Use of Residual Dipolar Couplings as Restraints in Ab Initio Protein Structure Prediction

Turkan Haliloglu1
Andrzej Kolinski2,3
Jeffrey Skolnick2
1Polymer Research Center and Chemical Engineering Department, Bogazici University, Bebek 80815, Istanbul, Turkey
2Buffalo Center of Excellence in Bioinformatics, 901 Washington St., Ste. 300, Buffalo, NY 14203
3Faculty of Chemistry, Warsaw University, Pasteura 1, 02-093 Warsaw, Poland

Received 23 May 2003; accepted 24 July 2003

Abstract: NMR residual dipolar couplings (RDCs), in the form of the projection angles between the respective internuclear bond vectors, are used as structural restraints in the ab initio structure prediction of a test set of six proteins. The restraints are applied using a recently developed SICHO (SIde-CHain-Only) lattice protein model that employs a replica exchange Monte Carlo (MC) algorithm to search conformational space. Using a small number of RDC restraints, the quality of the predicted structures is improved as reflected by lower RMSD/dRMSD (root mean square deviation/distance root mean square deviation) values from the corresponding native structures and by the higher correlation of the most cooperative mode of motion of each predicted structure with that of the native structure. The latter, in particular, has possible implications for the structure-based functional analysis of predicted structures. © 2003 Wiley Periodicals, Inc. Biopolymers 70: 548–562, 2003

Keywords: ab initio structure prediction; dynamic modes; residual dipolar coupling; SICHO model

Correspondence to: Jeffrey Skolnick; email: skolnick@buffalo.edu
Contract grant sponsor: NIH; Contract grant number: GM-37408
Contract grant sponsor: BU Research; Contract grant number: 00HAS02D-00HAS03 (T.H.)
Contract grant sponsor: DPT Project; Contract grant number: 01K120280 (T.H.)
Contract grant sponsor: EA-TUBA-GEHIP.2001-1-1 (T.H.)
© 2003 Wiley Periodicals, Inc.

548
INTRODUCTION

Methods that would allow for the more rapid determination of protein structure are of great importance, both from the viewpoint of traditional structural biology as well as for structural genomics projects. NMR residual dipolar couplings, RDCs, are of particular interest for this purpose, as they provide considerable structural information through their dependence on the orientation of an intermolecular vector relative to an order frame\(^1\) and offer the advantage that this data can be collected in a relatively short time period. The introduction of the RDC methodology has increased the scope of the problems that could be addressed by NMR spectroscopy. The identification of conformational changes, the relative orientations of domains and intermolecular complexes, studies of rapid recognition of homologous protein folds, and submillisecond timescale dynamics have all shown considerable progress due to the development of this methodology.\(^2\) It is clear that RDCs can improve the quality of the determined\(^3\) structures by providing long-range orientational restraints; that is, a set of global structural restraints that nicely complement those typically obtained from other data such as nuclear Overhauser enhancements (NOEs), scalar couplings, and chemical shifts.\(^1,4,5\) RDCs are also sensitive to internal motions; yet most of the existing structural refinement protocols for analyzing dipolar data implicitly assume that internal motions are either absent or are uniform and axially symmetric in nature.\(^6–8\) In a recent work,\(^9\) it was demonstrated that an \textit{ab initio} structure method combined with an extensive set of RDC restraints can provide a general method for the structure prediction of a variety of protein folds (proteins up to 125 residues). The effect of the completeness of the data set on the algorithm performance was also investigated in the latter study, and small decreases in accuracy and precision were observed for the structures studied when 30% of the data were randomly removed. However, it is not obvious to what extent this method can be generalized to larger proteins with a much less complete data set.

Experimental RDCs can be used to define the orientation of a vector with respect to an alignment tensor, which is affected either by paramagnetic properties of the molecule or solvation in liquid crystal media.\(^4,5,10–12\) In structure calculations, RDCs can be used for the optimization of the orientation of bond vectors with respect to the orientation of the external alignment or susceptibility tensor.\(^6,7,13–17\) Although the size of the alignment tensor can be derived from the distribution of the experimental RDCs, its orientation with respect to the coordinate system of a molecule is unknown at the beginning of the structure prediction; this could cause convergence problems in the folding process due to the multiple-minima problem.\(^18\) the solution of which requires complex sampling protocols. Thus, a method\(^19,20\) that is independent of the orientation of the alignment tensor with respect to the molecule is employed in the present study to transform the RDCs into the projection angles between the internuclear bond vectors for which the RDCs are measured. This approach could be effectively used at the beginning of the folding procedure or at any stage of the folding process. Another method of rapid generation of protein structures from dipolar coupling data was presented recently.\(^21\) This method employs dipolar coupling constraints in the form of a simple elliptical equation. Here, in this article, a test set of six proteins is used to investigate the effect of RDC restraints on the quality of the predicted structures generated by an \textit{ab initio} structure prediction algorithm based on the recent low-resolution protein model—namely, the SIde-CHain-Only (SICHO) model, where conformational space is sampled by a multiple copy simulated annealing MC algorithm.\(^22,23\) A single set having a small number of RDCs is employed. As the goal is to examine the contribution of RDC restraints alone, NOE restraints are not incorporated, although this could be readily done.

MODEL AND METHOD

Theory

The residual dipolar coupling, \(D_{ij}\), measured between coupled nuclei \(i\) and \(j\), provides geometric information relative to the common alignment frame of the form\(^7\)

\[
D_{ij}(\alpha, \beta) = D_{ij}(3\cos^2\beta - 1) + 3/2D_{ij}\cos^2\beta \sin^2\beta
\]  \(1\)

Here, \(D_i\) and \(D_j\) are, respectively, the axial and perpendicular components of the alignment tensor and can be determined from the experimental RDCs (powder pattern of couplings),\(^6\) \(\{\alpha, \beta\}\) is the vector orientation relative to this tensor.

Equations that allow us to use RDCs as restraints without the need to define the orientation of the alignment tensor can be derived from Eq. (1). This is done by calculating the projection angles \(\phi_p\) between all pairs of internuclear vectors \(i\) and \(j\) for which the dipolar coupling data are measured. There is a continuum of \(\alpha\beta\) pairs for each set of RDC data. These lead to the assignment of a set of \(\phi_p\) (considering mirror reflections symmetry along each axis as well) for each pair of vectors \(i\) and \(j\). In the simulations, the possible range for the angle \(\phi_p\) is no longer allowed to be in the whole interval from 0 to \(\pi\). Instead, two general possi-
Potentials and Simulations

A harmonic type of potential is implemented to impose the allowed ranges for the projection angles between a set of internuclear vectors to be satisfied during the simulation, in addition to the other terms of the potential of generic and residue specific short-/long-range interactions used in the lattice simulation algorithm based on the SICHO protein model.

SICHO\textsuperscript{22,23} is a lattice protein model that only uses one explicit interaction center per residue located at the side chain center of mass. These interaction centers are restricted to an underlying simple cubic lattice. The positions of the alpha carbons are estimated from the side chain coordinates used to map the approximate orientations of vectors associated with RDCs.

The RDC could depend on different bond vectors found in the structure (NH, C\textsuperscript{\alpha}N, C\textsuperscript{\alpha}H\textsuperscript{\alpha}, C\textsuperscript{\alpha}C). Here, the incorporated RDCs are for amide NH bond vectors. In a low-resolution model, the explicit description of an NH vector is not possible without a reverse mapping of the low-resolution structure back to its full atomistic representation; thus, it is not efficiently performed during the simulation. However, it is possible to have an approximate description for each NH vector by associating it with a perpendicular vector to the backbone. This vector is the normal vector to the plane described by the two successive virtual bonds adjacent to the \( \alpha \)-carbon center of the residue associated with the respective NH vector (i.e., by the cross-product of the successive two virtual bonds, each of which connects the two successive \( \alpha \)-C atoms). To see the error introduced by this description of an NH vector, the projection angles between the respective NH bond pairs are calculated for the native structure in atomic resolution from the Brookhaven Protein Data Bank (PDB) and compared with the corresponding angles calculated from the approximate description of the vectors in the low-resolution representation of the native structure. This error remains within 20–30\(^\circ\) for the structures of our test set, which is also within the angular resolution of the SICHO lattice protein model. Thus, the allowed range for the projection angles between the respective vectors calculated using Eq. (1) is also within the resolution of the model.

In the present work, a test set of six proteins (one blind) was used to test the effects of the RDC restraints on the quality of the predicted structures by \textit{ab initio} simulations. These structures are: acyl carrier protein (ACP, PDB code: 1ACP), Rubredoxin (PDB code: 1brf), RNA Binding Domain (NS1, PDB code: 1ns1), Nodulation Protein F (Nodf, PDB code: 1nodf), Carbohydrate Recognition Domain (Crd, PDB code: 1a3k), and Brc domain. The RDCs of each structure were obtained from private communications.\textsuperscript{24} The simulations were carried out for both cases, with and without restraints, and the results for each are presented and discussed under the corresponding subsections in Results and Discussion.

The quality of the predicted structures is evaluated from the viewpoint of two measures of fold quality. One analysis of the improved quality of the predicted structures relative to the absence of RDC information is the decrease in coordinate, RMSD, and distance, dRMSD, values of the predicted structure from the native structure. The implementation of the restraints during the folding process could drive the folding of the structure to the right topology by restricting the conformational space accessible to the structure (as in the case of ACP, see below). The orientational order imposed by the RDCs might also lead to improvements in the slowest dynamic mode shape of the structure that represents the most cooperative mode of motion and characterizes the global dynamic behavior of the structure. However, as shown below, the RMSD, which is a rather global type of structural measure, does not necessarily significantly reflect this improvement (as in the case of Rubredoxin). To address this issue, the Gaussian Network Model (GNM)\textsuperscript{25} is used for the latter analysis.

The GNM uses the topology of residue-residue contacts to model the proteins as an elastic network with uniform single parameter harmonic potentials between the \( \alpha \) carbons of contacting residue pairs. Adopting a harmonic interaction potential means assuming that the residues are undergoing the Gaussianly distributed fluctuations about these mean-positions. Using the GNM, the dynamics of a biomolecular system can be decomposed into a collection of internal motions of different frequencies with a procedure similar to normal mode analysis. The slowest modes with the lowest frequencies refer to the most cooperative motions involving the entire structure.

These dominant modes of motion give information about the molecular dynamics relevant to biological function that occurs on the global scale.\textsuperscript{26–29} Thus, the predicted structure ideally should be able to display the correct dynamic modes of motions that are inherent in its native packing density. The prediction of nativelike fluctuations could be particularly important if one wants to design a structure-based method to resolve some aspects of functional properties. In previous studies,\textsuperscript{26–29} it was suggested that the minima in the global mode shapes generally coincide with those residues acting as hinges, and the same regions are also usually observed to be correlated with (or juxtaposed to) biologically active sites, such as catalytic sites in enzymes. Maxima, on the other hand, correspond to segments distinguished by their enhanced mobilities and are often implicated in substrate recognition.

In this article, the slowest mode shape for each predicted structure with and without RDC restraints is calculated and compared to that of the native structure. A calculated linear correlation coefficient \( r \) of the slowest mode shape of the predicted best structure with that of the native (considering shifting of up to five residues for optimum matching between the two curves) is taken into account. We will also
examine the corresponding RMSD/dRMSD from the native to assess the improvement obtained by the incorporation of a small number of RDCs in the absence of any other experimental data (see Table I). The best structure is taken as the centroid of the cluster that has the lowest RMSD relative to native (except for Brct, which is a blind prediction). Furthermore, the functional implication of the fluctuations in the slowest mode is examined.

### Table I

<table>
<thead>
<tr>
<th>Protein</th>
<th>Without RDC restraints</th>
<th>With RDC restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Acyl carrier protein, ACP (1ACP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run</td>
<td>RMSD (Å)</td>
<td>(n*/n)</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>10.35</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>10.30</td>
<td>2/2</td>
</tr>
<tr>
<td>3</td>
<td>10.29</td>
<td>2/2</td>
</tr>
<tr>
<td>4</td>
<td>10.29</td>
<td>1/2</td>
</tr>
<tr>
<td>Average</td>
<td>10.31</td>
<td></td>
</tr>
<tr>
<td>b) Rubredoxin (1brf)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run</td>
<td>RMSD (Å)</td>
<td>(n*/n)</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>4.88</td>
<td>1/4</td>
</tr>
<tr>
<td>2</td>
<td>5.19</td>
<td>1/3</td>
</tr>
<tr>
<td>3</td>
<td>5.56</td>
<td>1/6</td>
</tr>
<tr>
<td>4</td>
<td>4.94</td>
<td>2/5</td>
</tr>
<tr>
<td>6</td>
<td>5.79</td>
<td>1/4</td>
</tr>
<tr>
<td>c) RNA binding domain, NS1 (1ns1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run</td>
<td>RMSD (Å)</td>
<td>(n*/n)</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>8.38</td>
<td>1/3</td>
</tr>
<tr>
<td>d) Nodulation Protein F, Nodf (1nodf)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Run</td>
<td>RMSD (Å)</td>
<td>(n*/n)</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>6.53</td>
<td>2/2</td>
</tr>
<tr>
<td>2</td>
<td>6.43</td>
<td>1/2</td>
</tr>
<tr>
<td>Average</td>
<td>6.48</td>
<td></td>
</tr>
</tbody>
</table>

*RMSD/dRMSD (Å) of the centroid of the best cluster for five predicted structures. n* and n show the rank of the best cluster and the total number of clusters, respectively. r is the correlation coefficient between the slowest mode shape of the predicted and native structures (a window of five residues is used for optimum matching of the two curves). The results with RDCs are marked.

$dRMSD = \frac{1}{N_{pair}} \sum_{ij} (r_{ij}^a - r_{ij}^b)^2/2$; $a$ and $b$ refer the two structures and $N_{pair}$ is the number of residue pairs, $ij$. 

The functional implication of the fluctuations in the slowest mode is examined.
RESULTS AND DISCUSSION

ACP

ACP, having 77 residues, is an essential cofactor in the biosynthesis of fatty acids in many reactions that require acyl transfer steps. It has mainly three α-helices (α1: Ile3-Leu15, α2: Leu37-Asp51, α3: Val65-His75) excluding small α-helical segment: Glu57-Phe62). The number of experimental RDCs used here is 24 (transformed to 24 × 23/2 orientational restraints). These restraints are associated with three α-helices of the structure and thus provide information about the relative orientation of each with respect to the other.

The projection angles $\phi_{ij}$ for a set of NH bond pairs $i$ and $j$ of ACP calculated from a set of measured RDCs are depicted in Figure 1(a). The RMSD/dRMSD from native values for all runs with/without RDC restraints and the corresponding correlation coefficients $r$ of the slowest mode shapes with that of the native are summarized in Table I(a).

For the best case, the RMSD/dRMSD improves from 10.29/5.40 Å to 6.59/4.16 Å with use of the RDC restraints. Figure 1(b) depicts the ribbon diagram of the best predicted structure with the RDC restraints superimposed with the native structure (PDB code: 1ACP). For comparison, the ribbon diagram of the predicted structure without the RDC restraints (with RMSD of 10.29 Å) is displayed in Figure 1(c). It could be noted that the alignment of the secondary structural units is improved by the incorporation of orientational restraints into the simulations, which, in turn, leads to a significant improvement in the prediction as reflected by its RMSD/dRMSD from native. In the case with RDC, the results suggest that the incorporation of distance restraints associated with the C-terminus might easily lead to the prediction of a better structure with even lower RMSD.

The mean-square fluctuations of the α-carbons in the first slowest mode for the best predicted structures with and without the RDC restraints compared to the
corresponding dynamic mode of the native structure are depicted in Figure 2(a) and (b), respectively. The correlation coefficient $r$ of the slowest mode shape increases, on the average, from 0.14 to 0.58, with RDCs (Table Ia).

The slowest mode shape of the native structure depicts that the region of $\alpha_3$ in the middle of the structure appears to be able to display high amplitude fluctuations (Figure 2a). It is interesting to note that the phoshopantetheine prosthetic group is attached to Ser36 (the site is a conserved Asp-Ser-Leu motif), which resides in the N-terminus of the latter helix.

X-ray crystallographic studies on Butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. The analysis of the molecular surface of ACP pictures a plastic hydrophobic cavity in the vicinity of Ser36, which is expanded in one crystal form and contracted in another crystal form, implying that the protein has adopted this conformation after delivery of substrate into the active site of a partner enzyme. The latter region to which Ser36 is attached can be recognized by enhanced fluctuations in the slowest mode of the predicted structure with the RDCs as well; yet, the mode shape seems to shift.
from that of the native by two-five residues. If we exclude the misfolded C-terminus, the maximum in the predicted slowest mode shape with the RDCs coincides with the segment comprising Ser36.

The dynamic model for the structure of ACP in solution based on two-dimensional NMR data suggests\(^3\) that the helices \(\alpha_1\) and \(\alpha_3\) move in a concerted fashion, although both remain in similar conformations, whereas helix disruption occurs in \(\alpha_2\) in some conformers. The fast exchange of amide protons of residues 41 and 42 and some other residues closer to the middle in \(\alpha_2\) compared to the other amide proton in other helices indicates instability of the middle helix. The enhanced mobility of \(\alpha_2\) is also reflected by fluctuations of the respective region in the slowest mode of both native structure and predicted structure with the RDCs (excluding the misfolded C-terminus). The helix disruption in some conformers might couple with a larger conformational change of the region of Ser36 associated with the delivery of the substrate, as discussed above. In the same dynamic model, Phe28 and Phe50 were found to violate a different set of restraints in each conformer; this may promote the existence of slower (more cooperative) internal motions of the structure. These two residues correspond to the two hinge-points, as preceding and succeeding the loop of Ser36, in the native slowest mode. The two hinge points—one at around Asn24 (yet, shifted by four residues from Phe28) and the other at Phe50—appear in the predicted slowest mode with the RDCs as well.

On the other hand, ACP has a hydrophobic cleft between \(\alpha_2\) and \(\alpha_3\) in which an acyl chain can lie.\(^3\) The secondary structure analysis of the latter region by two-dimensional \(^1\)H-NMR spectroscopy suggests a hinge region at residues Thr63 and Thr64. Furthermore, a number of hydrophobic residues line the contact area between these helices (Phe50, Ile54, Ala59, Val65, Ala68, and Tyr71), which could provide a site for acyl chain stabilization.\(^6\) These appear with relatively restrained fluctuations in the slowest modes of both native and predicted structures with RDCs (the acyl chain is not considered in the GNM calculations; thus, its presence with ACP should lead to more restrained fluctuations).

**Rubredoxin**

Rubredoxin, containing 53 residues, is an electron transfer protein whose structure has three \(\alpha\)-helices (\(\alpha_1\): Pro19-Asn21, \(\alpha_2\): Phe29-Glu31, \(\alpha_3\): Lys45-Glu47) and three small \(\beta\)-strands (\(\beta_1\): Lys2-Cys5, \(\beta_2\): Ile11-Asp13, \(\beta_3\): Phe48-Lys50). The number of RDCs employed is 15–24 [(15 \(\times\) 14/2) – (24 \(\times\) 23)/2 orientational restraints]. The restraints imposed are associated not only with these helical regions, but with other regions of the structure as well.

The results from the present simulations are summarized in Table I(b). Figure 3(a) and (b) depict the ribbon diagrams of the predicted structures of a representative case of Rubredoxin with RDC (RMSD from native of 4.34 Å; the number of RDCs is 22) and without RDC (RMSD from native of 4.88 Å) restraints, respectively. The mean-square fluctuations of the \(\alpha\)-carbon atoms along the slowest mode of the predicted structures in comparison with that of the native state structure (PDB code: 1brf) are presented in Figure 3(c) (with RDC restraints) and 3(d) (without RDC restraints), respectively. Although the RMSD values of the two structures do not differ significantly, the use of RDC restraints leads to the prediction of a structure that can display more nativelike fluctuations (the average correlation coefficient \(r\) increases from 0.46 without RDCs to 0.8 with RDCs).

As can be seen from Figure 3(c) and (d), the shape of the native slowest mode shows that there are basically three loops—the highly flexible middle loop Glu14-Leu32, which is separated from the other two by the two hinges, or minimum fluctuating regions around Tyr12 and Asp34. The active site of the Rubredoxin contains an iron that is coordinated by the sulfurs of four conserved cysteine residues (residues Cys5, Cys8, Cys38, and Cys41), which reside on the latter two associated loops that are the most conserved regions of the protein.\(^3\) Resolution crystal structures (1.5 Å) and molecular dynamics simulations of oxidized and reduced Rubredoxin from Clostridium pasteurianum suggest\(^4\) that a gating mechanism caused by the conformational change of Leu41, a nonpolar side chain, allows transient penetration of water molecules, which increases the polarity of water molecules and also provides a source of protons. Prior to this, expansion of the Fe-S cluster and concomitant contradiction of the NH...S hydrogen bonds lead to greater electrostatic stabilization of the negative charge in this region; this involves the breathing motion of Val8 and Val44. These structural rearrangements upon reduction suggest specific mechanisms by which electron transfer reactions of Rubredoxin should be facilitated. This may explain the mobility of the two associated loops in the slowest mode shape, which coordinate the Fe: all the latter residues appear around the maximum of the two loops in the respective regions in the native slowest mode, whereas all others, excluding Gly9, appear in the same positions in the predicted slowest mode shape with RDCs. For the highly flexible multiple turn region comprising residues Glu14-Leu32, reflected by the largest mean-
FIGURE 3 (a,b) Ribbon diagrams of predicted structures of Rubredoxin with RDC (RMSD from native of 4.34 Å) and without RDC (RMSD from native of 4.88 Å) restraints, respectively. (c,d) Mean-square fluctuations of C^\text{\textalpha} atoms in the slowest mode in comparison with that of the native state, respectively [(i) and (ii) refer to, respectively, the native and predicted structures].
square fluctuations between the latter two loops along the slowest mode shape of both native and predicted structures with RDCs, the amide exchange experiments\textsuperscript{35} promote a model of solvent exposure having a subglobal cooperative conformational opening for this region. On the other hand, the significantly lower exchange rates preceding and following the latter region suggest that this segment is constrained at either hand by the less flexible binding region. Indeed, this corresponds to two hinge points in the respective regions reflected by two minima in the slowest mode shapes of both the native and the predicted structures with RDCs; yet, the predicted mode shape appears slightly shifted (by two-five residues). The distribution of fluctuations in the predicted structure without RDCs does not imply any of these observations, as seen in the respective figure. The RDC restraints correct the slowest mode shape of the structure in comparison to that without the RDCs by shifting the minimum fluctuating regions closer to those points observed in the native state (Figure 3c,d). The latter results imply that the restrained/enhanced mobility of these loops has an implication for its function, and the RDCs promote the recovery of the nativelike distribution of these fluctuations.

This analysis is more challenging here, as there are differences between the dynamic modes of even relatively low RMSD structures. Ideally, one expects that as the RMSD values become lower, the mode shapes should be very similar between the structures with similar RMSD values. Nevertheless, these results imply that a structure with an RMSD value of about 4–5 Å does not necessarily have the correct mode shapes. This, on the other hand, emphasizes the significance of another measure of structural quality in structure prediction calculations.

RNA Binding Domain (NS1)

NS1 is a 73 residue RNA-binding/dimerization domain. The structure has three α-helices (α\textsubscript{1}: Thr5-Asp24, α\textsubscript{2}: Pro31-Thr49, α\textsubscript{3}: Ile54-Lys70). The RDC data are mainly associated with the NH bonds of the three α-helices in the structure. The total number of RDCs employed here is 24 (24 × 23/2 orientational restraints).

The ribbon diagrams of the native (PDB code: 1ns1) and the predicted (with RDCs, they have a RMSD/dRMSD of 7.85/5.73 Å; while without RDCs, RMSD/dRMSD is 8.4/6.18 Å) structures are depicted in Figure 4. Here, while the C-terminus of the structure is not correctly folded with or without the RDCs, the RDC restraints appear to contribute to the adjustment of the relative orientations of the α-helices.

Figure 5 displays the first slowest mode shape of the predicted structures with the RDC (a) and without the RDC (b) restraints. The correlation of the mode’s shape with native is higher for the case with the RDCs (\( r = 0.91 \) and \(-0.06, \) respectively). This observation can be expected, as the structure with the RDCs has a lower RMSD in comparison to the one without the RDCs; however, we have already shown with Rubredoxin that this does not necessarily happen.

As shown in Figure 5, the distribution of the native fluctuations along the slowest mode displays striking extensive mobility for the segment from residues Val22 to Asn53, which comprises the middle α-helix, α\textsubscript{2}. It was suggested\textsuperscript{36} from the distributions of basic residues and conserved salt bridges of dimeric NS1 that the face containing antiparallel helices 2 and 2’ forms a novel arginine-rich nucleic acid binding motif. Arg38 is absolutely required for binding, and Lys41 makes a strong contribution to the affinity of binding; those residues in each of the latter antiparallel helices contact the phosphate backbone of the RNA target.\textsuperscript{37} Arg38 and Lys41 appear at the maximum of the native mode shape in the respective region. The predicted mode shape with the RDCs reflects this extensively moving segment from residues Gly10 to Leu50, with the maximum being shifted by about six residues.

The results depict improvement in the predicted fluctuations with RDCs, yet a shift in the slowest mode shapes appears for a predicted structure with relatively high RMSD. However, the biologically active unit is a dimer, and we have simulated only the monomer; this may partially rationalize the relatively poor results that we have obtained.

Nodf

Nodf has 35 residues (the residue indexes range from 1–86 due to missing residues). It has three α-helices (α\textsubscript{1}: Leu5-Val17, α\textsubscript{2}: Asp46-Leu58, α\textsubscript{3}: Val76-Gly86). The number of RDCs employed is 15 (15 × 14/2 orientational restraints). The results from the simulations are presented in Table I(d).

The ribbon diagrams of the best predicted structures with the RDCs (RMSD/dRMSD is 5.33/4.62 Å) and without the RDCs (RMSD/dRMSD is 6.43/5.21 Å) superimposed on the native structure (PDB code: 1fh1) are displayed in Figure 6(a) and (b), respectively. Figure 6(c) displays the mean-square fluctuations of α–carbon atoms for the native state and for the predicted structures with/without RDC restraints, respectively. The calculation with the RDC restraints has both lower RMSD/dRMSD values and a slowest mode shape that is more highly correlated with the
native state; the average $r$ is 0.82 and 0.22 with the RDCs and without the RDCs, respectively. The C-terminus appears to be misfolded without the RDCs.

Nodf has a high level of homology with ACP, especially around the prosthetic group attachment site. The phosphopantetheine prosthetic group is attached to Ser45 (the site is a conserved Asp-Ser-Leu motif), which is Ser36 in ACP. Residue Leu46 begins the corresponding residue index is 14 in Figure 6(a) and (b) (the indexes are readjusted because of the missing residues between the $\alpha$-helices). The loops are missing in our prediction and in the PDB structure as well. Thus, the functional site should be in the loop preceding $\alpha_2$. Functional analysis of an interspecies chimera of ACPs indicates a specialized domain for protein recognition, and Nodf is a specialized ACP whose specific features are encoded in the C-terminal region of the protein. Both the predicted structure with RDCs and the native state structure have the minimum fluctuating region in the C-terminal region $\alpha_2$, and have the active site on the preceding arm of that hinge point; yet, the native state promotes higher fluctuation for the active site region. The missing loops may partially explain the latter. Nevertheless, the fluctuations of the C-terminal region of the structure from 43–93 (starting from 14 to 37 in our index), that is, the functional domain of Nodf, are predicted with the RDCs closer to native compared to those of the predicted structures without the RDCs. It was suggested that, as in the case of ACP, Nodf has a hydrophobic cleft where the acyl chain can lie between $\alpha_2$ and $\alpha_3$. There is a hinge region between the two latter $\alpha$-helices where the slowest modes of both the native and the predicted structure with the RDCs promote this behavior (the region around Glu24-Asn27) with highly restrained fluctua-
This behavior for acyl chain stabilization is also observed in the other ACP structures from several sources. The structure Crd contains 137 residues (PDB code: 1a3k) and is not folded below 10 Å either with or without RDC restraints. It is the only structure with all beta sheets in our test set. The number of RDCs employed is 24. Incorporation of NOE restraints might help for folding. Nevertheless, here we aim to see exclusively the contribution of the RDC restraints to the quality of the predicted structures; thus, NOE restraints are not employed in any of the cases.

**Crd**

The structure Crd contains 137 residues (PDB code: 1a3k) and is not folded below 10 Å either with or without RDC restraints. It is the only structure with all beta sheets in our test set. The number of RDCs employed is 24. Incorporation of NOE restraints might help for folding. Nevertheless, here we aim to see exclusively the contribution of the RDC restraints to the quality of the predicted structures; thus, NOE restraints are not employed in any of the cases.

**Brct Domain**

The Brct domain contains 92 residues, and these simulations constitute a blind prediction. The number of RDCs employed here is 25. A close structure in the PDB to the one predicted here is the Brct domain from...
DNA-repair protein Xrccl from Homo Sapiens (PDB code: 1cdz; residue number of 96) with 15.8% sequence identity. Figure 7(a) and (b) depicts the ribbon diagrams of the predicted structures, obtained from the centroids of the first cluster (ranked based on the energy), with and without RDC restraints, superimposed on 1cdz (RMSD of 5.14 Å with RDCs and RMSD of 7.21 Å without RDCs, respectively). If we look at the ribbon diagrams, the relative position of N-terminus of the structure is different for the two cases—with and without RDCs. The simulations with varying seed numbers were repeated for the case with RDCs, and the first cluster centroid of each consistently gave the structures with an RMSD of ≈5.14–5.65 Å (with the total number of clusters varying from one to four) from 1cdz. On the other hand, if we look
at the second cluster (the second lowest energy cluster centroid of the predicted structures without RDCs), its RMSD is 5.38 Å from 1cdz.

As Brct was a blind prediction, a relatively higher number of restraints from threading were incorporated compared to the other structures simulated here. This may partly explain why we observe that the first cluster centroid (ranked based on the energy) is the best structure in all the runs with the RDCs restraints.

**CONCLUSION**

The results suggest that RDC restraints lead to higher quality structures in *ab initio* prediction, as assessed by both structural and dynamic measures; yet, unlike in the previous work, a small number of RDCs is employed. The predicted structures should be evaluated from the view of both structural and dynamic measures, with the intricate relationship between the two, in connection with the biological function, being of utmost importance.

Ideally, the dynamic characteristics of the predicted structure should converge to those of the native as the native state is approached. The fluctuations and quality of the structure are strongly coupled as the fluctuations reflect the quality of the structure and are expected to improve as the quality of the structure is improved. However, it is not obvious for which range of RMSD/dRMSD values should this happen. The RMSD is a global measure of the similarities between the structures and may not reflect the differences in the local packing densities that inherently affect the dynamic behavior of the structures. The present results show that dynamic characteristics, described here in terms of the shape of the so-called global (or dominant) collective mode, between predicted structures with an RMSD from native around 4–5 Å and closer may not necessarily be similar, as for the Rubredoxin. The inclusion of RDCs corrects the shape of slowest mode to that of the native, as observed by the analysis of the centroid of the clusters (given in Table I) and the analysis of the individual conformations in the clusters. This has implications for structure prediction calculations, as the objective is to select the best structure closest to the native state with the right packing density that dictates the right dynamic modes of functional significance.

The results demonstrate that the predicted structures with the RDCs yield a distribution of motional fluctuations whose patterns can catch functionally important conformational changes as well as the approximate positions of the functionally important residues. This suggests that we can still extract functionally important information from low- to moderate-resolution structural models.

Regarding the implementation and use of RDCs, the following comments can be made:

![FIGURE 7](a,b) Ribbon diagrams of the predicted structures (first cluster centroids/lowest energy) (dark gray) with RDC restraints (RMSD from native of 5.14 Å) and without RDC restraints (RMSD from native of 7.21 Å), respectively, superimposed on the structure 1czd (light gray).
1. From the relative RDC values of the respective bonds, one has an approximate idea about the corresponding projection angles. If the bond pairs have high RDC values of the same sign, these vectors are close to the collinear. On the other hand, if the pairs have large but opposite RDC values, the vectors are normal or close to normal to each other. In a low-resolution model, such as is employed in the present work, those sets of RDC values that would give unambiguous projection angles within the resolution of the present chain model are preferred.

2. If the data are well associated with secondary units, it should be possible to use RDCs as restraints on the orientation of such secondary structural elements rather than the individual bond vectors; that is, instead of the projection angles of the vectors being associated with helices or groups of residues, the projection angles of segments could be obtained to align the secondary structural elements with respect to each other. Then, rapid characterization of quaternary structures and classification of tertiary fold could be feasible.

3. In some cases, there is a need to have more restraints to be able to fold the structure with or without RDCs (as was the case of 1a3k). It is obviously important to be able to implement the restraints from the beginning of the folding simulation; however, it would be useful as well to concentrate on the refinement of already folded structures. For the latter, it could be worthwhile to consider an all-atomic model/potential with RDC restraints, as refinement could be done in a more accurate manner with the explicit treatment of bond vectors of the structure associated with RDC data, if the structure is not misfolded. Such refinement could be particularly significant with RDC restraints in determining the relative orientations of the domains in multidomain structures and identifying the orientation of the ligands to the substrate in the complex systems with a small effort.

4. Larger sets of RDC restraints may lead to higher quality structures with lower RMSD. Therefore, the analysis of the effect of the number of RDCs on the quality of the predicted structures is important. However, besides the number, the positions of the restraints within the structure may affect the quality of the structures. A consideration of topologically different structures and implementation of the same number of restraints, but associated with different parts of the structure for each given case, would complement the latter analysis.

5. The process of deriving sequence specific RDCs may yield chemical shifts sensitive to the secondary structure. The incorporation of the chemical shift data together with RDCs can further improve the prediction of the fold. In a recent study by Rohr et al., the incorporation of both chemical shift and RDC data was shown to allow the best results in their protocol.

We thank Drs. James Prestegard, Gaetano Montelione, and Fang Tian for providing us with the RDC data used in this work. This research was supported in part by NIH Grant No. GM-37408 of the Division of General Medical Sciences of the National Institutes of Health. Useful discussions with W. Tian are gratefully acknowledged.

REFERENCES

24. Aramani, J.; Montelione, G. (RDCs of ACP, Nodf, NS1, Brc Domain, and Crd); Tian, F. (RDCs of Rubredoxin), private communications.