JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

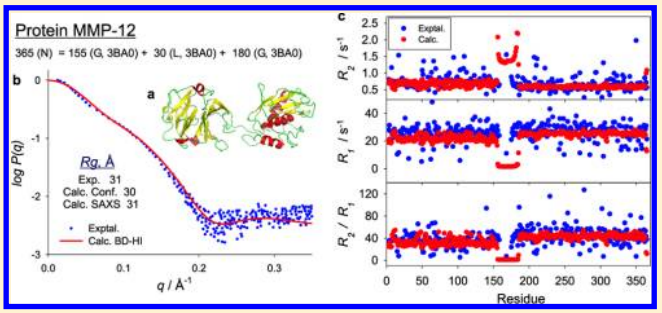
Prediction of Hydrodynamic and Other Solution Properties of Partially Disordered Proteins with a Simple, Coarse-Grained Model

D. Amorós, A. Ortega, and J. García de la Torre*

Departamento de Química Física, Facultad de Química, Universidad de Murcia, 30071 Murcia, Spain

Web-Enhanced Feature

ABSTRACT: The possibility of validating structures of intrinsically disordered proteins against solution properties is a goal that would be most helpful in the understanding of their function. We have devised a scheme for the prediction of solution properties of partially disordered proteins that comprise one or more ordered domains, along with flexible tails or linkers. A very simple, coarse-grained, residue-level model, which is easily parametrized using available structural information, along with previously developed tools for the simulation of solution conformation and dynamics, allows the prediction of properties like sedimentation coefficients, relaxation times, and X-ray or neutron scattering. This is demonstrated for a variety of partially disordered proteins, for which well-characterized solution properties are very accurately evaluated, with predictions falling in most cases within experimental errors.



INTRODUCTION

Over the past decade, the importance of intrinsically disordered (ID) proteins[1−3]—which deviate from the usual paradigm that related the function essentially to a quasirigid, folded, globular structure[4,5]—as manifested in a variety of (normal and pathological) physiological aspects, has become widely evident. The characterization of the conformational and dynamic behavior of intrinsically disordered proteins is crucial to understanding their role in such processes. The behavior in dilute solution, although it certainly differs from the environment that the protein experiences *in vivo*, is of great relevance; indeed, like in the case of globular or filamentous proteins and other biomacromolecules, solution properties determined by hydrodynamic techniques, NMR spectroscopies or radiation scattering, are essential to characterizing the structure and behavior of the individual molecules.[6] For that purpose, one needs computational methodologies that relate the solution properties of proteins to their structure.

In the case of quasirigid, globular proteins, the structure−property relationship has evolved from the classical theoretical approaches based on elementary models[7] (long rods for fibrous proteins, revolution ellipsoids for globular ones) to very detailed (even atomic) models and advanced treatments of the hydrodynamics of rigid particles, like—among others—those involved in our public-domain programs, HYDROPRO[8] and HYDRONMR[9] (including their most recent versions[10,11]).

Although the rigid-body hydrodynamics could be somehow modified to account for the conformational flexibility proper of disordered proteins,[12] simple and efficient tools like those mentioned above would be desirable for treating the long-time-scale features of their overall and internal hydrodynamics. Recently, we have made methodological advances for Monte Carlo (MC) and Brownian dynamics (BD) simulation of general, flexible models that have been implemented in our public-domain simulation program SIMUFLEX.[13] These methods could be employed for the prediction of solution properties and dynamics of proteins with any degree of flexibility, provided that a sufficiently realistic model, but still amenable to efficient computation, is used. Coarse-grained models, with nonatomic but sufficient structural detail, are most valuable for these purposes[14] and can be easily adapted for dynamic simulations with SIMUFLEX.[13] Our recent work on dendrimers (which are macromolecular constructs with some similarity to proteins) has demonstrated such a possibility.[15]

Our aim in this work has been to demonstrate the capability of a minimal model, with only one bead per amino acid residue, as proposed by previous workers,[16−18] to predict the overall conformational and dynamic properties, and even features of the internal dynamics of a particular kind of intrinsically disordered (ID) proteins, which we have denoted as partially disordered. While in many ID proteins the disorder is distributed along most of the polypeptide chains, with regions of local, not extensive, or even temporal order, there are some others, like the five cases considered in this paper, which can be regarded as composed of large globular domains joined by fairly long flexible linkers. Apart from fine deviations from this scheme, the overall solution properties of such kinds of proteins are dominated by the overall flexibility of the linker, that notorious influence of hydrodynamic interaction (HI) both in

Received: October 31, 2012
Published: January 15, 2013

the linker flexible chains[6,19] as well as in the globular domains.[18] Such important effects are properly accounted for in our SIMUFLEX methodology,[13] which allows conformational and dynamic prediction from simple coarse-grained models. Thus, joining these two paradigms, and with the help of a variety of other bioinformatics tools, we have devised a scheme that allows the prediction of solution properties of partially flexible proteins. We demonstrate how this scheme allows for accurate prediction (in many cases, within experimental errors) of a variety of solution properties of a number of such proteins, although we must also have caution that this scheme may not be valid for other ID proteins in which the disorder is distributed in a more complex manner, or for the prediction of more dynamically complex phenomena, like the kinetics and pathways of folding.

MODELS, POTENTIALS, AND PARAMETERS

The essential aspects of the calculation of the solution properties considered in this work and the description of the computational tools employed in the calculations are summarized in the Methods section at the end of this paper. Here, we just describe the physical aspects of the models employed for the proteins and the information employed to assign their parameters. We emphasize our attempts to (a) employ extremely simple models and (b) avoid, in most cases, adjusting the parameters as to fit properties, employing instead—in the spirit of a coarse-graining scheme—existing information available from previous (and usually more molecularly detailed) models of proteins.

Flexible Chains. For flexible regions, we adopt a simple model, with one identical element for each residue. Beads are joined by $C^{\alpha}-C^{\alpha}$ virtual bonds, having as equilibrium length the distance between neighbor alpha-carbon (C^{α}) atoms, b_e = 3.8 Å. A Hookean potential

$$V_b = 1/2K_b(b - b_e)^2 \tag{1}$$

with a sufficiently large value of the constant, K_b = 100 kcal/mol/Å^2, ensures that the RMS fluctuation of the instantaneous bond length b with respect to b_e is small, of only 0.14 Å.

In the spirit of maximum simplicity, for the angle between bonds we adopt a quadratic potential:

$$V_{\alpha} = 1/2K_{\alpha}(\alpha - \alpha_e)^2 \tag{2}$$

A statistical study by Kleywegt[20] of the bond angle and dihedrals between C^{α} atoms in coil regions of a set of about 600 nonhomologous protein structures[21] is available. According to this study, 97% of the internal angles between two successive C^{α} virtual bonds are in the range 75−150°. Apart from some detailed features in the distribution, and in terms of a Gaussian distribution corresponding to the quadratic potential in eq 2, this corresponds to α_e = 112° as the midpoint in the range and a constant that we set as K_{α} = 3.5 kcal/mol/rad^2. Furthermore, Kleywegt[20] presented statistical data of the torsion angles in the virtual chain, showing a distribution which, although nonuniform, is spread over the whole range 0−360°. Thus, for the shake of simplicity, we did not include torsional potential in flexible chains.

Excluded volume between nonbonded elements can be easily treated in Monte Carlo simulations with a hard-spheres (HS) potential, where $V_{EV}(r_{ij}) = 0$ if the interelement distance $r_{ij} > r_{HS}$ or infinite otherwise. We considered the hard-sphere radius r_{HS} as an adjustable parameter, fitting simulation results of R_g against a set of experimental data for 21 urea- or GuHCl-denaturated proteins, finding an optimum fit for r_{HS} = 2.25 Å with a rms percent deviation of less than 5%. For the hydrodynamic radius of the amino acid residues, we take the very same value that we[19] obtained in rigid-body (RB) calculations of residue-level models of globular proteins, σ = 6.1 Å. MC-RB simulation (see the Methods section) of hydrodynamic radii (corresponding to either D or $[\eta]$), using this parameter, resulted in a typical deviation of calculated results from experimental ones of about 6%. A more detailed report of our study on the flexible proteins chain, discussing the conformation and hydrodynamic behavior of denaturated proteins, is reported elsewhere.[22] Within the context of this paper, the model, potential, and parameters for flexible chains are just intended, and seem sufficiently appropriate, to represent the linker of tail regions in partially disordered proteins.

When flexible chains are subjected in this work to Brownian dynamics simulation with a hydrodynamic interaction (BD-HI), the discontinuous HS potential is replaced by a continuous repulsive potential, inspired in the Gō-like model (see below):

$$V_{EV}^{(i,j)} = \varepsilon_2(\sigma_2/r_{ij})^{12} \tag{3}$$

Using σ_2 = 4 Å as in the Gō-like model, and adjusting ε_2 = 2.0 kcal/mol, the BD-HI simulations predicted solution properties practically coincident with those from MC-RB simulations with the HS potential.

Structured Domains. In order to determine and maintain fixed the structured, globular, quasi-rigid domains, in the simulations of our coarse-grained C^{α}-only model, we follow the lead of other authors, employing the so-called Gō-like model.[23,24] The starting point is the atomic structure, given in a PDB-formatted file, which is based on the concept of *native* contacts, which are pairs of residues that are essential to maintaining the ordered structure. This concept is implemented in the CSU (Contacts of Structural Units) program described in ref 25.

Short range potentials are as follows. For bonds, the potential, $V(b_i)$, is given by eq 1, and parameters are the same as for the flexible chains. For bending angles, again a quadratic potential, $V(\alpha_i)$, given by eq 2, is employed, but now the equilibrium value $\alpha_{e,i}$ is that determined from the atomic (PDB) structure. The Gō-like model includes a torsional potential, V_{ϕ}:

$$V_{\phi}(\phi_i) = K_{\phi}^{(1)}[1 - \cos(\phi_i - \phi_{0i})] + K_{\phi}^{(3)}[1 - \cos 3(\phi_i - \phi_{0i})] \quad i = 1, ..., N - 3 \tag{4}$$

in which the equilibrium dihedral angles $\phi_{0i}^{(e)}$ are again those found in the PDB structure. In eq 2 and eq 4, we employ the same constants as in previous usages of these potentials,[23,24] K_{α} = 40 kcal/mol/rad^2, $K_{\phi}^{(1)}$ = 1.0 kcal/mol, and $K_{\phi}^{(3)}$ = 0.5 kcal/mol.

Nonbonded interactions, for elements separated along the polypeptide chain by more than three bonds, are as follows. Distinction is made between native and non-native contacts. In the latter case, for the globule we use the same purely repulsive potentials and parameters as for the flexible chains (i.e., the hard-spheres potential or that in eq 3). For a native contact, again following other authors, we adopt an attractive/repulsive potential given by a modified 12−10 Lennard-Jones equation:[23,24]

$$V_{nb}^{native}(i,j) = \varepsilon_1\left[5\left(\frac{\sigma_1}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_1}{r_{ij}}\right)^{10}\right] \quad (5)$$

with an energy constant ε_1 = 4.0 kcal/mol, σ_1 being equal to the distance between the C_α atoms in the PDB structure, and a cutoff distance equal to $3\sigma_1$.

RESULTS

Proof-of-Concept: A Chimeric Protein. Before proceeding with application to real proteins, the methodology was essayed for a chimeric construct with 156 amino acids (aa), composed by four domains. Domain I is an N-terminal ordered, globular domain, whose structure was that of the immunoglobulin-binding domain of protein G from *Streptococcus* (1PGA.pdb[26]), with 56 aa (1PGA). Domain II is a disordered, flexible linker with 44 glycine residues. Domain III is another ordered, α-helical domain with 25 aa whose structure was that of the antimicrobian peptide pleurocidin (1Z64.pdb[27]). Finally, domain IV is a C-terminal flexible tail with 31 glycine residues. An instantaneous conformation of our chimeric protein is displayed in Figure 1. The placement and number of residues of

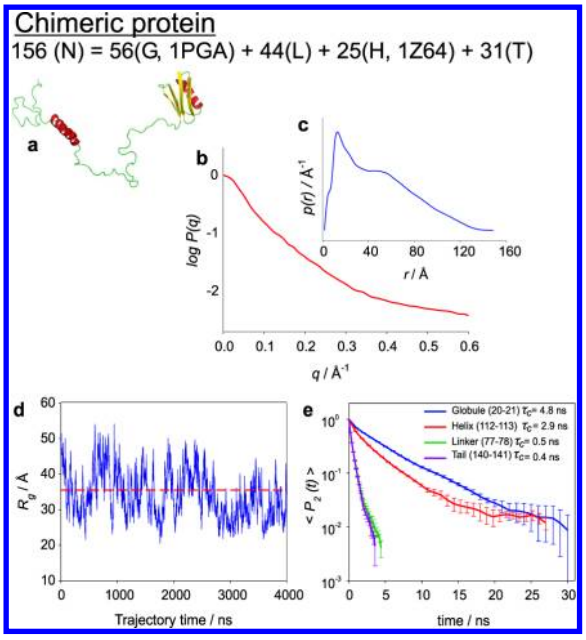


Figure 1. (a) Instantaneous conformation of the chimeric protein employed in tests, showing the four separate domains. (b,c) Results for scattering, namely normalized scattering intensity (form factor), $P(q)$, and distribution of distances, $p(r)$. (d) Time evolution of the radius of gyration during a 4 μs trajectory. (e) $\langle P_2(t)\rangle$ function for various $C_\alpha-C_\alpha$ virtual-bond vectors: (1) 20−21, within the globule; (2) 112−113, within the helix; (3) 77−78, by the midst of the linker; (4) 140−141, by the midst of the tail. A WEO showing a 0.5 μs trajectory movie of the chimeric protein in avi format is available.

the successive domains is indicated as 156(N) = 56(G,1PGA) + 44(L) + 25(H) + 31(T) where G, H, L, T, and N indicate the number of residues in rigid globular (with indication of the PDB code), rigid helical, flexible linker, flexible tail, and total number, respectively. This notation will be used hereafter.

Four BD-HI trajectories of 4 μs each were carried out. A movie displaying the Brownian motion of the protein during such long time (Web Enhanced Features) demonstrates that the rigid structure of the globular and helical domains are maintained during such a long time, with the linker and tail regions fluctuating as expected for flexible domains. We even simulated the 1PGA globule alone, in a process in which the Gō potentials that hold the globular structure are switched off and after a while are switched on again, seeing in the movie how the protein unfolds and folds back, recovering the very same initial structure.

Values of instantaneous R_g and other properties fluctuated accordingly. The time or conformational average of the properties, which would correspond to the observable quantities, were R_g = 34.50 ± 0.03 Å, D_t = (7.07 ± 0.01) × 10^{-7} cm^2/s, s = 2.87 ± 0.04 S, and $[\eta]$ = 16.1 ± 0.1 cm^3/g. We note the small statistical uncertainty, obtained as the RMS deviation of the results for each trajectory, which confirms that the typical trajectory length is long enough. We can also follow the reorientational dynamics from the time function $\langle P_2(t)\rangle$ for some characteristic vectors, from which relaxation and correlation times could be extracted. We also determine the angular dependence of scattering intensity $P(q)$ and the distribution of distances, $p(r)$. (Again, we refer the reader to the Methods section for a description of properties and computational methods.) Particularly noteworthy are the features presented by $p(r)$, like the peak and shoulder, associated with the two rigid areas, and the long tail due to the flexible regions. Figure 1 displays in detail all of these observations.

Protein S4. S4 is a chimeric protein devised to link catalytic domains for cellulose degradation: the cohesin domains of *Clostridium cellolotycum* (1G1K.pdb) and *C. cellolotycum* (1OHZ.pdb) joined by a 50 amino acid linker. Available small-angle scattering results by Hammel et al.[28] are compared to the computational prediction in Figure 2. The R_g computed

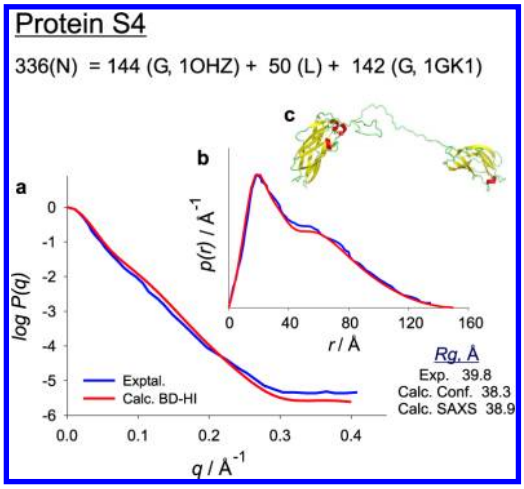


Figure 2. Protein S4: (a,b) SAXS results, scattering intensities $P(q)$ and distribution of distances $p(r)$; (c) snapshot in an instantaneous conformation. A WEO showing a 1 μs trajectory movie of S4 in avi format is available.

results (either as averages of instantaneous R_g's, or from the Guinier-plot analysis of very-low-angle $P(q)$) are remarkably close to the experimental value. Even more remarkable is the agreement in the full $p(r)$ distribution, which describes the conformational statistics over all the length scales. Indeed, not only are our numerical results coincident with the experimental results in the case of R_g but also the experimental maximum distance reported by these authors, r_{max} = 150 ± 8 Å, and our

computed result are practically identical, while a complex modeling work carried out by those workers produced a quite deviate value.

Protein ZipA. ZipA, which has 328 aa, is a membrane-anchored protein by an N-terminal domain, followed by an unstructured region, and a globular C-terminal (FtsZ-binding) domain comprising 142 residues. Scission of the first (transmembrane) produces a soluble protein, ZipA(23−328), whose sedimentation coefficient has been found to be 2.2 S. Thus, the protein experimentally studied, and simulated in our work, has 306 aa, with 163 in an N-terminal tail and 143 in the C-terminal globule, whose atomic structure is available (1F47.pdb,[29]). For more details, see Ohashi et al.[30]

We carried out BD-HI simulations of 15 μs, from which the diffusion coefficient D is calculated from the Brownian trajectories from the Einstein equation (Figure 3), and with the values M = 36474 Da and $\bar{\nu}$ = 0.732 cm^3/g our simulation predicted $s_{20,w}$ = 2.1 S, in excellent agreement with the experimental value.

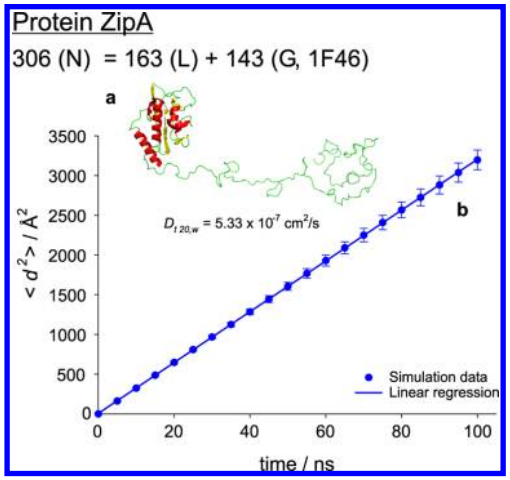


Figure 3. (a) Snapshot of protein ZipA in an instantaneous conformation. (b) Time-dependence of RMS displacement of the center of mass, to which the Einstein equation is applied to determine the diffusion coefficient, D. A WEO showing a 1 μs trajectory movie of ZipA in avi format is available.

Protein X from Sendai Virus. Protein X (PX) is the RNA-binding domain of phosphoprotein P, which is one of the two components of the RNA-polymerase of Sendai virus. It consists of two well-defined domains: a disordered N-terminal region with 42 aa and a C-terminal globule (1R4G.pdb,[31]) of 53 aa. The protein was cloned and purified, including 15 histidines at the N-terminal, so that the unstructured tail contains 56 residues. For more details, see Blanchard et al.[31] A snapshot of an instantaneous conformation is displayed in Figure 4.

This protein has been characterized by scattering—both SAXS and SANS—and NMR relaxation. As it has only one globule and one tail, like ZipA, the construction of its starting structure for the simulation was simple. Four trajectories, each of 10 μs with 20 000 recorded conformations, were generated. The result obtained for the radius of gyration from the coordinates of the C$^\alpha$ beads in the model was practically identical in the four runs, R_g = 26.7 ± 0.2 Å, and agrees very well with the results from SAXS (29.7 Å)[32] and SANS (25 ± 1 Å).[31] Not only this single scattering result but all of the scattering diagrams were in very good agreement with the experimental data. Figure 4 displays the $P(q)$ and $p(r)$, where there is concordance over the whole range of q and r (including the so-called D_{max}, at which $p(r)$ fades away, of about 90 Å). We have performed the calculation of $P(q)$ and $p(r)$ both from the positions of the model C$^\alpha$ elements and also from a putative all-atoms model, constructed from the C$^\alpha$ chain by means of program PULCHRA[33] (see Methods section). Figure 4 shows how the agreement, which is already acceptable for the C$^\alpha$ model, improves even more after the atomic reconstruction. We also show the scattering intensities in the form of a Kratky plot ($q^2P(q)$ vs q) which—in addition to confirming the quality of our prediction—shows the typical trend of a particle with a globular region and a flexible tail.

PX has been also thoroughly studied by NMR relaxation; ^{15}N relaxation dispersion measurements have been reported.[34] Experimental relaxation rates for the residues within the globule are $R_1 \approx 1.5$ s^{-1} and $R_2 \approx 14 \pm 2$ s^{-1}, with a correlation time of 7.3 ns. Our results for the $\langle P_2(t)\rangle$ for the successive virtual bonds within the globule showed a biexponential decay (see Methods) with two relaxation times. The small one, of less than 1 ns, is assigned to the fast internal motion, while the slow, more relevant component, assigned to overall tumbling, is found to be τ_c = 6.6 ± 1.3 (RMS and SD), which compares well with the experimental value. It is interesting to compare the correlation time τ_c of the whole PX with that predicted for the globule alone (without the tail), which as predicted by HYDROPRO[8] is 4.9 ns.

Protein MMP12. Matrix metalloproteases (MMPs) are involved in a number of extracellular processes, like the degradation of components of the extracellular matrix. Most of them are composed of two globular domains joined by a linker. Such is the case for MMP-12, which comprises the structured domains CAT, with 157 aa, and HPX, with 184 aa, with a rather small linker of only 15 residues; see Figure 5.

In contrast with the previously presented proteins that contain long unstructured regions, the linker in MMP-12 is quite short. Indeed, an atomic structure of the whole protein is available (3BA0.pdb,[35]). The conformational variability is not as extensive as in the other proteins studied in the present work, so that a shorter simulation, of only 10 μs, was sufficient. Scattering and NMR relaxation measurements are available for MMP-12.[35] For the radius of gyration, the value calculated either from the coordinates of the model elements, 30.3 ± 1.4 Å, or from the calculated Guinier plot, 31.3 Å, deviates from the experimental SAXS value, 25 ± 1 Å nearly 20%. Certainly, the agreement is not as good as that found for the radius of gyration or the hydrodynamic coefficients in other proteins. Nonetheless, when judging such deviations, one should keep in mind that, in addition to some small uncertainty in the experimental data itself (say, up to 5%), there is a great variety of factors (structure of globules, modeling of linkers, computational aspects, etc.) involved in the predictions, which can become quite appreciable in some instances. Nonetheless, the comparison with the full SAXS diagrams (calculated and experimental, Figure 5) is reassuring. Also, the computed scattering intensity follows very closely the experimentally observed trend.

Regarding NMR relaxation, the fact that this protein is predominantly structured, yet flexible but—because of the short linker—with a hinge-like kind of flexibility, we attempted a calculation of the T1 and T2 relaxation times, to be compared with available experimental data. The comparison is displayed in Figure 5. The computed relaxation time fluctuates along the

dx.doi.org/10.1021/ct300948u | J. Chem. Theory Comput. 2013, 9, 1678−1685

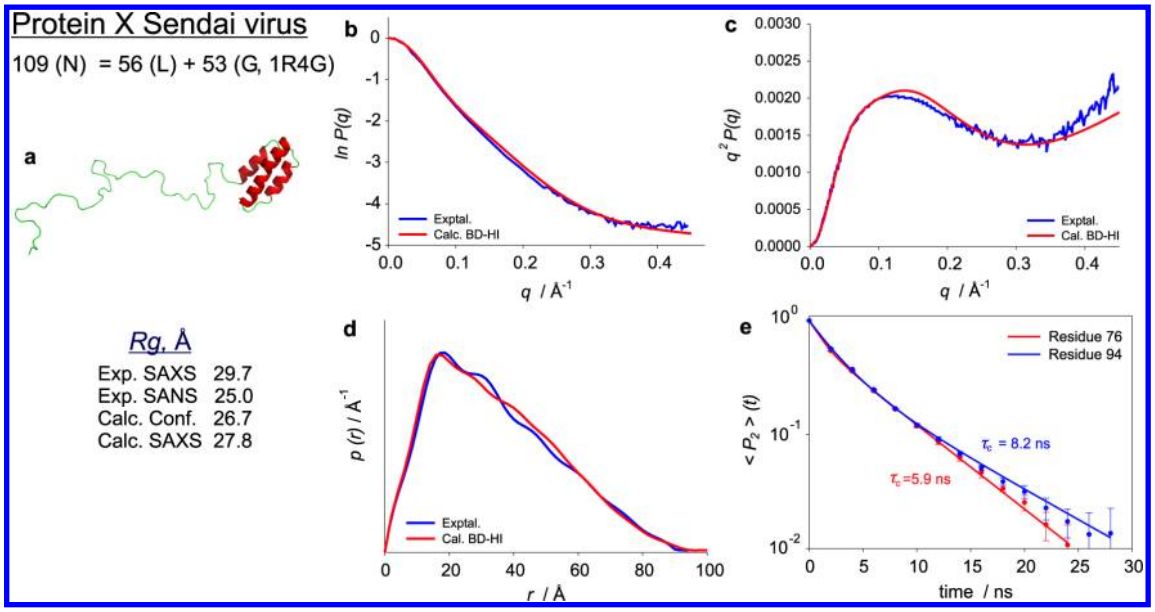


Figure 4. (a) Snapshot of protein X in an instantaneous conformation. (b,c) Scattering intensities, $P(q)$ plot and Kratly plot. (d) $p(r)$ distribution. (e) Time correlation function $\langle P_2(t)\rangle$ for two vectors having rather different orientations within the globular domain, with τ_c values that bracket the whole set of residue-specific values. A WEO showing a 2 μs trajectory movie of PX in avi format is available.

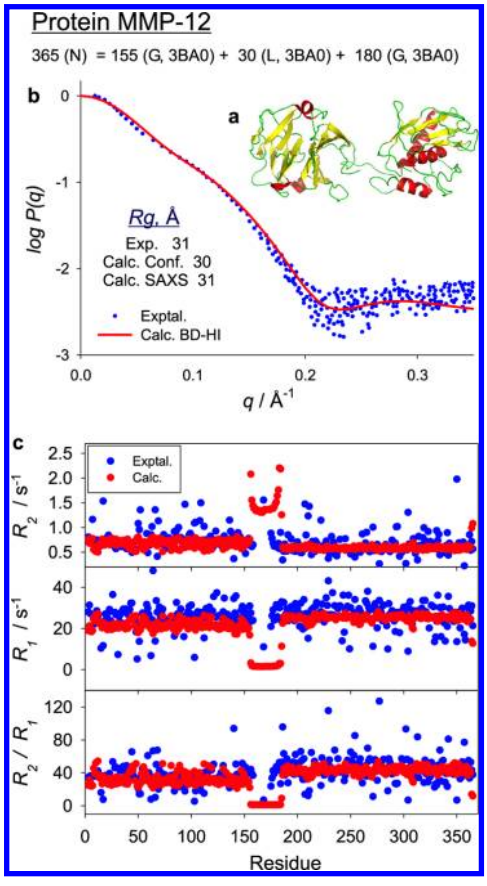


Figure 5. (a) Snapshot of protein MMP12 in an instantaneous conformation. (b) SAXS intensities. (c) NMR relaxation times. A WEO showing a 1 μs trajectory movie of MMP-12 in avi format is available.

sequence less than the experimental ones, but the agreement in the baselines (average over sequence) in both globules is very satisfactory.

In their SAXS and NMR relaxation study, Bertini et al.[35] were able to show that the crystallographic structure of MMP-12 is not representative of its solution conformation, on the basis of the disagreement between the results from those techniques and those predicted for the rigid structure. Particularly informative was the mismatch between the NMR relaxation results and the predictions from the rigid-body HYDRONMR[9] predictions. These authors were able to attribute the discrepancies to the flexibility of the linker, matching the SAXS results with an ensemble of random conformations of the linker. Our methodologies have enabled not only the reproduction of the SAXS results but also matching the dynamic NMR observations.

Protein BTK. Bruton's tyrosine kinase (BTK) is a protein that helps modulate for development of B lymphocytes and therefore is regarded as a potential target for the treatment of lymphoma. It is composed of 659 aa, arranged in the following (from N-t to C-t) globular domains: PH (169 aa, 1BTK.pdb[36]), SH3 (55 aa, 1AWX.pdb[37]), SH2 (96 aa, 2GE9.pdb[38]), and kinase (KIN, 258 aa, 1K2P.pdb,[39]). Between them there are three linkers of 50, 7, and 21 aa, respectively (see Figure 6).

Experimental results are available from SAXS, namely $P(q)$, $p(r)$ (Figure 6), and R_g = 50 Å,[40,41] and also for the sedimentation coefficient, $s_{20,w}$ = 3.3 S.[40] Owing to the large size of BTK, BD-HI simulations were not feasible; instead we employed the BD-noHI-RB, which is rigorous for scattering predictions and adequate for overall hydrodynamics (see the Methods section). Starting from independent conformations, six trajectories of 10 μs were conducted, each providing 10^5 conformations, and the results are the conformational averages of the 6×10^5 conformations. We found R_g = 49 Å, which practically matches the experimental data, although some differences are noticed in $P(q)$ and $p(r)$ (see Figure 6). The calculated $s_{20,w}$ = 3.9 S is about 10% above the experimental value, and the agreement for the full SAXS curve is just semiquantitive, as shown in Figure 6 (although in this case the concordance for R_g is very satisfactory). Like in the above commented comparison of the R_g values of MMP-12, one

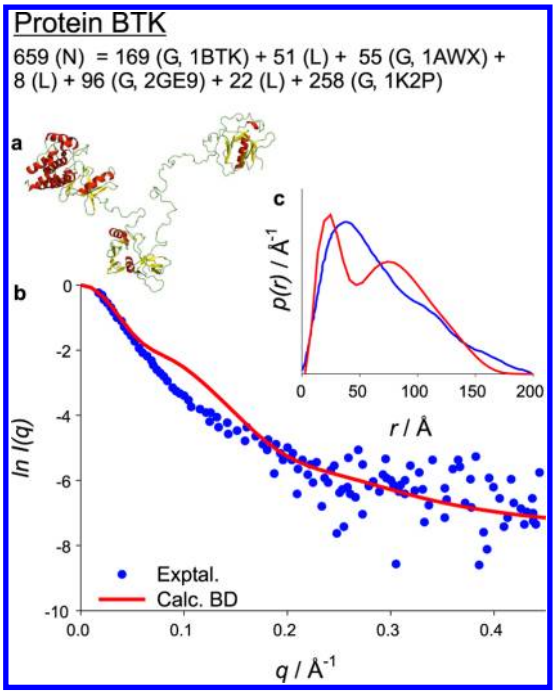


Figure 6. (a) Snapshot of protein BTK in an instantaneous conformation. (b) Scattering intensities: experimental data and calculated values. (c) Distance distribution, experimental (blue dots) and calculated (red line). A WEO showing a 15 μs trajectory movie of BTK in avi format is available.

should keep in mind, when judging the quality of the predictions, the multiple aspects like simulation procedures, computation of properties, etc., and in the particular case of BTK, its extraordinary structural complexity of this protein, with three different globular domains and two linkers. The schematic modeling procedure, in which domains are quasirigid globules of very flexible linkers, may be responsible for the deviations in the SAXS diagram at intermediate angles.

METHODS

The solution properties that we can consider in this work are both conformational and hydrodynamic. Among the former are the radius of gyration, R_g, and the whole angular variation of scattering intensity (usually small-angle X-ray or neutron scattering, SAXS or SANS, respectively), $P(q)$, where $q = (4\pi/\lambda)\sin(\theta/2)$ is the scattering variable corresponding to scattering angle θ and wavelength λ, and P is the normalized ($P(0) = 1$) scattering intensity. Overall hydrodynamic coefficients are the sedimentation coefficient s, translational diffusion coefficient D, and intrinsic viscosity $[\eta]$. The internal, reorientational dynamics of segments within the molecular model are expressed by the function $\langle P_2(t)\rangle = \langle(3\cos^2\theta(t) - 1)/2\rangle$, where $\theta(t)$ is the angle subtended by the orientations of the segment vector in two instants separated by time t, and the average $\langle ...\rangle$ is over the possible choices of the initial instant. A multicomponent analysis of $\langle P_2(t)\rangle$, fitted to a multiexponential decay, can yield a series of relaxation times $\tau_1 > \tau_2 > ...$ For dipoles within the globular regions, a two-exponential fit provides the model-free (Lipari-Szabo[42,43]) relaxation times: the fast one, $\tau_2 \equiv \tau_{int}$, and the most relevant one, the longest relaxation time, would correspond to the correlation time $\tau_1 \equiv \tau_c$. Eventually, the multiexponential $\langle P_2(t)\rangle$ can be recast into a multi-Lorentzian $J(\omega)$ from which NMR relaxation times T_1 and T_2 could be obtained.

The Monte Carlo (MC) simulation is a most adequate approach to predicting conformational properties, like R_g, $P(q)$, and $p(r)$. The overall hydrodynamic coefficients, s, D, and $[\eta]$ can also be obtained by the so-called Monte Carlo rigid-body (MC-RB) procedure,[44−46] in which properties are evaluated for each conformation as if it were a rigid particle, using bead model procedures, like the advanced methodology implemented in the HYDRO++ program.[47] The final result for each property is the average over all the individual values. This method has been tested to yield results that agree very well with experimental data for long, fully flexible polymer chains[48,49] and for semiflexible wormlike chains.[19] The MC-RB methodology for a model whose force field includes a number of features is implemented in our computer program MONTEHYDRO.[50]

For the description of internal dynamics and the time-course of conformational changes or fluctuations, in the time scale where these aspects are most significant, an adequate technique—the most adequate one when dealing with coarse-grained models—is Brownian dynamics (BD) simulation, which must include hydrodynamic interaction (HI) effects, which are in general compulsory for dynamics simulations of flexible polymers, and whose striking influence on protein dynamics has been demonstrated by Frembgen-Kesner and Elcock.[18,51] The inconvenience of Brownian dynamics with hydrodynamic interaction (BD-HI) is that it is computationally intensive, so it will be employed just for those properties for which the MC-RB procedure is not applicable. It is also noteworthy that when HI is neglected in Brownian dynamics, the procedure is still valid to sample the conformational space, so that such (BD-noHI) an approach can also be employed—paraphrasing Rossky et al.[52]—as a "smart" (alternative) Monte Carlo method. Thus, rigid-body calculations for instantaneous conformations can be also carried out in a BD-noHI-RB procedure. A significant development for Brownian dynamics simulation is our public-domain SIMUFLEX software,[13] that implements a quite general mechanic model which can be adapted to represent macromolecules with any kind of flexibility. This software package contains the programs BROWFLEX (which includes the same force field as MONTEHYDRO) for generation of the Brownian trajectories and ANAFLEX for their analysis, like that required for the evaluation of $\langle P_2(t)\rangle$, or the obtention of a rigorous value of D from the RMS displacement of the center of mass according to Einstein equation $\langle d^2\rangle = 6Dt$.

Diffusion and sedimentation coefficients, $D_{20,w}$ and $s_{20,w} = D_{20,w}M(1 - \bar{\nu}\rho)/RT$, were calculated under standard conditions (20°, water) using $T = 293$ K for the absolute temperature and $\eta_0 = 0.010$ poise and $\rho = 1.00$ cm^3/g for the solvent viscosity and density. Molecular weights, M, and specific volume, $\bar{\nu}$, were obtained from the sequence using the program SEDNTERP.[53]

Initial conformations of the partially disordered proteins were constructed from the structures of the globular domains, as specified by PDB files, and the known length (and eventually the sequence) of the linkers and tails, which were attached to the globules using the program MODELLER,[54] which adds putative chains to an incomplete atomic structure. In the case of tails, generating a plausible conformation from any of the terminal residues of the globule was a trivial task. For the generation of linkers, program VMD[55] was first employed to place the globules with adequate orientations and reasonable distance between them, so that MODELLER would, next,

dx.doi.org/10.1021/ct300948u | J. Chem. Theory Comput. 2013, 9, 1678−1685

construct a putative linker between the two globular domains as if it were an unresolved loop.

The program CSU[25] was employed to detect the so-called native contacts in the globular domains, thus determining the nonbonded part of the Gō-like model. This information was employed by an in-house program to construct the topology of the whole protein and the force-field specifications, which were cast in the form of the input files of BROWFLEX or MONTEHYDRO. As indicated above, these programs themselves or the ANAFLEX ancillary tool were in charge of calculating the radius of gyration and the hydrodynamic coefficients. Other in-house programs were written for the calculation of the scattering functions $P(q)$ and $p(r)$ from the coarse-grained bead chain. In some instances, we attempted a more detailed, atomistic calculation, which was carried out building an all-atoms reconstruction by means of program PULCHRA,[33] to which programs CRYSOL/GNOM[56] were applied. Finally, the calculations of T_1 and T_2 from $\langle P_2(t)\rangle$ were carried out with another in-house program, based on parts of our HYDRONMR code.[9]

Usually, several independent runs of BROWFLEX were done with different, independent starting conformations. Any property was calculated for each trajectory, so that the standard deviation of the set of values gives an estimate of the statistical error (which should be small if the trajectory is sufficiently long). In all cases, we verified that such is the case, with statistical errors on the order of, or smaller than, the typical errors of the experimental results.

CONCLUDING REMARKS

In this work we have demonstrated how a simple coarse-grained model, with just one element per amino acid residue,[16,17] suffices to predict accurately the solution properties of partially disordered proteins, in spite of the complexity of their structures, which include both rigid, ordered domains and flexible, disordered regions. The success of such an approach for single-valued properties, like the sedimentation coefficient (or, equivalently, the hydrodynamic radius), the overall correlation time, which is the primary result from NMR relaxation, and the radius of gyration that determines the limiting very-low-angle SAXS/SANS intensities, relies on a proper representation of the shape of the rigid domains and the conformational flexibility of the flexible ones. The former aspect is accomplished with basis on the previous experience on modeling rigid protein structures with the HYDROPRO[11] and HYDRONMR tools,[9,10] and the latter is achieved thanks to the possibilities for conformational and dynamic simulation offered by the methodologies embedded in SIMUFLEX,[13] and the careful choice of the flexible-model parameters. This choice has been made mainly from available molecular information, avoiding in most cases adjustable parameters. It is also shown that this mesoscale is even able to predict the medium-angle region of the scattering diagrams and the distribution of inter-residue distances. This is so because these results do not depend essentially on atomic details and can be described with an adequate computational scheme and a properly parametrized residue-level model.

We also remark on aspects of our computational scheme. The proper representation of hydrodynamic interaction, which is essential in the description of proteins dynamics,[18] as well as the implementation of new pieces for simulation (model potentials specific for this kind of proteins) and analysis (local dynamics) in SIMUFLEX[13] are core features of that scheme, but not the only ones. Furthermore, we have devised a systematic protocol that employs a number of a variety of available, public-domain tools (MODELER,[54] VMD,[55] CSU,[25] and CRYSOL/GNOM[56]). Furthermore, we have written ancillary, interfacing codes that provide input and analyze SIMUFLEX output using those tools. As judged from the successful predictions achieved in the present work, it is hoped that the simple model, treated with this systematic computational protocol will be helpful in the study of this structurally complex and notably relevant kind of protein.

COMPUTER PROGRAMS

New versions of our public-domain programs MONTEHYDRO and SIMUFLEX, which include potentials and forces needed for the present calculations, are available on our Web site http://leonardo.inf.um.es/macromol/.

ASSOCIATED CONTENT

Web-Enhanced Features

WEOs are available in the HTML version of the paper, including the following: a movie in avi format showing the time evolution of the chimeric protein, a movie in avi format showing the time evolution of S4, a movie in avi format showing the time evolution of ZipA, a movie in avi format showing the time evolution of PX, a movie in avi format showing the time evolution of MMP-12, and a movie in avi format showing the time evolution of BTK.

AUTHOR INFORMATION

Corresponding Author

*Phone: 34 968 367426. Fax: 34 968 364148. E-mail: jgt@um.es.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was performed within a Grupo de Excelencia de la Región de Murcia (grant 04531/GERM/06). Support provided by grant CTQ-2009-06831 from Ministerio de Ciencia e Innovacion including FEDER funds. Final revision and publication in 2013 was under grant CTQ2012-33717 from Ministerio de Ciencia y Competitividad. A.O. acknowledges a postdoctoral fellowship from Fundación CajaMurcia, and D.A. is recipient of a predoctoral fellowship from MEC. Computing time was provided by Fundación Parque Cientifico de Murcia.

REFERENCES

(1) Tompa, P. *Trends Biochem. Sci.* **2002**, *27*, 527−533.
(2) Fink, A. *Curr. Opin. Struct. Biol* **2005**, *15*, 35−41.
(3) *Structure and Function of Intrinsically Disordered Proteins*; Tompa, P., Ed.; CRC Press: Boca Raton, FL, 2010.
(4) Wright, P.; Dyson, H. *J. Mol. Biol.* **1999**, *293*, 321−331.
(5) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607−3622.
(6) Serdyuk, I. N.; Zaccai, N. R.; Zaccai, J. *Methods in Molecular Biophysics*; Cambridge University Press: New York, 2008; Vol. *1*, p 1.
(7) van Holde, K.; Johnson, W.; Ho, P. *Principles of Physical Biochemistry*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 1998.
(8) García de la Torre, J.; Huertas, M.; Carrasco, B. *Biophys. J.* **2000**, *78*, 719−730.
(9) García de la Torre, J.; Huertas, M.; Carrasco, B. *J. Magn. Reson.* **2000**, *147*, 138−146.
(10) Ortega, A.; Garcia de la Torre, J. *J. Am. Chem. Soc.* **2005**, *127*, 12764−12765.

(11) Ortega, A.; Amorós, D.; Garcia de la Torre, J. *Biophys. J.* **2011**, *101*, 892−898.
(12) Bae, S. H.; Dyson, H.; Wright, P. *J. Am. Chem. Soc.* **2009**, *131*, 6814−6821.
(13) Garcia de la Torre, J.; Hernandez Cifre, J. G.; Ortega, A.; Rodriguez Schmidt, R.; Fernandes, M.; Sánchez, H. E. P.; Pamies, R. *J. Chem. Theory Comput.* **2009**, *5*, 2606−2618.
(14) *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G., Ed.; CRC Press: Boca Raton, FL, 2010.
(15) del Rio Echenique, G.; Rodriguez Schmidt, R.; Freire, J.; Hernandez Cifre, J. G.; García de la Torre, J. *J. Am. Chem. Soc.* **2009**, *131*, 8548−8556.
(16) Tozzini, V.; Rocchia, W.; Cammon, J. M. *J. Chem. Theory Comput.* **2006**, *2*, 667−673.
(17) Tozzini, V.; McCammon, J. A. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2009.
(18) Frembgen-Kesner, T.; Elcock, A. *J. Chem. Theory Comput.* **2009**, *5*, 242−256.
(19) Amorós, D.; Ortega, A.; García de la Torre, J. *Macromolecules* **2011**, *44*, 5788−5797.
(20) Kleywegt, G. J. *J. Mol. Biol.* **1997**, *273*, 371−376.
(21) Hobohm, U.; Sander, C. *Protein Sci.* **1994**, *3*, 522−524.
(22) Amorós, D.; Ortega, A.; Garcia de la Torre, J. In preparation.
(23) Clementi, C.; Nyemeyer, H.; Onuchic, J. *J. Mol. Biol.* **2000**, *298*, 937−53.
(24) Frembgen-Kesner, T.; Elcock, A. H. *J. Chem. Theory Comput.* **2009**, *5* (2), 242−256.
(25) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. *Bioinformatics* **1999**, *15*, 327−332.
(26) Gallagher, T.; Alexander, P.; Bryan, P.; Gilliland, G. *Biochemistry* **1994**, *33*, 4721−4729.
(27) Syvitsky, R.; Burton, I.; Mattatall, N.; Douglas, S.; Jakeman, D. *Biochemistry* **2005**, *44*, 7282−7293.
(28) Hammel, M.; Fierobe, H.-P.; Czjzek, M.; Kurkal, V.; Smith, J. C.; Bayer, E. A.; Finet, S.; Reveceur-Bréchot, V. *J. Biol. Chem.* **2005**, *280*, 38562−38568.
(29) Mosyak, L.; Zhang, Y.; Glasfeld, E.; Haney, S.; Stahl, M.; Seehra, J.; Somers, W. *EMBO J.* **2000**, *19*, 3179−3191.
(30) Ohashi, T.; Hale, C.; de Boer, P.; Erikson, H. *J. Bacteriol.* **2002**, *184*, 4313−4315.
(31) Blanchard, L.; Tarbouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W. H.; Marion, D. *Virology* **2004**, *319*, 201−211.
(32) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17002−17007.
(33) Rotkiewicz, P.; Skolnick, J. *J. Comput. Chem.* **2008**, *29*, 1460−1465.
(34) Houben, K.; Blanchard, L.; Blackledge, M.; Marion, D. *Biophys. J.* **2007**, *93*, 2830−2844.
(35) Bertini, I.; Calderone, V.; Fragai, M.; Jaiswal, R.; Luchinat, C.; Melikian, M.; Mylonas, E.; Svergun, D. I. *J. Am. Chem. Soc.* **2008**, *130*, 7011−7021.
(36) Hyvonen, M.; Saraste, M. *EMBO J.* **1997**, *16*, 3396−3404.
(37) Hansson, H.; Mattsson, P.; Allard, P.; Haapaniemi, P.; Vihinen, M.; Smith, C.; Hard, T. *Biochemistry* **1998**, *37*, 2912−2924.
(38) Huang, K.-C.; Cheng, H.-T.; Pai, M.-T.; Tzeng, S.-R.; Cheng, J.-W. *J. Biomol. NMR* **2006**, *36*, 73−78.
(39) Mao, C.; Zhou, M.; Uckun, F. *J. Biol. Chem.* **2001**, *276*, 41435−41443.
(40) Marquez, J. A.; Smith, C. I.; Petoukhov, M. V.; Surdo, P. L.; Mattsson, P. T.; Knekt, M.; Westlund, A.; Scheffzek, K.; Saraste, M.; Svergun, D. I. *EMBO J.* **2003**, *22*, 4616−4624.
(41) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656−5664.
(42) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(43) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559−4570.
(44) Zimm, B. *Macromolecules* **1980**, *13*, 592−602.
(45) Hagerman, P.; Zimm, B. *Biopolymers* **1981**, *20*, 1481−1502.
(46) García de la Torre, J.; Jiménez, A.; Freire, J. *Macromolecules* **1982**, *15*, 148−154.
(47) García de la Torre, J.; del Rio Echenique, G.; Ortega, A. *J. Phys. Chem. B* **2007**, *111*, 955−961.
(48) García de la Torre, J.; López Martinez, M. C.; Tirado, M.; Freire, J. *Macromolecules* **1984**, *17*, 2715−2722.
(49) García Bernal, J.; Tirado, M.; García de la Torre, J. *Macromolecules* **1991**, *24*, 593−598.
(50) García de la Torre, J.; Ortega, A.; Perez Sánchez, H. E.; Hernandez Cifre, J. G. *Biophys. Chem.* **2005**, *116*, 121−128.
(51) Frembgen-Kesner, T.; Elcock, A. H. *Biophys. J.* **2010**, *99*, L75−L77.
(52) Rossky, P.; Doll, J.; Friedman, H. *J. Chem. Phys.* **1978**, *69*, 4628−4633.
(53) Laue, T. M.; Shah, B. D.; Ridgeway, T. M.; Pelletier, S. L. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science*; Harding, S.; Rowe, A.; Horton, J., Eds.; Royal Society of Chemistry: Cambridge, U. K., 1992; pp 90−125.
(54) Eswar, N.; Marti-Renom, M. A.; Webb, B.; Madhusudhan, M. S.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. *Curr. Protoc. Bioinf.* 2006, *15*, 5.6.1−5.6.30.
(55) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(56) Svergun, D. I.; C., C. B.; Koch, M. H. J. *J. Appl. Crystallogr.* **1995**, *28*, 768−773.