M.; Kolinski, M.; Kolinski, A.; Kmiecik, S. Coarse-grained simulations of membrane insertion and folding of small helical proteins using the CABS model. *J. Chem. Inf. Model.* **2016**, *56*, 2207–2215, doi:10.1021/acs.jcim.6b00350.

43.    Blaszczyk, M.; Jamroz, M.; Kmiecik, S.; Kolinski, A. CABS-fold: Server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Res.* **2013**, *41*, W406-11, doi:10.1093/nar/gkt462.

44.    Kmiecik, S.; Jamroz, M.; Kolinski, M. Structure prediction of the second extracellular loop in G-protein-coupled receptors. *Biophys. J.* **2014**, *106*, 2408–16, doi:10.1016/j.bpj.2014.04.022.

45.    Kolinski, A.; Bujnicki, J. M. Generalized protein structure prediction based on combination of fold-recognition with de novo folding and evaluation of models. *Proteins Struct. Funct. Genet.* **2005**, *61*, 84–90, doi:10.1002/prot.20723.

46.    Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43*, W419–W424, doi:10.1093/nar/gkv456.

47.    Blaszczyk, M.; Kurcinski, M.; Kouza, M.; Wieteska, L.; Debinski, A.; Kolinski, A.; Kmiecik, S. Modeling of protein-peptide interactions using the CABS-dock web server for binding site search and flexible docking. *Methods* **2016**, *93*, 72–83, doi:10.1016/j.ymeth.2015.07.004.

48.    Ciemny, M. P.; Kurcinski, M.; Kozak, K.; Kolinski, A.; Kmiecik, S. Highly flexible protein-peptide docking using cabs-dock. *Methods Mol. Biol.* **2017**, *1561*, 69–94, doi:10.1007/978-1-4939-6798-8_6.

49.    Ciemny, M. P.; Kurcinski, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. Modeling EphB4-EphrinB2 protein-protein interaction using flexible docking of a short linear motif. *Biomed. Eng. Online* **2017**, *16*, 71, doi:10.1186/s12938-017-0362-7.

50.    Kurcinski, M.; Blaszczyk, M.; Ciemny, M. P.; Kolinski, A.; Kmiecik, S. A protocol for CABS-dock protein-peptide docking driven by side-chain contact information. *Biomed. Eng. Online* **2017**, *16*, 73, doi:10.1186/s12938-017-0363-6.

51.    Jamroz, M.; Kolinski, A.; Kmiecik, S. Protocols for efficient simulations of long-time protein dynamics using coarse-grained CABS model. *Methods Mol. Biol.* **2014**, *1137*, 235–50, doi:10.1007/978-1-4939-0366-5_16.

52.    Dawid, A. E.; Gront, D.; Kolinski, A. Coarse-Grained Modeling of the Interplay between Secondary Structure Propensities and Protein Fold Assembly. *J. Chem. Theory Comput.* **2018**, *14*, 2277–2287, doi:10.1021/acs.jctc.7b01242.

53.    Feig, M.; Rotkiewicz, P.; Kolinski, A.; Skolnick, J.; Brooks, C. L. Accurate reconstruction of all-atom protein representations from side-chain-based low-resolution models. *Proteins Struct. Funct. Genet.* **2000**, *41*, 86–97, doi:10.1002/1097-0134(20001001)41:1<86::AID-PROT110>3.0.CO;2-Y.

54.    Yang, L.; Song, G.; Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929, doi:10.1529/biophysj.106.095927.

55.    Ma, J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **2005**, *13*, 373–380, doi:10.1016/j.str.2005.02.002.

56.    Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chem. Rev.* **2010**, *110*, 1463–1497, doi:10.1021/cr900095e.

57.    Hayward, S.; De Groot, B. L. Normal modes and essential dynamics. *Methods Mol. Biol.* **2008**, *443*, 89–106, doi:10.1007/978-1-59745-177-2_5.

58.    Case, D. A. Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4*, 285–290, doi:10.1016/S0959-440X(94)90321-2.

59.    Toussi, C. A.; Soheilifard, R. A better prediction of conformational changes of proteins using minimally connected network models. *Phys. Biol.* **2017**, *13*, 066013, doi:10.1088/1478-3975/13/6/066013.

60.    Fuglebakk, E.; Tiwari, S. P.; Reuter, N. Comparing the intrinsic dynamics of multiple

protein structures using elastic network models. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 911–922, doi:10.1016/j.bbagen.2014.09.021.

61.    Kim, M. H.; Lee, B. H.; Kim, M. K. Robust elastic network model: A general modeling for precise understanding of protein dynamics. *J. Struct. Biol.* **2015**, *190*, 338–347, doi:10.1016/j.jsb.2015.04.007.

62.    Van Wynsberghe, A. W.; Cui, Q. Interpreting Correlated Motions Using Normal Mode Analysis. *Structure* **2006**, *14*, 1647–1653, doi:10.1016/j.str.2006.09.003.

63.    Rueda, M.; Chacón, P.; Orozco, M. Thorough Validation of Protein Normal Mode Analysis: A Comparative Study with Essential Dynamics. *Structure* **2007**, *15*, 565–575, doi:10.1016/j.str.2007.03.013.

64.    Go, N.; Noguti, T.; Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci.* **1983**, *80*, 3696–3700, doi:10.1073/pnas.80.12.3696.

65.    Brooks, B.; Brooks, B.; Karplus, M.; Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A* **1983**, *80*, 6571–6575, doi:10.1073/pnas.80.21.6571.

66.    Lu, M.; Ma, J. The role of shape in determining molecular motions. *Biophys. J.* **2005**, *89*, 2395–2401, doi:10.1529/biophysj.105.065904.

67.    Howe, P. W. Principal components analysis of protein structure ensembles calculated using NMR data. *J. Biomol. NMR* **2001**, *20*, 61–70, doi:10.1023/A:1011210009067.

68.    Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. L. Close Correspondence between the Motions from Principal Component Analysis of Multiple HIV-1 Protease Structures and Elastic Network Modes. *Structure* **2008**, *16*, 321–330, doi:10.1016/j.str.2007.12.011.

69.    Bahar, I.; Atilgan, A. R.; Erman, B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **1997**, *2*, 173–181, doi:10.1016/S1359-0278(97)00024-2.

70.    Tirion, M. M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905–1908, doi:10.1103/PhysRevLett.77.1905.

71.    Ruvinsky, A. M.; Kirys, T.; Tuzikov, A. V; Vakser, I. A. Structure fluctuations and conformational changes in protein binding. *J. Bioinform. Comput. Biol.* **2012**, *10*, 1241002, doi:10.1142/S0219720012410028.

72.    Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217, doi:10.1002/jcc.540040211.

73.    Bahar, I.; Wallqvist, A.; Covell, D. G.; Jernigan, R. L. Correlation between native-state hydrogen exchange and cooperative residue fluctuations from a simple model. *Biochemistry* **1998**, *37*, 1067–1075, doi:10.1021/bi9720641.

74.    Atilgan, A. R.; Durell, S. R.; Jernigan, R. L.; Demirel, M. C.; Keskin, O.; Bahar, I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* **2001**, *80*, 505–515, doi:10.1016/S0006-3495(01)76033-X.

75.    Eyal, E.; Yang, L.-W.; Bahar, I. Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics* **2006**, *22*, 2619–2627, doi:10.1093/bioinformatics/btl448.

76.    Doruker, P.; Jernigan, R. L.; Bahar, I. Dynamics of large proteins through hierarchical levels of coarse-grained structures. *J. Comput. Chem.* **2002**, *23*, 119–127, doi:10.1002/jcc.1160.

77.    Kurkcuoglu, O.; Jernigan, R. L.; Doruker, P. Mixed levels of coarse-graining of large proteins using elastic network model succeeds in extracting the slowest motions.

*Polymer (Guildf).* **2004**, *45*, 649–657, doi:10.1016/j.polymer.2003.10.071.

78. Mishra, S. K.; Jernigan, R. L. Protein dynamic communities from elastic network models align closely to the communities defined by molecular dynamics. *PLoS One* **2018**, *13*, e0199225, doi:10.1371/journal.pone.0199225.

79. Orellana, L.; Yoluk, O.; Carrillo, O.; Orozco, M.; Lindahl, E. Prediction and validation of protein intermediate states from structurally rich ensembles and coarse-grained simulations. *Nat. Commun.* **2016**, *7*, 12575, doi:10.1038/ncomms12575.

80. Poma, A. B.; Li, M. S.; Theodorakis, P. E. Generalization of the elastic network model for the study of large conformational changes in biomolecules. *Phys. Chem. Chem. Phys.* **2018**, *20*, 17020–17028, doi:10.1039/c8cp03086c.

81. Putz, I.; Brock, O. Elastic network model of learned maintained contacts to predict protein motion. *PLoS One* **2017**, *12*, e0183889, doi:10.1371/journal.pone.0183889.

82. Takada, S. Coarse-grained molecular simulations of large biomolecules. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130–137, doi:10.1016/j.sbi.2012.01.010.

83. Potestio, R.; Pontiggia, F.; Micheletti, C. Coarse-grained description of protein internal dynamics: An optimal strategy for decomposing proteins in rigid subunits. *Biophys. J.* **2009**, *96*, 4993–5002, doi:10.1016/j.bpj.2009.03.051.

84. Kenzaki, H.; Koga, N.; Hori, N.; Kanada, R.; Li, W.; Okazaki, K. I.; Yao, X. Q.; Takada, S. CafeMol: A coarse-grained biomolecular simulator for simulating proteins at work. *J. Chem. Theory Comput.* **2011**, *7*, 1979–1989, doi:10.1021/ct2001045.

85. Krüger, D. M.; Ahmed, A.; Gohlke, H. NMSim web server: Integrated approach for normal mode-based geometric simulations of biologically relevant conformational transitions in proteins. *Nucleic Acids Res.* **2012**, *40*, W310-316, doi:10.1093/nar/gks478.

86. Li, H.; Chang, Y. Y.; Lee, J. Y.; Bahar, I.; Yang, L. W. DynOmics: Dynamics of structural proteome and beyond. *Nucleic Acids Res.* **2017**, *45*, W374-380, doi:10.1093/nar/gkx385.

87. Zheng, W.; Brooks, B. R.; Hummer, G. Protein conformational transitions explored by mixed elastic network models. *Proteins Struct. Funct. Bioinforma.* **2007**, *69*, 43–57, doi:10.1002/prot.21465.

88. Tekpinar, M.; Zheng, W. Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins Struct. Funct. Bioinforma.* **2010**, *78*, 2469–2481, doi:10.1002/prot.22755.

89. Hollup, S. M.; Salensminde, G.; Reuter, N. WEBnm@: A web application for normal mode analyses of proteins. *BMC Bioinformatics* **2005**, *6*, 52–60, doi:10.1186/1471-2105-6-52.

90. López-Blanco, J. R.; Aliaga, J. I.; Quintana-Ortí, E. S.; Chacón, P. IMODS: Internal coordinates normal mode analysis server. *Nucleic Acids Res.* **2014**, *42*, W271-276, doi:10.1093/nar/gku339.

91. Lindahl, E.; Azuara, C.; Koehl, P.; Delarue, M. NOMAD-Ref: Visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* **2006**, *34*, W52-56, doi:10.1093/nar/gkl082.

92. Suhre, K.; Sanejouand, Y.-H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **2004**, *32*, W610-614, doi:10.1093/nar/gkh368.

93. Li, H.; Chang, Y. Y.; Yang, L. W.; Bahar, I. iGNM 2.0: The Gaussian network model database for biomolecular structural dynamics. *Nucleic Acids Res.* **2016**, *44*, 415–422, doi:10.1093/nar/gkv1236.

94. Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **2011**, *27*, 1575–1577, doi:10.1093/bioinformatics/btr168.

95.     Zimmermann, M. T.; Kloczkowski, A.; Jernigan, R. L. MAVENs: Motion analysis and visualization of elastic networks and structural ensembles. *BMC Bioinformatics* **2011**, *12*, 264–271, doi:10.1186/1471-2105-12-264.

96.     Eyal, E.; Lum, G.; Bahar, I. The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics* **2015**, *31*, 1487–1489, doi:10.1093/bioinformatics/btu847.

97.     Derreumaux, P.; Mousseau, N. Coarse-grained protein molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 025101, doi:10.1063/1.2408414.

98.     Gu, R. X.; Ingólfsson, H. I.; De Vries, A. H.; Marrink, S. J.; Tieleman, D. P. Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. *J. Phys. Chem. B* **2017**, *121*, 3262–3275, doi:10.1021/acs.jpcb.6b07142.

99.     Gniewek, P.; Kolinski, A.; Jernigan, R. L.; Kloczkowski, A. Elastic network normal modes provide a basis for protein structure refinement. *J. Chem. Phys.* **2012**, *136*, 195101, doi:10.1063/1.4710986.

100.    Gniewek, P.; Kolinski, A.; Jernigan, R.; Kloczkowski, A. ANM Normal Modes Show the Directions for Protein Structure Refinement. *Biophys. J.* **2012**, *102*, 25a.

101.    Schindler, C. E. M.; de Vries, S. J.; Zacharias, M. iATTRACT: Simultaneous global and local interface optimization for protein-protein docking refinement. *Proteins Struct. Funct. Bioinforma.* **2015**, *83*, 248–258, doi:10.1002/prot.24728.

102.    Park, H.; DiMaio, F.; Baker, D. CASP11 refinement experiments with ROSETTA. *Proteins* **2016**, *84*, 314–322, doi:10.1002/prot.24862.

103.    Schröder, G. F.; Levitt, M.; Brunger, A. T. Deformable elastic network refinement for low-resolution macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2014**, *D70*, 2241–2255, doi:10.1107/S1399004714016496.

104.    Tirion, M. M.; Ben-Avraham, D. Atomic torsional modal analysis for high-resolution proteins. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2015**, *91*, 032712, doi:10.1103/PhysRevE.91.032712.

105.    Takada, S.; Kanada, R.; Tan, C.; Terakawa, T.; Li, W.; Kenzaki, H. Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc. Chem. Res.* **2015**, *48*, 3026–3035, doi:10.1021/acs.accounts.5b00338.

106.    Mirjalili, V.; Noyes, K.; Feig, M. Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. *Proteins Struct. Funct. Bioinforma.* **2014**, *82*, 196–207, doi:10.1002/prot.24336.

107.    Mirjalili, V.; Feig, M. Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. *J. Chem. Theory Comput.* **2013**, *9*, 1294–1303, doi:10.1021/ct300962x.

108.    Tama, F.; Miyashita, O.; Brooks, C. L. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* **2004**, *337*, 985–999, doi:10.1016/j.jmb.2004.01.048.

109.    Stansfeld, P. J.; Sansom, M. S. P. From coarse grained to atomistic: A serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.* **2011**, *7*, 1157–1166, doi:10.1021/ct100569y.

110.    Singharoy, A.; Teo, I.; McGreevy, R.; Stone, J. E.; Zhao, J.; Schulten, K. Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *Elife* **2016**, *5*, e16105, doi:10.7554/eLife.16105.

111.    Feig, M.; Mirjalili, V. Protein structure refinement via molecular-dynamics simulations: What works and what does not? *Proteins* **2016**, *84*, 282–92, doi:10.1002/prot.24871.

112.    Ghavami, A.; Veenhoff, L. M.; Van Der Giessen, E.; Onck, P. R. Probing the disordered domain of the nuclear pore complex through coarse-grained molecular

dynamics simulations. *Biophys. J.* **2014**, *107*, 1393–1402, doi:10.1016/j.bpj.2014.07.060.

113. Schuyler, Adam D. Jernigan, R. L.; Qasba, Pradman K. Ramakrishnan, B.; Chirikjian, G. S. Conformational transition paths are computed by combination and interpolation of normal modes of both struttures. *Proteins* **2009**, *74*, 760–776, doi:10.1007/978-1-62703-658-0_9.

114. Mahajan, S.; Sanejouand, Y. H. On the relationship between low-frequency normal modes and the large-scale conformational changes of proteins. *Arch. Biochem. Biophys.* **2015**, *567*, 59–65, doi:10.1016/j.abb.2014.12.020.

115. Chu, J.-W.; Voth, G. A. Coarse-Grained Free Energy Functions for Studying Protein Conformational Changes: A Double-Well Network Model. *Biophys. J.* **2007**, *93*, 3860–3871, doi:10.1529/biophysj.107.112060.

116. Zheng, W.; Tekpinar, M. Analysis of Protein Conformational Transitions Using Elastic Network Model. In *Protein Dynamics*; Springer, **2013**; pp. 159–172.

117. Tama, F.; Sanejouand, Y.-H. Conformational change of proteins arising from normal mode calculations. *Protein Eng. Des. Sel.* **2001**, *14*, 1–6, doi:10.1093/protein/14.1.1.

118. Frappier, V.; Najmanovich, R. J. A Coarse-Grained Elastic Network Atom Contact Model and Its Use in the Simulation of Protein Dynamics and the Prediction of the Effect of Mutations. *PLoS Comput. Biol.* **2014**, *10*, e1003569, doi:10.1371/journal.pcbi.1003569.

119. Venkatramani, R.; Sarkar, R.; Kotamarthi, H. C.; Koti, A. S. R. Assessing Limitations of Elastic Network Models in Describing Equilibrium Protein Flexibility and Extensions to Predict Non-Equilibrium Unfolding Dynamics of Proteins. *Biophys. J.* **2014**, *106*, 412A, doi:https://doi.org/10.1016/j.bpj.2013.11.2317.

120. Kurkcuoglu, O.; Doruker, P.; Sen, T. Z.; Kloczkowski, A.; Jernigan, R. L. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. *Phys. Biol.* **2008**, *5*, 046005, doi:10.1088/1478-3975/5/4/046005.

121. Kurkcuoglu, O.; Kurkcuoglu, Z.; Doruker, P.; Jernigan, R. L. Collective dynamics of the ribosomal tunnel revealed by elastic network modeling. *Proteins Struct. Funct. Bioinforma.* **2009**, *75*, 837–845, doi:10.1002/prot.22292.

122. Kurkcuoglu, O.; Turgut, O. T.; Cansu, S.; Jernigan, R. L.; Doruker, P. Focused functional dynamics of supramolecules by use of a mixed-resolution elastic network model. *Biophys. J.* **2009**, *97*, 1178–1187, doi:10.1016/j.bpj.2009.06.009.

123. Zhang, Y.; Kolinski, A.; Skolnick, J. TOUCHSTONE II: A new approach to ab initio protein structure prediction. *Biophys. J.* **2003**, *85*, 1145–1164, doi:10.1016/S0006-3495(03)74551-2.

124. Roy, A.; Kucukural, A.; Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010**, *5*, 725–738, doi:10.1038/nprot.2010.5.

125. Liwo, A.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. Computational techniques for efficient conformational sampling of proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134–139, doi:10.1016/j.sbi.2007.12.001.

126. Czaplewski, C.; Karczyńska, A.; Sieradzan, A. K.; Liwo, A. UNRES server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics. *Nucleic Acids Res.* **2018**, *46*, W304–W309, doi:10.1093/nar/gky328.

127. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819–834, doi:10.1021/ct700324x.

128. Tang, X.; Wang, Z.; Lei, T.; Zhou, W.; Chang, S.; Li, D. Importance of protein flexibility on molecular recognition: modeling binding mechanisms of aminopyrazine

inhibitors to Nek2. *Phys. Chem. Chem. Phys.* **2018**, *20*, 5591–5605, doi:10.1039/c7cp07588j.

129. Kong, X.; Sun, H.; Pan, P.; Zhu, F.; Chang, S.; Xu, L.; Li, Y.; Hou, T. Importance of protein flexibility in molecular recognition: a case study on Type-I1/2 inhibitors of ALK. *Phys. Chem. Chem. Phys.* **2018**, *20*, 4851–4863, doi:10.1039/c7cp08241j.

130. Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci.* **2018**, *115*, E4758-4766, doi:10.1073/pnas.1800690115.

131. Dietzen, M.; Zotenko, E.; Hildebrandt, A.; Lengauer, T. On the applicability of elastic network normal modes in small-molecule docking. *J. Chem. Inf. Model.* **2012**, *52*, 844–856, doi:10.1021/ci2004847.

132. Kar, P.; Feig, M. Hybrid All-Atom/Coarse-Grained Simulations of Proteins by Direct Coupling of CHARMM and PRIMO Force Fields. *J. Chem. Theory Comput.* **2017**, *13*, 5753–5765, doi:10.1021/acs.jctc.7b00840.

133. Karczyńska, A. S.; Mozolewska, M. A.; Krupa, P.; Giełdoń, A.; Liwo, A.; Czaplewski, C. Prediction of protein structure with the coarse-grained UNRES force field assisted by small X-ray scattering data and knowledge-based information. *Proteins Struct. Funct. Bioinforma.* **2018**, *86*, 228–239, doi:10.1002/prot.25421.

134. Chen, Y.; Roux, B. Enhanced Sampling of an Atomic Model with Hybrid Nonequilibrium Molecular Dynamics—Monte Carlo Simulations Guided by a Coarse-Grained Model. *J. Chem. Theory Comput.* **2015**, *11*, 3572–3583, doi:10.1021/acs.jctc.5b00372.

135. Rodrigues, J. P. G. L. M.; Bonvin, A. M. J. J. Integrative computational modeling of protein interactions. *FEBS J.* **2014**, *281*, 1988–2003, doi:10.1111/febs.12771.

136. Spiga, E.; Degiacomi, M. T.; Dal Peraro, M. New strategies for integrative dynamic modeling of macromolecular assembly. *Adv. Protein Chem. Struct. Biol.* **2014**, *96*, 77–111, doi:10.1016/bs.apcsb.2014.06.008.

137. Webb, B.; Viswanath, S.; Bonomi, M.; Pellarin, R.; Greenberg, C. H.; Saltzberg, D.; Sali, A. Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **2018**, *281*, 1988–2003, doi:10.1002/pro.3311.

Supplementary Materials: none

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

| | |
|---|---|
| MD | molecular dynamics |
| CG | coarse-grained |
| ENM | elastic network model |
| NMA | normal mode analysis |
| MC | Monte Carlo |
| AA | all-atom |
| CABS | C$\alpha$, C$\beta$, Side chain model |
| SURPASS | Single United Residue per Pre-averaged Secondary Structure fragment |

| | |
|---|---|
| REMC | replica exchange Monte Carlo |
| PCA | principle component analysis |
| PCs | principle components |
| GNM | Gaussian network model |
| PDB | Protein Data Bank |
| ANM | anisotropic network model |
| GPCR | G protein-coupled receptor |
| RMSF | root mean square fluctuations |
| NMR | nuclear magnetic resonance |
| UNRES | united residue model |
| GEN | generalized elastic network model |
| AFM | atomic force microscope |