# Reorientational Contact-Weighted Elastic Network Model for the Prediction of Protein Dynamics: Comparison with NMR Relaxation

Dengming Ming* and Rafael Brüschweiler†
*Computer and Computational Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545; and
†National High Magnet Field Laboratory, Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32310

ABSTRACT A new model for the prediction of protein backbone motions is presented. The model, termed reorientational contact-weighted elastic network model, is based on a multidimensional reorientational harmonic potential of the backbone amide bond vector orientations and it is applied to the interpretation of dynamics parameters obtained from NMR relaxation data. The individual energy terms are weighted as a function of the intervector distances and by the contact strengths of each bond vector with respect to its local environment. Correlated reorientational motional properties of the bond vectors are obtained by means of normal mode analysis. Application to a set of proteins with known three-dimensional structures yields good to excellent agreement between predicted and experimental NMR order parameters presenting an improvement over the local contact model. The reorientational eigenmodes of the reorientational contact-weighted elastic network model method provide direct information on the collective nature of protein backbone motions. The dominant eigenmodes have a notably low collectivity, which is consistent with the behavior found for reorientational eigenmodes from molecular dynamics simulations.

## INTRODUCTION

Over recent years, there has been considerable interest in the characterization of protein dynamics because of its importance for protein function (1–3). As increasing amounts of experimental data on dynamics are becoming available, new simplified structure-based approaches have recently emerged that allow a more quantitative prediction of protein motions (4–12). Nitrogen-15 NMR relaxation spectroscopy is now a standard tool for probing reorientational dynamics of N-H bonds along the backbone on the picosecond to nanosecond timescale (13,14) Motional amplitudes of N-H vectors can be extracted from the relaxation data in the form of Lipari-Szabo $S^2$ order parameters that take values between 0 (highly mobile) and 1 (highly restricted) (15,16). For further interpretation of $S^2$ order parameters a variety of approaches is available that have different degrees of complexity. NMR order parameters can be directly related to motional amplitudes of physical motional models, such as diffusion in a cone (17) or GAF-type motion (18). Alternatively, order parameters can be back-calculated from molecular dynamics (MD) simulations providing a rigorous test of the underlying molecular force field, or they can be computed from normal mode and quasiharmonic analysis offering useful information about correlation effects (19–21). Correlated reorientational motions are uncovered by reorientational eigenmode dynamics analysis (22).

A simple and computationally highly efficient method for the prediction of NMR order parameters from a static structure has been introduced recently (11). It expresses an N-H order parameter as an analytical function of the contact strengths between the amide proton $H_i^N$ and the carbonyl oxygen $O_{i-1}$ of the preceding amino acid and the surrounding protein atoms. A modified version of this contact model (CM) was successfully developed for side-chain methyl order parameters (12). Despite their simplicity, these contact models permit for many proteins a remarkably accurate prediction of NMR order parameters. In fact, the quality of the prediction often exceeds the one obtained from MD simulations.

A limitation of the contact model is its paucity of motional correlation information. In this article we address this limitation by introducing a generalization that incorporates contacts in a way that motional correlation effects are included. The method is a hybrid between the original contact model (11) and the Gaussian network model (GNM) or elastic network model (ENM) (4–9), which are popular tools for studying the coarse-grained dynamics of biomolecules. The GNM model has been applied to predict $^{15}N$ relaxation parameters of lysozyme (23) and adenylate kinase (24).

## METHODS

We consider a harmonic potential $V$ that restores the relative orientations between pairs of interatomic N-H vectors $\vec{r}_i, \vec{r}_j$:

$$V = \frac{\gamma}{2}\sum_{i<j}\sigma_{ij}C_iC_j\left[(\cos\theta_{ij} - \cos\theta_{ij}^0)^2 + (|\sin\theta_{ij}| - |\sin\theta_{ij}^0|)^2\right], \quad (1)$$

where $\theta_{ij}$ is the angle between the two vectors with equilibrium value $\theta_{ij}^0$ determined from the protein's PDB coordinates, $\cos\theta_{ij} = (\vec{r}_i \cdot \vec{r}_j)/|\vec{r}_i||\vec{r}_j|$ and $|\sin\theta_{ij}| = |\vec{r}_i \times \vec{r}_j|/|\vec{r}_i||\vec{r}_j|$, and $\gamma$ determines the total interaction strength. Because the length of $\vec{r}_i$ does not affect the potential energy, all $|\vec{r}_i|$ are set to 1. $\sigma_{ij} = \exp(-r_{ij}/r_{eff,l})$ where $r_{ij}$ is the distance between the hydrogen atoms belonging to internuclear vectors $i$ and vector $j$, and $r_{eff,l}$ is the effective interaction range between vectors. The contact term $C_i$ reflects the number

*Submitted August 2, 2005, and accepted for publication January 13, 2006.*

Address reprint requests to Prof. Rafael Brüschweiler, National High Magnetic Field Laboratory & Dept. of Chemistry and Biochemistry, Florida State University, 1800 E. Paul Dirac Dr., Tallahassee, FL 32310. Tel.: 850-644-5173; Fax: 850-644-1366; E-mail: bruschweiler@magnet.fsu.edu.



0006-3495/06/05/3382/07 $2.00

doi: 10.1529/biophysj.105.071902

and strength of steric contacts experienced by vector $i$ with its surrounding heavy atoms (11,12). For backbone N-H vector $i$,

$$C_{\mathrm{i,NH}} = \sum_{\mathrm{k}} \exp(-r^{\mathrm{O}}_{\mathrm{i-1,k}}/r_{\mathrm{eff,s}}) + 0.8\,\exp(-r^{\mathrm{H}}_{\mathrm{i,k}}/r_{\mathrm{eff,s}}), \quad (2)$$

where $r^{\mathrm{O}}_{\mathrm{i-1,k}}$ is the distance between the carbonyl oxygen of amino acid i−1 to heavy atom $k$ and $r^{\mathrm{H}}_{\mathrm{i,k}}$ is the distance between the amide proton H and heavy atom $k$. The sum ranges over all heavy atoms $k$ that do not belong to amino acids $i$ and $i-1$. The distance $r_{\mathrm{eff,s}}$, which is set to 1 Å, defines the interaction range of steric contacts. The larger the contact term, the more restricted is the local mobility of the bond vector and, in general, the higher will be the corresponding $S^2$ order parameter and vice versa. The contact contribution from the carbonyl oxygen atom of the preceding amino acid $i-1$, which is part of the same peptide plane as amide bond vector $i$, has been identified as an important determinant of the mobility of this bond vector (11).

Thus, the potential function $V$ of Eq. 1 is centered about the native structure and has the following characteristics. It consists of a sum of quadratic terms in the deviations of all pairwise vector orientations. Each term is weighted with the product of the exponential term $\sigma_{\mathrm{ij}} = \exp(-r_{\mathrm{ij}}/r_{\mathrm{eff,l}})$, emphasizing vector pairs that are in close vicinity, and the contact terms $C_{\mathrm{i}} \cdot C_{\mathrm{j}}$ of the two vectors. The restoring force is thereby scaled proportionally to the local mobility of either of the two vectors. This model is termed reorientational contact-weighted elastic network model (rCENM). The potential function $V$ is justified by the good agreement with an effective potential derived from quasiharmonic analysis of an MD trajectory as described below.

To derive dynamic properties from the rCENM model, the second-derivative matrix (Hessian matrix) of the potential energy function $V/\gamma$ of Eq. 1 is analytically constructed with respect to the Cartesian displacements of the $H^{\mathrm{N}}$ atoms in the $\cos\theta_{\mathrm{ij}}$ and $|\sin\theta_{\mathrm{ij}}|$ terms, whereas the $\sigma_{\mathrm{ij}}C_{\mathrm{i}}C_{\mathrm{j}}$ terms are treated as constants. The Hessian matrix is subsequently diagonalized yielding normalized eigenvectors (eigenmodes) $\vec{q}_{\mathrm{k}}$ with eigenvalues $\lambda_{\mathrm{k}}$. Because the eigenvalues $\lambda_{\mathrm{k}}$ belong to the Hessian matrix of $V/\gamma$ rather than of $V$, they have units of $\mathrm{Å}^{-2}$. For a system consisting of $N$ vectors, there are $2N-3$ nonzero modes. The $N+3$ modes with zero eigenvalues include $N$ local bond stretching modes (because bond stretching does not affect $V$) and three rigid-body rotational modes (because they leave the relative vector orientations and thereby $V$ invariant). The thermally averaged product $\Delta_\alpha\Delta_\beta$ of the rotational Cartesian displacements $\Delta_\alpha$ and$\Delta_\beta$ ($\alpha,\beta = x,y,z$) of the N-H vector $i$ is determined by

$$\langle\Delta_\alpha\Delta_\beta\rangle^{(\mathrm{i})} = \frac{k_{\mathrm{B}}T}{\gamma}\sum_{\mathrm{k}=1}^{2\mathrm{N}-3}\frac{q^{(\mathrm{i})}_{\mathrm{k},\alpha}q^{(\mathrm{i})}_{\mathrm{k},\beta}}{\lambda_{\mathrm{k}}}, \quad (3)$$

where the sum goes over all nonzero eigenvalues, $k_{\mathrm{B}}$ is Boltzmann's constant, and $T$ is the absolute temperature. For N-H vector $i$ the $S^2$ order parameter can be approximated by (25,26)

$$S^2_{\mathrm{i}} \cong \left\{1 + \frac{15}{2}\sum_{\alpha,\beta=1}^{3}\zeta_\alpha\zeta_\beta\langle\Delta_\alpha\Delta_\beta\rangle^{(\mathrm{i})} - \frac{3}{2}\sum_{\alpha=1}^{3}\langle\Delta_\alpha\Delta_\alpha\rangle^{(\mathrm{i})}\right\}^2, \quad (4)$$

where $\zeta_\alpha$ are the direction cosines of the internuclear vector $\vec{r}_{\mathrm{i}}$. It can be shown that the sum in Eq. 4, which includes the $\zeta_\alpha\zeta_\beta$, terms vanishes making the autocorrelation terms $\Delta_\alpha\Delta_\alpha$ the only dynamic contributions to $S^2_{\mathrm{i}}$. To confine the order parameters of Eq. 4 to a meaningful range between 0 and 1, the hyperbolic tangent function is applied and an offset parameter $b$ is subtracted, which accounts for zero-point vibrational effects (11,19):

$$S^2_{\mathrm{i}} \cong \left\{1 - \frac{3}{2}\tanh\left(\frac{k_{\mathrm{B}}T}{\gamma}\sum_{\mathrm{k}=1}^{2\mathrm{N}-3}\sum_{\alpha=1}^{3}\frac{q^{(\mathrm{i})}_{\mathrm{k},\alpha}q^{(\mathrm{i})}_{\mathrm{k},\alpha}}{\lambda_{\mathrm{k}}}\right)\right\}^2 - b. \quad (5)$$

Equation 5 depends on the choice of the interaction range parameter $r_{\mathrm{eff,l}}$, the interaction strength $\gamma$, and the offset parameter $b$, all of which are optimized using a fitting procedure that compares the predicted order parameters with the corresponding experimental values.

For each normal mode $\vec{q}_{\mathrm{k}}$, a collectivity $\kappa_{\mathrm{k}}$ can be defined, which is a measure for the ''delocalization'' of a mode corresponding to the percentage of bond vectors that are significantly modulated by mode $k$ (27):

$$\kappa_{\mathrm{k}} = \frac{100\%}{N}\exp\left\{-\sum_{\mathrm{i}=1}^{\mathrm{N}}p^{(\mathrm{i})}_{\mathrm{k}}\ln p^{(\mathrm{i})}_{\mathrm{k}}\right\}, \quad (6)$$

where $p^{(\mathrm{i})}_{\mathrm{k}} = \sum_{\alpha=\mathrm{x,y,z}} q^{(\mathrm{i})2}_{\mathrm{k},\alpha}$. In addition, an effective collectivity can be defined for each bond vector by averaging the collectivities of all modes weighted with their motional contributions to that bond vector:

$$\kappa^{\mathrm{eff}}_{\mathrm{i}} = \sum_{\mathrm{k}=1}^{2\mathrm{N}-3}\frac{p^{(\mathrm{i})}_{\mathrm{k}}}{\lambda_{\mathrm{k}}}\kappa_{\mathrm{k}}\Big/\sum_{\mathrm{k}=1}^{2\mathrm{N}-3}\frac{p^{(\mathrm{i})}_{\mathrm{k}}}{\lambda_{\mathrm{k}}}. \quad (7)$$

In the following section, Eq. 5 is applied to a set of protein structures determined by x-ray crystallography or by NMR that are deposited in the Protein Data Bank (PDB) and for which experimental N-H $S^2$ order parameters are available. For NMR structures the results are averaged over the first 10 models in the corresponding PDB file. For x-ray structures, which typically do not include hydrogen atoms, missing $H^{\mathrm{N}}$ hydrogen atoms are placed at their standard positions using MMTK software (28).

## RESULTS AND DISCUSSION

### Calibration

For the optimization of the three rCENM parameters (interaction range $r_{\mathrm{eff,l}}$, interaction strength $\gamma$, and offset $b$) Eq. 5 was applied to nine proteins with structures deposited in the PDB (Table 1). The three parameters were varied on a grid to optimize agreement between experimental and predicted N-H $S^2$ order parameters. Parameter optimization was pursued by minimizing the sum of the $\chi^2$ values, $\Sigma(\chi^2)$, which are the squares of the differences between the experimental and the predicted $S^2$ values, and by maximizing the sum of the correlation coefficients, $\Sigma(r_{\mathrm{j}})$, where $r_{\mathrm{j}}$ is Pearson's correlation coefficient between predicted and experimental $S^2$ values for protein $j$. As shown in Fig. 1, the sum $\Sigma(r_{\mathrm{j}})$ has a shallow maximum between $r_{\mathrm{eff,l}} = 3.0$ and $r_{\mathrm{eff,l}} = 10.0\,\mathrm{Å}$, whereas the sum $\Sigma(\chi^2)$ exhibits a more pronounced minimum at $r_{\mathrm{eff,l}} = 6.6$ Å. At this minimum, the optimal values for $\gamma$ and $b$ are $\gamma = 17.86\,k_{\mathrm{B}}T$ and $b$=0.045, which are the values used in the following analysis. The $\chi^2$ values and correlation coefficients $r_{\mathrm{j}}$ for all proteins analyzed in this study are compiled in Table 1.

### Validation of rCENM potential by MD

Although the rCENM potential of Eq. 1 is physically motivated (see previous section), its specific form is empirical. We therefore examined the validity of the potential by comparison with a reorientational quasiharmonic analysis of a 5-ns MD simulation of ubiquitin in explicit solvent (22). The reorientational motional correlation between vectors $i$ and $j$ was calculated according to

$$C_{\mathrm{ij}} = \langle\vec{r}_{\mathrm{i}}\cdot\vec{r}_{\mathrm{j}}\rangle - \langle\vec{r}_{\mathrm{i}}\rangle\cdot\langle\vec{r}_{\mathrm{j}}\rangle, \quad (8)$$

where $\vec{r}_{\mathrm{i}}$ is the orientation of the N-H vector $i$ and $\langle\rangle$ stands for the average over 1000 MD snapshots. On the other hand,

**TABLE 1 Comparison between experimental and predicted backbone N-H $S^2$ order parameters of nine proteins using Eq. 5 with $\gamma = 17.86\ k_BT$, $r_{eff,l} = 6.6$ Å, $b = 0.045$**

| Protein* | PDB entry | Resolution (Å) | $r^\dagger$ | | $\chi^{2\ddagger}$ | |
|---|---|---|---|---|---|---|
| | | | rCENM | CM | rCENM | CM |
| Ubiquitin | 1UBQ | 1.8 | 0.97 | 0.96 | 0.14 | 0.13 |
| Lysozyme | 1JEF | 1.8 | 0.68 | 0.70 | 0.35 | 0.39 |
| Interleukin-4 | 1HIK | 2.6 | 0.81 | 0.81 | 0.97 | 0.96 |
| Calmodulin | 2BBN | NMR | 0.74 | 0.75 | 0.77 | 0.81 |
| Calbindin | 4ICB | 1.6 | 0.72 | 0.66 | 0.18 | 0.22 |
| βARK1 PH domain | 1BAK | NMR | 0.84 | 0.61 | 1.71 | 3.70 |
| Frenolicin acyl carrier protein | 1OR5 | NMR | 0.87 | 0.89 | 0.71 | 0.74 |
| Ribonuclease HI | 1RBU | 1.8 | 0.75 | 0.74 | 1.15 | 1.15 |
| Ketosteroid isomerase | 8CHO | 2.3 | 0.78 | 0.59 | 0.93 | 1.22 |

*Experimental protein backbone N-H $S^2$ order parameters were taken from the references given in parentheses: ubiquitin (29), lysozyme (30), interleukin-4 (31), calmodulin (32), calbindin (33), βARK1 PH domain (34), frenolicin acyl carrier protein (35), ribonuclease HI (36), and ketosteroid isomerase (37).
†Linear correlation coefficient between experimental and predicted $S^2$.
‡Total $\chi^2 = \sum_j (S^2_{j,calc} - S^2_{j,exp})^2$.

the correlation strength can also be calculated using the normal modes of the rCENM potential $V$

$$\tilde{C}_{ij} = \sum_{k=1}^{2N-3} \sum_{\alpha=x,y,z} \frac{q_{k,\alpha}^{(i)} q_{k,\alpha}^{(j)}}{\lambda_k}. \quad (9)$$

The correlation coefficient between the 73 × 73 matrix elements $C_{ij}$ and $\tilde{C}_{ij}$ is $r = 0.84$. This shows that the rCENM potential is closely related to the quasiharmonic reorientational molecular mechanics potential providing a physical rationale for its usage in this context.

## Reorientational vector dynamics

The normal modes calculated from the rCENM provide direct insight into reorientational motions experienced by the backbone N-H bond vectors. A vibrational frequency can be computed for each mode according to $\omega_k/2\pi c = \sqrt{(\gamma\lambda_k/m_H)}/(2\pi c)$ where $m_H$ is the hydrogen mass and $c$ is the speed of light. For lysozyme at 300 K, for example, the smallest and largest eigenvalues $\lambda_k$ are 0.54 and 12.9 $Å^{-2}$, which correspond to vibrational frequencies 259 $cm^{-1}$ and 1267 $cm^{-1}$, respectively, that have periods of the order of 0.1 ps. Due to their partially collective character the modes are expected to be overdamped by friction with the solvent thereby increasing their effective correlation times possibly into the picosecond range (38).

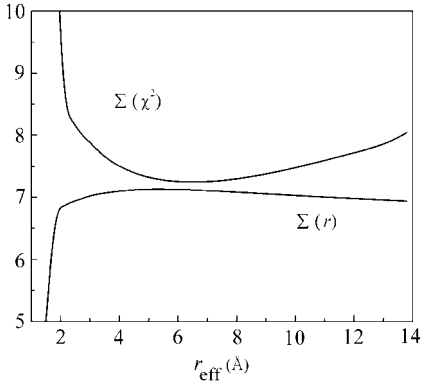


FIGURE 1 Optimization of rCENM parameters $r_{eff,l}$, $\gamma$, $b$ in Eqs. 1 and 5 for the prediction of N-H $S^2$ order parameters. $\Sigma(\chi^2)$ is the sum of $\chi^2 = \sum_j (S^2_{j,calc} - S^2_{j,exp})^2$ and $\Sigma(r) = \sum_i r_i$ is the sum of the correlation coefficients between experimental and predicted $S^2$ values for nine proteins (see also Table 1): ubiquitin, lysozyme, interleukin-4, calmodulin, calbindin, βARK1 PH domain, frenolicin acyl carrier protein, ribonuclease HI, and ketosteroid isomerase.

Because the reorientational potential includes the local contact $C_i$ of each vector $i$, it is interesting to know how $C_i$ is related to the total fluctuation amplitude of vector $i$, expressed as $D_i = \sum_{\alpha=x,y,z} \langle \Delta_\alpha \Delta_\alpha \rangle^{(i)}$ using Eq. 3. The $C_i$ and $D_i$ terms computed for the N-H vectors of lysozyme are anticorrelated with a correlation coefficient $r = -0.81$, which indicates that large (small) contact terms are generally accompanied by small (large) amplitude motions, consistent with the original contact model (11).

The main advantage of the rCENM over the CM is that information can be gained on correlated dynamics. A scalar measure of the collective character of a mode is its collectivity $\kappa$ (Eq. 6), the percentage of bond vectors that are significantly affected by this mode (27). A low (high) $\kappa$-value indicates that the corresponding mode has a localized (delocalized) character. Fig. 2 shows the collectivity versus $\lambda_k^{-1}$ for four proteins: lysozyme, ubiquitin, calmodulin, and calbindin. All four distributions are similar with the largest amplitude modes (small $\lambda_k$) and lowest amplitude modes (large $\lambda_k$) displaying a relatively localized character and the intermediate modes exhibiting a larger variation of collectivity. This behavior is remarkably similar to the one previously reported for reorientational eigenmodes extracted from all-atom MD simulations (see Figs. 2 of Brüschweiler and Lienin (39) and Showalter and Hall (40)) and qualitatively different from the behavior of quasiharmonic eigenmodes of backbone $C^\alpha$ atoms (Fig. 1 of Prompers and Brüschweiler (41)).

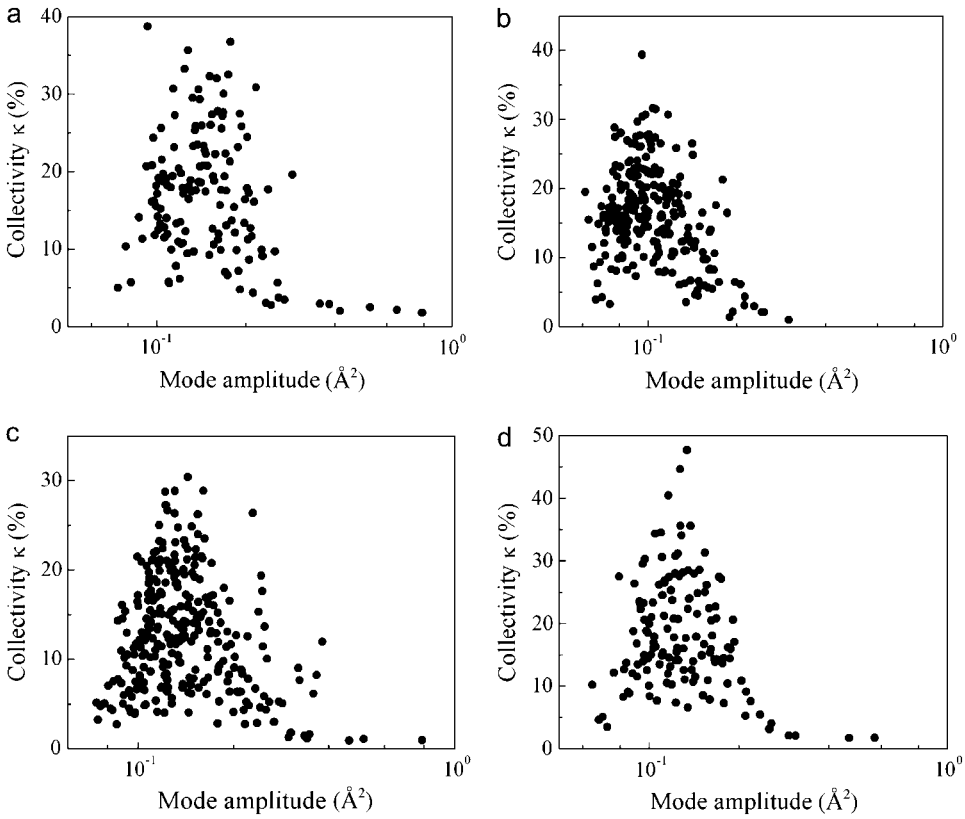


FIGURE 2 Mode collectivities $\kappa$ (Eq. 6) versus amplitudes $\lambda_k^{-1}$ of the reorientational eigenmodes determined from rCENM for (*a*) ubiquitin, (*b*) lysozyme, (*c*) calmodulin, and (*d*) calbindin.

## N-H order parameters

The comparison between experimental and back-calculated N-H $S^2$ order parameters is summarized in Table 1. For the nine proteins rCENM produces an average correlation coefficient of $\langle r \rangle = 0.80 \pm 0.08$ with respect to experimental order parameters, which is an improvement over the CM ($r = 0.75 \pm 0.12$). A residue-by-residue display of experimental and back-calculated N-H $S^2$ order parameters is given in Fig. 3 for ubiquitin, lysozyme, calbindin, and $\beta$ARK1 PH domain.

A high correlation coefficient between experimental and predicted order parameters of $r = 0.97$ is found for ubiquitin, which is almost the same as the one obtained with the contact model ($r = 0.96$) (11). For lysozyme, the agreement ($r = 0.68$) is somewhat lower than for the other proteins, but better than the prediction using the GNM model (23) after omission of a few outliers this study reported $r = 0.60$ using the x-ray structure 193L and $r = 0.65$ using the NMR structure 1HWA. When the order parameters of residues Ser-85 ($S^2_{exp} = 0.55$, $S^2_{calc} = 0.88$), located in the loop between the first $3_{10}$-helix and the C helix, and Asn-103 ($S^2_{exp} = 0.52$, $S^2_{calc} = 0.82$), located in the loop preceding the D-helix, are excluded, the rCENM method yields $r = 0.79$. The relaxation data of these two residues, which show the largest discrepancy between experiment and prediction, require treatment by the extended model-free approach (42) involving internal motions on two separate internal timescales. The order parameters of these residues sharply differ from those of their sequential neighbors.

For human interleukin-4 the rCENM successfully predicts order parameters in several regions with enhanced flexibility including the AB and CD loops as well as the N- and C-terminal regions despite the relatively low resolution of the x-ray structure (2.6 Å). The model also reproduces the rigid behavior of the helical regions with the exception of a systematic offset by ~0.05 toward lower values. Calbindin $D_{9k}$ contains two helix-loop-helix EF-hand motifs joined by a linker loop. The prediction successfully identifies the different regions according to their mobility as well as the most mobile linker region. Some of the larger discrepancies include Gly-18 ($S^2_{exp} = 0.85$, $S^2_{calc} = 0.72$) in the $Ca^{2+}$ loop I, Leu-40 ($S^2_{exp} = 0.71$, $S^2_{calc} = 0.84$) in the linker region, and Ile-73 ($S^2_{exp} = 0.68$, $S^2_{calc} = 0.84$) in helix IV.

The predictions for calmodulin and $\beta$ARK1 PH domain are based on their NMR structural ensembles. The table lists correlations and $\chi^2$ values between the average $S^2$ predictions for the first 10 NMR structures and the experiment. For calmodulin, the prediction based on the NMR structures performs clearly better than that based on the x-ray structure (result not shown), especially in the highly mobile central linker region that connects the two domains (32). This region forms an $\alpha$-helix in the crystal structure that exhibits numerous local contacts, which causes larger weight factors $C_iC_j$ in the potential of Eq. 1 thereby reducing the reorientational fluctuations of these vectors. The overall result is very

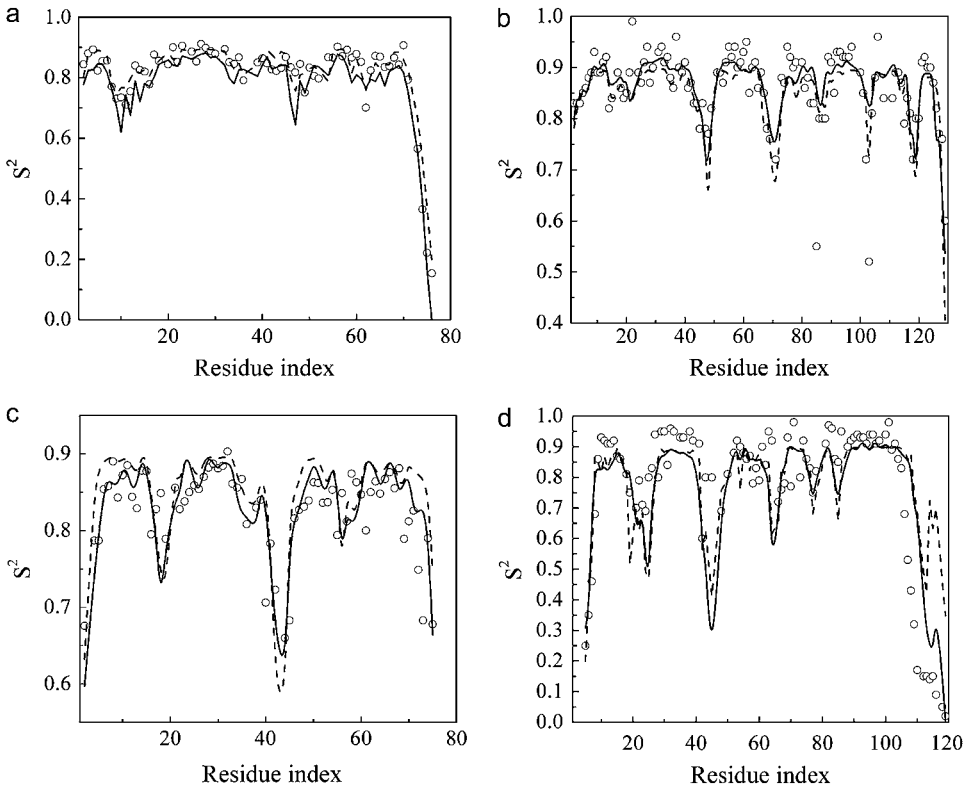


FIGURE 3 Experimental (○), rCENM (*solid lines*), and contact model (*dashed lines*) backbone N-H $S^2$ order parameters using Eq. 5 with $\gamma = 17.86\,k_BT$, $r_{eff,l} = 6.6$ Å, $b = 0.045$ of (*a*) ubiquitin, (*b*) lysozyme, (*c*) calbindin, and (*d*) βARK1 PH domain.

similar to that of the contact model (11). The βARK1 PH domain is composed of a β-barrel that packs against an α-helix attached to a long terminal loop. The rCENM prediction gives good agreement with the experimental data, except for certain loop regions including loops at β1/β2, β3/β4, β7/α where the calculated order parameters $S^2_{calc}$ are systematically lower than $S^2_{exp}$. It is interesting to note that for these regions the rCENM-derived values are well comparable with those from a thoroughly analyzed MD simulation (34): both calculations yield the lowest order parameters at the β3/β4 loop and at the termini. As observed with the contact model (11), the prediction is sensitive to the structure used and averaging of $S^2$ values predicted over the NMR ensemble can improve the agreement with the experiment. For both proteins, the correlation between prediction and experiment varies between $r = 0.63$ and 0.85 for individual NMR structures.

The prediction for frenolicin acyl carrier protein yields the second highest correlation coefficient ($r = 0.87$). Except for the termini, the largest discrepancy occurs in the flexible loop (Ala-16–Ala-25) between the first two helices. The calculated $S^2_{calc}$ values are systematically larger than the experimental values, although they both show a minimum in this region.

The overall performance of the rCENM prediction for *Escherichia coli* ribonuclease HI is good. It however underestimates motion in the two highly mobile loops from Cys-13 to Gly-15 and from Gly-123 to His-127; for example, for Gly-15 $S^2_{exp} = 0.48$ vs. $S^2_{calc} = 0.80$, for Gly-123 $S^2_{exp} = 0.46$ vs. $S^2_{calc} = 0.84$, and for Gly-126 $S^2_{exp} = 0.55$ vs. $S^2_{calc} = 0.79$. These two loops are located relatively close to the active site of the enzyme and take part in a cleft-like depression into which the substrate binds. There are two very sharp local minima in the $S^2_{exp}$ profile, whereas the predicted profile is much smoother, resembling the situation in lysozyme.

The ketosteroid isomerase is a homodimeric enzyme with both subunits consisting of three α-helices and a six-stranded mixed β-pleated sheet that forms a conical barrel. With the exception of the C-terminus, the average of predicted order parameters is 0.83, which is 0.1 less than the average of experimental values. This may indicate that this protein is substantially more rigid than the other proteins analyzed here. Except for this systematic shift, the prediction performs well, especially in the region of the large β-sheet.

We also examined predictions for other proteins. Poor prediction results are obtained for FK506 binding protein (43). The major structure of this protein is a twisted β-sheet packing against a small helix. The experimental data show a rather uneven variation of the order parameters among all the secondary structural elements of this protein, suggesting an unusually complex dynamic behavior. Except for a few small regions such as the C-terminus, the prediction appears to have no significant positive correlation with the experimental measurements. In some loop regions such as loops $\beta_1/\beta_2$, $\beta_2/\beta_3$, and $\beta_4$/helix, the predicted variation of order parameters with residue index is even reversed compared to the experiment.

To better understand the lack of agreement between the rCENM and experiment, we performed an MD simulation of

the FK506 binding protein in explicit solvent using CHARMM (44). The simulation started from the x-ray structure (PDB code 1FKJ) embedded in a box of water containing 5208 water molecules and a trajectory was run for 2 ns at 298 K with snapshots stored every picosecond. The N-H order parameters were computed from the trajectory, after each snapshot was aligned with respect to the snapshot at 1 ns, using the expression $S^2 = N^{-2}\sum_{i,j}(3(\vec{e}_i \cdot \vec{e}_j)^2 - 1)/2$, where $\vec{e}_i$ is the normalized N-H bond vector of snapshot $i$ and $N = 2000$ is the total number of snapshots. Comparison of the experimental and MD-derived order parameters is $r = 0.24$, which is similarly low as the correlation found between rCENM and experiment ($r = 0.15$).

Discrepancies between experiment and prediction can have three distinct origins and a combination thereof. It can be due to i), the inadequacy of the rCENM model, ii), errors in the experimental order parameters, and iii), inaccuracies of the available protein structure or a protein structure that is not representative for the situation in solution at which the relaxation data were collected. Failure by both rCENM and MD to explain the experimental data may indicate that the origin of the discrepancy is due to ii or iii.

The collective nature of the rCENM motions tends to smoothen the $S^2$ profiles along the protein's primary sequence. Therefore, the presence of spikes in experimental order parameters, as found for lysozyme and ribonuclease HI, can pose a challenge for the rCENM that cannot be easily accommodated by the single-minimum potential function of the type of Eq. 1. These order parameters may reflect local jump motions of the peptide bonds among two (or more) energy minima, which are not captured by harmonic potentials. Such effects could be addressed in a refined potential at the expense, however, of introducing additional fit parameters. Discrepancies between rCENM results and experimental data should help to identify protein regions that exhibit more complicated dynamic behavior that may warrant more detailed investigations, both experimentally and computationally.

The amide-bond specific effective collectivities, $\kappa_{eff}$, drawn in Fig. 4 as a dashed line, lie typically at ~10–20%. Low order parameters are often accompanied by reduced $\kappa_{eff}$ values reflecting the dominance of a few low collectivity modes for these amide vectors, which is fully consistent with the high amplitude low collectivity modes of Fig. 2. By contrast, bond vectors with high order parameters generally exhibit higher effective collectivities.

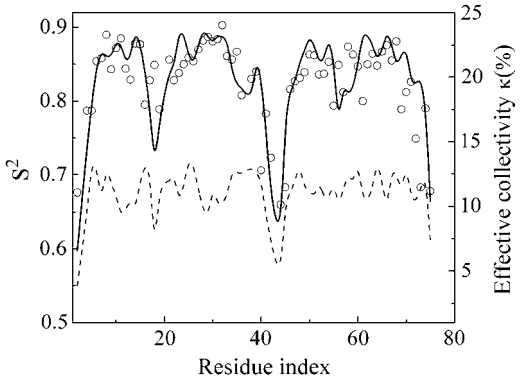


FIGURE 4 Effective collectivities $\kappa_{eff}$ (see Eq. 7) (*dashed line*) of the N-H bond motions as a function of residue number for calbindin.

## CONCLUSION

The reorientational elastic network model described here is a collective extension of the previously proposed contact model. The model expresses bond vector reorientational motions in terms of a multidimensional harmonic potential that encodes correlated motional behavior. The quality of the prediction is good to excellent for most of the proteins that were analyzed here. Although computationally slightly less efficient, the rCENM model presents an improvement over the local contact model predictions by 7% on average. The introduction of collective motions significantly improves the correlation between prediction and experiment in several cases, most notably for calbindin, βARK1 PH domain, and ketosteroid isomerase. The protein regions that show most improvements are often the ones that exhibit high mobility involving several consecutive residues, such as the terminal and extended loop regions, which are precisely the regions where collective motions are expected to play a prominent and potentially functionally relevant role.

This work was supported by the National Science Foundation (grant MCB-0211512).

## REFERENCES

1. Frauenfelder, H., S. G. Sligar, and P. G. Wolynes. 1991. The energy landscapes and motions of proteins. *Science*. 254:1598–1603.
2. Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol*. 9:646–652.
3. Benkovic, S. J., and S. Hammes-Schiffer. 2003. A perspective on enzyme catalysis. *Science*. 301:1196–1202.
4. Tirion, M. M. 1996. Large amplitude elastic protein motions from a single parameter, elastic analysis. *Phys. Rev. Lett*. 77:1905–1908.
5. Haliloglu, T., I. Bahar, and B. Erman. 1997. Gaussian dynamics of folded proteins. *Phys. Rev. Lett*. 79:3090–3093.
6. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins*. 33:417–429.
7. Ming, D., Y. Kong, S. J. Wakil, J. Brink, and J. Ma. 2002. Domain movements in human fatty acid synthase by quantized elastic deformational model. *Proc. Natl. Acad. Sci. USA*. 99:7895–7899.
8. Tama, F., and C. L. Brooks III. 2002. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol*. 318:733–747.
9. Delarue, M., and Y. H. Sanejouand. 2002. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. *J. Mol. Biol*. 320:1011–1024.
10. Halle, B. 2002. Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. USA*. 99:1274–1279.
11. Zhang, F., and R. Brüschweiler. 2002. Contact model for the prediction of NMR N-H order parameters in globular proteins. *J. Am. Chem. Soc*. 124:12654–12655.

12. Ming, D., and R. Brüschweiler. 2005. Prediction of methyl side-chain dynamics in proteins. *J. Biomol. NMR.* 29:363–368.
13. Palmer, A. G. 2004. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* 104:3623–3640.
14. Brüschweiler, R. 2003. New approaches for the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr. Opin. Struct. Biol.* 13:175–183.
15. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* 104:4546–4559.
16. Lipari, G., and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.* 104:4559–4570.
17. Lipari, G., and A. Szabo. 1980. Effect of librational motion on fluorescence depolarization and nuclear-magnetic resonance relaxation in macromolecules and membranes. *Biophys. J.* 30:489–506.
18. Bremi, T., and R. Brüschweiler. 1997. Locally anisotropic polypeptide backbone dynamics by NMR relaxation. *J. Am. Chem. Soc.* 119:6672–6673.
19. Brüschweiler, R. 1992. Normal modes and NMR order parameters in proteins. *J. Am. Chem. Soc.* 114:5341–5344.
20. Brüschweiler, R., and D. A. Case. 1994. A collective NMR relaxation model applied to protein dynamics. *Phys. Rev. Lett.* 72:940–943.
21. Case, D. A. 2002. Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.* 35:325–331.
22. Prompers, J. J., and R. Brüschweiler. 2001. Reorientational eigenmode dynamics: a combined MD/NMR relaxation analysis method for flexible parts in globular proteins. *J. Am. Chem. Soc.* 123:7305–7313.
23. Haliloglu, T., and I. Bahar. 1999. Structure-based analysis of protein dynamics: comparison of theoretical results for hen lysozyme with X-ray diffraction and NMR relaxation data. *Proteins.* 37:654–667.
24. Temiz, N. A., E. Meirovitch, and I. Bahar. 2004. Escherichia coli adenylate kinase dynamics: comparison of elastic network model modes with mode-coupling $^{15}$N-NMR relaxation data. *Proteins.* 57: 468–480.
25. Henry, E. R., and A. Szabo. 1985. Influence of vibrational motion on solid-state line shape and NMR relaxation. *J. Chem. Phys.* 82: 4753–4761.
26. Brüschweiler, R., and D. A. Case. 1994. Collective NMR relaxation model applied to protein dynamics. *Phys. Rev. Lett.* 72:940–943.
27. Brüschweiler, R. 1995. Collective protein dynamics and nuclear-spin relaxation. *J. Chem. Phys.* 102:3396–3403.
28. Hinsen, K. 2000. The molecular modeling tool kit: a new approach to molecular simulations. *J. Comput. Chem.* 21:79–85.
29. Tjandra, N., S. E. Feller, R. W. Pastor, and A. Bax. 1995. Rotational diffusion anisotropy of human ubiquitin from $^{15}$N NMR relaxation. *J. Am. Chem. Soc.* 117:12562–12566.
30. Buck, M., J. Boyd, C. Redfield, D. A. MacKenzie, D. J. Jeenes, D. B. Archer, and C. M. Dobson. 1995. Structural determinants of protein dynamics: analysis of $^{15}$N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. *Biochemistry.* 34:4041–4055.
31. Radfield, C., J. Boyd, L. J. Smith, and C. M. Dobson. 1992. Loop mobility in a four-helix-bundle protein: $^{15}$N NMR relaxation measurements on human Interleukin-4. *Biochemistry.* 31:10431–10437.
32. Barbato, G., M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax. 1992. Backbone dynamics of calmodulin studied by $^{15}$N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry.* 31:5269–5278.
33. Kördel, J., N. J. Skelton, M. Akke, A. G. I. I. I. Palmer, and W. J. Chazin. 1992. Backbone dynamics of calcium-loaded calbindin $D_{9k}$ studied by two-dimensional proton-detected $^{15}$N NMR spectroscopy. *Biochemistry.* 31:4856–4866.
34. Pfeiffer, S., D. Fushman, and D. Cowburn. 2001. Simulated and NMR-derived backbone dynamics of a protein with significant flexibility: a comparison of spectral densities for the $\beta$ARK1 PH domain. *J. Am. Chem. Soc.* 123:3021–3036.
35. Li, Q., C. Khosla, J. D. Puglisi, and C. W. Liu. 2003. Solution structure and backbone dynamics of the holo form of the frenolicin acyl carrier protein. *Biochemistry.* 42:4648–4657.
36. Yamasaki, K., M. Saito, M. Oobatake, and S. Kanaya. 1995. Characterization of the internal motions of *Escherichia coli* ribonuclease HI by a combination of $^{15}$N-NMR relaxation analysis and molecular dynamics simulation: examination of dynamic models. *Biochemistry.* 34:6587–6601.
37. Yun, S., D. S. Jang, D. Kim, K. Y. Choi, and H. C. Lee. 2001. $^{15}$N NMR relaxation studies of backbone dynamics in free and steroid-bound $\Delta^5$-3-ketosteroid isomerase from *Pseudomonas testosteroni*. *Biochemistry.* 40:3967–3973.
38. Go, N., T. Noguti, and T. Nishikawa. 1983. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc. Natl. Acad. Sci. USA.* 80:3696–3700.
39. Brüschweiler, R., and S. F. Lienin. 2000. Characterization of collective and anisotropic reorientational protein dynamics. *Phys. Rev. Lett.* 84: 5439–5442.
40. Showalter, S. A., and K. B. Hall. 2002. A functional role for correlated motion in the N-terminal RNA-binding domain of human U1A protein. *J. Mol. Biol.* 322:533–542.
41. Prompers, J. J., and R. Brüschweiler. 2002. Dynamic and structural analysis of isotropically distributed ensembles. *Proteins.* 46:177–189.
42. Clore, G. M., A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, and A. M. Gronenborn. 1990. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic resonance relaxation of proteins. *J. Am. Chem. Soc.* 112:4989–4991.
43. Cheng, J., C. A. Lepre, S. P. Chambers, J. R. Fulghum, J. A. Thomson, and J. M. Moore. 1993. $^{15}$N NMR relaxation studies of the FK506 binding protein: backbone dynamics of the uncomplexed receptor. *Biochemistry.* 32:9000–9010.
44. Brooks, B. R., R. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* 4:187–217.