JCTC Journal of Chemical Theory and Computation

Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition

Xavier Periole,[†,‡,§] Marco Cavalli,[†] Siewert-Jan Marrink,[‡] and Marco A. Ceruso*[,†]

Department of Chemistry and Biochemistry and Institute for Macromolecular Assemblies, The City College of New York, 160 Convent Ave, New York, New York 10031, and Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received April 29, 2009

Abstract: Structure-based and physics-based coarse-grained molecular force fields have become attractive approaches to gain mechanistic insight into the function of large biomolecular assemblies. Here, we study how both approaches can be combined into a single representation, that we term ELNEDIN. In this representation an elastic network is used as a structural scaffold to describe and maintain the overall shape of a protein and a physics-based coarse-grained model (MARTINI-2.1) is used to describe both inter- and intramolecular interactions in the system. The results show that when used in molecular dynamics simulations ELNEDIN models can be built so that the resulting structural and dynamical properties of a protein, including its collective motions, are comparable to those obtained using atomistic protein models. We then evaluate the behavior of such models in (1) long, microsecond time-scale, simulations, (2) the modeling of very large macromolecular assemblies, a viral capsid, and (3) the study of a protein−protein association process, the reassembly of the ROP homodimer. The results for this series of tests indicate that ELNEDIN models allow microsecond time-scale molecular dynamics simulations to be carried out readily, that large biological entities such as the viral capsid of the cowpea mosaic virus can be stably modeled as assemblies of independent ELNEDIN models, and that ELNEDIN models show significant promise for modeling protein−protein association processes.

Introduction

Computational modeling of molecular mechanisms of biological processes is a challenging task. It requires models that can reproduce accurately not only the structural and the dynamical properties of all molecular entities involved (i.e., protein receptors, protein effectors, ligands, lipid molecules, aqueous environment) but also the transient intermolecular interactions in which these entities engage and that modulate their various functional states. This task is often complicated further by the size of the biological systems involved and by the time scales over which these functional processes occur.[1]

One way to circumvent these challenges, without sacrificing on the resolution at which the biological system is represented, is to take advantage of experimentally available information.[2−4] For example, the existence of gain-of-function mutants[5,6] within the Gα subunit of the heterotrimeric G protein, transducin, was exploited. It enabled detailed atomistic insights to be obtained via classical molecular dynamics simulations, about the allosteric structural changes that accompany the activation of the G protein

* To whom correspondence should be addressed. Tel: +1-212-650-6035. Fax: +1-212-650-6107. Email: mceruso@sci.ccny.cuny.edu.
† The City College of New York.
‡ University of Groningen.
§ Current address: Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands.

10.1021/ct9002114 CCC: $40.75
Published on Web 08/19/2009

by its cognate transmembrane receptor rhodopsin. This was possible without modeling the receptor, the lipid environment, or their intermolecular interactions with transducin explicitly.[7] Another way to address these challenges is to make use of algorithms that enhance conformational sampling (for recent reviews see[1,8−14]). But the size of the systems that can be studied with such techniques is as limited as for the more classical approaches such as atomistic molecular dynamics simulations.

An approach that has attracted a great deal of interest has been to develop simplified molecular models to reduce the number of degrees of freedom that need to be taken into account.[1,10,15−18] This approach is particularly advantageous because it permits to increase the time-scale of the simulation and the size of the molecular system simultaneously. The challenge for these simplified or coarse-grained (CG) molecular models is to achieve an accurate description of the free energy surface. Transferability of the CG model is also a challenge. Ideally one would like to have a CG model readily applicable in a variety of molecular contexts.[1]

To date, CG models have been developed for a variety of biomolecules including lipids,[19−22] proteins[10,23−28] and DNA.[29−31] Typically a CG model groups atoms into single interaction centers. The degree of coarse-graining can vary from 2−6 atoms to the whole macromolecule and such force fields are usually parametrized following a knowledge-based or a physics-based approach (see refs 1, 10, and 18 for a description of recent CG models). From the latter category, the MARTINI force field,[21,22] compatible with the so-called GROMOS philosophy,[32] is based on the parametrization of a large library of building blocks against experimental thermodynamic data. This approach is particularly valuable because a number of biological phenomena (e.g., protein folding, peptide-membrane binding, or protein−protein association) depend on the degree to which the constituent groups partition between polar and nonpolar environments. In addition because the building blocks can be combined to construct virtually any molecule,[22,28] they offer an immediate solution to the transferability challenge. Finally, another advantage of the MARTINI force field is that its degree of coarse-graining (∼residue-level) is sufficiently detailed to maintain a close tie with experimental approaches that probe the involvement of particular residues in functional mechanisms. The MARTINI model has already been used in a number of protein studies, including the self-assembly of rhodopsins[33] and the gating of mechanosensitive[34] and voltage gated membrane channels.[35]

Elastic Network (EN) models represent another form of coarse-graining (for recent reviews see[36−38]). EN models were first introduced by Tirion[39] as an alternative to classical atomistic normal-mode analysis. By taking the native structure as the minimum of the free energy EN models eliminated the need of an often-costly minimization procedure, but in so doing, EN models introduced an intrinsic bias toward the initial experimental structure. In an EN model, the structure of a macromolecule is described as a network of point masses connected to one another with springs when the distance between the point masses is less than a predefined cutoff distance (R_C). In the simplest form of EN model, the values of the spring force constant, K_{SPRING}, and the cutoff, R_C, are taken to be the same throughout the network. These two parameters, characterize the network, that is, its rigidity and its extent. A number of variants have been proposed throughout the years[40−42] and the effects of various schemes for building the network of connected residues and setting the spring force constants have been evaluated in a number of studies.[43−46] Recently, EN models have attracted a lot of interest because in combination with the rotational-translational block approximation[47−50] the size of the biological systems that can be studied has been increased dramatically (e.g., see refs 51 and 52). But most importantly, the EN approach has shown a remarkable ability to reproduce a number of biologically relevant dynamical properties of macromolecules.[36,38]

EN models have been combined with atomistic (AT) and CG molecular force fields. When used in combination with AT models the main focus has been to enhance or to guide conformational sampling.[53−57] When used in combination with CG models the main goal has been to maintain the structure of the modeled biomolecule. Bond and co-workers have used both a classical EN scaffold ($R_C = 0.7$ nm and $K_{SPRING} = 1000$ kJ mol^{-1} nm^{-2}) and hydrogen-bond based harmonic distance restraints (force constant 1000 kJ mol^{-1} nm^{-2}), when modeling membrane proteins and membrane peptides, respectively.[58,59] Similarly, Periole et al.[33] used a combination of sequential ($C\alpha_i \rightarrow C\alpha_{i+4}$, $C\alpha_i \rightarrow C\alpha_{i+20}$, and $C\alpha_i \rightarrow C\alpha_{i+30}$ with $K = 9250$ kJ mol^{-1} nm^{-2}) and distance-based (between elements of secondary structure with $K = 1250$ kJ mol^{-1} nm^{-2}) harmonic restraints when studying the self-assembly of rhodopsin in membrane models. While a qualitative agreement between AT and CG simulations was found in these CG studies, the effect of using such restraints on the structure and the dynamics of the model remains to be determined.

In the present study, we investigate the possibility of combining both structure-based and physics-based CG models into a unique representation. Specifically, we characterize the structural and dynamical consequences of combining an EN model with the MARTINI force field.[22,28] For the sake of simplicity we focus only on EN scaffolds characterized by two unique parameters: R_C and K_{SPRING}. We use three model proteins, the B1 domain of protein G,[60] the src-SH3 domain,[61] and the villin headpiece subdomain,[62] respectively representing the $\alpha + \beta$ structural class, the all-β structural class, and the all-α structural class, to determine whether optimal and possibly universal values for the EN parameters, R_C and K_{SPRING}, can be identified. The aim is to have the combined EN-CG protein model, referred to as an ELNEDIN protein model from now on, reproduce quantitatively the structural and dynamical properties of the same protein simulated with an atomistic force field. In addition, we evaluate the behavior of an ELNEDIN model in the context of long (microsecond time-scale) molecular dynamics simulations, and large macromolecular assemblies (the capsid of the cowpea mosaic virus[63]). Finally, we evaluate the ability of an ELNEDIN model to be used for the study of protein−protein association processes. This last test was

carried out in the context of the homodimeric four-helix bundle repressor of primer protein, ROP.[64,65]

Methods

Protein Systems. A total of five protein systems were studied. The villin headpiece subdomain (PDB[66] entry 1YRF[62]), the D48G mutant of the α-spectrin SH3 domain (PDB entry 1BK2[61]), and the B1 domain of protein G (PDB entry 1PGB[60]) were used for the comparison of ELNEDIN and atomistic models. The capsid of the cowpea mosaic virus (PDB entry 1NY7[63]), which consists of 60 copies of two proteins containing 190 and 369 residues respectively, was used to evaluate the capacity of ELNEDIN to handle large macromolecular assemblies. And finally a mutant of the homodimeric four-helix bundle repressor of primer protein,[64] with an Ala residue inserted on either side of D31 (PDB entry 1RPO[65]) was used for protein−protein association evaluation.

Atomistic Simulations. All molecular dynamics (MD) simulations were performed using the GROMACS simulation package.[67] Simulations using an atomistic (AT) representation were based on the united-hydrogen GROMOS-43a1 force field[68] for the protein and the SPC water model for the solvent.[69] Each system (see below) was solvated in a rectangular box. The minimum distance between the protein and the edges of the box was initially set to 1.0 nm. The systems were simulated at constant pressure (1 bar) and constant temperature (300 K). Both the temperature and pressure were maintained close to their target values using the Berendsen weak coupling algorithm[70] with time constants $\tau_T = 0.1$ ps and $\tau_P = 1$ ps, for the temperature and pressure respectively. A twin-range cutoff (1.0−1.4 nm) was used for the nonbonded interactions. Interactions within the short-range cutoff (1.0 nm) were evaluated every time step (2 fs), whereas interactions within the long-range cutoff (1.4 nm) were updated every 10 steps together with the pair-list. To correct for the truncation of electrostatic interactions beyond the long-range cutoff a Reaction-Field[71] correction was applied ($\varepsilon = 78$). Bond lengths were constrained using the LINCS algorithm[72] for the protein and the SETTLE algorithm[73] for the water.

After solvation each system was energy-minimized and the solvent relaxed for 10 ps with position restraints (1000 kJ mol^{-1} nm^{-2}) applied to all heavy atoms of the protein. When necessary explicit counterions were added to ensure electroneutrality of the simulation box, and the resulting neutralized system was energy-minimized again. The system was then simulated for 10 ps at constant pressure and temperature, with the restraining potentials applied on Cα atoms only. Finally the system was simulated for 100 ns without any restraints. The last 60 ns were used in the analyses.

ELNEDIN Models and Simulations. The current version of ELNEDIN is based on the version 2.1 of the MARTINI molecular force field.[22,28] However, because we chose to use the position of the Cα atom to place the backbone bead, instead of that of the center of mass of the −N, Cα, C, O− main chain atoms, as in MARTINI, several minor modifications were introduced. These modifications, which only involve changes in the bonded interactions and the structural mapping of aromatic residues from AT to coarse-grained (CG) representation, are detailed in the Supporting Information section. The nonbonded interactions described in MARTINI 2.1 were not modified.

The EN scaffold component of the ELNEDIN models was built only across the CG backbone beads (i.e., Cα atoms). Two backbone beads were linked by a spring with force constant, K_{SPRING}, only if the distance between them in the experimental structure was less than a predefined cutoff value, R_C, and if they were separated by at least two positions in the protein sequence (see Supporting Information for details on how bonded interactions between sequential, $i \rightarrow i + 1$ and $i \rightarrow i + 2$ residues were described). For a given ELNEDIN model, the values of R_C and K_{SPRING} are identical for all pairs of backbone beads. The equilibrium length of a given spring was set to the experimentally observed distance between the two Cα atoms that it connects.

In the simulations using an ELNEDIN model, the temperature and pressure were treated as in the AT simulations, with $\tau_T = 0.5$ ps and $\tau_P = 1.2$ ps. The nonbonded interactions were treated with a switch function from 0.0 to 1.2 nm for the Coulomb interactions, and from 0.9 to 1.2 nm for the Lennard-Jones interactions, conform the standard MARTINI protocol. The integration time-step was set to 20 fs and the neighbor list was updated every 5 steps.

Each ELNEDIN system was solvated in a cubic box with a minimum of 1.2 nm between any protein bead and the edge of the box. After energy-minimization with position restraints (1000 kJ mol^{-1} nm^{-2}) applied to all protein beads, a 50 ps MD simulation using a 1 fs time-step (real times) with the same position restraints was used to relax both the solvent molecules and the protein in the force field. The system was further relaxed with a 1 ns long MD run with a 20 fs time-step and position restraints on the protein "backbone" beads. Finally each system was simulated for production without any restraints. The length of the various production runs is reported in the text.

Note that because of the smoothing of the energy surface in the CG model the time scales are generally faster. Typically a standard conversion factor of 4 is used, corresponding to the effective speed up factor in the diffusion dynamics of CG water compared to real water taking into account that the CG water represents 4 real waters.[21] The CG simulation times reported in the Results are thus effective times (4 × simulation-time) noted with an asterisk (*), unless otherwise stated.

Comparison of ELNEDIN and Atomistic Models. To evaluate the effect of the properties of the EN scaffold on the structural and dynamical properties of a protein modeled using an ELNEDIN representation, the values of R_C and K_{SPRING} were varied systematically in the range 10−10000 kJ mol^{-1} nm^{-2} and 0.6−1.2 nm, respectively. AT simulations were used as benchmark. The comparison of ELNEDIN and AT simulations was based on four structural and dynamical quantities. These quantities were computed from the equilibrated portions of the MD trajectories. The last 60 ns of 100 ns-long MD simulations were used in the case of AT runs,

and the last 15 ns (60 ns*) of 20 ns-long MD simulations for the CG runs. The degree of motion and fluctuation is time-scale dependent, it is therefore important that the length of the trajectories used in the comparison is the same for both models. The four quantities were (1) the time average of the Cα root-mean-square deviation, RMSD, (2) the root-mean-square deviation per residue, RMSD_res, (3) the root-mean-square fluctuation per residue, RMSF_res, and (4) the essential subspace (first ten eigenvectors of the covariance matrix of positional fluctuations; see below and references[7,74−77]). The RMSD and RMSD_res quantify the global and local structural deformation from the experimental structure, while the RMSF and essential subspace characterize the local fluctuations (deviation from the mean) and the direction of the large-amplitude fluctuations of the biomolecule, respectively.

Four similarity indices were defined. The ΔRMSD index was defined as the absolute difference between the average values of RMSD of the protein in the two approaches

$$\Delta\mathrm{RMSD} = |\langle \mathrm{RMSD}\rangle^{\mathrm{AT}}_{\mathrm{last60ns}} - \langle \mathrm{RMSD}\rangle^{\mathrm{ELNEDIN}}_{\mathrm{last60ns*}}|$$

The index of similarity for comparing RMSD_res and RMSF_res values obtained in CG and AT models was defined as

$$\Delta\mathrm{RMSX\_res} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{RMSX\_res}_i^{\mathrm{AT}} - \mathrm{RMSX\_res}_i^{\mathrm{ELNEDIN}})^2}$$

with RMSX representing either RMSD or RMSF, and N the number of residues in a protein.

And finally, the similarity between the essential subspaces obtained from AT and CG models was quantified by computing the root-mean-square inner-product (RMSIP) between the first 10 first eigenvectors in each simulation[76−79]

$$\mathrm{RMSIP} = \sqrt{\frac{1}{10}\sum_{i=1}^{10}\sum_{j=1}^{10}(\eta_i^{\mathrm{AT}}\cdot\eta_j^{\mathrm{ELNEDIN}})^2}$$

where η_i^{AT} and $\eta_j^{\mathrm{ELNEDIN}}$ are the ith and jth eigenvectors obtained from the AT and ELNEDIN simulation, respectively. For a given system and simulation the essential subspace was computed by diagonalizing the covariance matrix of positional fluctuations $[C_{ij}]_{i,j\in\{1,\ldots,3N\}^2}$ whose elements are given by

$$C_{ij} = \langle (q_i - \langle q_i\rangle)\cdot(q_j - \langle q_j\rangle)\rangle$$

where q_i is one of the Cartesian coordinates of one of the Cα atoms in the molecule and $\langle q_i\rangle$ is the corresponding average value over the ensemble of configurations considered for analysis.

Results and Discussion

Components of an ELNEDIN Model. Figure 1 illustrates the various components that make up an ELNEDIN model. The rationale underlying ELNEDIN is to combine a structure-based coarse-grained model, such as an elastic network,[39] with a physics-based coarse-grained (CG) molecular force field to represent a protein. The elastic network acts as a structural scaffold while the force field directs intermolecular interactions. It should be noted that Tozzini and McCammon have followed a similar rationale using Morse potentials to introduce a bias toward the native structure and to supplement a statistics-based force field.[24] Here, we use a simple two-parameter, R_C/K_{SPRING}, elastic network. In addition to this elastic network scaffold, each amino acid is geometrically represented and "typified" (Lennard-Jones and Coulomb potential parameters) according to the MARTINI force field.[22,28] (Slight modifications to the latest version of the MARTINI force field were needed, see Methods, these changes are described in detail in the Supporting Information). It is important to note that backbone beads linked with springs do not interact with each other via nonbonded potentials (Lennard-Jones and Coulomb) because bonded beads are excluded from the nonbonded interaction lists. Thus, the choice of values for R_C and K_{SPRING} will directly affect the extent to which either bonded or nonbonded potentials contribute to the internal dynamics of a given protein model.

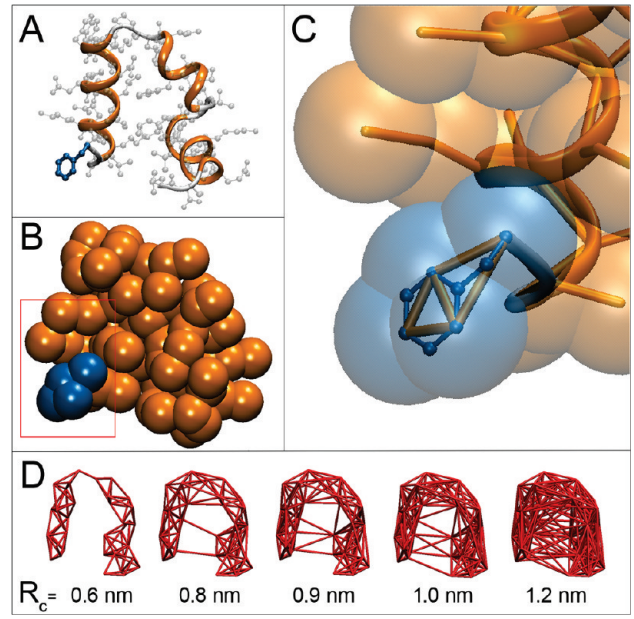

Figure 1. Components of the ELNEDIN model of the villin headpiece subdomain. (A) Ribbon and ball and stick representations of the villin-headpiece subdomain (PDB entry 1YRF[62]). The N-terminal phenylalanine residue is highlighted in blue. (B) CPK representation of a coarse-grained model of the villin-headpiece subdomain based on the modified MARTINI force field. The N-terminal phenylalanine is highlighted in blue. (C) Close-up view of the N-terminal phenylalanine with details of its all atom (blue) and CG (gold) bonding network. (E) Five elastic network scaffolds of the villin-headpiece subdomain built (from left to right) with R_C = 0.6, 0.8, 0.9, 1.0, and 1.2 nm, respectively. All graphics were created using the visualization software VMD.[97]

Influence of the Properties of the Elastic Network on the Structure and Dynamics of a Protein. To evaluate the influence of the scaffold parameters, R_C and K_{SPRING}, on the structure and the dynamics of a protein we first focused on the B1 domain of protein G,[60] a protein whose conformational

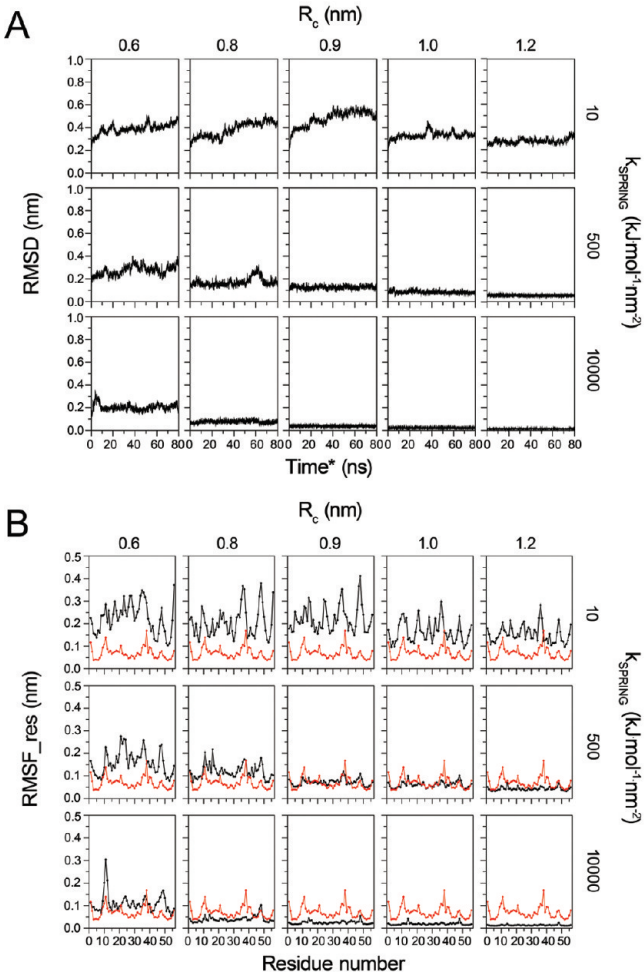

Figure 2. Effect of K_{SPRING} and R_C values on the structure and dynamics of the B1 domain of protein G. (A) Root-mean-square deviation from the experimental structure as a function of effective time. (B) Root-mean-square fluctuation of the backbone bead of each residue as a function of residue number (black curve). The root-mean-square fluctuation of the Cα atoms calculated from an all-atom MD simulation trajectory is shown in red to illustrate the similarities and the differences between the two approaches.

properties we have studied in detail previously.[77] Fifteen different scaffolds were built by varying R_C from 0.6 to 1.2 nm, and K_{SPRING} from 10 to 10000 kJ mol^{-1} nm^{-2}.

The main results of these computational experiments are summarized in Figure 2. The root-mean-square deviation (RMSD) of the backbone beads with respect to their crystallographic position plotted as a function of time showed that the global deformation of the protein decreases when increasing both R_C and K_{SPRING} values (Figure 2A). The deformation of the protein can exceed 0.5 nm with a flexible and undersized scaffold (small values of R_C and K_{SPRING}, upper left panel Figure 2A) and can be as low as 0.01 nm with the stiffest and most extended variant (large values of R_C and K_{SPRING}, lower right panel Figure 2A). Intermediate deformations are observed not only with intermediate values of R_C and K_{SPRING} but also with combinations of short/strong and long/weak values for R_C and K_{SPRING}, respectively. This behavior indicates that R_C and K_{SPRING} compensate each other to maintain the overall structure of the protein.

A similar behavior is observed when monitoring the root-mean-square fluctuation of each residue (RMSF_res) with respect to its average position (Figure 2B). However, it can also be seen that the pattern of the residue-fluctuations (distribution of peaks and valleys across the sequence) varies significantly in some cases between one set of R_C and K_{SPRING} parameters and another. These variations indicate that notwithstanding the compensating effect of the scaffold parameters in relation to overall deformation and amplitude of fluctuations, accurate fluctuation patterns that would agree with experimental B-factors or residue fluctuations computed from atomistic (AT) simulations may only be achieved for specific combinations of the scaffold parameters. This effect can be observed in Figure 2B by comparing the AT (red) and CG (black) fluctuations that show that only the parameters sets $R_C = 0.9$ or 1.0 nm with $K_{SPRING} = 500$ kJ mol^{-1} nm^{-2} provide reasonable overlap between the CG and AT models. To address this issue more in depth, we carried out a systematic comparison between several structural and dynamical properties computed from MD simulations using AT and ELNEDIN representations using three distinct proteins each representing a different structural class.

Comparison of Structural and Dynamical Properties Computed from MD Simulations using AT and ELNEDIN Representations. The B1 domain of protein G, the villin headpiece subdomain, and the α-spectrin SH3 domain, were used for the comparison of AT and ELNEDIN models. These proteins belong to different structural classes: the B1 domain belongs to the $\alpha + \beta$ class, the villin headpiece is an all-α protein, and the SH3 domain belongs to the all-β class. The small size of these proteins allowed relatively long (~100 ns) MD simulations to be performed in an acceptable amount of time with an atomistic (AT) representation. It is important to note that limiting ourselves to small proteins does not affect the generality of our study because ultimately we do not intend to scaffold large proteins with a single EN but with discontinuous or independent elastic networks instead (see below).

For each protein, MD simulations using an AT representation were carried for 100 ns, and MD simulations using an ELNEDIN representation were carried out for 20 ns (corresponding to an effective time of 80 ns*). The parameters for the EN scaffold were varied systematically with R_C (nm) ∈ {0.6, 0.8, 0.9, 1.0, 1.2} and K_{SPRING} (kJ mol^{-1} nm^{-2}) ∈ {10, 50, 100, 200, 500, 1000, 2000, 5000, 10000}. Thus for each protein a total of 45 MD simulations were performed using an ELNEDIN representation.

Four physical quantities were computed from each MD trajectory: (1) the time-average root-mean-square deviation (RMSD) of the backbone beads (Cα atoms), which quantifies the global deformation of the protein with respect to the experimental model, (2) the root-mean-square deviation of the backbone beads per residue (RMSD_res) which quantifies the structural deformation (deviation from the initial structure) of each amino acid, (3) the root-mean-square fluctuation of the backbone beads per residue (RMSF_res) which measures the fluctuation (deviation with respect to the mean position) of each residue, and (4) the large-amplitude collective motions of each protein system. The collective

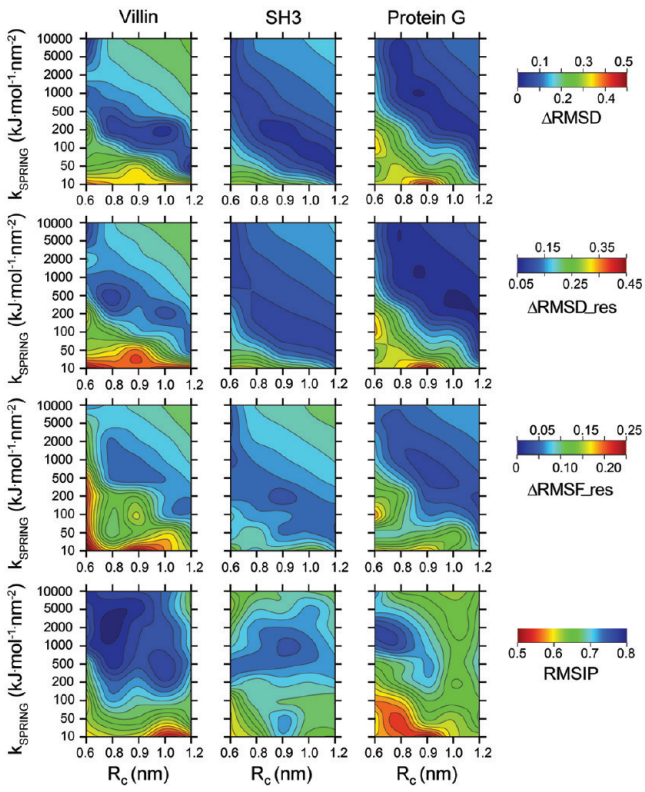

Figure 3. Comparing ELNEDIN and AT representations. The values of ΔRMSD, ΔRMSD_res, ΔRMSF_res, and RMSIP are reported for each of the three model proteins. For all four indices of similarity the color-coding ranges from red to blue, so that regions of low similarity between ELNEDIN and AT models always appear in red, while regions of high similarity are colored in blue. Note that low values for ΔRMSD, ΔRMSD_res, ΔRMSF_res indicate high similarity but that low RMSIP values indicate low similarity.

motions were computed by essential dynamics analysis.[74,75] The comparison between ELNEDIN and AT was quantified using four similarity indices: ΔRMSD, ΔRMSD_res, ΔRMSF_res, and RMSIP (see Methods). The values of these indices are reported on 2D contour maps in R_C/K_{SPRING} space (Figure 3). The values of the indices are color-coded so that the similarity between the ELNEDIN and AT simulations increased from red to blue for all four indices considered. Note that for ΔRMSD, ΔRMSD_res, and ΔRMSF_res the lower the value (minimum = 0 nm) the better the agreement between ELNEDIN and AT models; but for RMSIP the higher the value (maximum = 1) the better the correlation between ELNEDIN and AT models. Similarity indices below ~0.15 nm (cyan to blue) were considered good for ΔRMSD and ΔRMSD_res. The corresponding cutoff values for ΔRMSF_res and RMSIP were set at ~0.075 nm or below, and 0.75 or above, respectively.

The 2D contour maps for the ΔRMSD, ΔRMSD_res and ΔRMSF_res indices confirmed the compensatory relationship between R_C and K_{SPRING}. Indeed, for these three indices, the EN scaffolds which provide the best agreement between ELNEDIN and AT models (blue regions in the corresponding panels) are clearly distributed diagonally across each of the corresponding 2D contour map (Figure 3). This is the case for all three proteins and suggests that the compensatory behavior is independent of the structural class of the protein. This result is consistent with the work of Tirion who reported a similar inverse relationship between these two parameters in order to maximize the agreement between EN and classical normal-mode analysis.[39] It is noteworthy, however, that the width of the diagonal region within which quantitative agreement (cyan to blue contours) between ELNEDIN and AT is achieved, varies from one protein to the other, suggesting that optimal values for R_C and K_{SPRING} may depend on the specific protein.

Such a protein specific behavior is even more marked in the case of the RMSIP index. For this index, the region of R_C and K_{SPRING} values for which the ELNEDIN model agrees with the AT model is confined to a specific perimeter, which is not diagonal (i.e., the effects of R_C and K_{SPRING} do not compensate one another) and whose extent and shape differs for each protein (compare the blue regions across bottom three panels in Figure 3). The reason why a protein specific behavior appears readily in the RMSIP index is that the computation of this latter index is more sensitive to the variation of the measured property (in this case the positional fluctuation) across the sequence than the other indices. Indeed, in the case of the ΔRMSD, ΔRMSD_res and ΔRMSF_res indices these sequence variations are averaged out, and therefore not as readily detectable, although present (see the distribution of peaks and valleys of the residue fluctuations of the B1 domain of protein G in Figure 2B). Thus, taken together, these results suggest that while the compensatory mode in which R_C and K_{SPRING} influence the structural and dynamical properties may be independent of the specific protein being modeled, the region in the R_C/K_{SPRING} space that gives the best quantitative agreement between ELNEDIN and AT models is protein specific.

Nevertheless, a region of R_C/K_{SPRING} space in which ELNEDIN models provide adequate quantitative structural and dynamical agreement with AT models could be delineated (data not shown). A search for a consensus set of parameters across the various indices and protein systems revealed that values within 0.8 and 1.0 nm for R_C and ranging from 500 to 1000 kJ mol^{-1} nm^{-2} for K_{SPRING} could provide adequate quantitative agreement with atomistic simulations. These values are close to the values of R_C = 0.7 nm and K_{SPRING} = 1000 kJ mol^{-1} nm^{-2} used by Bond and co-workers.[58,59] The values are also within the range used in typical EN applications,[39−41,43−45,72,80,81] which range from 0.7 to 1.6 nm for R_C and from 200 to 4000 kJ mol^{-1} nm^{-2} for K_{SPRING}. It is important to note that in EN-based normal-mode analysis the spring force constant is a parameter that is usually adjusted a posteriori in order to match the amplitude of the fluctuations with experimental B-factors for example. In that case the value of the spring force constant only affects the amplitude of the motions, not their directions. EN models are therefore often referred to as single parameter (R_C) models. This is not the case for an ELNEDIN model. Both R_C and K_{SPRING} contribute to the accuracy of the model. This is because, as stated earlier, beads connected via springs do not interact with each other via nonbonded potentials. Thus an increase (respectively decrease) in the value of R_C will

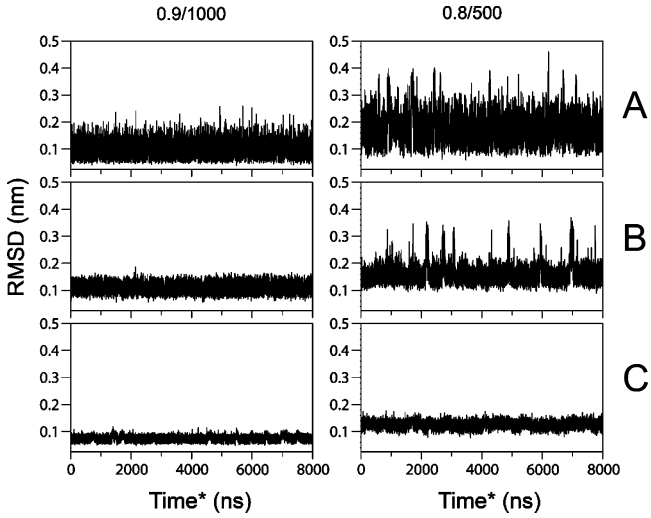

Figure 4. Long time-scale simulations using two distinct scaffolds. The RMSD vs time* of the three protein models relative to their respective experimental structure during long ELNEDIN simulations is shown. (A) The villin headpiece subdomain. (B) The B1 domain of protein G. (C) the src SH3 domain. The values of the R_C (nm) and K_{SPRING} (kJ mol^{-1} nm^{-2}) constants used in the simulations are indicated above the panels: left panel simulations use a 0.9/1000 R_C/K_{SPRING} combination, while the right panel simulations use a 0.8/500 combination.

modify the potential energy in two ways: (1) it will reduce (respectively increase) the number of bead pairs that can interact via nonbonded potentials and (2) increase (respectively reduce) the number of bead pairs that will interact via harmonic springs (K_{SPRING}). This interplay between the elastic network and the nonbonded potentials of the CG force field has the appealing consequence of introducing anisotropy in the description of the protein, akin to that sought in anisotropic network models,[41,46,81] despite the use of an isotropic network.

Long Time-Scale Behavior of ELNEDIN Models. To evaluate the behavior of ELNEDIN models during long MD simulations, each of the three test proteins, the villin headpiece subdomain, the B1 domain of protein G, and the α-spectrin SH3 domain, was simulated for 8 μs*. Two distinct combinations of spring force constant and cutoff radius were used: (i) $R_C = 0.9$ nm, $K_{SPRING} = 1000$ kJ mol^{-1} nm^{-2} (0.9/1000) and (ii) $R_C = 0.8$ nm, $K_{SPRING} = 500$ kJ mol^{-1} nm^{-2} (0.8/500). The time series of the RMSD of each protein from its experimental structure are shown in Figure 4. Overall, the plots indicate that in all cases the native fold is well preserved (average RMSD value ≤0.2 nm). This is expected since we are using an EN scaffold for that very purpose. But the plots also reveal how the quality of the scaffold might influence the ability of a given system to experience transient structural transitions during a simulation. This effect is clear in the case of the villin headpiece subdomain and the B1 domain of protein G, which experience more conformational transitions when modeled with the 0.8/500 scaffold than with the 0.9/1000 one. Visual inspection of the trajectories indicated that these conformational changes involved changes in orientation of secondary structure elements. These transitions had a lifetime on the order of 10−100 ns, and the protein always returned to the native state. This behavior was not as marked for the SH3 domain in which only very small amplitude conformational transitions were observed. The smaller amplitude of the conformational transitions in the SH3 system may be due to its higher degree of compactness with respect to the other systems (e.g., for a given cutoff value SH3 has the largest average number of springs per residue followed by protein G and then villin). Taken together, these data indicate that despite the intrinsic bias that an EN scaffold might introduce toward the native structure it does not preclude changes in the relative orientation of secondary structure elements. However the ability to observe such transitions depends on the choice of EN parameters.

Table 1. Correlation Coefficients between NMR- and CG-Based Residue Fluctuations[a]

R_C/K_{SPRING}	Villin headpiece subdomain	B1 domain of protein G[d]	SH3 domain
0.8/500	0.71[b] − 0.85[c]	0.73	0.75[e]−0.86[f]
0.9/1000	0.80[b]	0.62	0.68[e]

[a] Only Cα atoms were used to compute the fluctuations. [b] The NMR ensemble of structures was taken from entry 2JMO[98] in the PDB. Note that this entry contains a fluorinated residue. [c] The correlation coefficient with respect to the fluctuations computed from a 4 μs*-long simulation with $R_C/K_{SPRING} = 0.8/500$ of the wild-type NMR structure (PDB entry 1VII[99]) is 0.85. [d] The NMR ensemble structures was taken from entry 1GB1[100] in the PDB. [e] The ensemble of NMR structures was taken from entry 1AEY[101] in the PDB. This entry corresponds to the wild-type sequence protein and not the mutant sequence used here. [f] The correlation coefficient with respect to the fluctuations computed from a 4 μs*-long simulation with $R_C/K_{SPRING} = 0.8/500$ of the wild-type structure (PDB entry 1SHG[102]) is 0.86.

To ascertain that the dynamic behavior observed for each protein is inline with experimental data we compared the residue-based fluctuations obtained from each long-time-scale simulation to those computed from the NMR ensemble of the cognate structure deposited in the PDB. The correlation coefficients between the two sets of values are reported in Table 1. Overall the data shows good agreement with experimentally derived fluctuations with correlation coefficients ranging from 0.62 to 0.86. These values indicate that the pattern of fluctuations across the sequence obtained from the simulation was very similar to that obtained from the ensemble of structures satisfying the NMR restraints. Note that the level of agreement between the experimental and ELNEDIN data depends, to a large extent, on the quality of the underlying atomistic simulations from which the optimal choice of EN parameters was inferred.

Application of ELNEDIN to Large Macromolecular Assemblies. We chose to model the viral capsid of the Cowpea Mosaic Virus (CPMV PDB entry 1NY7[63]). The viral capsid of CPMV is a highly symmetrical assembly consisting of 60 copies of two proteins having 190 and 369 residues, respectively. This assembly is almost spherical in shape with a ∼26 nm diameter (Figure 5A). After solvation, the system contained 268,883 CG beads, which would correspond to 2,852,940 atoms if an atomistic representation were used.

Since a single EN scaffold might adversely affect the dynamics of the viral capsid, for example, by interfering with

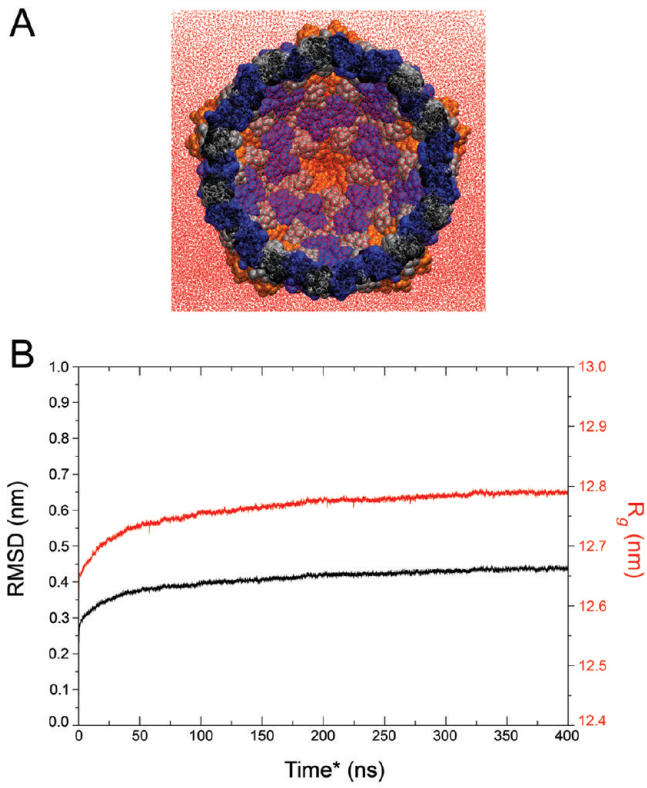

Figure 5. Modeling the Cowpea Mosaic virus (CPMV). (A) The viral capsid of CPMV is shown sliced across the middle. The red dots are solvent molecules; the S viral protein is shown in orange and the L viral protein in gray and blue for aesthetic reasons only. (B) Plot of the RMSD from the initial configuration and radius of gyration (R_g) of the capsid vs time*.

changes in the relative orientations of the protein domains with respect to each other, EN scaffolds were built for each protein domain separately. But each EN scaffold was built using the same values of R_C and K_{SPRING}, that is, 0.9 nm for R_C and 500 kJ mol^{-1} nm^{-2} for K_{SPRING}. Note that beads in distinct proteins interact according to the nonbonded terms of the MARTINI force field. The overall ELNEDIN model was simulated for 400 ns* at 1 atm and 300 K.

The time series of the RMSD and the radius of gyration (R_g) are shown in Figure 5B. Both parameters indicate that the CPMV capsid is structurally stable. The overall RMSD, which remains below ~0.5 nm throughout the simulation, is remarkably small considering the size of the system. Moreover, the viral capsid did not show any symptom of collapse as has been observed for similar systems using an atomistic force field[82] or a more approximate level of coarse-graining than the one used here.[83] This suggests that large macromolecules can be modeled effectively as assemblies of independent ELNEDIN models. One reason for this apparent success might be that maintaining the structural integrity of each subunit independently may have contributed favorably to the overall stability of the viral capsid. This effect may need to be investigated further if reliable mechanistic inferences are sought concerning CPMV. However, this is beyond the scope of the current study. Finally, it is interesting to note that it took about 50 ns for the viral capsid to reach a reasonably stable value for the RMSD and R_g, suggesting that for such large systems the period of relaxation is significantly longer than for smaller proteins.

Application of ELNEDIN to Protein−Protein Association. As a last application, we evaluated the ability of ELNEDIN models to be used for the study of protein−protein association processes. A mutant (PDB entry 1RPO[65]) of the repressor of primer protein (ROP) was used for this test. ROP is a homo dimeric four-helix bundle,[64] each monomer consisting of two antiparallel α-helices. ELNEDIN models were prepared for the native dimer configuration as well as for three other configurations in which the monomers were placed at 0.5 nm (TX = 0.5), 1.0 nm (TX = 1.0), and 1.5 nm (TX = 1.5) from each other by simple translation in the direction perpendicular to the plane defined by the monomer−mononer interface. The relative orientation of the two monomers was not modified. It would have added the difficult task of sampling the conformational space, which is beyond the scope of the present test. For each starting configuration, native, TX = 0.5, TX = 1.0, and TX = 1.5, five independent MD simulations (400 ns*) were carried out by using different sets of initial velocities. Two slightly distinct EN scaffolds were tested: one with R_C = 0.9 nm and K_{SPRING} = 500 kJ mol^{-1} nm^{-2} (0.9/500), and the other with R_C = 1.0 nm and K_{SPRING}= 1000 kJ mol^{-1} nm^{-2} (1.0/1000). Note that while each monomer model used the same R_C and K_{SPRING} values for the EN scaffold, each monomer possessed its own separate scaffold and was thus free to move independently from the other monomer during the MD simulation.

The MD simulations of the native dimer configuration showed that the four-helix bundle could adopt two distinct conformations. These are marked native 1 and native 2 in Figure 6A. The native 2 conformation differs from the native 1 (experimental structure) in the degree by which the monomers are tilted with respect to each other. The significance of this second orientation is not clear but it is noteworthy that similar but less pronounced changes have also been observed in atomistic simulations of ROP.[78] The two native conformations were observed for both EN scaffolds. When simulating the translated systems formation of either the native 1 or native 2 states was taken as a successful reassembly event.

The results of the various monomer association tests were as follows. Monomers that were placed 0.5 nm apart (TX = 0.5 runs) were always able to reassemble into a native structure, and this was true for both EN scaffolds. TX = 1.0 systems were also able to reassemble into a native structure but in this case the more flexible scaffold performed better than the more rigid one: 5 out of 5 successful reassemblies were obtained for the 0.9/500 scaffold vs 2 out of 5 for the 1.0/1000 scaffold. The same trend but with less success overall was observed for the TX = 1.5 systems: 2 out of 5 runs produced a native structure when using the 0.9/500 scaffold vs only 1 out of 5 runs produced a native structure when using the 1.0/1000 scaffold. Runs that failed to produce a native structure after 400 ns* of MD simulation all showed that the monomers had assembled into non-native states burying sometimes as much surface area as the native structure (data not shown). No attempt was made to study

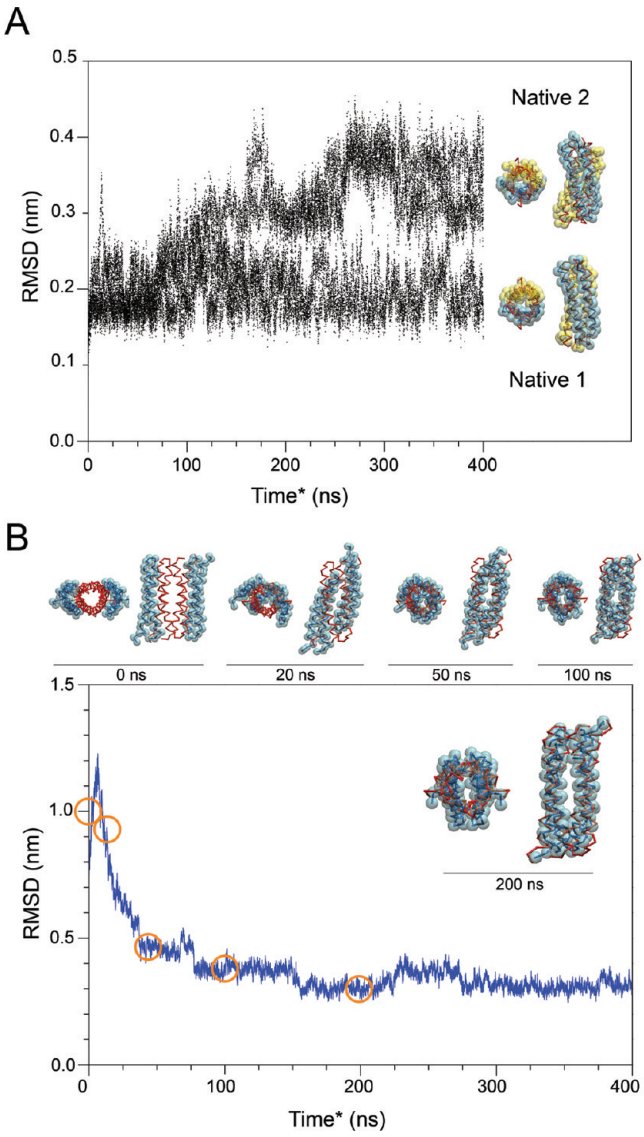

Figure 6. Modeling association of ROP monomers. (A) RMSD plots vs time for the simulations of the native dimer reveal two stable conformational states. (B) Structural intermediates observed in the course of one of the protein association runs and corresponding RMSD trace. The individual monomers (in blue) were initially placed at a distance of 1.5 nm from each other. The trace of the experimental structure is shown in red. In these representations the solvent is omitted for clarity.

these non-native systems, which would be considered false positives in a docking experiment, in more detail.

Taken together, these results show that the quality of the scaffold affects the ability of the monomers to reassemble into a native structure. This is not surprising since the stability of a given structure (be it tertiary[84] or quaternary in this case) is likely to arise from a balance between local/internal and long-range interactions. It is noteworthy that the more successful scaffold is also the more flexible one (0.9/500 vs 1.0/1000). This finding is consistent with the recognized importance of flexibility in molecular recognition[85−87] and underlines the necessity to include an accurate description of a protein's internal dynamics when attempting to predict protein−protein interactions.[88−93]

Visual inspection of the reassembly process in the various successful runs revealed the interesting fact that the sequence of conformational rearrangements that followed the initial encounter and led to the final native structure was not unique. In fact, reassembly proceeded in a different manner in each case. Such multiplicity of pathways has been observed by others[94,95] and has led to the hypothesis that protein association like protein folding also proceeds on a funnel-shaped free-energy landscape.[96] No effort was made to analyze these pathways further since many more simulations would be needed to establish statistically significant results. Nevertheless, for illustration purposes only, one such pathway is depicted in Figure 6B where the time series of the RMSD value of the complex relative to the experimental structure is shown together with representative structures at the different stages of the association process. The figure shows that the two monomers rapidly came into contact with each other. The packing of the core is out-of-register (see side-view at 20 ns*) and is shifted laterally (see top-view) with respect to the native state (shown in red). Following this encounter the process of association took ∼130 ns* and could be described as one monomer slithering along the length of the other monomer until the stable native interface (native 2) was reformed. The complex then remained stable for the following 250 ns* of the simulation with an RMSD value of 0.3 nm from the experimental structure.

Overall the result of these, albeit limited, docking experiments are extremely encouraging and suggests that the combination of a structure-based and a physics-based CG model such as ELNEDIN might provide a valuable approach to the difficult task of predicting protein−protein interfaces and association processes.

Conclusion

We have investigated the possibility to combine a structure-based and physics-based coarse-grained model into a unique representation. An elastic network was mixed with the MARTINI 2.1 force field for the purpose of carrying out molecular dynamics simulations of biological systems. The effect of varying the properties of the EN on the quality of the model was studied systematically. The results of these computational studies show that it is possible to identify appropriate values for the EN parameters, R_C and K_{SPRING}, such that the resulting ELNEDIN model is capable to reproduce simultaneously the global and local deformations of a protein, its residue fluctuations, and its large-amplitude collective motions, as observed in atomistic models. Although, the optimal values for R_C and K_{SPRING} depend on the specific protein studied, we find that values ranging from 0.8 to 1.0 nm for R_C and from 500 to 1000 kJ mol^{-1} nm^{-2} for K_{SPRING} can provide adequate quantitative agreement with atomistic simulations. The results also show that ELNEDIN models are stable enough to allow microsecond time-scale molecular dynamics simulations to be carried out readily. In some cases, transient structural changes corresponding to changes in orientation of elements of secondary structure can be observed during these simulations. But it should be noted that because of the inherent structural bias of EN toward

the reference configuration, the model cannot be expected to produce conformational changes akin to those necessary for a protein to fold. Finally, we find that large biological entities such as the viral capsid of the cowpea mosaic virus can be stably modeled as assemblies of independent ELNEDIN models, and that ELNEDIN models show significant promise for modeling protein association processes.

Abbreviations. AT, atomistic; CG, coarse-grained; EN, elastic network; MD, molecular dynamics; NMR, nuclear magnetic resonance; RMSD, root-mean-square deviation; RMSD_res, root-mean-square deviation per residue; RMSF_res root-mean-square fluctuation per residue; RMSIP, root-mean-square inner-product. CPMV, cowpea mosaic virus; ROP, repressor of primer.

Acknowledgment. The authors acknowledge funding from the CUNY Research Foundation and a CUNY Incentive Collaborative Grants to MAC, as well as from The Netherlands Organization for Scientific Research (NWO) for SJM and XP. The authors also thank the CUNY High performance computing facility (CUNY HPC) and The Netherlands National Computing Facilities (NCF) for allocation of computing time.

Supporting Information Available: Description of the protocol used to extract the parameters for angle bending and bond stretching potentials (other than elastic network bonds) used in ELNEDIN models, description of the structural mapping (from AT to CG) of aromatic residues implemented in ELNEDIN models, and the complete set bonded parameters for amino acid residues as used in this study (note that the nonbonded parameters were taken without modification from the MARTINI force field). This set of parameters should not be considered as a new version of the MARTINI force field but as an alternative to MARTINI-2.1. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tozzini, V. Coarse-Grained Models for Proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150.

(2) de Vlieg, J.; Berendsen, H. J. C.; van Gunsteren, W. F. An NMR-Based Molecular-Dynamics Simulation of the Interaction of the Lac Repressor Headpiece and Its Operator in Aqueous-Solution. *Proteins* **1989**, *6*, 104–127.

(3) Paci, E.; Vendruscolo, M.; Dobson, C. M.; Karplus, M. Determination of a Transition State at Atomic Resolution from Protein Engineering Data. *J. Mol. Biol.* **2002**, *324*, 151–163.

(4) van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. Molecular Simulation as an Aid to Experimentalists. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.

(5) Marin, E. P.; Krishna, A. G.; Sakmar, T. P. Rapid Activation of Transducin by Mutations Distant from the Nucleotide-Binding Site - Evidence for a Mechanistic Model of Receptor-Catalyzed Nucleotide Exchange by G Proteins. *J. Biol. Chem.* **2001**, *276*, 27400–27405.

(6) Marin, E. P.; Krishna, A. G.; Sakmar, T. P. Disruption of the α5 Helix of Transducin Impairs Rhodopsin-Catalyzed Nucleotide Exchange. *Biochemistry* **2002**, *41*, 6988–6994.

(7) Ceruso, M. A.; Periole, X.; Weinstein, H. Molecular Dynamics Simulations of Transducin: Interdomain and Front to Back Communication in Activation and Nucleotide Exchange. *J. Mol. Biol.* **2004**, *338*, 469–481.

(8) Liwo, A.; Czaplewski, C.; Oldziej, S.; Scheraga, H. A. Computational Techniques for Efficient Conformational Sampling of Proteins. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134–139.

(9) Christen, M.; van Gunsteren, W. F. On Searching in, Sampling of, and Dynamically Moving through Conformational Space of Biomolecular Systems: A Review. *J. Comput. Chem.* **2008**, *29*, 157–166.

(10) Ayton, G. S.; Noid, W. G.; Voth, G. A. Multiscale Modeling of Biomolecular Systems: In Serial and in Parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.

(11) Elber, R. Long-Timescale Simulation Methods. *Curr. Opin. Struct. Biol.* **2005**, *15*, 151–156.

(12) Tai, K. Conformational Sampling for the Impatient. *Biophys. Chem.* **2004**, *107*, 213–220.

(13) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.

(14) Schlick, T. Time-Trimming Tricks for Dynamic Simulations: Splitting Force Updates to Reduce Computational Work. *Structure* **2001**, *9*, R45–53.

(15) Flory, P. J. Statistical Thermodynamics of Random Networks. *Proc. R. Soc. A* **1976**, *351*, 351–380.

(16) Levitt, M. A Simplified Representation of Protein Conformations for Rapid Simulation of Protein Folding. *J. Mol. Biol.* **1976**, *104*, 59–107.

(17) Chan, H. S.; Dill, K. A. Intrachain Loops in Polymers—Effects of Excluded Volume. *J. Chem. Phys.* **1989**, *90*, 492–509.

(18) Voth, G. A., Ed. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: New York, 2009.

(19) Smit, B.; Hilbers, P. A. J.; Esselink, K.; Rupert, L. A. M.; van Os, N. M.; Schlijper, A. G. Computer Simulations of a Water/Oil Interface in the Presence of Micelles. *Nature* **1990**, *348*, 624–625.

(20) Saiz, L.; Klein, M. L. Computer Simulation Studies of Model Biological Membranes. *Acc. Chem. Res.* **2002**, *35*, 482–489.

(21) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108*, 750–760.

(22) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The Martini Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

(23) Zuckerman, D. M. Simulation of an Ensemble of Conformational Transitions in a United-Residue Model of Calmodulin. *J. Phys. Chem. B* **2004**, *108*, 5127–5137.

(24) Tozzini, V.; McCammon, J. A. A Coarse Grained Model for the Dynamics of Flap Opening in HIV-1 Protease. *Chem. Phys. Lett.* **2005**, *413*, 123–128.

(25) Tozzini, V.; Rocchia, W.; McCammon, J. A. Mapping All-Atom Models onto One-Bead Coarse-Grained Models: General Properties and Applications to a Minimal Polypeptide Model. *J. Chem. Theory Comput.* **2006**, *2*, 667–673.

(26) Pizzitutti, F.; Marchi, M.; Borgis, D. Coarse-Graining the Accessible Surface and the Electrostatics of Proteins for Protein−Protein Interactions. *J. Chem. Theory Comput.* **2007**, *3*, 1867–1876.

(27) Basdevant, N.; Borgis, D.; Ha-Duong, T. A Coarse-Grained Protein−Protein Potential Derived from an All-Atom Force Field. *J. Phys. Chem. B* **2007**, *111*, 9390–9399.

(28) 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.

(29) Tan, R. K.-Z.; Petrov, A. S.; Devkota, B.; Harvey, S. C. Coarse-Grained Models for Nucleic Acids and Large Nucleoprotein Assemblies. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: New York, 2009; Chapter 15, pp 225−235.

(30) Tepper, H. L.; Voth, G. A. A Coarse-Grained Model for Double-Helix Molecules in Solution: Spontaneous Helix Formation and Equilibrium Properties. *J. Chem. Phys.* **2005**, *122*, 124906.

(31) Knotts, T. A. T.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. A Coarse Grain Model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901–084912.

(32) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The Gromos Force-Field Parameter Sets 53a5 and 53a6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(33) Periole, X.; Huber, T.; Marrink, S. J.; Sakmar, T. P. G Protein-Coupled Receptors Self-Assemble in Dynamics Simulations of Model Bilayers. *J. Am. Chem. Soc.* **2007**, *129*, 10126–10132.

(34) Yefimov, S.; van der Giessen, E.; Onck, P. R.; Marrink, S. J. Mechanosensitive Membrane Channels in Action. *Biophys. J.* **2008**, *94*, 2994–3002.

(35) Treptow, W.; Marrink, S. J.; Tarek, M. Gating Motions in Voltage-Gated Potassium Channels Revealed by Coarse-Grained Molecular Dynamics Simulations. *J. Phys. Chem. B* **2008**, *112*, 3277–3282.

(36) Bahar, I.; Rader, A. J. Coarse-Grained Normal Mode Analysis in Structural Biology. *Curr. Opin. Struct. Biol.* **2005**, *15*, 586–592.

(37) Ma, J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* **2005**, *13*, 373–380.

(38) Tama, F.; Brooks, C. L. Symmetry, Form, and Shape: Guiding Principles for Robustness in Macromolecular Machines. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 115–133.

(39) Tirion, M. M. Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.* **1996**, *77*, 1905.

(40) Bahar, I.; Kaplan, M.; Jernigan, R. L. Short-Range Conformational Energies, Secondary Structure Propensities, and Recognition of Correct Sequence-Structure Matches. *Proteins* **1997**, *29*, 292–308.

(41) Doruker, P.; Atilgan, A. R.; Bahar, I. Dynamics of Proteins Predicted by Molecular Dynamics Simulations and Analytical Approaches: Application to α-Amylase Inhibitor. *Proteins* **2000**, *40*, 512–524.

(42) Song, G.; Jernigan, R. L. An Enhanced Elastic Network Model to Represent the Motions of Domain-Swapped Proteins. *Proteins* **2006**, *63*, 197–209.

(43) Kundu, S.; Melton, J. S.; Sorensen, D. C.; Phillips, G. N., Jr. Dynamics of Proteins in Crystals: Comparison of Experiment with Simple Models. *Biophys. J.* **2002**, *83*, 723–732.

(44) Sen, T. Z.; Feng, Y.; Garcia, J. V.; Kloczkowski, A.; Jernigan, R. L. The Extent of Cooperativity of Protein Motions Observed with Elastic Network Models Is Similar for Atomic and Coarser-Grained Models. *J. Chem. Theory Comput.* **2006**, *2*, 696–704.

(45) Kondrashov, D. A.; Van Wynsberghe, A. W.; Bannen, R. M.; Cui, Q.; Phillips, G. N. Protein Structural Variation in Computational Models and Crystallographic Data. *Structure* **2007**, *15*, 169–177.

(46) Lyman, E.; Pfaendtner, J.; Voth, G. A. Systematic Multiscale Parameterization of Heterogeneous Elastic Network Models of Proteins. *Biophys. J.* **2008**, *95*, 4183–4192.

(47) Durand, P.; Trinquier, G.; Sanejouand, Y. H. A New Approach for Determining Low-Frequency Normal Modes in Macromolecules. *Biopolymers* **1994**, *34*, 759–771.

(48) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. *Proteins* **2000**, *41*, 1–7.

(49) 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.

(50) Li, G.; Cui, Q. Analysis of Functional Motions in Brownian Molecular Machines with an Efficient Block Normal Mode Approach: Myosin-II and Ca^{2+}-ATPase. *Biophys. J.* **2004**, *86*, 743–763.

(51) Mitra, K.; Schaffitzel, C.; Shaikh, T.; Tama, F.; Jenni, S.; Brooks, C. L.; Ban, N.; Frank, J. Structure of the E-Coli Protein-Conducting Channel Bound to a Translating Ribosome. *Nature* **2005**, *438*, 318–324.

(52) Tama, F.; Brooks, C. L. 3rd Diversity and Identity of Mechanical Properties of Icosahedral Viral Capsids Studied with Elastic Network Normal Mode Analysis. *J. Mol. Biol.* **2005**, *345*, 299–314.

(53) He, J. B.; Zhang, Z. Y.; Shi, Y. Y.; Liu, H. Y. Efficiently Explore the Energy Landscape of Proteins in Molecular Dynamics Simulations by Amplifying Collective Motions. *J. Chem. Phys.* **2003**, *119*, 4005–4017.

(54) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. Nonlinear Elasticity, Proteinquakes, and the Energy Landscapes of Functional Transitions in Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12570–12575.

(55) Tatsumi, R.; Fukunishi, Y.; Nakamura, H. A Hybrid Method of Molecular Dynamics and Harmonic Dynamics for Docking of Flexible Ligand to Flexible Receptor. *J. Comput. Chem.* **2004**, *25*, 1995–2005.

(56) Miller, B. T.; Zheng, W.; Venable, R. M.; Pastor, R. W.; Brooks, B. R. Langevin Network Model of Myosin. *J. Phys. Chem. B* **2008**, *112*, 6274–6281.

(57) Zacharias, M. Combining Elastic Network Analysis and Molecular Dynamics Simulations by Hamiltonian Replica Exchange. *J. Chem. Theory Comput.* **2008**, *4*, 477–487.

(58) Bond, P. J.; Sansom, M. S. Insertion and Assembly of Membrane Proteins Via Simulation. *J. Am. Chem. Soc.* **2006**, *128*, 2697–2704.

(59) Bond, P. J.; Holyoake, J.; Ivetac, A.; Khalid, S.; Sansom, M. S. P. Coarse-Grained Molecular Dynamics Simulations of Membrane Proteins and Peptides. *J. Struct. Biol.* **2007**, *157*, 593–605.

(60) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. L. Two Crystal Structures of the B1 Immunoglobulin-Binding Domain of Streptococcal Protein G and Comparison with NMR. *Biochemistry* **1994**, *33*, 4721–4729.

(61) Martinez, J. C.; Pisabarro, M. T.; Serrano, L. Obligatory Steps in Protein Folding and the Conformational Diversity of the Transition State. *Nat. Struct. Biol.* **1998**, *5*, 721–729.

(62) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. High-Resolution X-Ray Crystal Structures of the Villin Headpiece Subdomain, an Ultrafast Folding Protein. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 7517–7522.

(63) Lin, T. W.; Chen, Z. G.; Usha, R.; Stauffacher, C. V.; Dai, J. B.; Schmidt, T.; Johnson, J. E. The Refined Crystal Structure of Cowpea Mosaic Virus at 2.8 Angstrom Resolution. *Virology* **1999**, *265*, 20–34.

(64) Banner, D. W.; Kokkinidis, M.; Tsernoglou, D. Structure of the Cole1 Rop Protein at 1.7 Å Resolution. *J. Mol. Biol.* **1987**, *196*, 657–675.

(65) Vlassi, M.; Steif, C.; Weber, P.; Tsernoglou, D.; Wilson, K. S.; Hinz, H. J.; Kokkinidis, M. Restored Heptad Pattern Continuity Does Not Alter the Folding of a Four-Alpha-Helix Bundle. *Nat. Struct. Biol.* **1994**, *1*, 706–716.

(66) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(67) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(68) van Gunsteren, W. F.; Daura, X.; Mark, A. E. Gromos Force Field. *Encyclopaedia Comput. Chem.* **1998**, *2*, 1211–1216.

(69) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interactions Models for Water in Relation to Protein Hydration. In Intermolecular Forces, Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1981; pp 331−342.

(70) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(71) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A Generalized Reaction Field Method for Molecular Dynamics Simulations. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(72) Hinsen, K. Analysis of Domain Motions by Approximate Normal Mode Calculations. *Proteins* **1998**, *33*, 417–429.

(73) Miyamoto, S.; Kollman, P. A. Settle−An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

(74) Garcia, A. E. Large-Amplitude Nonlinear Motions in Proteins. *Phys. Rev. Lett.* **1992**, *68*, 2696–2699.

(75) Amadei, A.; Linssen, A. B.; Berendsen, H. J. C. Essential Dynamics of Proteins. *Proteins* **1993**, *17*, 412–425.

(76) Amadei, A.; Ceruso, M. A.; Di Nola, A. On the Convergence of the Conformational Coordinates Basis Set Obtained by the Essential Dynamics Analysis of Proteins' Molecular Dynamics Simulations. *Proteins* **1999**, *36*, 419–424.

(77) Ceruso, M. A.; Amadei, A.; Di Nola, A. Mechanics and Dynamics of B1 Domain of Protein G: Role of Packing and Surface Hydrophobic Residues. *Protein Sci.* **1999**, *8*, 147–160.

(78) Ceruso, M. A.; Grottesi, A.; Di Nola, A. Effects of Core-Packing on the Structure, Function, and Mechanics of a Four-Helix-Bundle Protein Rop. *Proteins* **1999**, *36*, 436–446.

(79) Ceruso, M. A.; Grottesi, A.; Di Nola, A. Dynamic Effects of Mutations within Two Loops of Cytochrome C551 from Pseudomonas Aeruginosa. *Proteins* **2003**, *50*, 222–229.

(80) Haliloglu, T.; Bahar, I.; Erman, B. Gaussian Dynamics of Folded Proteins. *Phys. Rev. Lett.* **1997**, *79*, 3090.

(81) 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.

(82) Freddolino, P. L.; Arkhipov, A. S.; Larson, S. B.; McPherson, A.; Schulten, K. Molecular Dynamics Simulations of the Complete Satellite Tobacco Mosaic Virus. *Structure* **2006**, *14*, 437–449.

(83) Arkhipov, A.; Freddolino, P. L.; Schulten, K. Stability and Dynamics of Virus Capsids Described by Coarse-Grained Modeling. *Structure* **2006**, *14*, 1767–1777.

(84) Ceruso, M. A.; Weinstein, H. Structural Mimicry of Proline Kinks: Tertiary Packing Interactions Support Local Structural Distortions. *J. Mol. Biol.* **2002**, *318*, 1237–1249.

(85) Yamniuk, A. P.; Vogel, H. J. Calmodulin's Flexibility Allows for Promiscuity in Its Interactions with Target Proteins and Peptides. *Mol. Biotechnol.* **2004**, *27*, 33–57.

(86) Basdevant, N.; Weinstein, H.; Ceruso, M. Thermodynamic Basis for Promiscuity and Selectivity in Protein-Protein Interactions: PDZ Domains, a Case Study. *J. Am. Chem. Soc.* **2006**, *128*, 12766–12777.

(87) Bakan, A.; Lazo, J. S.; Wipf, P.; Brummond, K. M.; Bahar, I. Toward a Molecular Understanding of the Interaction of Dual Specificity Phosphatases with Substrates: Insights from Structure-Based Modeling and High-Throughput Screening. *Curr. Med. Chem.* **2008**, *15*, 2536–2544.

(88) Deremble, C.; Lavery, R. Macromolecular Recognition. *Curr. Opin. Struct. Biol.* **2005**, *15*, 171–175.

(89) May, A.; Zacharias, M. Accounting for Global Protein Deformability During Protein−Protein and Protein−Ligand Docking. *Biochim. Biophys. Acta* **2005**, *1754*, 225–231.

(90) McCammon, J. A. Target Flexibility in Molecular Recognition. *Biochim. Biophys. Acta* **2005**, *1754*, 221–224.

(91) Bonvin, A. M. J. J. Flexible Protein−Protein Docking. *Curr. Opin. Struct. Biol.* **2006**, *16*, 194–200.

(92) Totrov, M.; Abagyan, R. Flexible Ligand Docking to Multiple Receptor Conformations: A Practical Alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178–184.

(93) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. Principles of Flexible Protein−Protein Docking. *Proteins Struct. Funct. Gen.* **2008**, *73*, 271–289.

(94) Camacho, C. J.; Vajda, S. Protein Docking Along Smooth Association Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10636–10641.

(95) Camacho, C. J.; Weng, Z.; Vajda, S.; DeLisi, C. Free Energy Landscapes of Encounter Complexes in Protein−Protein Association. *Biophys. J.* **1999**, *76*, 1166–1178.

(96) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Protein Sci.* **2000**, *9*, 10–19.

(97) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(98) Cornilescu, G.; Hadley, E. B.; Woll, M. G.; Markley, J. L.; Gellman, S. H.; Cornilescu, C. C. Solution Structure of a Small Protein Containing a Fluorinated Side Chain in the Core. *Protein Sci.* **2007**, *16*, 14–19.

(99) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. NMR Structure of the 35-Residue Villin Headpiece Subdomain. *Nat. Struct. Biol.* **1997**, *4*, 180–184.

(100) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. A Novel, Highly Stable Fold of the Immunoglobulin Binding Domain of Streptococcal Protein G. *Science* **1991**, *253*, 657–661.

(101) Blanco, F. J.; Ortiz, A. R.; Serrano, L. ^{1}H and ^{15}N NMR Assignment and Solution Structure of the SH3 Domain of Spectrin: Comparison of Unrefined and Refined Structure Sets with the Crystal Structure. *J. Biomol. NMR* **1997**, *9*, 347–357.

(102) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. Crystal Structure of a Src-Homology 3 (SH3) Domain. *Nature* **1992**, *359*, 851–855.

CT9002114