

Predicting conformational entropy of bond vectors in proteins by networks of coupled rotators

Anne Dhulesia,^{1,a)} Geoffrey Bodenhausen,^{1,2} and Daniel Abergel^{1,b)}

¹Département de Chimie, Associé au CNRS, Ecole Normale Supérieure, 24, rue Lhomond, 75005 Paris, France

²Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne, BCH 1015 Lausanne, Switzerland

(Received 21 April 2008; accepted 21 July 2008; published online 5 September 2008)

In this article, a formal expression for the conformational entropy of a bond vector in a protein is derived using the networks of coupled rotators model for the description of internal dynamics. Analytical relationships between NMR order parameters and conformational entropies are derived, and the possibility to extract the latter from NMR experiments is discussed. These results are illustrated in the case of the calcium-binding protein calbindin. © 2008 American Institute of Physics. [DOI: 10.1063/1.2969809]

INTRODUCTION

It has long been assumed that the function of a protein is embodied in its three dimensional structure. This simple picture tends to be gradually replaced by a more dynamic view according to which internal motions represent an essential ingredient of protein function,¹ so that a protein “structure” should be rather viewed as a superposition of interconverting conformations. The time scales of the dynamical processes that occur in proteins span several orders of magnitude. Motions with characteristic time constants longer than microseconds can be related to important biological processes, such as enzymatic catalysis, ligand binding, and protein-protein and protein-nucleic acid interactions. Alternatively, fast internal motions (on a subnanosecond time scale) corresponding to rotations, vibrations, and librations of chemical bonds can be related to conformational entropy. Nuclear magnetic resonance (NMR) spectroscopy represents a unique tool to study protein dynamics, as it allows one to probe events on the pico- to millisecond time scale, on a residue-per-residue basis, via the measurement of spin relaxation rates.^{1–3} One of the most challenging issues in the study of protein dynamics is therefore to relate NMR relaxation rates to a quantitative description of internal motions. This comprises two different aspects, kinetic and equilibrium, and includes the prediction of relevant correlation functions and conformational entropies. Because internal dynamics are constrained by a potential, the internal correlation functions that are relevant to spin relaxation tend to a nonvanishing limit at long times, known in the NMR community as the *squared order parameter* and denoted by S^2 , which is a measure of the motional amplitude of the bond vectors.⁴ Establishing a formal relationship between site-specific order parameters S^2 determined from experimental NMR relaxation measurements and conformational entropies of individual bond vectors is a challenging

task. Most approaches are based on the assumption that all vectors are independent,^{5,6} which affords in this case a simple link between conformational entropies of the individual vectors and of the whole protein, but may seem rather unrealistic. We have therefore investigated the possibility of using networks of coupled rotators⁷ (NCRs) to predict conformational entropies and derive their relationships with the order parameters.

RESULTS

As detailed elsewhere,⁷ in the NCR approach the protein is represented by an ensemble of vectors that are parallel to chemical bonds and assumed to undergo wobbling motions about their average direction. Each pair of network vectors ($\mathbf{u}_i, \mathbf{u}_j$) is coupled by a potential U_{ij} that tends to restore the instantaneous angle θ_{ij} between them to its equilibrium value θ_{0ij} , given by the three dimensional protein structure determined by X ray or NMR. These pair potentials U_{ij} depend on local atomic densities ρ_i and ρ_j , as well as on a scaling constant κ_0 , common to all vectors in the protein,^{7–9}

$$U_{ij} = -\rho_i \rho_j \kappa_0 k T P_2(\cos(\theta_{ij} - \theta_{0ij})), \quad (1)$$

where $P_2(x)$ is the second Legendre polynomial. Because proteins are densely packed molecules, the bond vector motions are assumed to be restricted. The potentials U_{ij} can then be expanded up to the second order in terms of the components (X_i, Y_i) and (X_j, Y_j) of \mathbf{u}_i and \mathbf{u}_j in their respective local reference frames where the z axes are parallel to the average bond vectors $\langle \mathbf{u}_i \rangle$ and $\langle \mathbf{u}_j \rangle$. In the NCR model, the coupled vectors diffuse in the potential $U = \sum_{i < j} U_{ij}$ and are driven by stochastic Langevin forces. The dynamics are described by rotational Langevin equations,¹⁰ which, due to the above approximations, allow one to determine the evolution of various averages of the components of the vectors \mathbf{u}_i .⁷ These can then be used to predict NMR observables. Of central importance in what follows, the order parameter S_i^2 , defined as the limit of the internal autocorrelation function $C_i^1(t)$ of the bond vector \mathbf{u}_i , can be expressed as.^{7,11}

^{a)}Present address: Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, UK.

^{b)}Author to whom correspondence should be addressed. Electronic mail: daniel.abergel@ens.fr.

$$S_i^2 = \lim_{t \rightarrow \infty} C_i^t(t) = 1 - 3(\langle X_i^2 \rangle + \langle Y_i^2 \rangle). \quad (2)$$

More generally, as only the second moments of the components X_i, Y_i of the coupled vectors are nonzero,⁷ the equilibrium statistical information is entirely described by the covariance matrix \mathbf{M} ,

$$\mathbf{M} = \begin{pmatrix} \langle X_1^2 \rangle & \langle X_1 Y_1 \rangle & \dots & \langle X_1 X_N \rangle & \langle X_1 Y_N \rangle \\ \langle X_1 Y_1 \rangle & \langle Y_1^2 \rangle & \dots & \langle Y_1 X_N \rangle & \langle Y_1 Y_N \rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle Y_N X_1 \rangle & \langle Y_N Y_1 \rangle & \dots & \langle Y_N X_N \rangle & \langle Y_N^2 \rangle \end{pmatrix}. \quad (3)$$

In this case, the probability distribution function $p(\mathbf{X})$ of the variable $\mathbf{X} = [X_1, Y_1, \dots, X_N, Y_N]^t$ is Gaussian,¹²

$$p(\mathbf{X}) = \frac{\sqrt{\det \mathbf{H}}}{(2\pi)^N} e^{-\mathbf{X}^t \mathbf{H} \mathbf{X} / 2}, \quad (4)$$

where $\mathbf{H} = \mathbf{M}^{-1}$ is the *Hessian* matrix. Thus, Eq. (4) corresponds to a Boltzmann distribution with the average potential $\mathbf{X}^t \mathbf{H} \mathbf{X} / 2$ and the partition function $\mathcal{Z} = (2\pi)^N / \sqrt{\det \mathbf{H}}$.

We shall denote by $S_{i,\text{NCR}}^e$ the conformational entropy associated with the vector \mathbf{u}_i , the motion of which is described by the variables (X_i, Y_i) . By definition, $S_{i,\text{NCR}}^e$ is obtained from the probability distribution $p(X_i, Y_i)$ of the components (X_i, Y_i) ,

$$S_{i,\text{NCR}}^e = -k \int p(X_i, Y_i) \log p(X_i, Y_i) dX_i dY_i, \quad (5)$$

where the marginal distribution function $p(X_i, Y_i)$ is obtained by integrating $p(\mathbf{X})$ over all variables other than X_i and Y_i . For simplicity, we consider the first vector, $i=1$,

$$p(X_1, Y_1) = \int p(\mathbf{X}, Y) dX^{N-1} dY^{N-1}, \quad (6)$$

$$p(X_1, Y_1) = \frac{\sqrt{\det \mathbf{H}}}{(2\pi)^N} \int dX^{N-1} dY^{N-1} e^{-(1/2)\mathbf{X}^t \mathbf{H} \mathbf{X}}, \quad (7)$$

with the shorthand notations $dX^{N-1} = dX_2 \cdots dX_N$ and $dY^{N-1} = dY_2 \cdots dY_N$. Standard properties of partial Gaussian integrals give

$$p(X_1, Y_1) = \frac{1}{2\pi} \sqrt{\frac{\det \mathbf{H}}{\det W_0^{(X_1 Y_1)}}} e^{-(1/2)[X_1 Y_1] U_1^{(2)} [X_1 Y_1]^t}, \quad (8)$$

where the matrices $W_0^{(X_1 Y_1)}$ and $U_1^{(2)}$ result from the block decomposition of \mathbf{H} ,

$$\mathbf{H} = \begin{pmatrix} U_0^{(X_1 Y_1)} & V^{(X_1 Y_1)} \\ V^{(X_1 Y_1)t} & W_0^{(X_1 Y_1)} \end{pmatrix}, \quad (9)$$

$U_0^{(X_1 Y_1)}$ is a 2×2 matrix that connects X_1 and Y_1 , and we define $U_1^{(2)}$ as

$$U_1^{(2)} = U_0^{(X_1 Y_1)} - V^{(X_1 Y_1)} (W_0^{(X_1 Y_1)})^{-1} V^{(X_1 Y_1)t}. \quad (10)$$

Thus, $U_1^{(2)}$ is the effective diffusion potential of vector \mathbf{u}_1 , the

conformational entropy of which, $S_{1,\text{NCR}}^e$, can be derived from Eq. (5),

$$-\frac{S_{1,\text{NCR}}^e}{k} = \log \left(\frac{\sqrt{\det U_1^{(2)}}}{2\pi} \right) - \frac{U_{11}^{(2)}}{2} \langle X_1^2 \rangle - \frac{U_{22}^{(2)}}{2} \langle Y_1^2 \rangle - \frac{U_{12}^{(2)} + U_{21}^{(2)}}{2} \langle X_1 Y_1 \rangle, \quad (11)$$

where the $U_{ij}^{(2)}$ are the matrix elements of the symmetric matrix $U_1^{(2)}$, with $U_{12}^{(2)} = U_{21}^{(2)}$. Obviously, the same result can be obtained for any vector of the network by changing indices 1 to i in Eqs. (5)–(8) and by proper row and column permutations in Eq. (9).

Equation (11) defines the relationship between the covariance matrix and the conformational entropy $S_{1,\text{NCR}}^e$ of the vector \mathbf{u}_1 in the framework of the NCR model. Therefore, when amide NH vectors are included in the network, which is required to predict ¹⁵N relaxation rates, the NCR model affords a means of predicting local conformational entropy, *based on the sole knowledge of the molecular structure*. The averages $\langle X_1^2 \rangle$, $\langle Y_1^2 \rangle$, and $\langle X_1 Y_1 \rangle$ that appear in Eq. (11) can be expressed in terms of the matrix elements of the effective potential $U_1^{(2)}$,

$$\langle X_1^2 \rangle = \frac{U_{22}^{(2)}}{\det U_1^{(2)}}, \quad \langle Y_1^2 \rangle = \frac{U_{11}^{(2)}}{\det U_1^{(2)}}, \quad (12)$$

$$\langle X_1 Y_1 \rangle = - \frac{\sqrt{\det U_1^{(2)}} (U_{12}^{(2)} + U_{21}^{(2)})}{(U_{22}^{(2)} U_{11}^{(2)} - (U_{12}^{(2)} + U_{21}^{(2)})^2)^{3/2}}.$$

These relations can be simplified by going into the frame of the eigenvectors of $U_1^{(2)}$, leading to a new simpler formulation of the conformational entropy of the bond vector,

$$-\frac{S_{1,\text{NCR}}^e}{k} = \log \left(\frac{\sqrt{\det U_1^{(2)}}}{2\pi} \right) - 1. \quad (13)$$

The connection between Lipari–Szabo order parameters S^2 , as obtained from NMR relaxation experiments^{4,13} or from residual dipolar couplings and conformational entropy associated with a NH bond vector has been the subject of many studies aiming at extracting thermodynamic quantities from NMR experiments.^{1,5,14–16} Since both order parameters and conformational entropies can be predicted by the NCR model, the relationship between them can be easily investigated. Thus, Eqs. (2), (11), and (13) determine the formal connection between $S_{1,\text{NCR}}^2$ and the conformational entropy $S_{1,\text{NCR}}^e$ associated with NH vectors.

To illustrate this point, we applied these formulas to the case of the protein calbindin.¹⁷ In all what follows, we used a NCR comprising all backbone NH and CO bond vectors that are within 7.5 Å of each other.¹⁸ The constant κ_0 is fixed to 3 for all pairwise potentials and the atomic density ρ_i for NH or CO bond vectors is determined by the number of atoms within 7.5 Å of the C and N atoms, respectively. The equilibrium values of the internuclear angles θ_{0ij} are taken

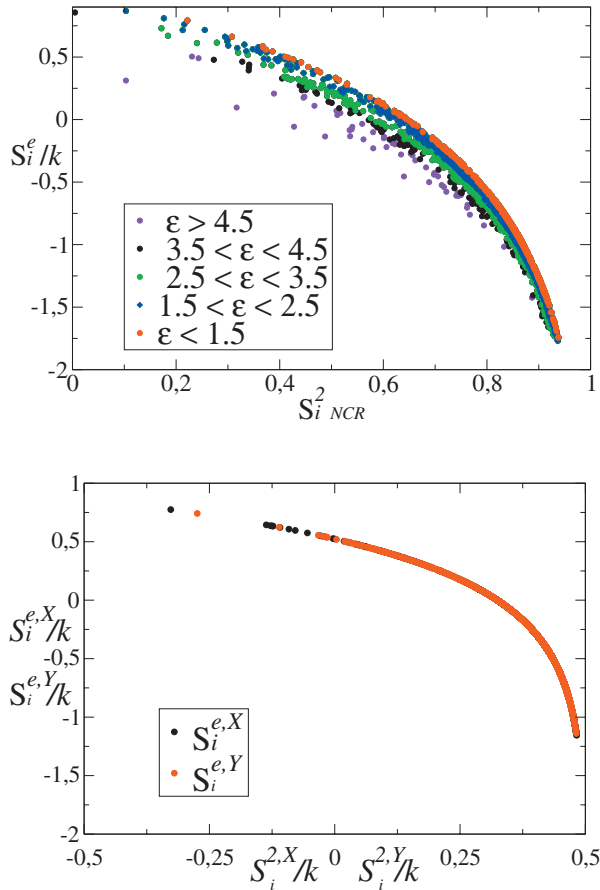


FIG. 1. (Color) Top: Plot of the conformational entropy $S_{i,NCR}^e$ vs $S_{i,NCR}^2$ color coded as a function of the anisotropy ϵ of the effective potential (see text for the definition of ϵ). Bottom: “Directional” entropies $S_i^{e,X}$ and $S_i^{e,Y}$ of amide NH bond vectors.

from the 33 (PDB entry 1CLB) and 10 (PDB entry 1B1G) NMR structures of the free and calcium bound states of calbindin, respectively:^{17,19} the dispersion among the structures within these ensembles leads to standard deviations in all quantities computed by the NCR model. As shown previously, this set of parameters permits the calculation of order parameters $S_{i,NCR}^2$ according to Eq. (2), with linear correlation coefficients relative to experimental data of 0.74 ± 0.08 and 0.71 ± 0.10 for the free and bound states, respectively.⁹ It also allows one to determine the covariance matrix M [Eq. (3)], which contains all the necessary information to compute the conformational entropy of the bond vectors, according to Eqs. (4)–(13). We will only present results relative to NH bond vectors, as no NMR experimental order parameters S^2 are currently available for CO bonds of calbindin to validate our calculated values. However, the same conclusion regarding the link between bond vector conformational entropy $S_{i,NCR}^e$ and order parameter $S_{i,NCR}^2$ holds for CO and NH bond vectors (data not shown), as it would for any type of bond vectors in a protein. The plot of $S_{i,NCR}^e$ versus $S_{i,NCR}^2$ is depicted in Fig. 1, where the scattered graph on the top shows a monotonously decaying trend. This is in accordance with the expectation that conformational entropy decreases as the order parameter increases toward its rigid limit $S_{i,NCR}^2 = 1$. However, in contrast to most earlier models,^{5,15,16} no one-to-one correspondence exists between the two quantities. The

graph reflects the fact that vectors with the same order parameter $S_{i,NCR}^2$ may have different conformational entropies, and conversely, vectors that have the same $S_{i,NCR}^e$ are not necessarily associated with the same order parameter. Obviously, neither quantity provides a complete characterization of the local disorder.

Alternatively, $S_{i,NCR}^e$ can be split into contributions along the two orthogonal directions of the eigenvectors of $U_i^{(2)}$, X_i' , and Y_i' ,

$$S_i^{e,X'} = -k \int p(X_i') \log p(X_i') = \frac{k}{2} [\log(2\pi \langle X_i'^2 \rangle) + 1], \quad (14)$$

$$S_i^{e,Y'} = -k \int p(Y_i') \log p(Y_i') = \frac{k}{2} [\log(2\pi \langle Y_i'^2 \rangle) + 1]. \quad (15)$$

Here, $p(X_i')$ and $p(Y_i')$ denote the marginal averages for the single variables X_i' and Y_i' that define the $N_i H_i$ vector displacements along the principal axes of the effective potential $U_i^{(2)}$. Equations (14) and (15) are analogous to the expression derived by Yang and Kay for a harmonic potential that depends on a single variable.¹⁶ Remarkably, it can be shown that the additivity relationship $S_{i,NCR}^e = S_i^{e,X'} + S_i^{e,Y'}$ holds in the eigenframe of the effective potential $U_i^{(2)}$. Equations (14) and (15) show that, in contrast to the above results, there is a simple one-to-one relationship between conformational entropy and the *pseudo-order parameter* along a single direction, defined as $S_i^{X'^2} = 1/2 - 3 \langle X_i'^2 \rangle$ (see Fig. 1). The order parameter $S_{i,NCR}^2$ is therefore directly related to the sum of the conformational entropies $S_i^{e,X'}$ and $S_i^{e,Y'}$ of the variables X_i' and Y_i' [see Eqs. (2), (14), and (15)],

$$S_{i,NCR}^2 = 1 - \frac{3}{2\pi} \left\{ \exp\left(2 \frac{S_i^{e,X'}}{k} - 1\right) + \exp\left(2 \frac{S_i^{e,Y'}}{k} - 1\right) \right\}. \quad (16)$$

However, $S_{i,NCR}^2 = S_{j,NCR}^2$ for two network vectors does not imply that $\langle X_i'^2 \rangle = \langle X_j'^2 \rangle$ and $\langle Y_i'^2 \rangle = \langle Y_j'^2 \rangle$, but only the less stringent relation $\langle X_i'^2 + Y_i'^2 \rangle = \langle X_j'^2 + Y_j'^2 \rangle$ [see Eq. (2)]. The degeneracy of the conformational entropy S_i^e is embodied in the fact that the effective potential depends on two independent variables, X_i and Y_i , and is characterized by a pair of eigenvalues and the associated pair of eigenvectors (a similar effect was observed by Yang and Kay when relating conformational entropy of individual bond vectors to a two-parameter diffusion-in-a-cone model¹⁶). In the NCR model, the local ellipticity of the effective potential is determined by the ratio of its largest to smallest eigenvalues and defines the local anisotropy ϵ . Whereas the order parameter is a quantity that does not bear any information about the motional anisotropy in the local (X_i, Y_i) plane, as attested by its invariance through a rotation about the z -axis of the local frame, this is not the case for $\langle X_i^2 \rangle$ and $\langle Y_i^2 \rangle$. Thus, two vectors have equal conformational entropies only when they have the same order parameters and when the anisotropies ϵ of their effective potentials are equal. Therefore, the variations of the local

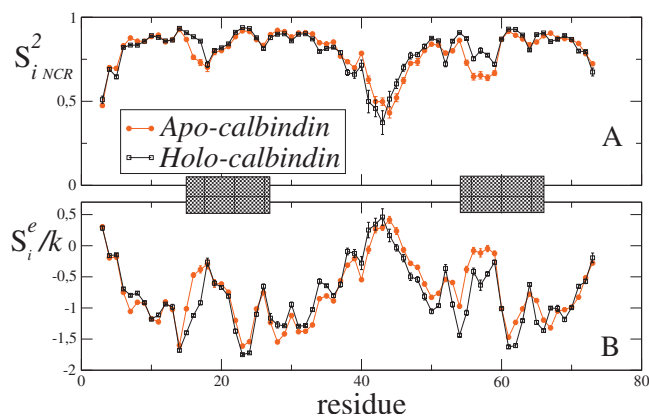


FIG. 2. (Color) Plot of the predicted order parameters $S_{i,NCR}^2$ (top) and conformational entropy $S_{i,NCR}^e$ (bottom) in the protein calbindin. Values for the *apo*- and *holo*-calbindin are depicted by filled circles and open squares. The calcium-binding sites (residues 14–27 and 54–65) are indicated by shaded boxes.

anisotropies ϵ from a network vector to another are responsible for the absence of a one-to-one correspondence between the bond vector conformational entropies and the order parameters. This is illustrated in Fig. 1, where the plot of $S_{i,NCR}^e$ versus $S_{i,NCR}^2$ is depicted with different colors for various anisotropies ϵ of the effective potential.

In general, the quest for an order parameter-conformational entropy relationship is motivated by the possibility to predict thermodynamic quantities, such as ΔG or ΔS , upon ligand binding or protein-protein interactions. Standard experimental techniques, such as isothermal calorimetry, allow one to determine global thermodynamic quantities, relative to the whole protein, and NMR relaxation brings the possibility of breaking down these quantities into residue contributions. Provided that the motions of the residues are decoupled, the partition function of the protein can be factorized into the product of the partition functions of individual vectors and the entropy contributions of each residue add up to the total entropy of the protein.^{5,16} Such a factorization becomes irrelevant to the NCR model, since the latter relies on couplings between all rotators throughout the protein. Nevertheless, as demonstrated above, since one can obtain the protein's partition function from the NCR covariance matrix, it is possible to calculate conformational entropies of individual residues, on the one hand, and the thermodynamical entropy of the protein, on the other hand. Of course, as expected, the $S_{i,NCR}^e$ do not add up to yield the thermodynamical entropy of the protein.

We applied the NCR approach to the calculation of conformational entropies of $N_i H_i$ vectors \mathbf{u}_i in the Ca^{2+} -bound¹⁷ and Ca^{2+} -free¹⁹ forms of calbindin.⁹ Results are depicted in Fig. 2(b) and compared to $S_{i,NCR}^2$ order parameters obtained in previous work, shown in Fig. 2(a). Observation of Fig. 2 reveals that both quantities provide similar dynamical pictures for most residues. Indeed, larger $S_{i,NCR}^2$ are associated with lower $S_{i,NCR}^e$ and vice versa. In particular, the rigidity gained in the Ca^{2+} binding loops²⁰ in *holo*-calbindin is reflected by an increase in $S_{i,NCR}^2$, so that $\Delta S_{i,NCR}^2 = S_{i,NCR}^2(\text{holo}) - S_{i,NCR}^2(\text{apo}) > 0$ (as previously observed⁹) and a decrease in conformational entropy, $\Delta S_{i,NCR}^e = S_{i,NCR}^e(\text{holo})$

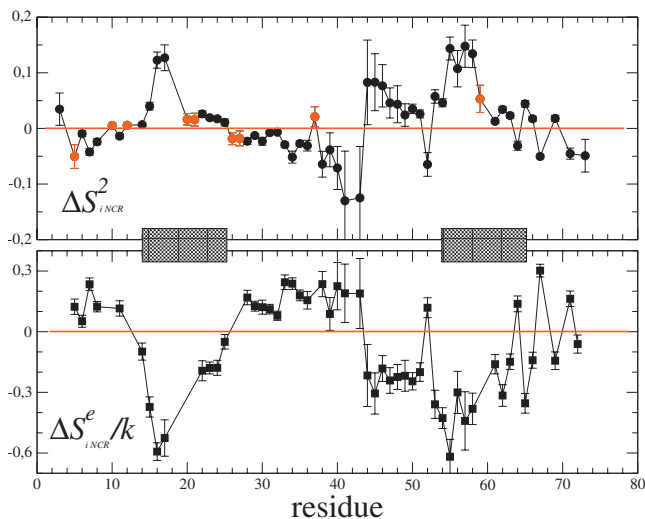


FIG. 3. (Color) Differences of $S_{i,NCR}^2$ and $S_{i,NCR}^e$ between *apo*- and *holo*-calbindin. Only differences larger than the error bars are plotted. The red circles indicate NH vectors for which $\Delta S_{i,NCR}^2 \neq 0$ and $\Delta S_{i,NCR}^e \approx 0$ (residues 5, 10, 12, 20, 21, 26, 27, 37, and 59).

$-S_{i,NCR}^e(\text{apo}) < 0$. The analysis of the differences of $S_{i,NCR}^2$ and $S_{i,NCR}^e$ between *apo*- and *holo*-calbindin shows that for some of the residues, $\Delta S_{i,NCR}^2 \neq 0$, whereas $\Delta S_{i,NCR}^e = 0$ (residues 5, 10, 12, 20, 21, 26, 27, 37, and 59) (Fig. 3).

CONCLUSION

In this article, we have established formal relationships between NMR order parameters and conformational entropies, based on a simple dynamical model of internal motions in a protein. Our results demonstrate that caution must be exercised when estimating local conformational entropy from experimental order parameters, as clearly seen from the lack of a one-to-one correspondence between the two quantities. We therefore suggest to use a theoretical framework such as the NCR model in order to relate order parameters to conformational entropies.

ACKNOWLEDGMENTS

This work has been supported by the European Union (Coordination Action ‘‘Focusing NMR on the machinery of life,’’ and the Integrated Infrastructure Initiative ‘‘EU-NMR’’), the Centre National de la Recherche Scientifique (CNRS), and the Agence Nationale de la Recherche (Contract No. ANR-05-BLAN-0255).

¹V. A. Jarymowicz and M. Stone, *Chem. Rev. (Washington, D.C.)* **106**, 1624 (2006).

²A. Mittermaier and L. E. Kay, *Science* **312**, 224 (2006).

³A. G. Palmer III and C. Lim, *Chem. Rev. (Washington, D.C.)* **104**, 3623 (2004).

⁴G. Lipari and A. Szabo, *J. Am. Chem. Soc.* **104**, 4546 (1982).

⁵M. Akke, R. Brüschweiler, and A. G. Palmer III, *J. Am. Chem. Soc.* **115**, 9832 (1993).

⁶L. Zidek, M. V. Novotny, and M. J. Stone, *Nat. Struct. Biol.* **6**, 1118 (1999).

⁷D. Abergel and G. Bodenhausen, *J. Chem. Phys.* **123**, 204901 (2005).

⁸D. Abergel and G. Bodenhausen, *J. Chem. Phys.* **121**, 761 (2004).

⁹A. Dhulesia, D. Abergel, and G. Bodenhausen, *J. Am. Chem. Soc.* **129**, 4998 (2007).

- ¹⁰W. T. Coffey, Y. P. Kalmykov, and J. T. Waldron, *The Langevin Equation* (World Scientific, Singapore, 1996).
- ¹¹V. A. Daragan and K. H. Mayo, *J. Phys. Chem. B* **103**, 6829 (1999).
- ¹²N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- ¹³G. Lipari and A. Szabo, *J. Am. Chem. Soc.* **104**, 4559 (1982).
- ¹⁴T. I. Igumenova, K. K. Frederick, and A. Joshua Wand, *Chem. Rev.* (Washington, D.C.) **106**, 1672 (2006).
- ¹⁵Z. Li, S. Raychaudhuri, and J. A. Wand, *Protein Sci.* **5**, 2645 (2006).
- ¹⁶D. Yang and L. E. Kay, *J. Mol. Biol.* **263**, 369 (1996).
- ¹⁷J. Kordel, N. J. Skelton, M. Akke, and W. J. Chazin, *J. Mol. Biol.* **231**, 711 (1993).
- ¹⁸B. Halle, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1275 (2002).
- ¹⁹N. J. Skelton, J. Kordel, and W. J. Chazin, *J. Mol. Biol.* **249**, 441 (1995).
- ²⁰M. Akke, N. J. Skelton, J. Kordel, A. G. Palmer, and W. J. Chazin, *Biochemistry* **32**, 9832 (1993).