JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Characterization of Protein Conformational Changes with Sparse Spin-Label Distance Constraints

G. Jeschke*

Lab. Phys. Chem., ETH Zürich, Wolfgang-Pauli-Strasse 10, CH-8093 Zürich, Switzerland

Supporting Information

ABSTRACT: The combination of site-directed spin labeling with pulse EPR distance measurements can provide a moderate number of distance constraints on the nanometer length scale for proteins in different states. By adapting an existing algorithm (Zheng, W.; Brooks, B. R. *Biophys. J.* **2006**, *90*, 4327) to the problem, we address the question to what extent conformational change can be characterized when the protein structure is known for one of the states, whereas only a sparse set of distance constraints between spin labels is available for the other state. We find that the type and general direction of the conformational change can be recognized, while the amplitude may be uncertain.

1. INTRODUCTION

Protein function often relies on transitions between different three-dimensional structures, for instance, for substrate binding in enzymes,[1] transmission of mechanical energy in motor proteins,[2] and translocation of substrates across the cell membrane in active transporters.[3,4] In some cases, the different conformations can be captured in crystal structures; an extensive list of examples was collected in the Database of Molecular Movements (www.molmovdb.org).[5] However, in many cases, it proves difficult or even impossible to find sets of crystallization conditions that stabilize each of the individual states involved in the function of a protein. Moreover, such crystallization conditions are often distinct from physiological conditions, owing to unusual pH values and, for membrane proteins, owing to the presence of harsh detergents and the absence of a lipid bilayer. Thus, there is a need for alternative characterization techniques that can be applied to proteins in more native-like environments and do not require crystallization. Spectroscopic techniques, such as NMR[6] and electron paramagnetic resonance (EPR) spectroscopy, are promising candidates.

Since most proteins are diamagnetic, application of EPR spectroscopy requires site-directed spin labeling (SDSL).[7] By measuring distances between spin labels in the 1−2 nm range by continuous-wave (cw)[8] techniques and in the 2−6 nm range by pulse EPR techniques,[9−11] qualitative and semiquantitative models for protein structures and conformational changes have been obtained.[12,13] However, a general approach for characterization of conformational changes from such long-range distance constraints is still missing.

The main obstacle in the development of such an approach is sparsity of the distance constraints,[14−18] which arises from the requirement to generate, express, spin label, and measure one double mutant of the protein for each constraint. Typical studies rely on 10−30 constraints, with about 100 constraints for a protein with several hundred residues probably being the current limit. It follows that conformational changes can be characterized only on a coarse-grained level, but even then substantial additional information is needed. Assumption of canonical bond lengths and bond angles, as for solving structures in X-ray crystallography and NMR spectroscopy, and local optimization based on molecular force fields are usually not sufficient to solve the problem. This creates a situation where it is difficult to estimate uncertainty of the models.

The problem is aggravated by the use of label-to-label distance constraints, which are related to backbone-to-backbone distance constraints by the generally unknown conformational distribution of the spin label.[19,20] Reasonably precise predictions of the mean distance can be obtained by rotamer library approaches.[21] An early version of this approach was successful in determining the relative arrangement of the protomers in the dimer of the sodium/proton antiporter NhaA of *Escherichia coli*.[21,22] This study assumed rigid-body motion of the protomers to avoid the problem of sparse constraints. Later, it was shown by electron microscopy that in the dimer a β-sheet domain was differently arranged with respect to the transmembrane core than in the X-ray structure of the monomer, probably owing to packing effects.[23] This exemplifies the requirement for flexible fitting methods.

Furthermore, a recent comparison between experimental distance measurements on proteins with known structures and rotamer library predictions for spin label conformations reveals that the predicted mean distances exhibit a standard deviation of about 3 Å from the experimental values.[24] Similar deviations are found for more elaborate predictions by force-field based molecular dynamics approaches.[20] Part of the deviation may result from differences in crystal and solution structure and the mean deviation may also be reduced by avoiding tight labeling sites.[25] However, detailed analysis of some of the larger deviations reveals that a substantial part of the uncertainty is due to poor estimates of interaction energies.[24,26] Hence, an estimate of model reliability and uncertainty should take distance errors of this magnitude into account.

Special Issue: Wilfred F. van Gunsteren Festschrift

Received: February 8, 2012
Published: April 13, 2012

A promising approach for flexible fitting of conformational changes with a moderate number of distance constraints was suggested by Zheng and Brooks.[27] This approach is based on a residue-level anisotropic elastic network model (ANM)[28] and restricts the protein motion to a small subset of the normal modes of the ANM. To overcome restriction to small-scale changes where the harmonic approximation inherent in the ANM is still valid, the Zheng/Brooks fit algorithm uses interactive reorientation of the modes. The distance constraints are combined with local pseudoenergy terms to set up a linear regression problem that is iteratively solved. To generate an optimum set of distance constraints, an auxiliary algorithm selects pairs of residues with maximum mutual linear independence with respect to the basis set of normal modes. Fits with a basis size of 10 modes and 10 $C^{\alpha}-C^{\alpha}$ distance constraints provided a good description of conformational changes for a set of 16 proteins with about 100 to 700 residues, assuming an error of the constraints up to 2 Å.[27]

Our initial tests with the pair selection algorithm showed that most distances are in the range accessible with pulse dipolar EPR spectroscopy experiments, such as the DEER[9,10] and DQ-EPR[11] sequence. The few remaining distances fall into the cw EPR range.[8] This match between the Zheng/Brooks algorithm and EPR distance measurements is caused by the match between the EPR length scale and protein dimensions.

Several open questions need to be addressed to develop this approach into a generally applicable methodology for SDSL EPR studies. First, closer analysis of the test set of proteins shows that almost all test cases correspond to highly collective domain hinge motions. Since the success of ANM in the description of conformational changes depends on the type of motion,[29] and in particular on its collectivity,[30] the approach needs to be tested on a more comprehensive set of conformational changes. Second, spin labels have to be integrated into the approach, and it needs to be tested how this influences uncertainty of the models, in particular, considering the uncertainty of predictions of the conformational distribution of the labels and the restriction of spin labeling to only a fraction of residues of the protein. Third, it must be assessed what information can and cannot be derived from models for conformational changes that are generated by such an approach. This work addresses all these questions.

The article is structured as follows. We start with introducing the methodology, including slight modifications in the original Zheng/Brooks algorithms. Then, we assess error sources in the approach step by step. First, we study to what extent conformational changes of different types can be covered by a small number of normal modes of an ANM. We compare normal modes based on only the starting structure with reoriented and completely recomputed modes. Second, we inquire to what extent this potential coverage of the conformational changes by the basis set of modes can be realized by fitting with a small number of exact $C^{\alpha}-C^{\alpha}$ distance constraints. Third, for those conformational changes that can be reasonably characterized by such $C^{\alpha}-C^{\alpha}$ distance constraints, we assess the decrease in coverage of the transition when using exact label-to-label distance constraints. We check whether this performance loss can be compensated by a larger ratio of constraints to normal modes. Fourth, we investigate the influence of random distance errors with a standard deviation of 3 Å on the quality of the results. Fifth, we test whether the main characteristics of conformational changes are revealed by the fits. We conclude with a discussion of possible future improvements.

2. METHODOLOGY

2.1. Modifications in the Zheng/Brooks Algorithm. The Zheng/Brooks algorithm requires the initial structure $\mathbf{R}_i$ of the transition and a small number of $C^{\alpha}-C^{\alpha}$ distance constraints for the final structure to compute the full final structure $\mathbf{R}_f$. The initial structure is used to compute a matrix $\mathbf{u}$ of normal modes of an ANM and the corresponding vector of eigenvalues λ by diagonalization of the Hessian matrix as defined by Bahar et al.[28]

This algorithm was initially implemented into the software package MMM[21] as described in the original publication.[27] Prompted by a recent discussion on advantages of distance-dependent force constants in ANM,[33,34] we now use force constants that scale with r^{-6}, where r is the distance between the two C^{α} atoms.[35] Furthermore, we have addressed the problem of local structure deformation by multiplying the force constants for direct neighbors (residues n and $n + 1$) and next neighbors (residues n and $n + 2$) in the peptide backbone by a factor of 10 000. This strongly constrains the corresponding distances. The $(n,n + 1)$ distance is fixed by peptide bond geometry and the $(n,n + 2)$ distance by the relative stability of the secondary structure. Note that recent parametrization of an ANM by fitting to molecular dynamics simulations also led to strongly enhanced force constants for $(n,n + 1)$ pairs and $(n,n + 2)$ pairs.[36] Such modification of the Hessian obviates the need for an energy minimization step by recruiting normal modes $\mathbf{u}_{B+7}$... $\mathbf{u}_{3N}$. This step was required in the original Zheng/Brooks algorithm to prevent local deformation. Furthermore, Zheng and Brooks added an elastic energy cost term for direct and next neighbors in the pseudoenergy function used for setting up the linear regression problem. We find that this term has to be retained despite using larger force constants in the network model.

Furthermore, we have implemented the option to run the algorithm with a repeated Hessian setup and diagonalization rather than normal mode reorientation. In the pair selection algorithm, we start with a correlation cutoff $C_{cutoff} = 0.3$ as Zheng and Brooks do, but we automatically increase this value in steps of 0.05 until the requested number of constraint pairs is found.

2.2. Integration of Spin Labels in the Zheng/Brooks Pair Selection Algorithm. On the initial structure, we first perform a spin labeling site scan[25] for methane thiosulfonate spin labels at a temperature of 175 K in MMM using the recently introduced rotamer library approach.[21] In this paper, we restrict labeling sites to the residue types Cys, Ile, Leu, Met, Ser, Thr, and Val, where labeling is least likely to cause structural or functional changes. The user of MMM can override this choice. Site pair selection has been implemented with a dedicated graphical user interface window in MMM. When this window opens, the list of successfully *in silico* labeled sites is separated into advisible and unadvisible labeling sites. This separation is based on the number of rotamers that do no clash with the protein. By default, we advise labeling only if at least 10 rotamers are populated, thus excluding restricted sites where errors in prediction of the mean label position are expected to be particularly large.[25] The user can override this value and can shift sites from the selected to the unselected list or *vice versa*. Such editing of the site list can be based on knowledge on binding sites, on biochemical results, and on the

success of spin labeling experiments and functional assays. In this work, we generally excluded sites at the two N terminal and two C terminal residues of the peptide chain.

The Zheng/Brooks pair selection algorithm is run on only the subset of selected labeling sites. Furthermore, only pairs are considered for which the label-to-label distance in the initial structure falls within a certain distance range. If cw EPR distances are allowed, the minimum of the distance range is 1.0 nm; otherwise it is 1.7 nm. In this work, we did not allow for distances in the cw EPR range. The default value for the upper limit, used throughout this work, is 5 nm. Depending on properties of the protein and sensitivity of the available spectrometer, this value can be changed. Finally, the user selects the number P of requested pairs. By default, the algorithm is run with a matching number of pairs and low-frequency modes $P = B$. This choice can also be overridden by the user. Initially a correlation cutoff of 0.30 is used. If the number of selected pairs is smaller than P, the correlation cutoff is increased in increments of 0.05 until P pairs are found or $C_{cutoff} = 0.75$ is reached. In this work, the number of requested pairs was always found with $C_{cutoff} \leq 0.70$.

For test purposes, a constraint file can be generated for a target structure (final structure of the transition). This requires that a PDB file of this structure also be loaded into MMM and that all selected labeling sites be resolved in the target structure. In this work, we have manually removed those residues from the list of allowed sites that were missing in the target structures. The constraint file lists chain identifiers and residue numbers as well as distances and standard deviations of the label-to-label distance distribution. Note that distances in this constraint file may be somewhat outside the requested distance ranges, as the target structure differs from the initial structure.

The motion-in-a-cone model[15,37] or clash analysis by MtsslWizard[24] could be considered as an alternative to the rotamer library approach. Among these models, the motion-in-a-cone model does not make predictions about the suitability of labeling sites and probably predicts the mean spin label position required in the fit algorithm (*vide infra*) less precisely. MtsslWizard appears to have the same precision for the mean position but sometimes fails to recognize sites where labeling is possible.[24] Therefore, we did not test these approaches in this context.

2.3. Integration of Spin Labels in the Zheng/Brooks Fit Algorithm. Spin labels are attached *in silico* to the initial protein structure by the rotamer library approach at the sites featuring in distance constraints. The population weighted average over all rotamers of the N−O bond midpoint coordinate is defined as the label coordinate. During the fit, the vector between the C^α atom of the labeled residue and the label coordinate moves together with a local residue frame. This is similar in spirit to mode reorientation in the Zheng/Brooks algorithm. Translation and rotation of the local residue frame during an incremental structural change is determined by least-squares coordinate superposition of the C^α atoms of residues $n - 2$... $n + 2$, where n is the labeling site. If one or two of these residues are missing in the structure, only the existing ones are superimposed. If less than three of these C^α atoms are present, the label coordinate is changed by the same vector as the C^α coordinate of residue n.

This coarse-grained description of label reorientation misses possible changes in the conformational distribution of the label and thus introduces some constraint error. Hence, even if exact distance constraints for the target structure are used and the C^α coordinates of the target structure could be matched exactly by the fit, some residual deviation for the distance constraints would be expected.

A dedicated graphical user interface window of MMM was implemented for fitting. On loading of a constraint file, the labels are automatically attached, and matching of the distance constraints is visualized with a traffic light color code in a wire model of the structure and in a constraint plot. After fitting, the wire model can be animated to visualize the conformational change.

2.4. Assessing Coverage of the Structural Change. To benchmark fit results, we start by superimposing the two vectors of $3N$ Cartesian C^α atom coordinates each, $\mathbf{R}_i = (x_{i1}, y_{i1}, z_{i1}; x_{i2}, y_{i2}, z_{i2}, ..., x_{iN}, y_{iN}, z_{iN})$ for the initial and $\mathbf{R}_f$ for the final structure by the translation and rotation that minimizes their root-mean-square deviation[31] (RMSD)

$$|\Delta\mathbf{R}| = |\mathbf{R}_f - \mathbf{R}_i| \tag{1}$$

The normal modes $\mathbf{u}_k$, which are the eigenvectors of the Hessian, are ordered by ascending eigenvalues λ_k. The coordinate change $\Delta\mathbf{R} = \mathbf{R}_f - \mathbf{R}_i$ can then be written as a linear combination of the normal modes

$$\Delta\mathbf{R} = \sum_k c_k \mathbf{u}_k \tag{2}$$

where the coefficients c_k are scalar products of $\Delta\mathbf{R}$ with the corresponding eigenvector $\mathbf{u}_k$. In the absence of special symmetry, $c_k = 0$ for $k = 1-6$, as the first six normal modes describe translation and rotation of the network and have eigenvalues of zero within numerical precision.

We now consider a reduced basis of B normal modes $\mathbf{u}_7$... $\mathbf{u}_{B+6}$. The linear combination of these B modes with unchanged coefficients c_7... c_{B+6} provides an estimate $\Delta\mathbf{R}_0$ of the coordinate change that has least-squares deviation from $\Delta\mathbf{R}$ among all of the possible linear combinations of these B modes. We define the fractional coverage of the conformational change by this basis of B modes as

$$f_B = |\Delta\mathbf{R}_0|/|\Delta\mathbf{R}| \tag{3}$$

In general, f_B increases with increasing basis size B.

In the context of the Zheng/Brooks fit algorithm, a larger coverage $f_{B,r}$ can be achieved by iteratively updating the modes by reorientation.[27] An estimate of the improved fractional coverage $f_{B,r}$ can be obtained by the following iterative procedure. For a single step, the coordinate change $\Delta\mathbf{R}_0$ is scaled by a factor s, so that in the first iterated structure $\mathbf{R}_1 = \mathbf{R}_i + s\Delta\mathbf{R}_0$, no C^α atom moves by more than 0.2 Å with respect to $\mathbf{R}_i$. Using these changed coordinates, the B normal modes are reoriented, and a new set of coefficients c_7... c_{B+6} is computed for the reoriented modes. These coefficients and modes provide a new estimate $\Delta\mathbf{R}_1$ for the remaining coordinate change. This cycle is iterated until the estimated final structure is converged to $\mathbf{R}_{f,r}$. The fractional coverage with mode reorientation is then given by $f_{B,r} = |\mathbf{R}_{f,r} - \mathbf{R}_i|/|\Delta\mathbf{R}|$.

The same algorithm can be applied to recomputation of the normal modes by setting up and diagonalizing a new Hessian for each intermediate structure $\mathbf{R}_j$. We define the fractional coverage $f_{B,d}$ with iterative diagonalization analogously. Furthermore, we define the fractional coverage f_{fit} for a structure $\mathbf{R}_{fit}$ obtained with the Zheng/Brooks fit algorithm as $f_{fit} = |\mathbf{R}_{fit} - \mathbf{R}_i|/|\Delta\mathbf{R}|$. These values allow for assessing which steps or aspects of the approach introduce deviations between the true and modeled conformational change. Furthermore,

Table 1. Characterization of the Set of Structure Pairs by Peptide Chain Length, C^α RMSD $|\Delta\mathbf{R}|$ between the Two Structures, Collectivity κ of the Coordinate Change, Number of Modes $B_{2\text{Å}}$ Required to Fit the Target within 2 Å C^α RMSD, Fractional Coverages f_{10} with 10 and f_{20} with 20 Modes, $f_{20,r}$ and $f_{20,d}$ with 20 Reoriented or Recomputed Modes, respectively, f_{fit} with a Fit by 20 C^α−C^α Distance Constraints and a Basis of 20 Modes, and Remaining C^α RMSD $|\Delta\mathbf{R}_{fit}|$ of the Fit

transition	size	$\|\Delta\mathbf{R}\|$ (Å)	type[a]	κ	$B_{2\text{ Å}}$	f_{10}	f_{20}	$f_{20,r}$	$f_{20,d}$	f_{fit}	$\|\Delta\mathbf{R}_{fit}\|$
1L5BA/1L5EA	101	6.51	d/s	0.694	31	0.509	0.643	0.705	0.792	0.377	4.06
1BUY/1EERA	166	4.08	l/a	0.354	123	0.226	0.312	0.358	0.402	<0	4.23
1CBU/1C9KB	180	3.11	f/s	0.254	17	0.254	0.376	0.351	0.412	0.208	2.46
1GRNA/1CF4A	184	5.51	d/n	0.362	183	0.116	0.317	0.339	0.429	0.156	4.65
1AKEA/4AKE4	214	7.13	d/h	0.481	170	0.432	0.486	0.444	0.595	0.472	3.77
1GGGA/1WDN	220	5.34	d/h	0.747	4	0.679	0.693	0.806	0.789	0.571	2.29
2LAO/1LST	238	4.70	d/h	0.723	2	0.762	0.771	0.872	0.857	0.596	1.90
2JLN/2X79	351	3.46	l/c	0.579	40	0.228	0.315	0.320	0.378	0.125	3.03
1OMP/1ANF	370	3.77	d/h	0.723	2	0.710	0.731	0.761	0.761	0.589	1.55
1EPS/1G6S	427	7.59	d/h	0.699	19	0.715	0.737	0.841	0.837	0.663	2.55
1BNCA/1DV2A	433	3.86	d/h	0.263	1	0.706	0.723	0.851	0.840	0.556	1.71
1AONA/1OELA	524	12.38	l/c	0.521	273	0.582	0.612	0.786	0.814	0.682	3.94
2HMIA/3HVTA	555	3.45	d/r	0.651	17	0.390	0.426	0.434	0.435	<0	5.64
1D6M/1I7DA	620	3.40	d/n	0.620	17	0.329	0.538	0.554	0.556	0.081	3.12
1LFH/1LFG	691	6.43	d/h	0.491	66	0.622	0.646	0.737	0.795	0.721	1.79
1B7TA/1DFKA	720	11.51	d/n	0.433	26	0.680	0.778	0.803	0.874	0.751	2.86
1N0VC/1N0U	819	13.61	d/h	0.707	208	0.603	0.700	0.777	0.818	0.749	3.41
1SU4/1T5S	994	13.55	l/c	0.636	588	0.513	0.616	0.762	0.783	0.656	4.67

[a]first character: d domain, f fragment, l larger than domain, second character: a allosteric, c complex, h hinge, n not classified, r partial refolding, s shear.

normalization to $|\Delta\mathbf{R}|$ allows for better comparison of the performance of the approach for a strongly different extent of the structural change.

To test the hypothesis that coverage is correlated with collectivity of the structural change, we apply the definition of normal mode collectivity κ by Brüschweiler[32] to the structural change vector:

$$\kappa = \exp\{-\sum_{m=1}^{3N} \Delta R_m{}^2 \log \Delta R_m{}^2\}/(3N) \quad (4)$$

2.5. Estimate of Uncertainty of the Fit. To the "exact" label-to-label distance constraints generated from the target structure, normally distributed pseudorandom numbers with a standard deviation of 3.0 Å are added, and a fit is performed with these erroneous constraints. The number of trials and the standard deviation of the pseudorandom error can be specified in the constraint file with the keyword ENSEMBLE. In this work, we performed 10 trials for each conformational change studied. During the fit, the approach to the target structure can be monitored for test purposes by including in the constraint file the keyword TARGET with the PDB identifier as an argument. The modified sets of distance constraints for all trials and diagnostic information on the fit are saved in a log file, provided in the Supporting Information together with example constraint files. The mean fractional coverage $\langle f_e \rangle$ and standard deviation σ_f for the ensemble are computed, and the structure with the fractional coverage $f_{fit,i}$ that is closest to $\langle f_e \rangle$ is defined as the representative structure of the ensemble.

2.6. Reverse Coarse-Graining. Structural models generated in this work are coarse and should not be discussed on an atomistic level. Nevertheless, fully atomistic coordinate files of the ensembles in PDB format are provided in the Supporting Information for convenient analysis and visualization with standard software. The full set of atom coordinates of the model is generated from the full set of coordinates of the template structure and the C^α atom coordinates of the fit. Local residue frames are defined in the same way as for spin label reorientation (*vide infra*), and all atom coordinates of a residue are subjected to the same transformation between the initial and final structure as the local residue frame. Note that this may distort peptide bond geometry and may lead to clashes between sidegroup atoms. Better models can be obtained with reconstruction from only the C^α trace by using maxsprout.[38,39]

One might expect that conformational distributions of spin labels computed from such a model of the final structure provide better estimates of the mean local-frame N−O bond coordinates. If so, a second fit starting from the reverse coarse-grained final structure of the first fit might improve the result. We have tested this proposition and found no significant improvement.

3. RESULTS

3.1. Coverage of Structural Changes by a Small Fraction of Normal Modes. In order to assess utility of the approach for a wide range of possible structural changes, we constructed a test set of 18 structure pairs with a C^α RMSD between 3 and 14 Å, between 100 and 1000 residues, and with different types of motion of fragments, domains, and subgroups of residues larger than domains. The classification of motions in the Database of Macromolecular Movements was adopted for this.[5] For comparison, we selected seven structure pairs that were also included in the test set of Zheng and Brooks, covering a broad range of protein size and C^α RMSD. We also added two structure pairs that were studied in the context of flexible fitting of small-angle X-ray scattering (SAXS) data based on normal modes.[40−43] Of the remaining nine structure pairs, eight were selected from the Database of Macromolecular Movements[5] with the purpose to add motions of as many types as possible and the only restriction that the C^α RMSD between the two structures must exceed 3 Å. The set was completed with a structure pair for the nucleobase−cation symporter Mhp1 in order to include a membrane protein. The set of structure pairs is listed in Table 1, where a fifth character of the

structure identifiers denotes the chain in multichain PDB files. Structure 1EPS is available only as a C^{α} trace, whereas *in silico* spin labeling requires a complete structure. The complete structure was generated with maxsprout.[38,39]

For four structure pairs, dependence of the absolute mode scaling coefficients $|c_k|$ on mode number $k - 6$ is shown in Figure 1 for the lowest 100 significant modes. The values were

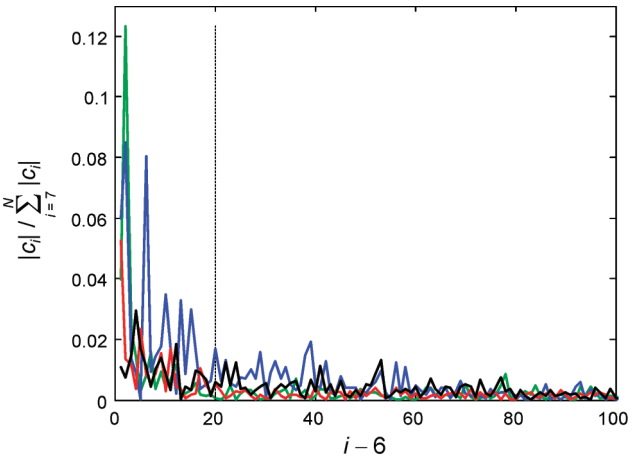

Figure 1. Relative contribution of low-frequency normal modes to conformational changes of four selected pairs of protein structures, 1L5B/1L5E (blue), 2JLN/2X79 (black), 1OMP/1ANF (green), and 1SU4/1T5S (red). Shown are the normalized absolute coefficients of modes 7−106 in the linear combination that reproduces the structural change. The vertical dashed line denotes the basis size of 20 modes.

normalized to the sum of $|c_k|$ for all modes. Clearly, the number of significant low-frequency modes does not correlate with the number of residues, as it does not differ much between the smallest proteins (1L5B/1L5E, 101 residues, blue) and the largest protein (1SU4/1T5S, 994 residues, red) in the set. The reduced relative contribution of the 20 lowest modes to the structural change for the larger protein is due to the much larger number of high-frequency normal modes, which all add very small but nonvanishing contributions. Dominance of a single mode is typical for simple hinge motions (1OMP/1ANF, green). For a similar size of the protein, distribution of the $|c_k|/\Sigma_k |c_k|$ may vary quite strongly with the type of motion, as is seen by comparing the plots for 1OMP/1ANF (370 residues, green) and 2JLN/2X79 (351 residues, black).

To characterize how well a basis set of slow normal modes covers the structural changes, we first determined the number $B_{2Å}$ of low-frequency normal modes of the ANM that is required to approximate the final structure with a C^{α} RMSD of 2 Å. This number varies strongly and tends to increase with increasing magnitude of the structural change $|\Delta\mathbf{R}|$ and decreasing collectivity κ of the motion (Table 1). Fractional coverage of the structural change by 10 modes (f_{10}) or 20 modes (f_{20}) varies much less and is anticorrelated with collectivity, but not with protein size or $|\Delta\mathbf{R}|$. In many cases, fractional coverage does not increase strongly when increasing basis size from 10 to 20 modes, in particular not for hinge motions and for some of the complex motions that are multiple hinge motions. However, in some cases, the improvement is significant. Thus, we study further issues with a basis size of 20 modes or with both basis sizes.

In about half of the cases, mode reorientation does not improve fractional coverage significantly, as can be seen by comparing the f_{20} and $f_{20,r}$ values in Table 1. However, in the remaining cases, the improvement exceeds 10%. In fact, in a few cases significant further improvement is obtained with repeated setup and diagonalization of the Hessian (compare $f_{20,r}$ and $f_{20,d}$). Although the latter approach is computationally much more expensive, we thus performed all further computations with repeated Hessian setup and diagonalization, as the computational expense is still negligible compared to the expense of acquiring experimental data. With this approach, fractional coverage of the conformational changes by the lowest 20 normal modes exceeds 1/3 irrespective of the type of motion. For hinge motions, fractional coverage exceeds 3/4, except for adenylate kinase (1AKE/4AKE), where it is approximately 0.6.

3.2. Fit Results with C^{α}−C^{α} Distance Constraints. For all conformational changes, we created 20 C^{α}−C^{α} distance constraints, using the Zheng/Brooks pair selection algorithm and fitted these constraints with the Zheng/Brooks fit algorithm using a basis of 20 modes. For the examples used by Zheng and Brooks,[27] our fits are slightly worse, except for the structure pair 1LFH/1LFG, where our fit with 20 constraints and 20 modes is better (C^{α} RMSD 1.8 instead of 2.7 Å). In the remaining cases, the differences are well below the expected uncertainty of structures that could be obtained from experimental restraints.

Domain hinge motions are reproduced with fractional coverage well above 0.5, except for the pair 1AKE/4AKE, where the coverage is slightly worse. In this case, relative reorientation of the two domains is more dramatic than in the remaining structures. Fractional coverage of approximately 2/3 is obtained for two of the complex motions of subgroups of residues larger than domains (1AON/1OEL and 1SU4/1T5S). Both of these conformational changes can be considered as a superposition of several hinge motions. Reorientation of the lever arm with respect to the motor domain in myosin (1B7T/1DFK) is also well reproduced with coverage of about 0.75.

In contrast, for the relative rearrangement of packed helices in the nucleobase-cation symporter Mhp1 (2JLN/2X79) fitting with distance constraints reduces the already mediocre coverage of 0.375 to only 0.125. Tama and Sanejouand reported earlier that low-frequency normal modes are less successful for predicting such rearrangements of helices inside densely packed regions than for predicting large-scale domain motion.[29] The GTP binding protein (1GRN/1CF4) is another example. For the shear motions 1CBU/1C9K and 1L5B/1L5E, we observe a reduction of the coverage by a factor of 2 when fitting with distance constraints.

In three cases, the Zheng/Brooks fit algorithm fails, although the 20 lowest normal modes formally account for 40−55% of the structural change. The least surprising of these cases is HIV-1 reverse transcriptase (2HMIA/3HVTA), where the conformational change involves partial refolding. Such a transition cannot be expected to proceed along instantaneous normal modes. The allosteric transition of erythropoietin (1BUY/1EER) and the unclassified domain motion of topoisomerase III (1D6M/1I7D) are the other two failures. The former case might also be classified as a rearrangement of helices within a densely packed region, while in the latter case the DNA binding domain rotates with respect to the remaining protein. These three structure pairs are excluded from further study, so that the test set for the SDSL EPR approach consists of 15 conformational changes.

Table 2. Characterization of Fits of Conformational Changes by Label-to-Label Distance Constraints, Including Number N_{sites} of Spin Labeling Sites, Fractional Coverages $f_{20,20}$ with 20 Constraints and 20 Modes, $f_{20,10}$ with 20 Constraints and 20 Modes, $f_{50,20}$ with 50 Constraints and 20 Modes, Mean Fractional Coverage $\langle f_e \rangle$ for an Ensemble of 10 Structures, Assuming 3 Å Standard Deviation of the Distance Constraints, Standard Deviation σ_f of the Fractional Coverage in the Ensemble, Mean C^α RMSD $\langle |\Delta R_{fit}| \rangle$ of the Fit from the Target, Standard Deviation $\sigma\Delta_R$ of This RMSD, and Mean RMSD $\langle \Delta r \rangle$ of the Distance Constraints after the Fit

transition	N_{sites}	$f_{20,20}$	$f_{20,10}$	$f_{50,20}$	$\langle f_e \rangle$	σ_f	$\langle \lvert\Delta R_{fit}\rvert \rangle$ (Å)	$\sigma_{\Delta R}$ (Å)	$\langle \Delta r \rangle$ (Å)
1L5BA/1L5EA	42	0.226	0.500	0.560	0.405	0.057	3.88	0.37	4.29
1CBU/1C9KB	33	<0	<0	<0					
1GRNA/1CF4A	44	0.017	<0	0.098					
1AKEA/4AKEA	46	0.405	0.251	0.459	0.389	0.059	4.36	0.42	4.22
1GGGA/1WDN	42	0.177	0.518	0.575	0.370	0.092	3.36	0.49	3.96
2LAO/1LST	39	0.589	0.707	0.630	0.422	0.072	2.72	0.34	3.78
2JLN/2X79	132	<0	0.075	0.256	0.143	0.061	2.96	0.21	3.84
1OMP/1ANF	49	0.172	0.520	0.559	0.248	0.215	2.84	0.81	3.73
1EPS/1G6S	88	0.567	0.400	0.597	0.460	0.182	4.10	1.38	4.75
1BNCA/1DV2A	74	0.140	0.603	0.545	0.451	0.096	2.12	0.37	4.07
1AONA/1OELA	138	0.008	0.644	0.628	0.541	0.060	5.68	0.74	5.32
1LFH/1LFG	123	0.339	0.173	0.518	0.299	0.157	4.51	1.01	4.14
1B7TA/1DFKA	156	0.293	0.158	0.606	0.415	0.103	6.73	1.19	4.39
1N0VC/1N0U	162	0.597	0.698	0.768	0.528	0.048	6.42	0.65	5.80
1SU4/1T5S	251	0.340	0.593	0.738	0.505	0.055	6.71	0.74	4.58

3.3. Fit Results with Label-to-Label Distance Constraints. Restrictions on labeling sites reduce the number of available network nodes for pair selection. This in turn requires larger correlation cutoff values in the pair selection algorithm to generate the same number of constraints. Generation of 20 C^α−C^α constraints for 20 modes succeeds with correlation cutoff values between 0.30 (more than 250 residues) and 0.35 (more than 150 residues), with the exception of the pair 1L5B/1L5E that requires C_{cutoff} = 0.40. To generate the same number of label-to-label constraints, C_{cutoff} must be increased by 0.05 to 0.30, depending on the protein. This causes a stronger linear dependence with respect to the basis of normal modes of the label-to-label vectors compared to the C^α−C^α vectors.

Furthermore, conformational distribution of the label in the initial and final structure will generally differ. This is not accounted for by reorientation of the C^α-label vector with the local residue frame. Hence, the distances computed for the final structure cannot be exactly realized even for a perfect fit of the C^α coordinates. As a result, the computed constraints feature a conformation error. We find that this conformation error leads to a more serious overfitting problem than is observed with C^α−C^α constraints. In the latter case, overfitting may occur as the mode basis does not allow for fitting the true C^α coordinates, so that an attempt to exactly fit the constraints leads to deformation of the structure compared to the best possible fit. For C^α−C^α distance constraints, the deformation rarely exceeds 0.4 Å C^α RMSD (typically 0.1−0.2 Å). With label-to-label constraints, overfitting is serious for many of the larger proteins. For 1BNC/1DV2, we find a deformation of 1.4 Å; for 1AON/1OEL, 4.8 Å; for 1LFH/1LFG, 1.2 Å; and for 1B7T/1DFK, 3.8 Å. This suggests a different convergence criterion, and we shall come back to that later.

Furthermore, we find that for some transitions that can be reproduced with reasonable coverage from C^α−C^α constraints, fits with 20 label-to-label constraints and a basis of 20 modes fail (see $f_{20,20}$ in Table 2). This prompted us to investigate whether influence of the conformation error and larger linear dependence of the constraints can be compensated by increasing the ratio of constraints to normal modes. Indeed, reducing the number of slow modes to B = 10 for the same number of 20 constraints improves fractional coverage for most but not all transitions (see $f_{20,10}$ in Table 2). For the pairs 1AKE/4AKE, 1EPS/1G6S, 1LFH/1LFG, and 1B7T/1DFK, the result becomes significantly worse. In most cases, similar fit quality as with 20 C^α−C^α constraints and B = 20 can be attained with 50 label-to-label constraints and B = 20 (compare f_{fit} in Table 1 and $f_{50,20}$ in Table 2). For 1L5B/1L5E, 2JLN/2X79, and 1SU4/1T5S, the 50 label-to-label constraints provide significantly better fits than 20 C^α−C^α constraints, whereas for 1LFH/1LFG and 1B7T/1DFK, fit quality is significantly worse. For the conformational changes 1CBU/1C9K and 1GRN/1CF4, fits fail even with 50 label-to-label constraints.

For all of the other transitions, we studied the influence of stochastic errors of the distance constraints on fit quality. For this, we added normally distributed pseudorandom numbers with a standard deviation of 3 Å to the exact label-to-label distance constraints for the target structure. For all structure pairs, 10 such sets of erroneous distance constraints were generated, and an ensemble of 10 models for the final structure was computed. To avoid serious overfitting, we employed a convergence criterion based on the RMSD $\langle \Delta r_0 \rangle$ of the distance constraints from the label-to-label distances in the initial structure and on the uncertainty of the distance constraints σ_c = 3 Å. From these values, we computed thresholds $t_1 = (3\sigma_c + \langle \Delta r_0 \rangle)/4$ and $t_2 = \sigma_c + \langle \Delta r_0 \rangle/8$. If the pseudoenergy convergence criterion of the Zheng/Brooks algorithm is not met before reaching the larger one of these two thresholds, the fit is stopped at the threshold. Fits were performed with 20 label-to-label distance constraints and a basis of 10 slow modes for all structure pairs for which $f_{20,20}$ or $f_{20,10}$ was at least 0.5. In the remaining cases, 50 constraints and a basis of 20 slow modes were employed.

In Table 2, the model ensembles are characterized by mean fractional coverage $\langle f_e \rangle$ of the conformational change, the standard deviation σ_f of fractional coverage, the mean RMSD $\langle |\Delta R_{fit}| \rangle$ of the final fitted structure from the target, the corresponding standard deviation $\sigma_{\Delta R}$, and the mean residual RMSD of the distance constraints $\langle \Delta r \rangle$ after fitting. In general

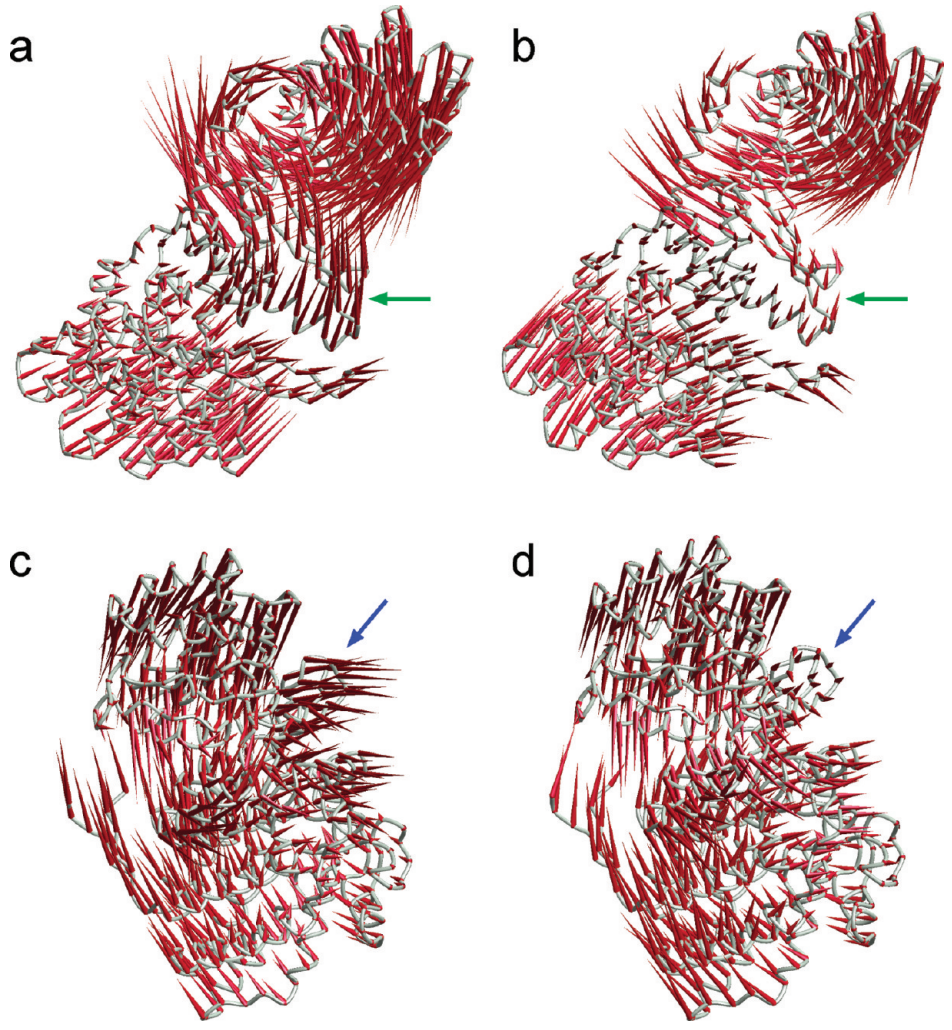

Figure 2. Visualization of the conformational changes between crystal structures 1AON and 1OEL (a,c) and between the crystal structure 1AON and a fit of the final state with 20 label-to-label distance constraints with an error of 3 Å (b,d). Gray coil models correspond to the initial state. Crimson cones point from C^{α} positions in the initial state to C^{α} positions in the final state. The view in a and b maximizes changes in the paper plane; the view in c and d is rotated by 90°. Arrows point to the strongest deviations of the fit.

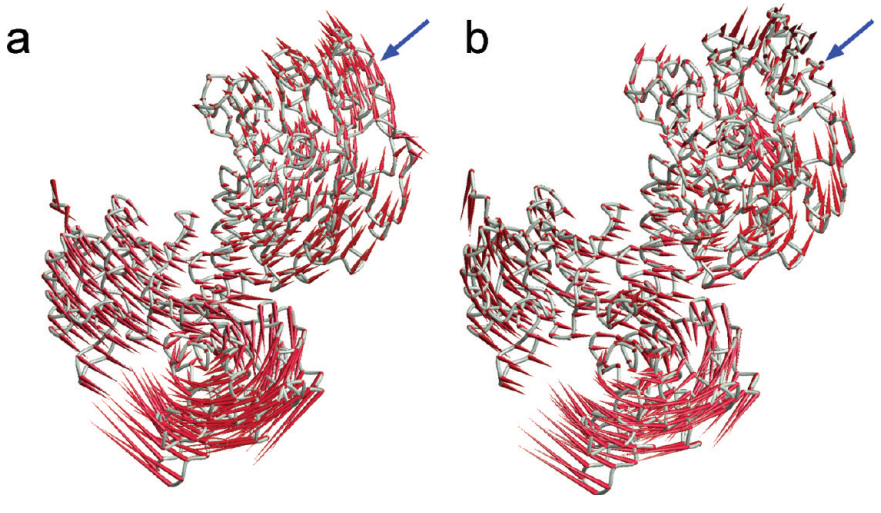

Figure 3. Visualization of the conformational changes between crystal structures 1LFH and 1LFG (a) and between the crystal structure 1LFH and a fit of the final state with 50 label-to-label distance constraints with an error of 3 Å (b).

$\langle f_e \rangle$ lies between 64 and 88% of the fractional coverage that is achieved with the same number of constraints and modes for exact distance constraints. The only exceptions are the transitions 1OMP/1ANF (44%), 2JLN/2X79 (56%), and

dx.doi.org/10.1021/ct300113z | J. Chem. Theory Comput. 2012, 8, 3854−3863

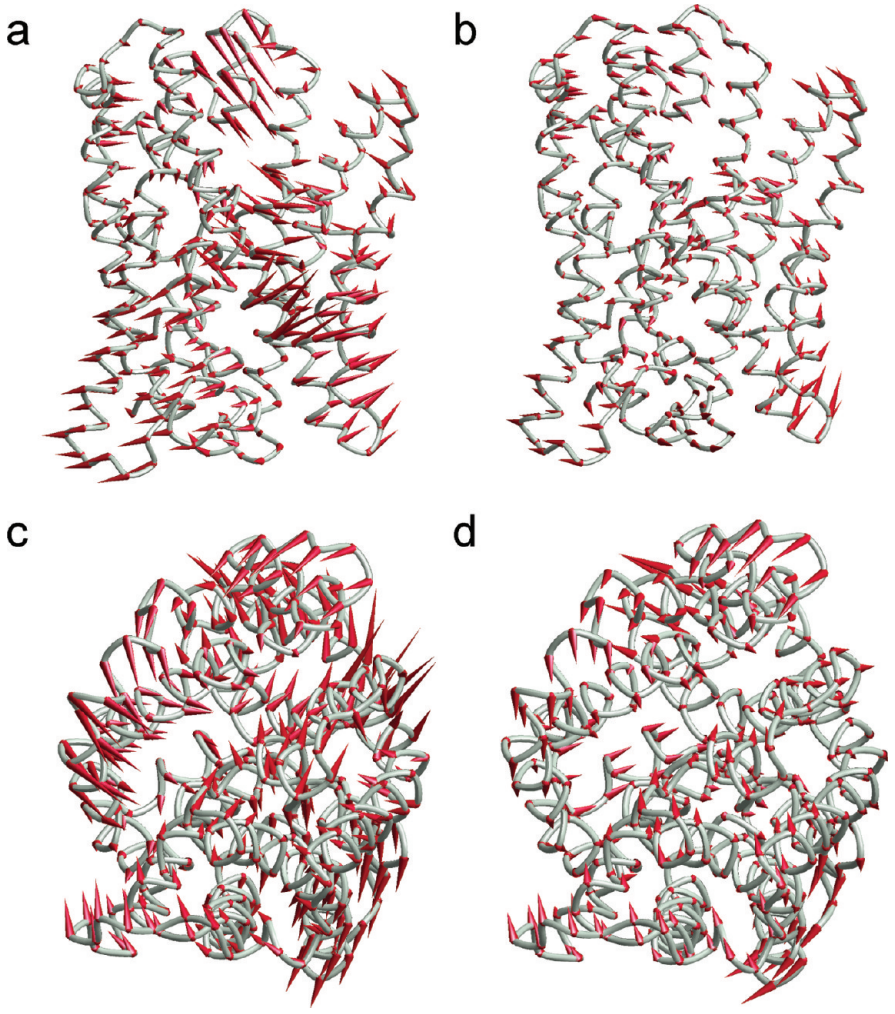

Figure 4. Visualization of the conformational changes between crystal structures 2JLN and 2X79 (a,c) and between the crystal structure 2JLN and a fit of the final state with 50 label-to-label distance constraints with an error of 3 Å (b,d). The view in a and b is approximately along the lipid bilayer plane; the view in c and d is approximately perpendicular to that plane.

1LFH/1LFG (58%). In the former two cases, the strong deterioration was expected, as the C^{α} RMSD between the initial and final structure only slightly exceeds the assumed RMSD of the label-to-label distance constraints.

In several cases, $\langle|\Delta\mathbf{R}_{fit}|\rangle$ is larger than 5 Å. The mean fractional coverage of the conformational changes only rarely exceeds 0.5. Are models of that quality of any use? To answer this question, we have visualized the conformational changes both from the two experimental structures and from the experimental initial structure and the fit. From each ensemble, we used the representative final structure, i.e., the one whose fractional coverage is closest to $\langle f_e\rangle$. Here, we discuss three cases corresponding to good, medium, and poor performance of the approach. The remaining visualizations can be found in the Supporting Information.

At the best mean coverage of $\langle f_e\rangle = 0.541$ encountered for the bacterial chaperonin GroEL (1AON/1OEL, see Figure 2), the main characteristics of relative domain motion are well represented in the fit. Apart from modest differences in amplitude and direction of the motion in some segments, the main deviations are seen for residues 371−397 (green arrows in Figure 2a,b) and residues 170−191 (blue arrows in Figure 2c,d), for which the model strongly underestimates amplitude of the motion.

As a case of medium performance, we consider the iron transport protein lactoferrin (1LFH/1LFG, see Figure 3) with $\langle f_e\rangle = 0.299$. Again, the main characteristics of the conformational change are well represented. The reduced coverage is mainly due to an underestimate of the amplitude of the motion for those parts of the protein that move most strongly with respect to each other. In addition, the fit does not capture the movement of a domain comprising residues 548−584 (blue arrow).

For the core helices of the nucleobase-cation symporter Mhp1 (2JLN/2X79, see Figure 4), the approach performs poorly with $\langle f_e\rangle = 0.143$. Although some features of the conformational change are captured, such model quality is hardly sufficient for reliable discussion of the conformational change. A similar case is encountered for the maltose binding protein (1OMP/1ANF, see Supporting Information). In all other cases, the visualizations allow for recognizing the main characteristics of the conformational change. In none of the cases is quality of the fit, as expressed by $\langle|\Delta\mathbf{R}_{fit}|\rangle$, sufficient to model side chain conformations with a confidence that would allow for discussion of the model on an atomistic level, for instance, in terms of ligand binding.

dx.doi.org/10.1021/ct300113z | J. Chem. Theory Comput. 2012, 8, 3854−3863

4. DISCUSSION

4.1. Limitations and Scope of the Approach. By restricting the conformational change to a moderately sized basis of slow normal modes, only a fraction of the coordinate change can be covered (see column $f_{20,d}$ in Table 1). For 20 modes, this fraction rarely exceeds 0.8 and may be smaller than 0.4. In general, we observe that mediocre fractional coverage by the basis (<0.6) is related to small collectivity of the conformational change and leads to problems in fitting with the Zheng/Brooks algorithm, even if exact C^α−C^α distance constraints are used (see column f_{fit} in Table 1). The approach failed in a number of cases where motion could not be classified as hinge or shear motion or as a superposition of several hinge or shear motions.

When going from C^α−C^α distance constraints to exact label-to-label constraints, fractional coverage for the same number of constraints and the same basis size further deteriorates, most probably due to the restricted choice of labeling sites and the error introduced by assuming rigid-body motion of the mean label position with a local residue frame. The situation can be improved by using an excess of constraints with respect to the basis size. In a few cases, sufficient coverage could only be achieved by increasing the number of constraints to 50, a number that is not prohibitive but would require substantial experimental effort. In many cases, useful results could be obtained with 20 constraints and a basis size of 10. For two of the three cases, where fractional coverage with 20 C^α−C^α distance constraints was below 0.25, even 50 label-to-label constraints were insufficient.

When assuming a realistic error of the label-to-label distance constraints (3 Å standard deviation), which is mainly due to errors in prediction of the mean label position with respect to the C^α atom, fractional coverage decreases to less than 0.6 even in the best cases. For structural changes with C^α RMSD of the final from the initial structure larger than 4 Å, the main characteristics of the conformational changes can still be inferred by visualization of the motion from the initial structure to the fit. In particular, with few exceptions, it is recognized which protein segments move relative to each other, and the general direction of this motion is reproduced. Mediocre fractional coverage is mainly due to wrong estimates for the amplitude of the motion, with underestimates being much more common than overestimates. Because of the large uncertainty in the amplitude of the motion, models obtained with this approach should be interpreted only qualitatively on the basis of visualizations and never in atomistic detail.

Failure of the approach is usually recognized by relatively minor differences between the initial structure and the fit, as seen in Figure 4c,d and in the Supporting Information for the structure pair 1OMP/1ANF. Nevertheless, reliability of the fits should be tested by repeating the procedure with only part of the available experimental constraints and by adding pseudorandom numbers to the constraints to simulate their uncertainty.[18]

4.2. Possible Improvements. The approach introduced here would greatly benefit from reduced uncertainty in predicting mean spin label positions with respect to the C^α coordinate. Empirical parametrization of label-protein interaction energies in the rotamer library approach[21] may soon become feasible, as more and more label-to-label distances in known structures are reported. The other main reason for limited precision of the approach is limited coverage of the coordinate change by a moderately sized basis of slow normal modes. As was recently shown in the example of flexible structure fitting to SAXS data,[43] basis size reduction can be avoided if the model is stabilized by further pseudoenergy terms. In this study, fits to theoretical, noiseless SAXS data for the pairs 1AKE/4AKE, 2LAO/1LST, 1ANF/1OMP, 1LFG/1LFH, and 1N0V/1N0U with a complete basis gave better results than were found in an earlier study with a reduced basis[42] and, in some of the cases, slightly better fits than were obtained with 20 exact C^α−C^α distance constraints in the present study. It remains to be seen whether such an approach is also feasible for a moderate number of label-to-label distance constraints and whether it extends to types of motion other than domain hinge motion.

Finally, the approach could be improved by better parametrization of the elastic network model. By definition of larger force constants for direct and next neighbors in the peptide chain we could avoid relaxation of the local structure along normal modes not belonging to the fit basis, as was necessary in the original Zheng/Brooks algorithm. This suggests that parametrization of elastic network models based on physical arguments[35] can improve the stability of the fit procedure.

5. CONCLUSION

The Zheng/Brooks algorithm for modeling protein conformational changes by iterative fitting of distance constraints can be adapted to SDSL EPR by treating the mean spin label coordinate as a fixed point in a local frame of the labeled residue. This approximation, together with the reduced choice of sites for selection of optimum distance constraints, causes some deterioration in fit quality, which can be partially compensated by increasing the ratio between the number of distance constraints and the size of the basis set of slow normal modes. The approach works best for domain hinge motion or superposition of several domain hinge motions and tends to fail for other types of motion if collectivity of the structural change is less than about 0.4. When a realistic uncertainty of the distance constraints of 3 Å and a realistic number of 20−50 constraints are assumed, the approach is applicable to conformational changes with an amplitude of at least 4 Å C^α RMSD. In these cases, directions of relative domain motion can be predicted quite well, whereas amplitudes are more uncertain. Combination of the Zheng/Brooks algorithm with SDSL EPR distance measurements in the nanometer range appears to be a promising method for characterizing the type and general direction of large-scale conformational changes in proteins. The approach described in this paper has been implemented as open-source software in version 2011.1 of MMM.[44]

■ ASSOCIATED CONTENT

Supporting Information

(1) Figures with visualization of further structural changes. (2) Ensembles of structural models obtained by fitting to label-to-label distance constraints, log files of the ensemble fits, and example constraint files in MMM format. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +41 44 632 5702. Fax: +41 44 633 1448. E-mail: gjeschke@ethz.ch.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The author thanks Christine Peter for the initial hint to elastic network models and Swiss National Fund for financial support (200020_132255).

■ REFERENCES

(1) Schramm, V. L. *Annu. Rev. Biochem.* **2011**, *80*, 703−732.
(2) Sablin, E. P.; Fletterick, R. J. *J. Biol. Chem.* **2004**, *279*, 15707−15710.
(3) Boudker, O.; Verdon, G. *Trends. Pharmacol Sci.* **2010**, *31*, 418−426.
(4) Forrest, L. R.; Kramer, R.; Ziegler, C. *Biochim. Biophys. Acta* **2010**, *1807*, 167−188.
(5) Gerstein, M.; Krebs, W. *Nucleic. Acids. Res.* **1998**, *26*, 4280−4290.
(6) Eisenmesser, E. Z.; Bosco, D. A.; Akke, M.; Kern, D. *Science* **2002**, *295*, 1520−1523.
(7) Hubbell, W. L.; Cafiso, D. S.; Altenbach, C. *Nat. Struct. Biol.* **2000**, *7*, 735−739.
(8) Rabenstein, M. D.; Shin, Y. K. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 8239−8243.
(9) Milov, A. D.; Salikhov, K. M.; Shirov, M. D. *Fiz. Tverd. Tela* **1981**, *23*, 975−982.
(10) Pannier, M.; Veit, S.; Godt, A.; Jeschke, G.; Spiess, H. W. *J. Magn. Reson.* **2000**, *142*, 331−340.
(11) Borbat, P. P.; Freed, J. H. *Chem. Phys. Lett.* **1999**, *313*, 145−154.
(12) Schiemann, O.; Prisner, T. F. *Q. Rev. Biophys.* **2007**, *40*, 1−53.
(13) Jeschke, G. *Annu. Rev. Phys. Chem.* **2012**, *63*, 419−446.
(14) Sale, K.; Faulon, J. L.; Gray, G. A.; Schoeniger, J. S.; Young, M. M. *Protein Sci.* **2004**, *13*, 2613−2627.
(15) Alexander, N.; Bortolus, M.; Al-Mestarihi, A.; Mchaourab, H.; Meiler, J. *Structure* **2008**, *16*, 181−195.
(16) Hirst, S. J.; Alexander, N.; Mchaourab, H. S.; Meiler, J. *J. Struct. Biol.* **2010**, *173*, 506−514.
(17) Kazmier, K.; Alexander, N. S.; Meiler, J.; Mchaourab, H. S. *J. Struct. Biol.* **2010**, *173*, 549−557.
(18) Hilger, D.; Polyhach, Y.; Jung, H.; Jeschke, G. *Biophys. J.* **2009**, *96*, 217−225.
(19) Borbat, P. P.; Mchaourab, H. S.; Freed, J. H. *J. Am. Chem. Soc.* **2002**, *124*, 5304−5314.
(20) Sale, K.; Song, L.; Liu, Y. S.; Perozo, E.; Fajer, P. *J. Am. Chem. Soc.* **2005**, *127*, 9334−9335.
(21) Polyhach, Y.; Bordignon, E.; Jeschke, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2356−2366.
(22) Hilger, D.; Polyhach, Y.; Padan, E.; Jung, H.; Jeschke, G. *Biophys. J.* **2007**, *93*, 3675−3683.
(23) Appel, M.; Hizlan, D.; Vinothkumar, K. R.; Ziegler, C.; Kuhlbrandt, W. *J. Mol. Biol.* **2009**, *386*, 351−365.
(24) Hagelueken, G.; Ward, R.; Naismith, J. H.; Schiemann, O. *Appl. Magn. Reson.* **2012**, *42*, 377−391.
(25) Polyhach, Y.; Jeschke, G. *Spectroscopy (Amesterdam, Neth.)* **2010**, *24*, 651−659.
(26) Lillington, J. E.; Lovett, J. E.; Johnson, S.; Roversi, P.; Timmel, C. R.; Lea, S. M. *J. Mol. Biol.* **2010**, *405*, 427−435.
(27) Zheng, W.; Brooks, B. R. *Biophys. J.* **2006**, *90*, 4327−4336.
(28) Bahar, I.; Lezon, T. R.; Bakan, A.; Shrivastava, I. H. *Chem. Rev.* **2010**, *110*, 1463−1497.
(29) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1−6.
(30) Yang, L.; Song, G.; Jernigan, R. L. *Biophys. J.* **2007**, *93*, 920−929.
(31) McLachlan, A. D. *J. Mol. Biol.* **1979**, *128*, 49−79.
(32) Brüschweiler, R. *J. Chem. Phys.* **1995**, *102*, 3396−3403.
(33) Yang, L.; Song, G.; Jernigan, R. L. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12347−12352.
(34) Hinsen, K. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, E128.
(35) Hinsen, K.; Petrescu, A. J.; Dellerue, S.; Bellissent-Funel, M. C.; Kneller, G. R. *Chem. Phys.* **2000**, *261*, 25−37.
(36) Orellana, L.; Rueda, M.; Ferrer-Costa, C.; Lopez-Blanco, J. R.; Chacón, P.; Orozco, M. *J. Chem. Theory Comput.* **2010**, *6*, 2910−2923.
(37) Hirst, S. J.; Alexander, N.; Mchaourab, H. S.; Meiler, J. *J. Struct. Biol.* **2011**, *173*, 506−514.
(38) Holm, L.; Sander, C. *J. Mol. Biol.* **1991**, *218*, 183−194.
(39) MaxSprout Tool for Reconsruction of 3D coordinates from C(alpha) trace. http://www.ebi.ac.uk/Tools/maxsprout/ (accessed April 2, 2012).
(40) Gorba, C.; Miyashita, O.; Tama, F. *Biophys. J.* **2008**, *94*, 1589−1599.
(41) Miyashita, O.; Gorba, C.; Tama, F. *J. Struct. Biol.* **2011**, *173*, 451−460.
(42) Gorba, C.; Tama, F. *Bioinform. Biol. Insights* **2010**, *4*, 43−54.
(43) Zheng, W.; Tekpinar, M. *Biophys. J.* **2011**, *101*, 2981−2991.
(44) ETH−EPR Software. http://www.epr.ethz.ch/software/index (accessed April 2, 2012).