Prediction of the rotational tumbling time for proteins with disordered segments

Sung-Hun Bae, **H. Jane Dyson**, and **Peter E. Wright**[*]
Department of Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037

Abstract

For well-structured, rigid proteins, the prediction of rotational tumbling time (τ_c) using atomic coordinates is reasonably accurate, but is inaccurate for proteins with long unstructured sequences. Under physiological conditions, many proteins contain long disordered segments that play important regulatory roles in fundamental biological events including signal transduction and molecular recognition. Here we describe an ensemble approach to the boundary element method that accurately predicts τ_c for such proteins by introducing two layers of molecular surfaces whose correlated velocities decay exponentially with distance. Reliable prediction of τ_c will help to detect intra- and inter-molecular interactions and conformational switches between more-ordered and less-ordered states of the disordered segments. The method has been extensively validated using 12 reference proteins with 14 to 103 disordered residues at the N- and/or C-terminus, and has been successfully employed to explain a set of published results on a system that incorporates a conformational switch.

Keywords

protein; disordered; unstructured; tumbling; rotational diffusion; boundary element

Introduction

Accurate prediction of rotational and translational diffusion coefficients of proteins is important for the interpretation of a wide variety of biophysical data. For ideal spherical particles of radius r in solution of viscosity η, the translational (D_T) and rotational ($D_R = 1/\tau_c$) diffusion coefficients are described by the Stokes-Einstein (SE) relations: $D_T = k_B T/(6\pi\eta r)$ and $D_R = k_B T/(8\pi\eta r^3)$. For globular proteins in aqueous solution, the SE relation overestimates τ_c by a nearly constant factor of 2;[1] an empirical correction can be applied such that a robust and accurate estimation of τ_c is made using the molecular weight and the consensus partial specific volume for globular proteins (0.73 cm^{3}g^{-1}). For globular proteins with arbitrary shape, τ_c can be predicted from atomic coordinates by two distinct but related methods. In the bead method[2-4], a molecule is modeled as a group of beads which covers the molecular surface or fills the volume occupied by the molecule. By contrast, the boundary element method uses a group of triangular patches for modeling the surface[5-7] (Figure 1). Despite this difference, both methods calculate various hydrodynamic properties of interest using a similar formalism,

which describes the hydrodynamic interaction between beads or surface patches. Although both methods predict the behavior of rigid proteins with reasonable accuracy[4,8], they are inaccurate for flexible proteins that have long unstructured sequences; these can be best described by an ensemble approach. Many proteins contain long disordered segments under physiological conditions. Such proteins play important regulatory roles in many fundamental biological events including signal transduction and molecular recognition[9,10]. Reliable prediction of τ_c for proteins containing disordered segments will help to interpret τ_c data from NMR relaxation, fluorescence depolarization, electric birefringence, and dielectric relaxation measurements and to detect intra- and inter-molecular interactions and conformational regulatory switches between ordered and disordered states of the disordered segments. Here we describe a new method that accurately predicts τ_c for proteins with disordered segments by assuming two layers of molecular surfaces whose correlated velocities decay exponentially with distance. This method was extensively validated with 12 reference proteins that have 14 to 103 disordered residues at the N- and/or C-terminus and has been successfully employed to explain published results on calmodulin and the protein kinase Syk.

Methods

Boundary Element Method

The boundary element method for calculating hydrodynamic properties was originally proposed by Youngren and Acrivos[5] and recently reviewed by Aragon[7]. The essence of this method is briefly restated here. In a sticky boundary condition, which is relevant to biomolecules in aqueous solution, the velocity field of flow, $\mathbf{v}(\mathbf{y})$ at position $\mathbf{y}$ in the fluid, is described by spatial integration of the Oseen hydrodynamic interaction tensor ($\mathbf{H}$) from a particle surface element (dS_x) which experiences surface stress force ($f(\mathbf{x})$) at position $\mathbf{x}$ in the fluid (eq. 1). This hydrodynamic interaction is inversely proportional to the distance of interaction ($|\mathbf{x}-\mathbf{y}|$) (eq. 2). In the eq 2, $\mathbf{I}$ is the identity matrix and η is the viscosity of solution.

$$\mathbf{v}(\mathbf{y}) = \int_S \mathbf{H}(\mathbf{x},\mathbf{y}) \cdot \mathbf{f}(\mathbf{x})\, dS_x \tag{1}$$

$$\mathbf{H}(\mathbf{x},\mathbf{y}) = \frac{1}{8\pi\eta\,|\mathbf{x}-\mathbf{y}|}\left[\mathbf{I} + \frac{(\mathbf{x}-\mathbf{y})\otimes(\mathbf{x}-\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^2}\right] \tag{2}$$

For N finite surface elements, the purely shape-dependent terms, $\mathbf{H}$ and dS can be rewritten as a 3N×3N matrix $\mathbf{G}$ in three-dimensional Cartesian coordinates (eqs. 3 and 4) and the previously unknown 3N×1 surface stress force matrix, $\mathbf{f}$, can be calculated if the inverted matrix of $\mathbf{G}$ and a 3N×1 velocity field matrix, $\mathbf{v}$ (eq. 5) are given.

$$\mathbf{v}(\mathbf{y}_k) = \sum_{j=1}^{N} \mathbf{G}_{kj} \cdot \mathbf{f}_j \tag{3}$$

$$\mathbf{G}_{kj} = \int_{\Delta_j} \mathbf{H}(\mathbf{x}, y_k)\, dS_x \tag{4}$$

$$\mathbf{f}_j = \sum_{k=1}^{N} \mathbf{G}_{jk}^{-1} \mathbf{v}(\mathbf{y}_k)$$

(5)

In the case of a rigid particle in which all the surface elements translate and rotate as a single entity, the velocity field matrix can be determined from any given translational ($\mathbf{v}_p$) and angular ($\boldsymbol{\omega}_p$) velocities of the whole body. The force and torque exerted by translation and rotation of the particle in the fluid are vector sums of the individual force and torque of the surface elements. They are also directly related to velocities $\mathbf{v}_p$ and $\boldsymbol{\omega}_p$ through friction tensors (3×3 matrices), $\mathbf{K}$ (eqs. 6 and 7). The rotational ($\mathbf{D}_{rr}$) and translational ($\mathbf{D}_{tt}$) diffusion tensors (3×3 matrices) are then obtained from these friction tensors at a given temperature (T) (eq. 8). On the left, the 6×6 matrix of the rotational diffusion tensors, $\mathbf{D}_{tt}$, $\mathbf{D}_{rr}$, and $\mathbf{D}_{tr}$ correspond to the translational effect alone, the rotational effect alone, and the coupling between them arising from the screw-like properties of particles, respectively[11] (eq. 8). The $\mathbf{D}_{rt}$ matrix is assumed equal to the transpose of the $\mathbf{D}_{tr}$ matrix. The same notations were used for the 6×6 matrix of friction tensors on the right (eq. 8).

$$\text{Force} = \sum_{j=1}^{N} \mathbf{f}_j(\mathbf{x})\,\Delta j = -\mathbf{K}_{tt} \cdot \mathbf{v}_p - \mathbf{K}_{tr} \cdot \omega_p$$

(6)

$$\text{Torque} = \sum_{j=1}^{N} \mathbf{x}_p \times \mathbf{f}_j(\mathbf{x})\,\Delta j = -\mathbf{K}_{rt} \cdot \mathbf{v}_p - \mathbf{K}_{rr} \cdot \omega_p$$

(7)

$$\begin{pmatrix} D_{tt} & D_{tr}^T \\ D_{tr} & D_{rr} \end{pmatrix} = k_B T \begin{pmatrix} K_{tt} & K_{tr}^T \\ K_{tr} & K_{rr} \end{pmatrix}^{-1}$$

(8)

The eigenvectors of $\mathbf{D}_{tt}$ and $\mathbf{D}_{rr}$ represent the principal axis of the translational and rotational diffusion tensors, respectively. The rotational correlation time (τ_c) is derived from the average (D_{iso}) of three eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) of the $\mathbf{D}_{rr}$ (eq. 9).

$$\tau_c = \frac{1}{6 D_{iso}} = \frac{1}{2(\lambda_1 + \lambda_2 + \lambda_3)}$$

(9)

Ensemble Generation

Ensemble conformers were generated using a previously described method[12] with the following modifications. Since deposited PDB coordinates do not generally include the disordered flexible parts, the coordinates of any flexible parts were modeled by using the molecular modeling toolkit (MMTK)[13] while retaining the coordinates of the structured part. Conformational heterogeneity of disordered segments was achieved by rotating backbone and side chain dihedral angles of the modeled template structure. From the junction between the rigid and flexible parts toward the Nor C-terminus, the molecular coordinates were rotated according to a pair of backbone dihedral angles (ϕ–ψ) and side chain dihedral angles (χ_1–χ_4) which were randomly selected from an amino acid specific dihedral angle library. The ϕ–ψ angle library was built from 500 low homology and high resolution x-ray structures[14]

(resolutions < 1.8 Å) excluding all residues in α-helices, β-sheets and turns determined by the DSSP[15]. The χ angles library was adopted from a published rotamer library[14]. Residues immediately preceding proline were treated as an additional amino acid type, due to the restricted local conformation[16]. Overlap of van der Waals radii was evaluated in each step and the random selection of dihedral angles was repeated for each residue up to 10 times until there was no van der Waals clash between flexible and rigid parts. If clashes remained, the random selection of dihedral angles was restarted from the previous residue. For calmodulin and the protein kinase Syk, which have two domains connected by a disordered region, simultaneous random selections of all flexible dihedral angles and the evaluation of overall van der Waals clash were repeated until the selected conformation causes no van der Waals clash.

Molecular Surface Calculation

Triangulated molecular surfaces were generated by the MSMS program[17] using 1.1 Å inflated van der Waals radii to account for the uniform hydration thickness of the protein molecules in aqueous solution[8]. In excess of 10,000 triangulated surfaces were calculated (as defined in .vert and .face output files of the MSMS program) and converted to 700-1200 surface elements by the COARSEN utility included in the GNU triangulated surface (GTS) library. The boundary of the rigid segment was determined by noting the steady decrease of $[^1H]$-^{15}N NOE and ^{15}N R_2 relaxation rates from the plateau values of the rigid segment. The experimental rotational correlation times (τ_c^{exp}), determined from ^{15}N R_2/R_1 ratios, were taken from the literature[12, 18-27].

Boundary Element Calculation

All calculations were performed using an in-house C program BE2 which reads the two layers of molecular surfaces, evaluates their correlated velocity factors with given γ and ε parameters (eq. 11), and feeds them into the standard boundary element method calculations to get rotational and translational diffusion tensors for flexible proteins. Numerical routines were adopted from the GNU scientific library (GSL). The implementation of the standard boundary element method in the program was extensively tested with fully rigid proteins (Table S1). The predicted tumbling times are in excellent agreement with the experimental data (Figure S1). For the calculation of flexible proteins, the number of instantaneous or flexible surface elements was 700 to 1200 and that of rigid surface elements was about 10,000, since computational cost mainly depends on the number of instantaneous boundary elements. To minimize possible artifacts due to boundary element modeling and coarsening procedures, extrapolation to an infinite number of surface elements was performed using multiple calculations with varying numbers of surface elements. There was no significant difference between the extrapolation from six calculations using 700, 800, 900, 1000, 1100, and 1200 surface elements and a single calculation using 800 surface elements (Table 2; Table S2 and Figure S2, method a-ext). The results from a single calculation using 800 surface elements are therefore presented in Table 1. In order to find optimal γ and ε parameters for a general prediction, absolute deviations of predicted tumbling times (τ_c^{pred}) from experimental data ($\tau_c$exp) were averaged over 1,000 ensemble structures (N=1,000) and 12 reference proteins (M=12) for each combination of γ and ε parameters (eqs. 10 and 11). A γ and ε set with the minimum mean deviation was chosen as the optimal set of γ and ε parameters.

$$\text{mean dev.} (\gamma,\varepsilon) = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} \left| \tau_c^{exp}(m) - \tau_c^{pred}(m,n,\gamma,\varepsilon) \right|}{MN}$$

(10)

Results and Discussion

Ensemble approach to the boundary element method

In the classical boundary element method for rigid particles, all the surface elements translate and rotate as a single entity and the velocity field matrix ($\mathbf{v}$) can be determined immediately if the translational ($\mathbf{v}_p$) and angular ($\boldsymbol{\omega}_p$) velocities of the whole body are given (eqs. 1-5). However, in the case of flexible molecules, surface patches related to the unfolded chain do not translate and rotate synchronously with other rigid patches. Consequently, the velocity field matrix is not directly related to the velocities $\mathbf{v}_p$ and $\boldsymbol{\omega}_p$ of the whole molecule, which makes it impossible in practice to derive definite values for D_R ($=1/\tau_c$) and D_T.

$$\mathbf{v}(\mathbf{y}_k) = \mathbf{v}(\mathbf{y}'_k) \, \exp\left(-\frac{\delta - \gamma}{\varepsilon}\right), (\delta \geq \gamma \geq 0, \varepsilon > 0) \tag{11}$$

To resolve this problem, a sub-boundary ("rigid surface") is introduced inside the real instantaneous molecular boundary (Figure 2), such that the rigid surface encloses a least set of common atoms within the heterogeneous ensemble structures. It is then hypothesized that the velocity of an instantaneous surface patch ($v(y_k)$) is correlated with that of the closest rigid surface patch ($v(y'_k)$) by some factor which decays with distance (δ) between two surface patches (eq. 11 and Figure 2, distance a). That is, disordered parts far from the rigid surface would be less likely to translate and rotate synchronously with the rigid part (Figure 1). As the δ in eq. 11, the distance to the closest rigid surface element (Figure 2, distance a in Å) provides a more accurate prediction of τ_c than alternatives such as a spatial distance (Figure 2, distance b in Å) or residue number distance (Figure 2, distance c in residue) to the beginning of the flexible part (Table S2 and Figure S2). For the calculations of the distances b and c, each surface element is correlated with a specific residue based on the atomic coordinates in the relevant member of the disordered ensemble. The conformational space accessible to the disordered part, especially close to the rigid surface, will be constrained not only by the degrees of freedom of dihedral angles or chain length but also by the shape of the molecular surface around the beginning of the flexible part. The distance to a closest surface element of the rigid portion of the protein is thus more appropriate for describing the velocity correlation.

Since our major interest is in the tumbling time of the rigid part in the presence of disordered segments, the ensemble of disordered segments that affect the tumbling of the rigid part can be treated as a mean viscous medium surrounding the rigid part, by analogy with the effective medium theory in polymer dynamics[28]. In the absence of the disordered segments, the velocity field decreases in inverse proportion to the viscosity of the solution and the distance of interaction ($|\mathbf{x}-\mathbf{y}|$), as described by the Oseen equation (eq. 2). On the other hand, in the presence of the disordered segments, the additional effective viscosity caused by the dragging force of the disordered ensemble disturbs the velocity field, resulting in an effect similar to the screening of the hydrodynamic interaction in a concentrated polymer solution. The velocity field decreases rapidly and the hydrodynamic interaction between two points separated by a distance that is larger than a certain characteristic length termed the hydrodynamic screening length[28] becomes negligible.

In our description of the perturbation of the velocity field by the disordered ensemble (eq. 11), ε is a modulator of the decay rate for velocity correlation and γ is a generous distance margin that in effect expands the given surface boundary and therefore compensates for the ambiguity in the definition of the rigid surface. More importantly, γ also accounts for the differences in the nature of the hydration between the rigid and disordered parts.

Hydration of the flexible surface

The classical boundary element method provides a unified description of specific volume, translational and rotational diffusion coefficients with a single universal hydration parameter for globular proteins[7,8]. Numerical calculations using beads and boundary element modeling assume a solvent excluded molecular surface surrounded by a uniform thickness of hydration. For both methods, an essentially identical consensus thickness of the hydration layer ($\approx$1.2 Å in the bead modeling study[29] and 1.1 ± 0.1 Å in the boundary element modeling study[8]) yields hydrodynamic properties in excellent agreement with experiment. Our implementation of the boundary element method[7] by using 700 to 1200 triangulated surface elements derived from atomic coordinates of crystallographic or NMR structures and a single uniform hydration layer thickness of 1.1 Å also gives excellent agreement with experimental data from 35 fully folded globular proteins whose experimental τ_c range from 2 ns to 125 ns (Table S1 and Figure S1).

On the other hand, the hydration thickness for an unfolded disordered protein chain is not as well understood in detail as for globular proteins. Since the molecular surface of the disordered protein chain is largely exposed to solvent, interactions with surrounding water molecules should be different for an unfolded chain and for a folded globular protein domain. Recent NMR relaxation studies have revealed that the hydration of an unfolded protein is significantly higher than that of a globular protein of similar size[30,31]. The activation energy for the dynamics of the most strongly bound part of the hydration layer is about 50% larger for an unstructured protein than for a globular protein[30]. Therefore, the effective hydration layer for the unfolded chain will be thicker than for the folded part, and the general assumption of a uniform hydration layer should be corrected for the unfolded chain. In our method, the thicker hydration layer for flexible boundary elements may be considered in the calculation of velocity field correlation by the γ parameter, which effectively reduces the distance between rigid and flexible surface elements.

Prediction of rotational correlation times

Applying our assumption of two surfaces and velocity correlation to the boundary element method, tumbling times were calculated with arrays of γ and ε values for 12 reference proteins which have 14 to 103 disordered segments at the N- and/or C-terminus, represented by 1,000 ensemble structures. The average deviation from the experimental tumbling time ($|\tau_c\text{pred}-\tau_c\text{exp}|$) for 1,000 ensemble structures was represented by a color code for each protein (Figure 3). Blue regions display good agreement between prediction ($\tau_c\text{pred}$) and experiment ($\tau_c\text{exp}$). A wide range of γ and ε parameters results in good agreement for proteins with relatively short disordered segments. In contrast, proteins with long disordered segments require a more restricted range of γ and ε parameters for the same degree of agreement. This is consistent with the expected behavior of proteins with short disordered segments, which should be more similar to rigid proteins than those with long disordered segments: any combination of γ and ε will produce a good result in the limit of a fully rigid protein. Surprisingly, there exists a consensus range of γ and ε parameters [γ:4-8 Å, ε:20-24 Å] that give good predictions for all proteins tested. The consensus range was obtained by averaging the absolute deviation ($|\tau_c\text{pred}-\tau_c\text{exp}|$) at each combination of γ and ε parameters over the 1,000 ensemble structures and 12 reference proteins (Table 2). Although γ and ε are not completely independent, a non-zero value of γ is necessary for minimum mean deviation from the experimental data. Since the main purpose of the parameter γ is to correct the hydration thickness of disordered and rigid surface boundaries in the calculation of the velocity correlation, γ should be interpreted only within a reasonable range. The $\tau_c\text{pred}$ obtained by using the consensus set of γ and ε parameters and an ensemble average of 1,000 structures gives $|\tau_c\text{pred}-\tau_c\text{exp}| \approx 0.6$ (ns) (Table 2), which is sufficiently accurate for most purposes.

That a consensus is obtained in this case indicates that the relation of the correlated velocities between the disordered chain and the rigid surfaces of any given member of ensemble is essentially the same for all proteins tested. More importantly, the average stochastic influence of the disordered segments on the overall rotational tumbling of the rigid part can be described in a systematic way, regardless of the amino acid sequence or length of the disordered segments.

In contrast to our method, application of the empirical SE relation[1] to proteins with disordered terminal segments always underestimates τ_c (Figure 4). This is because of the general preference of disordered segments for extended conformations; proteins containing disordered segments occupy a larger effective volume and therefore have lower density than rigid globular proteins of equivalent molecular weight. Since the SE relation is density-based, it is biased in this case. By contrast, application of the conventional boundary element method, where each ensemble structure is assumed to be rigid, always overestimates the τ_c because the effective volume of the molecule is smaller than expected due to the flexibility of the disordered chain. In addition, the resulting distribution of tumbling is substantially broader (blue curves in Figure 5), and the maximum of the distribution occurs at significantly greater τ_c values than predicted by our method or measured experimentally.

Conformational sampling and averaging of the ensemble

The conformation of the disordered segment in our method is represented by a set of ensemble structures generated by applying backbone ($\phi\phi-\psi$) and side chain dihedral angles (X_1-X_4) randomly selected from an amino acid specific coil library, rejecting conformations that exhibit steric clashes. Sets of ensemble structures generated from this coil library have provided good representations of the unfolded state of proteins in the interpretation of residual dipolar coupling (RDC) and small angle X-ray scattering (SAXS) data[12]: ensemble structures were generated by adding consecutive peptide planes and tetrahedral junctions in the inverse direction to the protein sequence, starting from the terminus of the folded domain and progressing to the terminus of the flexible domain. Apart from this minor difference in implementation of the coil library, the sampled conformational space or conformational preference for disordered segments should be essentially the same in the published version[12] as obtained by our ensemble generation method.

The mean value for conformation-dependent tumbling times can be obtained by an ensemble average according to a probability distribution of the conformations. In the disordered state of proteins, the free energy is dominantly affected by chain elasticity, excluded volume or steric exclusion and solvent interaction[32]. Since all ensemble structures generated in our method are supposed to have the same bond lengths, bond angles and excluded volume, the internal energy and thus probability for each ensemble structure can be assumed equal, and the conversion from one conformation to another might be expected to be faster than tumbling because of the zero or low energy barrier between conformations. In principle, the tumbling times in different conformations are averaged by two mechanisms, depending on the rate of interconversion between substates. If the interconversion is faster than the time scale of tumbling, all molecules would experience a similar mean tumbling time, even while averaging was occurring at an individual molecule level. On the other hand, if the conversion is slower than the time scale of tumbling, molecules would experience different tumbling times; in this case, the ensemble average would be applicable. When interconversion is rapid, which is the most likely case for a disordered ensemble, the averaging is performed over the time for an individual molecule, but it will be the same as the averaging performed over ensemble conformations if the duration of measurement or observation is long enough that a system accesses all the substates with equal probability. The various methods for measuring the rotational tumbling time, such as NMR relaxation, electric birefringence, fluorescence depolarization, or dielectric relaxation, differ in their sensitivity to motions on different time scales. For example, NMR relaxation

measurements (R_1 and R_2) are relatively insensitive to motions slower than the tumbling time, as indicated by long time scale molecular dynamics[33]. However, the bias in the comparison between the ensemble average and the NMR measurement should be negligible since the chance of interconversion on the time scale of tumbling would be very small. Furthermore, the calculated tumbling times for ensemble conformations are sufficiently homogeneous that the possible difficulties arising from the ensemble average under the influence of an ambiguous interconversion rate may be ignored (Figures 5 and 6). The distribution of tumbling times calculated for each ensemble structure is narrow, with its maximum occurrence found at the ensemble averaged mean tumbling time (Figure 5). This is consistent with the dominant influence on the tumbling of the rigid portion of the protein of the flexible surface elements near the rigid boundary, which, unlike boundary elements located further away from the interface, do not change their structure much between individual conformers in the ensemble of the flexible portion. In accordance with the narrow distribution of calculated tumbling times, a relatively small number of ensemble structures (~1,000) is sufficient to achieve practical convergence (Figure 6). Global parameters of a disordered ensemble, such as tumbling time, size, and shape, appear to require a smaller number of ensemble structures (1,000-2,000) for convergence than do local parameters such as bond vector orientations (measured by residual dipolar couplings), which require up to 50,000 ensemble structures for convergence[12].

Generalization for multi-domain proteins with a disordered linker region

The validation described above was carried out for proteins with N- and/or C-terminal disordered segments. However, the relation of velocity fields between the rigid and the instantaneous surfaces (eq. 11) should also be valid for multi-domain proteins connected by a disordered linker because no assumption was made on the conformation of the disordered chains to describe the relation: the domains can be regarded as special conformations formed by the disordered segments. In an ensemble for such a protein, each domain would retain its fold but overall conformation would fluctuate due to the heterogeneous conformations of the linker. For each ensemble structure, the τ_c of a specific domain can be calculated by defining the rigid surface to enclose the domain of interest and the instantaneous surface to enclose the whole molecule. Once the rigid and instantaneous surfaces are defined, the subsequent boundary element calculations are identical to those for the proteins with terminally disordered segments.

Two domains connected by a disordered linker affect the tumbling of each other. However, this coupling of tumbling becomes negligible as the length of the disordered linker increases beyond a certain length. The velocity field calculations adopting the residue number distance to the beginning of the flexible part as δ (Figure 2, distance c; eq. 11) show that two velocity fields separated by about 20 and 45 disordered residues would be correlated to about 36% ($\approx 1/e$) and 10%, respectively (Table S2 and Figure S2). In the disordered segment, each residue is at a different distance from the beginning of the flexible part and its contribution to the tumbling of the rigid part is additive, but diminishes as the distance increases. For example, τ_c of the rigid globular portion of the Syrian hamster prion protein is 7.4 ns (calculated) and the addition of 44 disordered residues increases the τ_c by 6.2 ns (measured), but addition of a further 59 disordered residues only increases the τ_c by another 2.4 ns (measured) (Table 1). Although the residue number distance (Figure 2, distance c) is not the best predictor for τ_c, two domains connected by a disordered interdomain linker of about 45 or more residues may be regarded as decoupled within the approximate limit of projected error (1.1 ns for the method using the distance c).

Application to calmodulin

To validate the application of the method to a protein containing rigid domains separated by a flexible region, we performed calculations for calmodulin (CaM). CaM plays a critical role in

Ca^{2+}-dependent signaling pathways and has two globular Ca^{2+} binding domains connected by a flexible linker[36,37]. The flexibility of the linker is essential for CaM to rearrange the Ca^{2+} binding domains and bind tightly to numerous target proteins. NMR relaxation data for the Ca^{2+}-ligated *Drosophila* CaM showed that the tumbling times for the N- and C-domains are 7.1 and 6.3 ns at 35°C, respectively; at 20°C the corresponding correlation times would be 10.4 and 9.2 ns for the N- and C-domains, respectively[37]. On the basis of the crystal structure of the Ca^{2+}-ligated *Drosophila* CaM (PDB id: 4CLN)[38], the calculated tumbling times for the full-length protein (residues 1-148; dumbbell-like overall conformation), and the isolated N-domain (residues 4-76) and C-domain (residues 82-147) are 14.1, 4.9, and 4.3 ns at 20°C, respectively. The measured tumbling times are significantly shorter than the tumbling time expected for the dumbbell-like conformation, but are much longer than expected for the isolated domains, suggesting that the tumbling of the N- and C-domains is partly decoupled due to the flexibility of the interdomain linker. However, it remains unclear whether the measured tumbling times are explained by complete or partial disorder of the linker. We therefore applied our method to estimate the τ_c of the CaM, assuming that the linker is completely disordered. An ensemble for the full-length CaM was generated with residues 1-3, 77-81, and 148 being completely disordered while retaining the N- and C-domain folds as in the crystal structure (number of ensemble structures 1,000). The boundary element calculations were performed separately for the N- and C-domains using the consensus parameter set [γ:4-8 Å, ε:20-24 Å]. For instance, the rigid surface was defined to enclose the N-domain when the τ_c is calculated for the N-domain. The tumbling times for the N- and C-domains calculated using our approach (10.1 and 9.5 ns at 20°C, respectively) are in excellent agreement with the experimental data, suggesting that the interdomain linker is completely disordered in the Ca^{2+}-ligated form of CaM.

Applications to an order-to-disorder switch

As an example of the utility of our prediction method, we describe its application to obtain insights into the structural basis for an order-to-disorder regulatory switch in the protein tyrosine kinase Syk[34]. The Syk kinase plays a important role in signaling through the antigen B cell receptor, binding the receptor through two tandem SH2 domains connected by a folded 45 amino acid interdomain A. Upon phosphorylation of Tyr-130 in the interdomain A, Syk dissociates from the receptor. On the basis of changes in the rotational tumbling times of the two SH2 domains in wild-type Syk (τ_c=19.2 ns, derived from NMR relaxation measurements) and in a Tyr130Glu mutant that mimics the phosphorylated form (τ_c=12.1 ns), Zhang et al. [34] suggested decoupling of the two SH2 domains in the phosphorylated Syk by a mechanism involving disordering of interdomain A. Predicted values of τ_c for the isolated SH2 domains (7.5 and 7.6 ns for the N- and C-terminal SH2 domains, respectively, estimated using a beads approximation) provided qualitative support for the mechanistic model[34]. However, since the predicted and measured t_c values differ so significantly, it is unclear whether the decreased rotational tumbling time in the Glu130 mutant results from complete or partial disordering of Syk. We therefore applied our method to estimate rotational correlation times in this system. Following the same procedures used for the 12 reference proteins, three protein models were constructed based on the crystal structure of Syk[35] (PDB id: 1A81): (i) full-length Syk containing both SH2 domains and the folded interdomain A (residues 2-262; rigid residues 11-260; number of ensemble structures 1,000); (ii) the N-terminal SH2 domain plus interdomain A (residues 2-159; rigid residues 11-114; number of ensemble structures 1,000); and (iii) interdomain A plus the C-terminal SH2 domain (residues 115-262; rigid residues 160-260; number of ensemble structures 1,000). In the full-length Syk (model (i)), interdomain A (residue 115-159) was assumed to be ordered and globular, while in models (ii) and (iii) the interdomain A was assumed to be completely disordered, leading to an assumption that the two SH2 domains are decoupled. The boundary element calculations were performed using the consensus parameter set [γ:4-8 Å, ε:20-24 Å]. The resulting τ_c for the full-length Syk, and

the N- and C-terminal SH2 domains with disordered interdomain A are 20.4 ns, 12.6 ns, and 11.3 ns, respectively. Similar boundary element calculations were performed with a full-length Syk ensemble in which the interdomain A is completely disordered while both SH2 domain conformations are retained (number of ensemble structures 1,000), so as to verify whether the 45 disordered residues of the interdomain A are enough to decouple the two domains. The resulting τ_c for the N- and C-terminal SH2 domains are 13.3 ns and 13.0 ns, respectively, which are close to the values based on the decoupling assumption made in the models (ii) and (iii), implying that the two SH2 domains are almost, if not completely, decoupled. These values are in excellent agreement with the experimental data in the limit of projected error of prediction (≈ 0.6 ns), which strongly supports the suggestion that the interdomain A behaves like a disordered chain and that the two SH2 domains are decoupled in the Tyr-130 phosphorylated state[34].

Concluding remarks

An ensemble approach to the boundary element method has been applied to the problem of molecular tumbling of proteins with long disordered segments. In the presence of disordered segments, the tumbling of the folded, rigid portion becomes slower than it would be for this part alone. The dragging effect can be described in a systematic way regardless of the sequence or length of the disordered segments, using the assumption of two surface boundaries and the specific relation between them. In general, the presence of disordered residues farther than a defined cutoff distance from the rigid part has a negligible effect on the tumbling of the rigid domain, although the disordered regions do contribute to the heterogeneity of the ensemble conformations. These observations explain the significant discrepancies between experimental data and values calculated based on the assumption that each ensemble conformation is rigid. Extension of the classical boundary element method or bead modeling into the problem of a disordered ensemble is largely limited by poor understanding of the motional coupling between the rigid and disordered segments and the hydrodynamic nature of the disordered chain. Nevertheless, the tumbling times predicted by our approach are in excellent agreement with experiment. This prediction method is therefore sufficiently accurate that it can be utilized to identify important biological events such as inter- and/or intra-molecular interactions between unfolded and folded regions of a protein and to probe the mechanistic basis of order-disorder switches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Halle B, Davidovic M. Proc. Natl. Acad. Sci U. S. A 2003;100:12135–12140. [PubMed: 14528004]

2. Kirkwood JG, Riseman J. J. Chem. Phys 1948;16:565–573.

3. Bloomfield VA. Science 1968;161:1212–1219. [PubMed: 4876948]

4. Garcia de la Torre J, Huertas ML, Carrasco B. J. Magn Reson 2000;147:138–146. [PubMed: 11042057]

5. Youngren GK, Acrivos A. J. Fluid Mech 1975;69:377–403.

6. Allison SA. Biophys. Chem 2001;93:197–213. [PubMed: 11804726]

7. Aragon S. J. Comput. Chem 2004;25:1191–1205. [PubMed: 15116362]

8. Aragon S, Hahn DK. Biophys. J 2006;91:1591–1603. [PubMed: 16714342]

9. Dyson HJ, Wright PE. Nat. Rev. Mol. Cell Biol 2005;6:197–208. [PubMed: 15738986]

10. Dunker AK, Brown CJ, Obradovic Z. Adv. Protein Chem 2002;62:25–49. [PubMed: 12418100]

11. Brenner H. J. Colloid Interface Sci 1967;23:407–436.

12. Bernado P, Blanchard L, Timmins P, Marion D, Ruigrok RW, Blackledge M. Proc. Natl. Acad. Sci. U. S. A 2005;102:17002–17007. [PubMed: 16284250]

13. Hinsen K. J. Comp. Chem 2000;21:79–85.

14. Lovell SC, Word JM, Richardson JS, Richardson DC. Proteins 2000;40:389–408. [PubMed: 10861930]

15. Kabsch W, Sander C. Biopolymers 1983;22:2577–2637. [PubMed: 6667333]

16. MacArthur MW, Thornton JM. J. Mol. Biol 1991;218:397–412. [PubMed: 2010917]

17. Sanner MF, Olson AJ, Spehner J-C. Biopolymers 1996;38:305–320. [PubMed: 8906967]

18. Ebert MO, Bae SH, Dyson HJ, Wright PE. Biochemistry 2008;47:1299–1308. [PubMed: 18177052]

19. Viles JH, Donne D, Kroon G, Prusiner SB, Cohen FE, Dyson HJ, Wright PE. Biochemistry 2001;40:2743–2753. [PubMed: 11258885]

20. Nicastro G, Margiocco P, Cardinali B, Stagnaro P, Cauglia F, Cuniberti C, Collini M, Thomas D, Pastore A, Rocco M. Biophys. J 2004;87:1227–1240. [PubMed: 15298925]

21. Hansen DF, Yang D, Feng H, Zhou Z, Wiesner S, Bai Y, Kay LE. J. Am. Chem. Soc 2007;129:11468–11479. [PubMed: 17722922]

22. Houben K, Blanchard L, Blackledge M, Marion D. Biophys. J 2007;93:2830–2844. [PubMed: 17586564]

23. Butterwick JA, Patrick LJ, Astrof NS, Kroenke CD, Cole R, Rance M, Palmer AG III. J. Mol. Biol 2004;339:855–871. [PubMed: 15165855]

24. Yao S, Headey SJ, Keizer DW, Bach LA, Norton RS. Biochemistry 2004;43:11187–11195. [PubMed: 15366928]

25. Zhukov I, Bayer P, Scholermann B, Ejchart A. Acta Biochim. Pol 2007;54:769–775. [PubMed: 17968439]

26. Stone MJ, Fairbrother WJ, Palmer AG III, Reizer J, Saier MH Jr. Wright PE. Biochemistry 1992;31:4394–4406. [PubMed: 1316146]

27. LiWang AC, Cao JJ, Zheng H, Lu Z, Peiper SC, LiWang PJ. Biochemistry 1999;38:442–453. [PubMed: 9890927]

28. Doi, M.; Edwards, SF. The theory of polymer dynamics. Clarendon Press; Oxford: 1986. p. 172-180.

29. Garcia de la Torre J. Biophys. Chem 2001;93:159–170. [PubMed: 11804723]

30. Bokor M, Csizmok V, Kovacs D, Banki P, Friedrich P, Tompa P, Tompa K. Biophys. J 2005;88:2030–2037. [PubMed: 15613629]

31. Csizmok V, Bokor M, Banki P, Klement E, Medzihradszky KF, Friedrich P, Tompa K, Tompa P. Biochemistry 2005;44:3955–3964. [PubMed: 15751971]

32. Dill KA, Shortle D. Annu. Rev Biochem 1991;60:795–825. [PubMed: 1883209]

33. Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, Arkin IT, Jensen MO, Xu H, Trbovic N, Friesner RA, Iii AG, Shaw DE. J Phys. Chem. B 2008;112:6155–6158. [PubMed: 18311962]

34. Zhang Y, Oh H, Burton RA, Burgner JW, Geahlen RL, Post CB. Proc. Natl. Acad. Sci U. S. A 2008;105:11760–11765. [PubMed: 18689684]

35. Futterer K, Wong J, Grucza RA, Chan AC, Waksman G. J Mol. Biol 1998;281:523–537. [PubMed: 9698567]

36. Tjandra N, Kuboniwa H, Ren H, Bax A. Eur. J Biochem 1995;230:1014–1024. [PubMed: 7601131]

37. Barbato G, Ikura M, Kay LE, Pastor RW, Bax A. Biochemistry 1992;31:5269–5278. [PubMed: 1606151]

38. Taylor DA, Sack JS, Maune JF, Beckingham K, Quiocho FA. J Biol. Chem 1991;266:21375–21380. [PubMed: 1939171]

39. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, Wright PE. Nature 2002;415:549–553. [PubMed: 11823864]

40. Gossert AD, Bonjour S, Lysek DA, Fiorito F, Wuthrich K. Proc. Natl. Acad. Sci U. S. A 2005;102:646–650. [PubMed: 15647363]

41. Bae S-H, Legname G, Serban A, Prusiner SB, Wright PE, Dyson HJ. manuscript in preparation. 2009

42. James TL, Liu H, Ulyanov NB, Farr-Jones S, Zhang H, Donne DG, Kaneko K, Groth D, Mehlhorn I, Prusiner SB, Cohen FE. Proc. Natl. Acad. Sci U. S. A 1997;94:10086–10091. [PubMed: 9294167]

43. Improta S, Politou AS, Pastore A. Structure 1996;4:323–337. [PubMed: 8805538]

44. Feng H, Zhou Z, Bai Y. Proc. Natl. Acad. Sci U. S. A 2005;102:5026–5031. [PubMed: 15793003]

45. Blanchard L, Tarbouriech N, Blackledge M, Timmins P, Burmeister WP, Ruigrok RW, Marion D. Virology 2004;319:201–211. [PubMed: 14980481]

46. Ishikawa K, Okumura M, Katayanagi K, Kimura S, Kanaya S, Nakamura H, Morikawa K. J Mol. Biol 1993;230:529–542. [PubMed: 8385228]

47. Headey SJ, Keizer DW, Yao S, Brasier G, Kantharidis P, Bach LA, Norton RS. Mol. Endocrinol 2004;18:2740–2750. [PubMed: 15308688]

48. Rak A, Kalinin A, Shcherbakov D, Bayer P. Biochem. Biophys. Res. Commun 2002;299:710–714. [PubMed: 12470636]

49. Chen Y, Case DA, Reizer J, Saier MH Jr. Wright PE. Proteins 1998;31:258–270. [PubMed: 9593197]

50. Li Y, Liu D, Cao R, Kumar S, Dong C, An J, Wilson SR, Gao YG, Huang Z. Proteins 2007;67:243–246. [PubMed: 17243149]
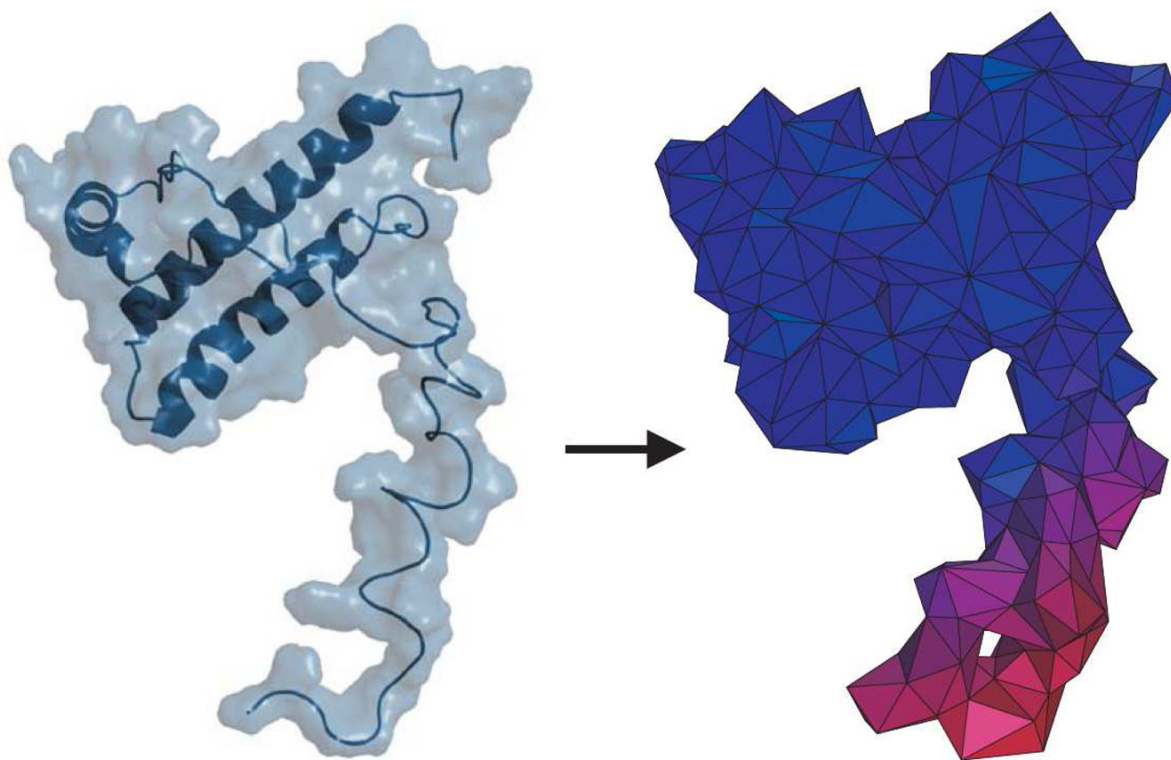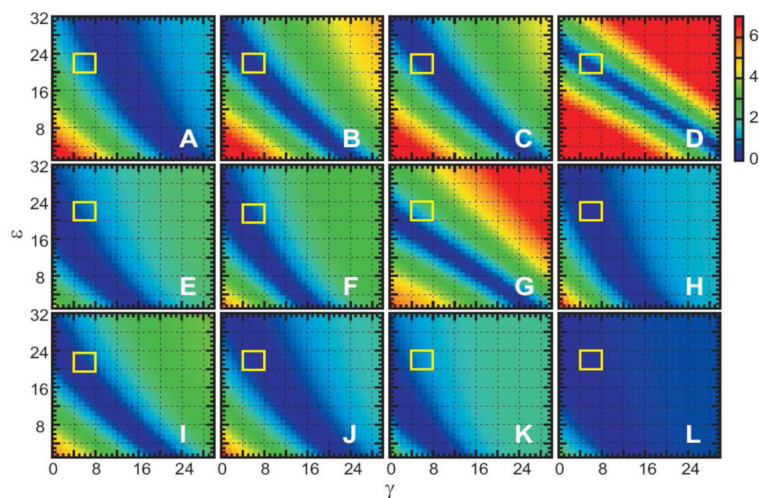
Figure 1.
Modeling the surface of an ensemble structure by the boundary elements. In the left, the ribbon diagram and the molecular surface are superimposed for an ensemble structure from the mouse prion protein (residues 89-230) which has 38 and 6 disordered residues at N- and C- terminus, respectively. In the right hand panel, the boundary elements (triangular patches) representing the molecular surface are illustrated. The factor for correlated velocity (eq. 11) is displayed on each surface patch by color gradient (blue:1→red:0). Note that the surface patches on the ordered part are blue and those on the disordered N terminus become increasingly red as their distances from the ordered part increase.
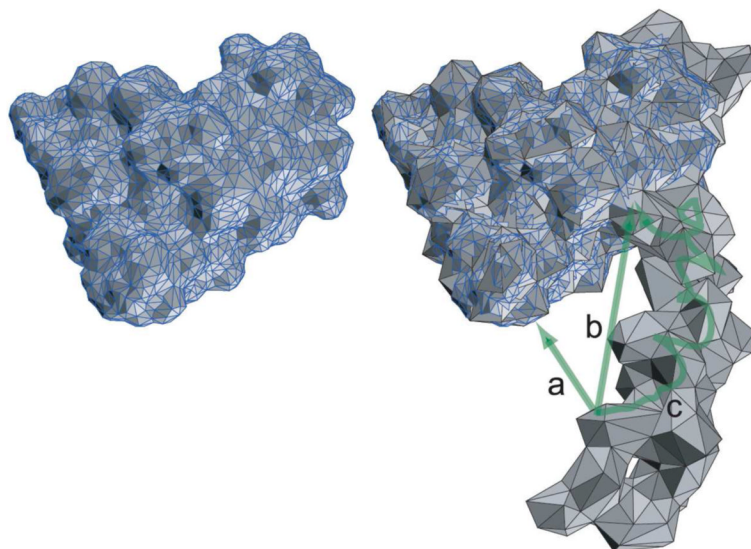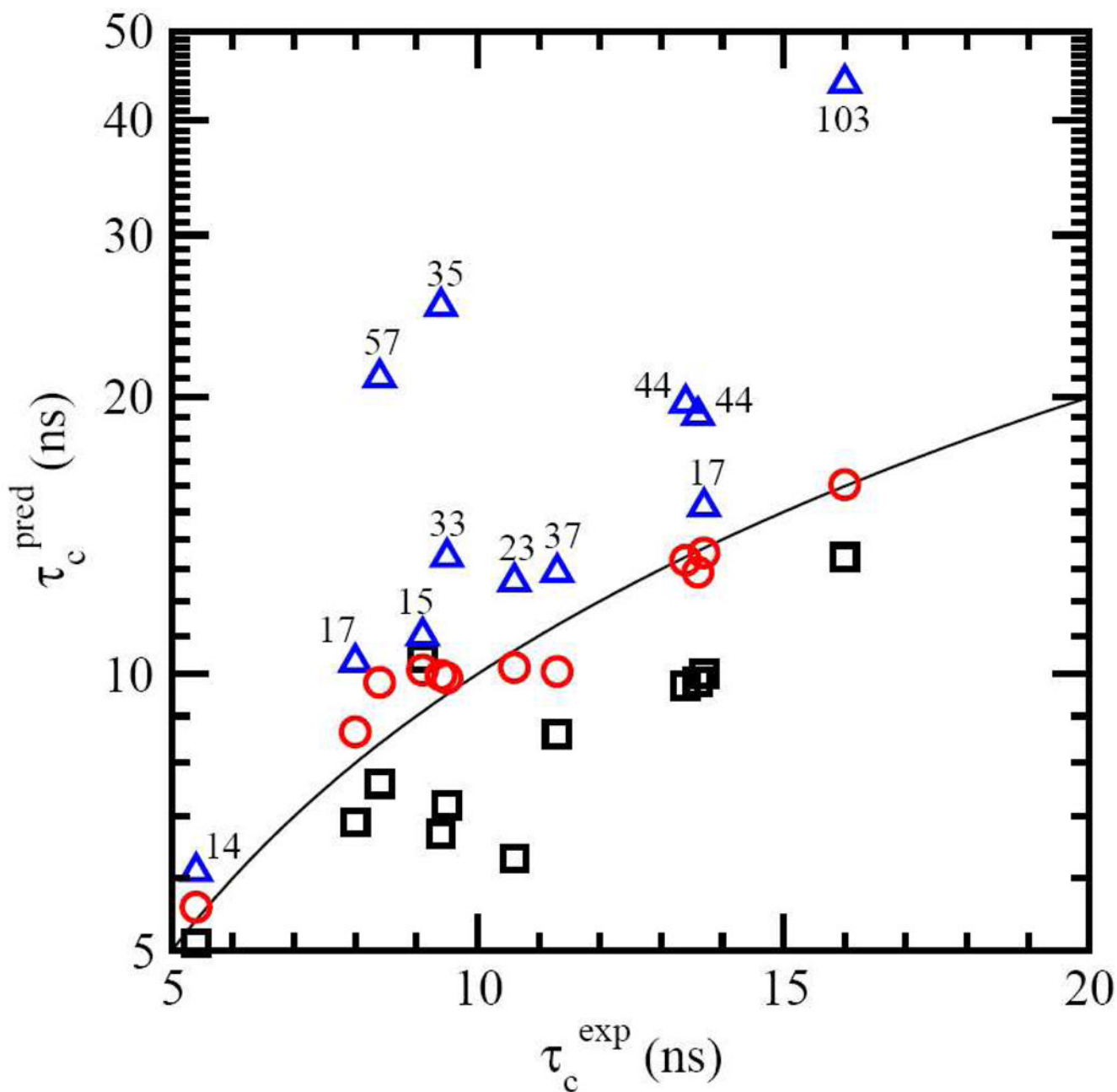
Figure 2.
Defining the rigid and instantaneous surfaces and their correlation distance (δ). A rigid surface (left, blue edges) which encloses a least set of common atoms within the heterogeneous ensemble structures and an instantaneous surface (right, black edges) which encloses one member of the ensemble structures superimposed on the rigid surface are illustrated for an ensemble structure of the mouse prion protein (residues 89-230). The molecular surfaces are modeled by a group of triangular patches for the boundary element calculations (2,400 and 1,200 triangular surface elements are displayed here for the rigid and instantaneous surfaces, respectively). The velocity of an instantaneous surface patch is hypothesized to be correlated with that of the closest rigid surface patch by some factor which decays with distance (δ) between two surface patches (eq. 11) (a). Alternatively, a spatial distance (b) or residue number distance (c) to the beginning of the flexible part may be used for the distance, δ. For the distances b and c, each surface element on both the rigid and instantaneous surfaces is correlated with a specific residue based on the atomic coordinates in the relevant member of the disordered ensemble.
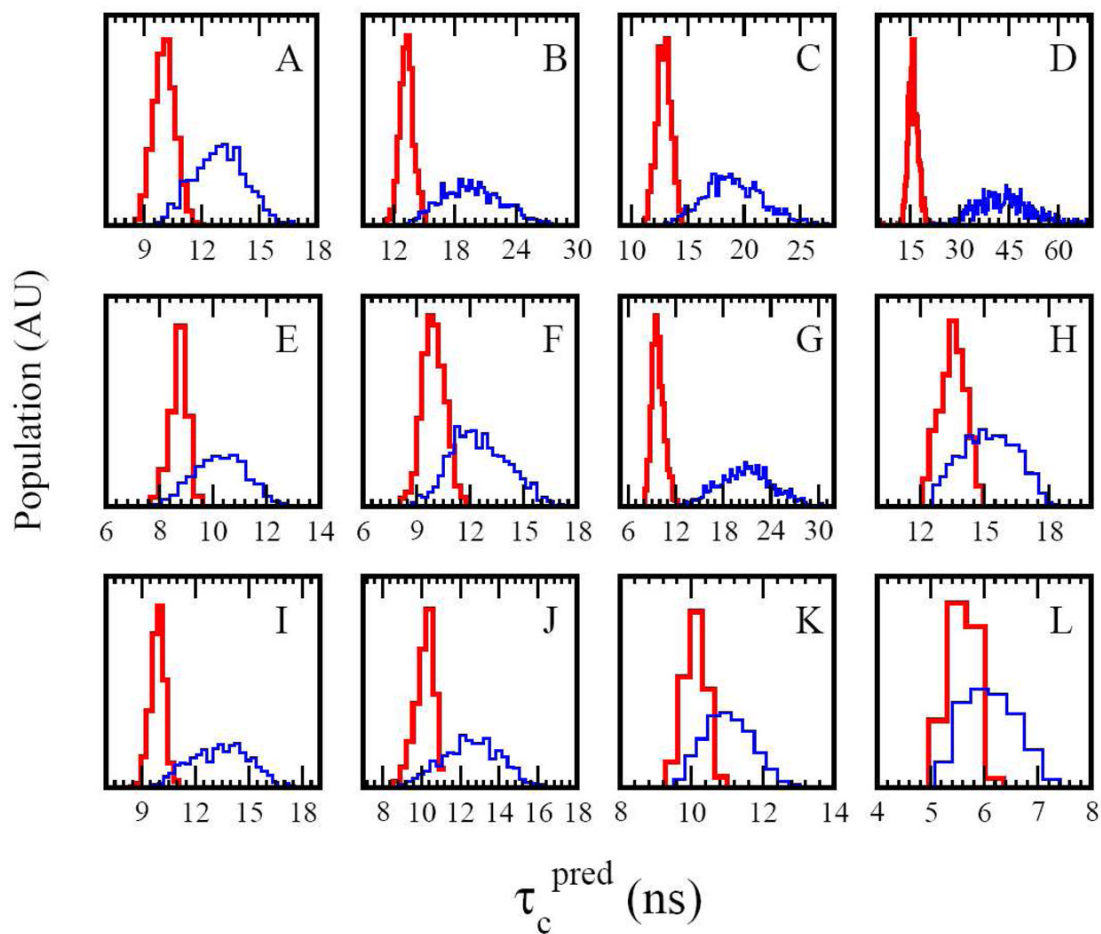
Figure 3.
Predictions of rotational correlation times ($\tau_c = D_R^{-1}$) with different sets of γ and ε parameters (eq. 11). $|\tau_c^{pred} - \tau_c^{exp}|$ is shown as color gradient (blue:0 ns → red:>7 ns). The consensus γ and ε parameters range is indicated by the yellow box. (A-L) 12 reference proteins, with the number of disordered residues indicated. (A) ACTR(1018-1088)/CBP(2059-2117):*37*, (B) MoPrP (89-230):*44*, (C) ShPrP(90-231):*44*, (D) ShPrP(29-231):*103*, (E) I27(1-103):*17*, (F) PUF1 (1-106):*35*, (G) PX(474-568):*57*, (H) ttRNH(1-166):*17*, (I) IGFBP-6(1-107):*33*, (J) Yfia (1-113):*23*, (K) IIAglc(1-162):*15*, (L) VMIP-II(1-74):*14*.
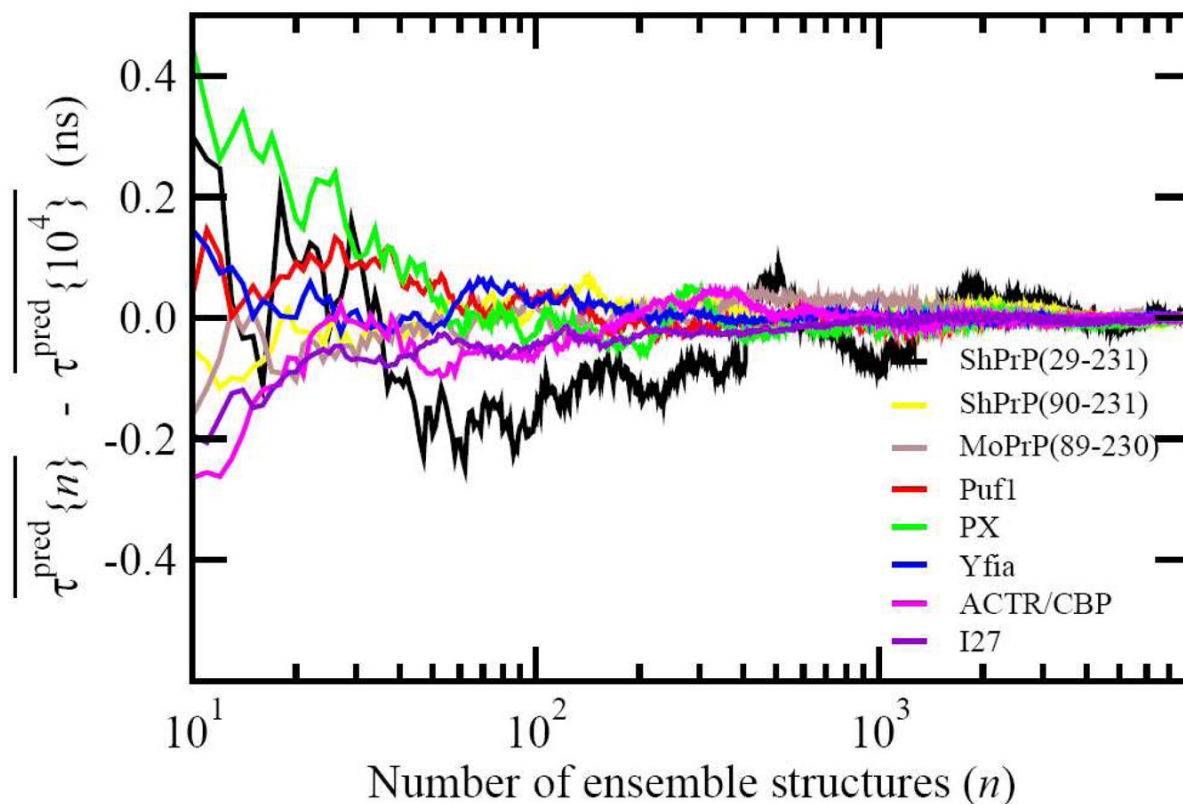
Figure 4.
Comparison of predicted and experimental correlation times. Predicted tumbling times ($\tau_c$pred) from the empirical Stokes-Einstein method (black square), conventional boundary element method with rigid assumption (blue triangle), and the method described in this study (red circle) are compared with the experimental data ($\tau_c$exp). The solid line represents perfect agreement between $\tau_c$pred and $\tau_c$exp. The number of disordered residues for each protein is indicated near the triangle symbol. A semi-log scale is used for the y-axis to cover the wide range of $\tau_c$pred.

Figure 5.
Distribution of the predicted rotational correlation times τ_c^{pred} for 1,000 ensemble structures. Red histograms display the distribution of τ_c^{pred} from the ensemble approach to the boundary element method using two layers of molecular surfaces and their correlated velocity relation. For comparison, the distribution of τ_c^{pred} from the conventional boundary element method assuming each ensemble structure is rigid is displayed in blue histograms. The x-axis shows the rotational correlation time in ns and the bin size is 0.3 ns in all histograms. The y-axis is in arbitrary units and represents the population at each τ_c^{pred} bin. (A) ACTR(1018-1088)/CBP (2059-2117), (B) MoPrP(89-230), (C) ShPrP(90-231), (D) ShPrP(29-231), (E) I27(1-103), (F) PUF1(1-106), (G) PX(474-568), (H) ttRNH(1-166), (I) IGFBP-6(1-107), (J) Yfia(1-113), (K) IIAglc(1-162), (L) VMIP-II(1-74).

Figure 6.
Convergence of the predicted rotational correlation time ($\tau_c$pred). The prediction was made from the ensemble approach to the boundary element method using two layers of molecular surfaces and their correlated velocity relation. The difference between the $\tau_c$pred averaged over n ensemble structures $\left(\overline{\tau_c^{pred}}\{n\}\right)$ and the $\tau_c$pred averaged over 10^4 ensemble structures $\left(\overline{\tau_c^{pred}}\{10^4\}\right)$ is plotted as a function of the number of ensemble structures (n). As the number of ensemble structures increases, the difference approaches toward zero, indicating convergence.

Table 1

Reference proteins for validation

Protein	Rigid part/ Disordered/ Total[a]	Exp[b]	SE[c]	BE2[d]	D_T[d]
(1) ACTR(1018-1088) / CBP (2059-2117)	(1043-1081/2059-2112) /37 / 130	11.3	8.6	10.1 (12.9)	11.3 (9.5)
(2) MoPrP(89-230)	127-224 / 44 / 142	13.4	9.7	13.1 (19.7)	10.8 (8.3)
(3) ShPrP(90-231)	128-225 / 44 / 142	13.6	9.8	12.7 (19.1)	10.9 (8.4)
(4) ShPrP(29-231)	128-225 / 103 / 203	16.0	13.4	16.1 (43.9)	11.9 (6.4)
(5) I27(1-103)	17-102 / 17 / 103	8.0	6.9	8.6 (10.3)	11.4 (10.2)
(6) PUF1(1-106)	25-95 / 35 / 106	9.4	6.7	9.9 (25.1)	11.3 (9.7)
(7) PX(474-568)	517-568 / 57 / 109	8.4	7.6	9.5 (21.0)	13.6 (8.1)
(8) ttRNH(1-166)	4-152 / 17 / 166	13.7	10.0	13.5 (15.2)	9.7 (9.0)
(9) IGFBP-6(1-107)	28-101 / 33 / 107	9.5	7.2	9.9 (13.4)	11.7 (9.4)
(10) Yfia(1-113)	3-92 / 23 / 113	10.6	6.3	10.1 (12.6)	11.1 (9.6)
(11) IIAglc(1-162)	15-161 / 15 / 162	9.1	10.4	10.1 (11.0)	10.5 (9.9)
(12) VMIP-II(1-74)	14-73 / 14 / 74	5.4	5.1	5.6 (6.1)	12.9 (12.1)

[a] Residues belonging to the rigid part, the number of the disordered residues including the unstructured leading sequence, and the number of total residues in the protein constructs are shown, respectively.

[b] Experimental rotational correlation times (τ_c^{exp}, ns) determined from N $^{15}R_2/R_1$ ratios were taken from literature. For comparison, these values were converted to the standard condition (20°C, water) assuming temperature dependent viscosity change is similar to water. (1) complex of p160 steroid receptor coactivator ACTR and general transcription activator CBP/p300 (pdb id: 1kbh[39], τ_c^{exp}: 8.5 ns at 31°C[18]; (2) Mouse prion protein (89-230) (pdb id: 1xyx[40], τ_c^{exp}: 11.7 ns at 25°C[41]; (3) Syrian hamster prion protein (90-231) (pdb id: 1b10[42], τ_c^{exp}: 10.5 ns at 30°C[19]; (4) Syrian hamster prion protein (29-231) (pdb id: 1b10[42], τ_c^{exp}: 12.3 ns at 30°C[19]); (5) immunoglobulin-like module from muscle protein titin plus a 14 leading sequence for histidine tag and factor Xa cleavage site (MHHHHHHSSIEGRG) (pdb id: 1tit[43], τ_c^{exp}: 8.0 ns at 20°C[20]); (6) redesigned apocytochrome b562 (pdb id: 1yzc[44], τ_c^{exp}: 10.8 ns at 15°C[21]); (7) Sendai virus phosphoprotein P plus a N-terminal 14 leading sequence for histidine tag and factor Xa cleavage site (MRGSHHHHHHIEGR) (pdb id: 1r4g[45], τ_c^{exp}: 7.3 ns at 25°C[12,22]); (8) *Thermus thermophilus* ribonuclease HI mutant (cystein-free) (pdb id: 1ril[46], τ_c^{exp}: 9.0 ns at 37°C[23]; (9) C-terminal domain of insulin-like growth factor (IGF) binding protein 6 plus 27 unstructured leading sequence (pdb id: 1rmj[47], τ_c^{exp}: 8.3 ns at 25°C[24]); (10) ribosome-associated cold shock response protein Yfia from *Escherichia coli* (pdb id: 1n3g[48], τ_c^{exp}: 8.8 ns at 27°C[25]); (11) *Bacillus subtilis* glucose permease IIA domain (pdb id: 1ax3[49], τ_c^{exp}: 6.2 ns at 35°C[26]); (12) viral macrophage-inflammatory protein-II (pdb id: 2fht[50], τ_c^{exp}: 4.7 ns at 25°C[27]).

[c] SE is the predicted rotational correlation time (τ_c^{pre}, ns) from the empirical Stokes-Einstein relation ($\tau_c^{-1} = k_BT/(8\pi\eta r^3)/2$) using the molecular weight and the consensus partial specific volume for globular proteins (0.73 cm³ g⁻¹).

[d] BE2 and D_T are the τ_c^{pred} and the predicted translational diffusion coefficient (10^{-7} cm² s⁻¹), respectively. They are obtained from the ensemble approach to the boundary element method using two layers of molecular surfaces and their correlated velocity relation which is described in this study. The values calculated from the conventional boundary element method, assuming that each ensemble structure is rigid, are shown in parentheses in the columns for BE2 and D_T.

Table 2

Mean deviations of predicted τ_c from the experimental data (ns)

		γ (Å)						
		0-4	4-8	8-12	12-16	16-20	20-24	24-28	28-32
	0-4	3.679	4.922	3.699	2.591	1.805	1.343	1.137	1.557
	4-8	4.791	3.584	2.481	1.650	1.209	1.062	1.492	2.012
	8-12	3.488	2.396	1.506	1.022	0.949	1.339	1.983	2.560
	12-16	2.378	1.454	0.833	0.798	1.227	1.902	2.488	2.997
ε (Å)	16-20	1.519	0.809	0.656	1.112	1.768	2.371	2.895	3.343
	20-24	0.911	**<u>0.582</u>**	0.991	1.602	2.211	2.751	3.220	3.621
	24-28	0.673	0.854	1.424	2.018	2.573	3.063	3.484	3.849
	28-32	0.847	1.237	1.788	2.363	2.873	3.317	3.702	4.033
	32-36	1.072	1.560	2.095	2.629	3.103	3.516	3.868	4.176

For optimal γ and ε parameters for a general prediction, absolute deviations of predicted tumbling times (τ_c^{pred}) from experimental data (τ_c^{exp}) were averaged over 1,000 ensemble structures and 12 reference proteins for each combination of γ and ε parameters (Table 1, eqs. 10 and 11). The minimum mean deviation indicated in bold and underlined is obtained with γ=4-8 Å and ε=20-24 Å.