RDC-assisted modeling of symmetric protein homo-oligomers

XU WANG,[1,2] SONAL BANSAL,[1,2] MEI JIANG,[3–5] AND JAMES H. PRESTEGARD[1,2]

[1]Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, USA
[2]Northeast Structural Genomics Consortium, University of Georgia, Athens, Georgia 30602, USA
[3]Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, New Jersey 08854, USA
[4]Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA
[5]Northeast Structural Genomics Consortium, Rutgers University, Piscataway, New Jersey 08854, USA

(RECEIVED December 8, 2007; FINAL REVISION March 3, 2008; ACCEPTED March 4, 2008)

Abstract

Protein oligomerization serves an important function in biological processes, yet solving structures of protein oligomers has always been a challenge. For solution NMR, the challenge arises both from the increased size of these systems and, in the case of homo-oligomers, from ambiguities in assignment of intra- as opposed to intersubunit NOEs. In this study, we present a residual dipolar coupling (RDC)-assisted method for constructing models of homo-oligomers with purely rotational symmetry. Utilizing the fact that one of the principal axes of the tensor describing the alignment needed for RDC measurement is always parallel to the oligomer symmetry axis, it is possible to greatly restrict possible models for the oligomer. Here, it is shown that, if the monomer structure is known, all allowed dimer models can be constructed using a grid search algorithm and evaluated based on RDC simulations and the quality of the interface between the subunits. Using the *Bacillus subtilis* protein YkuJ as an example, it is shown that the evaluation criteria based on just two sets of NH RDCs are very selective and can unambiguously produce a model in good agreement with an existing X-ray structure of YkuJ.

Keywords: NMR; residual dipolar coupling; homo-oligomer; YkuJ; computational modeling

The determination of the positions of atoms within a single polypeptide chain of a protein has been the primary focus of protein structure determination. However, the positioning of subunits in multi-subunit proteins is an equally important, but challenging, task. The challenge is particularly high for NMR-based structure determination of symmetric homo-oligomers. In these cases, a single set of resonances corresponding to the atoms in a monomer unit is seen. NOEs between these resonances can, therefore, correspond to either intra- or intersubunit contacts, resulting in ambiguities in assignment. While X-filtered NOE techniques allow detection and distinction between intra- and intersubunit NOEs (Ikura and Bax 1992; Folkers et al. 1994; Zwahlen et al. 1997), the techniques require preparation of mixed isotopically labeled samples, are often not very sensitive, and the detected NOEs are often few in number. Here, we describe a systematic approach to homo-oligomer structure determination that does not require the observation or use of intersubunit NOE constraints. It relies on residual dipolar couplings (RDCs) from well-defined portions of a monomer to determine alignment frame axes used in RDC measurement, and recognition of the fact that one of the principal axes of the alignment tensor must coincide with the rotational symmetry axis of the oligomer. Combining the resulting constraint on monomer orientation with simple scoring functions based on the predicted alignment and a residue-pair contact potential is demonstrated

Reprint requests to: James H. Prestegard, Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA; e-mail: jpresteg@ccrc.uga.edu; fax: (706) 542-4412.

Abbreviations: RDC, residual dipolar coupling; PEG, polyethylene glycol; NESG, Northeast Structural Genomics Consortium; PDB, Protein Data Bank.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.073395108.

to provide an accurate structure in the case of a homodimer recently targeted as a part of the Protein Structure Initiative (Wunderlich et al. 2004; Kuzin et al. 2005).

The frequency of occurrence of protein oligomers in biology is surprisingly large. From recent genomic and proteomic studies it is estimated that 60%–70% of the proteins in every genome are homo-oligomers (Goodsell and Olson 2000; Levy et al. 2006). Among these, homodimers are likely to be the most common; a survey of all Protein Data Bank (PDB) structures done by Levy et al. showed that close to half of all homo-oligomer structures in the PDB are homodimers (Levy et al. 2006). There are good reasons for the observed prevalence of oligomers. Oligomerization can contribute to protein stability, resistance to degradation, improved enzyme efficiency, elevated binding affinities, and novel regulatory mechanisms (Goodsell and Olson 2000; Ali and Imperiali 2005; Bachhawat et al. 2005; Levy et al. 2006). From an evolutionary point of view, it may be advantageous to achieve these benefits through complexes of smaller proteins, as opposed to large multi-domain proteins, because they are individually less prone to transcription or translation errors (Goodsell and Olson 2000; Ali and Imperiali 2005).

Despite the predicted prevalence of oligomer structures in biology, oligomer structures in the PDB are actually underrepresented. An analysis in 1996 revealed that oligomer structures make up less than 25% of all PDB structures (Jones and Thornton 1996). This percentage has risen in recent years, but considering that the predicted percentage in various genomes is above 50%, the challenges associated with oligomer structure determination are expected to increase. In meeting these challenges, it is important to seek methods, such as NMR, that can determine oligomer structures in solution, as proper contacts of weak oligomers may not be accurately represented by crystallography where crystal-packing forces can play an important role (Zhang et al. 1995).

In this study, the protein YkuJ from *Bacillus subtilis* will be used as a test case for this new approach to dimer structure determination. This protein was a target of the Northeast Structural Genomics (NESG) effort and was given the NESG target name SR360. Its structure was initially determined using X-ray crystallography, and the coordinates have been deposited in the PDB as 2FFG. YkuJ's function is as yet uncharacterized, but its mixed α/β fold, consisting of an antiparallel β-sheet with C- and N-terminal helices packed against one side, is similar to the *Pseudomonas* avirulence protein AvrPphB. Light scattering and sedimentation equilibrium studies have shown that YkuJ is a dimeric protein in solution with a K_d in the nM range. However, the X-ray structure of YkuJ is a tetrameric complex with two possible dimer combinations (Fig. 1). One of the objectives of our study was to determine which, if either, of these dimers is present in solution.

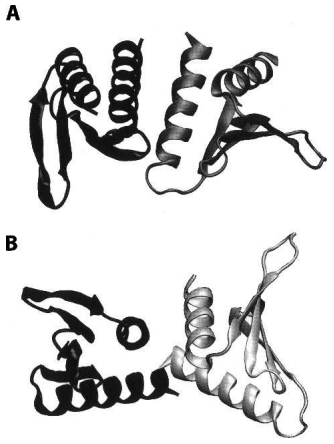

Figure 1. Schematic diagram of the two possible dimer models from the X-ray structure of YkuJ (SR360) (PDB access code 2FFG). (*A*) The dimer formed in the asymmetric unit. (*B*) The dimer formed by two molecules across the asymmetric unit. Subunit with the same shade represents the same molecule in unit cell.

The strategy proposed here combines NMR constraints from RDCs in weakly aligned systems with a computational approach for selecting the appropriate monomer–monomer interface. The use of RDCs as a source of orientational information has become routine in investigations of protein structure and dynamics (Bax and Grishaev 2005; Prestegard et al. 2005). In fact, several groups are working on protocols for solving protein structures using primarily RDC and chemical shift information rather than relying on NOE-derived distance restraints (Delaglio et al. 2000; Valafar and Prestegard 2003; Prestegard et al. 2005; Mayer et al. 2006). There have also been two particularly significant applications to the determination of a dimer structure that uses RDCs, which differ from the work presented here in that alignment axes are allowed to float during structure determination and a small number of experimental distance restraints are used to help determine structures (Bewley and Clore 2000; Rumpel et al. 2008). Procedures closer to those employed here have been previously used in the assembly of heterologous protein–protein complexes (Jain et al. 2004).

Structure determination from RDCs uses the angle dependence of the residual through-space coupling of individual pairs of nuclei, D_{ij}, as a constraint on the orientation of the ij vector relative to the axes of a principal alignment frame (Tjandra and Bax 1997):

$$D_{ij} = (D_a/r^3)[(3\cos^2(\theta) - 1) + (3/2)R\sin^2(\theta)\cos(2\phi)] \quad (1)$$

Here, θ and ϕ are polar angles of the internuclear vector in this frame, r is the vector length, and D_a and R are the

global axial and rhombicity parameters that characterize the molecular alignment. In the case of the $^{1}H–^{15}N$ RDCs, r is actually an amide bond length and can be considered fixed at 1.025 Å.

For the purpose of determining the symmetry axis of the oligomer, the monomer structure is assumed to be rigid, and the primary interest will be in the orientation of alignment axes in the molecular frame. This alignment is most conveniently found by writing equations for an entire set of RDCs in the form of Equation 2, where the $\theta_{k,l}$ are angles of the bond vector relative to the molecular frame and S_{kl} are elements of an order matrix. Solving for the S_{kl} by singular value decomposition and finding the transformation that diagonalizes the matrix gives the needed information on how the principal alignment frame is oriented in the molecular frame (Losonczi et al. 1999)

$$D_{ij} = (D_{max\,ij}/r^3)\sum_{kl} S_{kl}\cos(\theta_k)\cos(\theta_l) \quad (2)$$

Building a dimer structure under RDC constraints uses the fact that for a dimer with rotational symmetry, the C_{2V} axis must coincide with one of the axes of the principal alignment frame. This coincidence is actually independent of the order of the oligomerization and is true for all oligomers with rotational symmetry (Al-Hashimi et al. 2000, 2001; Bewley and Clore 2000). In fact, for oligomers with threefold or higher rotational symmetry, the alignment would be axially symmetric, and the nondegenerate axis would be parallel to the symmetry axis of the oligomer (Al-Hashimi et al. 2000). For dimers of C_{2V} symmetry there are actually multiple possible solutions, in that any one of the three principal frame axes can be chosen as the symmetry axis. This degeneracy can be removed by using data from two independent alignments, in which case only the axis corresponding to the C_{2V} is common to the two frames. With this axis identified, dimers can be generated by rotating the monomer by 180° about the C_{2V} axis. Assembling the dimer is now a problem involving only translation in the plane perpendicular to the symmetry axis. A search for possible dimer structures over points reached by translation in a plane can be accomplished easily using a grid search in which only structures on grid points with no van der Waals violations and some minimum contact surface are evaluated. There are more sophisticated approaches such as the FT-based docking method (Katchalskikatzir et al. 1992), but a simple grid search suffices for this initial presentation. One component of the evaluation score can come from a comparison of measured RDCs and RDCs predicted based on the trial dimer structure and a steric alignment model. A steric alignment prediction algorithm is embodied in the program PALES (Zweckstetter and Bax 2000). This assumes alignment is achieved simply by collision of asymmetric molecules with planar walls representing neutral bilayer fragments or other such structures. Since at least one of the alignment media used is usually bilayer-like, this component of the score is considered important.

Other score components can be taken from a wealth of procedures generated by investigators using purely computational approaches to oligomer structure prediction (Huang et al. 2005a; de Vries et al. 2006; Potluri et al. 2006). These include factors such as change in solvent-accessible surface, shape complementarity, van der Waals interaction energy, and residue-pairing potentials for interfacial residues (Moont et al. 1999). The simple residue-pairing potential will prove to be particularly useful. It also has the additional advantage in being less sensitive to knowledge of the exact side chain geometry at the interface.

Results

Alignment of YkuJ and determination of the symmetry axis orientation

YkuJ was aligned in both a 4% dispersion of an alkylated polyethylene (PEG) detergent, $C_{12}E_5$, and a 10% dispersion of Pf1 phage. RDCs from both alignment media fit the crystal structure well without additional refinement against the RDCs. Correlation plots of experimental versus back-calculated RDCs using the best set of alignment parameters are shown for the two media in Figure 2, A and B. The quality of the fit can be assigned a numeric value using a Q factor (Cornilescu and Bax 2000) in which zero is a perfect fit and 0.3 is typical of crystal structures of 2 Å resolution (Bax and Grishaev 2005). The Q factors for PEG RDCs and the phage RDCs were 0.29 and 0.28, respectively.

The extent of alignment for the two media differed by nearly a factor of two with principal order parameters of (1.87×10^{-4}, 3.44×10^{-4}, -5.32×10^{-4}) and (-3.77×10^{-4}, -7.81×10^{-4}, 1.16×10^{-3}) for the respective PEG and the phage media. More importantly, however, is the fact that the axes of the principal alignment frames were not in general coincident. For a dimer with C_{2V} symmetry, only the rotational symmetry axis should be coincident. This allows the axis corresponding to the C_{2V} axis to be identified. The coincidence of a single axis is easily shown by plotting the directions of principal order (Sxx, Syy, and Szz) from all allowed order tensor solutions on a Sauson–Flamsteed plot (Fig. 2C). The Sxx component from both media was the only axis that overlaps for solutions in the two alignment media. The *X*-axis of the best-fit PEG alignment tensor had an orientation of (−14°,−55°) in terms of longitude and latitude in the PDB coordinate frame. The *X*-axis of the best-fit phage

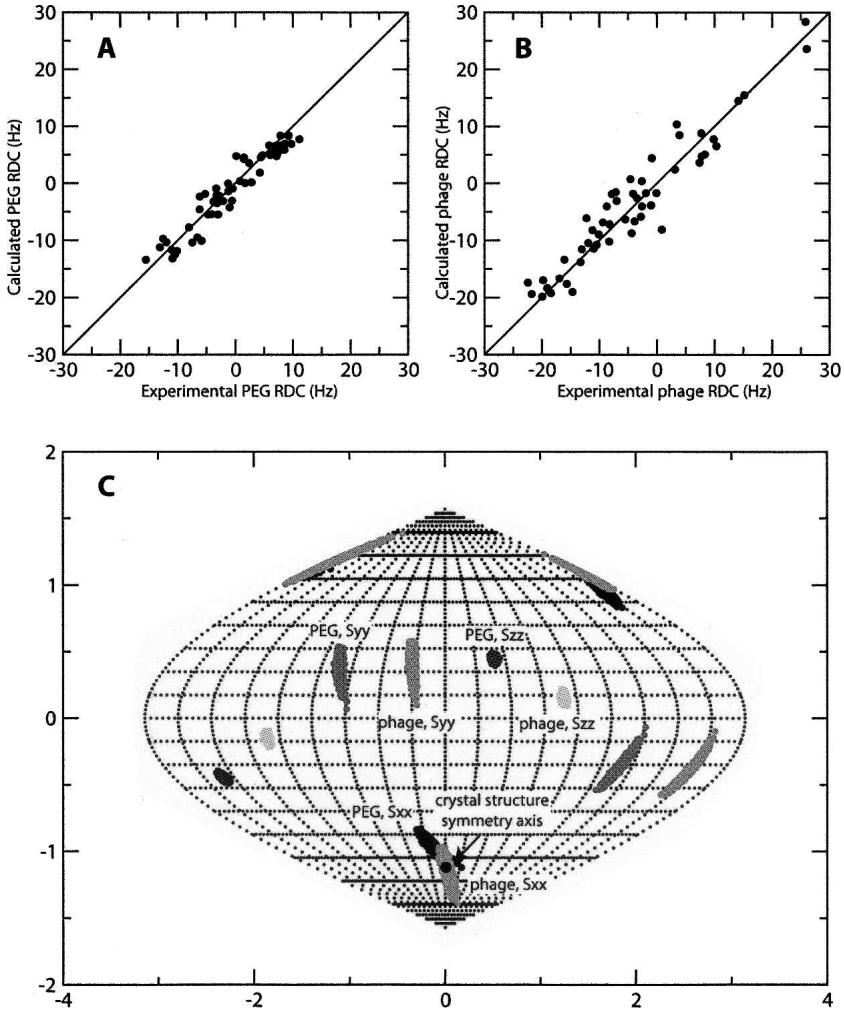

Figure 2. (*A*) Correlation between experimental RDCs collected in a PEG alignment medium and the back-calculated RDCs using dimer model B from the X-ray structure. (*B*) Correlation between experimental RDCs collected in phage alignment medium and the back-calculated RDCs using dimer model B from the X-ray structure. (*C*) Sauson–Flamsteed projection of the orientation of the PEG and phage alignment tensor principal axes frames in the PDB coordinate. Each horizontal grid represents 20° and each vertical grid represents 10°. (Black dot) Orientation of the crystal structure symmetry axis.

alignment tensor had an orientation of (6°,−64°), and the consensus was centered at (−2°,−60°). Using the X-ray structure, the true symmetry axis of dimer model B has an orientation of (1°,−64°), which is close to both the consensus axis as well as the axis specified by phage alignment (alignment in phage was stronger, and these measurements are likely to be more precise).

Dimer models from two sets of RDCs

To generate the dimer models, a grid search algorithm was applied using, as the symmetry axis, the *X*-axes of the PEG alignment tensor, the phage alignment tensor, and the consensus between the two. In this procedure, two monomeric subunits were produced by duplicating the single monomer oriented in the alignment tensor frame indicated by the RDC data and rotating the new copy by 180°. Dimer models were produced by fixing one of the subunits and translating the second subunit in the plane perpendicular to the symmetry axis in 1 Å steps along either one of the two axes perpendicular to the symmetry axis. At each step, maximum surface complementarity between the monomers was achieved by doing a short molecular dynamics (MD) run followed by energy minimization using the CHARMM22 force field. Measuring the shape complementarity of the models before and after molecular dynamics surface optimization showed that the shape complementarities of the models were improved significantly by the short MD simulation. Using the criteria of no intermolecular backbone closer than 4 Å and no intermolecular atomic distances < 2 Å,

~1000 models from a total of 4900 possible translational points initially explored in the grid search were determined as acceptable. The models were then evaluated using the measures described in Materials and Methods. Figure 3 illustrates schematically the algorithm used to generate the dimer models. The particular mix of programs used to generate the models is not critical, and it may well be possible to incorporate most steps in scripts run under a single structure determination program such as XPLOR-NIH (Schwieters et al. 2003).

In all three cases (phage, PEG, and consensus axis), the combined score of the Pearson correlation coefficient and residue-pairing potential identified a single cluster of models as good candidates (Fig. 4; see Materials and Methods for the definition of the combined score). Comparison with the known X-ray models showed that all three clusters represented models that were nearly identical to dimer model B from the X-ray structure, but each had a backbone RMSD of ~45 Å with respect to dimer model A. As expected based on the more precise data, the *X*-axis determined from phage RDC data generated models that were closest to dimer model B. The top 30 scoring models generated with the phage RDCs had an average backbone RMSD of 1.4 ± 0.5 Å relative to dimer model B. The top 30 models generated by the consensus *X*-axis had an average backbone RMSD of 1.6 ± 0.4 Å. The *X*-axis from PEG RDC data generated the worst models, but the average backbone RMSD for the top 30 models was still only 7 ± 1 Å when compared with dimer model B. Figure 5 shows the superimposition of the models with dimer B from the X-ray structure. It is worth emphasizing the importance of using the residual-pairing score as a selection criterion. Because of the globular nature of YkuJ, for each search there were often two clusters of models giving similarly high correlation between PALES-predicted RDCs and experimental RDCs. But, residue-pairing scores allowed the plausibility of each interface to be objectively evaluated, thus eliminating one cluster as a likely candidate.

Despite the general agreement between X-ray dimer structure B and the top scoring models, the local correlation between the combined score and deviation from the X-ray structure was not strong. Table 1 lists the combined score for the top 30 models in each search and their RMSD from dimer model B. In all three searches, the top-scoring model was not the closest match to the reference. The closest match received the sixteenth, fifth, and fifth highest score in the phage, PEG, and consensus searches, respectively. The combined score also did not correlate well with the quality of the model across different searches. This local insensitivity most likely stems from the fact that the combined score is a measure of global properties, like shape, and as such is not sensitive to subtle local differences between models. Smaller grid sizes (0.5 Å) were also tried during these searches. However, no significant increase in model quality was seen despite the fourfold increase in computational demand.

Discussion

The above results have shown that it is possible to construct a model of a dimer with only RDC information and knowledge of the correct subunit structure. In this particular case, the subunit structure was conveniently available from a crystal structure, providing an opportunity to validate the RDC-assisted approach, without worrying about the quality of NMR structures that could be contaminated by incorrect NOE interpretations. The crystal structure did, however, show a tetramer with two possible dimer structures. This provided the additional opportunity to allow the RDC-assisted procedure to select the dimer structure that appears to dominate in solution. The dimer pair B as shown in Figure 1B is clearly the better fit to RDCs, and using the consensus axis, the average RMSD between the top 30 RDC-assisted models and pair B is just 1.6 ± 0.4 Å. This is a value that would

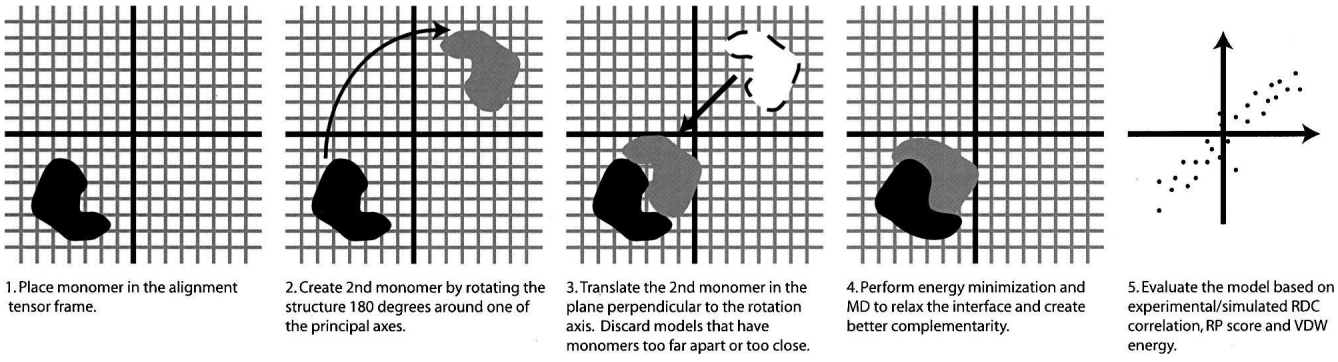

Figure 3. Algorithm for constructing all possible dimer models given the symmetry axis using a grid search algorithm.

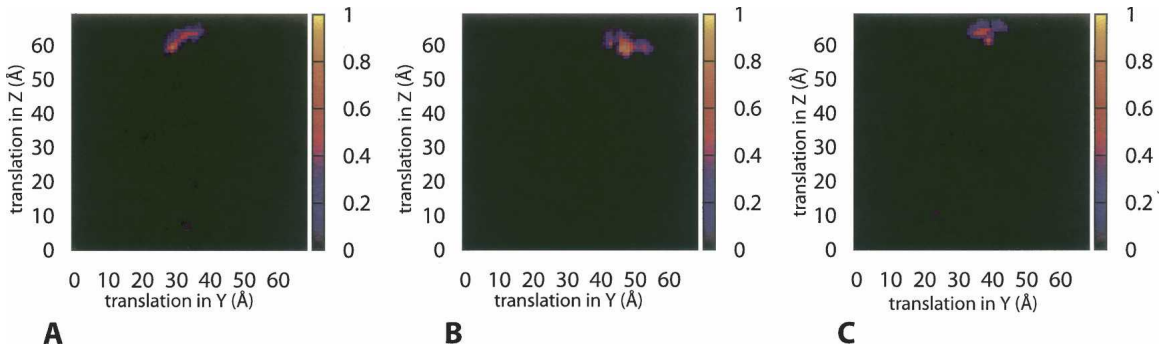

Figure 4. Surface plots of the combined score of dimer models generated using the *X*-axis of the PEG alignment tensor (*A*), the phage alignment tensor (*B*), and the consensus orientation (*C*) as the symmetry axis. During the search, the center of mass of one subunit is fixed at the point (35 Å, 35 Å). The translation refers to the coordinate of the center of mass of the second subunit. Models with VDW energy higher than the median value or RDC correlation <0.85 or residual pairing score <0 are given a score of 0. The score is indicated by the color of each grid point: (black) 0, (yellow) 1. The color gradient for the score is also shown on the *right* of each plot.

be comparable to the precision of most NOE-based NMR structures. In retrospect, the choice of dimer B over dimer A appears to be justified on another basis. The interface area for dimer A of the crystal structure is just 211 Å^2 while that for dimer B is 506 Å^2. Hence, the RDC-assisted methods appear to have selected the most stable dimer and produced a structure with an acceptable accuracy.

There are opportunities for extension of the methods described here. There is, for example, no reason that the RDC-assisted method should not work, after small adjustments in algorithm parameters, for a trimer or even a tetramer, if it is known that such structures possess the appropriate rotational symmetry. Also, it may be possible to proceed without two sets of nondegenerate RDCs. This expands the search space from one plane perpendicular to a single axis to three planes perpendicular to three axes. However, small amounts of additional data, or even just rudimentary knowledge of the properties of protein complexes, may allow proper choice among these alternatives. Recently, for example, some authors have made use of radius of gyration predictions coming from small-angle scattering measurements to refine multimer structures (Sun et al. 2004; Nakasako et al. 2005), and others have used paramagnetic agents to identify probable dimer interfaces (Yang et al. 1996; Jensen et al. 2004). Some of these additional sources might also substitute for our use of a back-calculated RDC score using PALES. While the RDC correlation scores have proved useful in the present case, there are cases where RDC predictions are erroneous. This is particularly the case when the interaction between alignment media and the protein is not entirely steric.

Perhaps the biggest obstacle to the general adoption of the approach we suggest is the availability of a correct monomer structure. Such structures can come from NMR, but structure determination of dimers using NMR is plagued by ambiguities in assigning NOEs to intra- as opposed to intermolecular atom pairs. By symmetry, corresponding atoms in different subunits must give rise to the same resonance, and this introduces an ambiguity in assignment. The conventional approach to resolving the ambiguity, using X-filtered NOEs, will of course, allow the correct monomer structure to be determined, but with sufficient intermolecular NOEs assigned, there may not be a need for RDC-assisted dimer structure determination. In cases where ambiguous NOEs exist (tight dimers), but X-filtered NOEs cannot be observed in

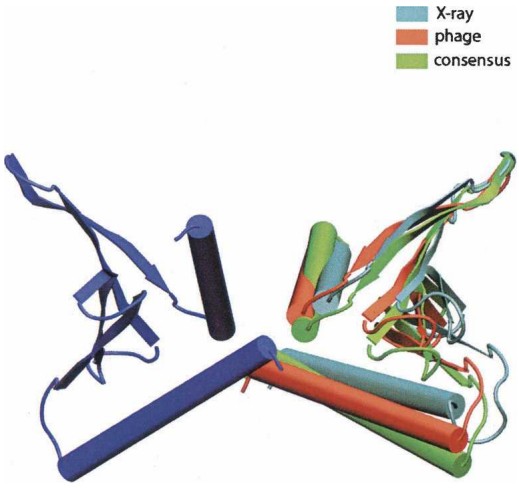

Figure 5. Superimposition of the best models generated by the phage alignment tensor and the consensus searches with dimer model B from the X-ray structure. Chain A in all models are superimposed and are represented by the blue subunit. The best model for each search is chosen by plotting the grid positions of the top 30 models in each search and choosing the model closest to the center. Phage alignment tensor is from grid position (47 Å, 59 Å). Consensus axis search model is from grid position (37 Å, 63 Å).

Table 1. *Combined score and backbone RMSD, relative to dimer model B of the X-ray structure, of the top 30 models generated*

PEG			Phage			Consensus		
Grid position	Score	RMSD	Grid position	Score	RMSD	Grid position	Score	RMSD
30,59	0.95	6.3	48,57	0.95	1.8	40,61	0.96	2.0
29,59	0.82	6.3	49,58	0.87	1.3	39,61	0.84	2.0
34,63	0.64	7.5	47,59	0.87	1.3	37,63	0.67	1.5
30,60	0.62	6.1	48,60	0.85	0.8	36,63	0.66	1.6
30,61	0.60	**6.0**	48,58	0.84	1.4	38,64	0.60	**1.2**
33,62	0.59	6.9	46,59	0.74	1.6	37,64	0.55	1.3
36,64	0.57	9.1	46,58	0.72	2.0	40,62	0.54	1.7
29,60	0.51	6.1	49,60	0.72	0.8	39,64	0.52	1.3
31,61	0.49	6.1	47,60	0.69	1.0	38,63	0.51	1.4
33,63	0.49	7.0	48,59	0.68	1.1	38,65	0.47	**1.2**
36,62	0.47	8.7	49,59	0.68	1.0	40,65	0.47	1.4
31,60	0.43	6.2	46,60	0.66	1.3	36,64	0.43	1.4
31,64	0.43	6.6	47,58	0.65	2.7	44,65	0.43	2.4
35,64	0.41	8.4	44,60	0.64	1.9	43,65	0.42	2.1
32,62	0.41	6.4	44,61	0.63	1.7	35,63	0.40	1.8
37,63	0.41	9.6	48,61	0.62	**0.7**	40,63	0.40	1.5
32,63	0.41	6.6	47,61	0.58	0.8	39,62	0.39	1.7
34,64	0.40	7.8	50,60	0.56	0.9	34,64	0.37	1.7
31,62	0.38	6.1	53,59	0.53	1.8	37,65	0.36	**1.2**
35,63	0.37	8.2	54,58	0.53	2.1	39,65	0.35	**1.2**
36,63	0.35	8.9	46,61	0.53	1.1	41,65	0.32	1.6
32,64	0.34	6.8	47,62	0.52	**0.7**	35,64	0.32	1.5
28,59	0.32	6.5	49,61	0.51	**0.7**	41,64	0.32	1.6
32,61	0.31	6.4	52,59	0.51	1.5	40,64	0.31	1.4
33,64	0.31	7.2	51,59	0.49	1.2	38,62	0.28	1.7
28,58	0.30	6.8	46,62	0.48	0.9	45,65	0.27	2.7
31,63	0.30	6.3	50,59	0.48	1.0	44,66	0.27	2.6
30,63	0.29	6.2	43,60	0.45	2.2	36,65	0.26	1.3
29,61	0.29	**6.0**	54,59	0.44	2.1	34,63	0.26	**1.2**
28,60	0.29	6.3	53,61	0.41	1.9	37,66	0.25	2.0

Models were generated by using, as the symmetry axis, the *X*-axis of the PEG alignment tensor, the phage alignment tensor, and the consensus axis. The models are listed in the order of descending score, and the lowest RMSD value in each search is highlighted. The backbone RMSD is measured from residue 5 to residue 79.

adequate numbers for sensitivity reasons, there is a potential problem with corrupt monomer structures (Nabuurs et al. 2006). This may not, however, be an insurmountable problem. Orientation of monomer units can be determined from any set of RDCs, coming from a reliable part of the monomer structure, that are adequate in number to determine an order tensor (in practice more than 15). The requirements would, of course, be an identification of reliable versus corrupt parts of a structure, and an ability to explore alternate conformations for the corrupt part during the grid search. In principle, observation of inconsistencies between observed and predicted NOEs may highlight suspect areas (Huang et al. 2005b). NOEs from these areas can be eliminated and new conformations for the corrupt parts can be found using RDC and backbone dihedral angle potentials in combination with molecular force fields. When such procedures are successful, models can be used to resolve intra- versus inter-NOE ambiguities and NOE data reintroduced to further improve structures.

One circumstance in which the application of RDC-assisted dimer structure determination may be less complex is in the case of weak dimers. These very often do not have good hydrophobic interface contacts and may actually have few ambiguous NOE problems. Structures determined as a monomer in these cases would be reliable and could be used with RDCs collected under dimer conditions to extend monomer structures to an accurate dimer structure. The only caveat is that conditions under which the fraction of the dimeric species is fairly high must be accessible. Test calculations on SR360 suggest that 5%–10% monomer could lead to as much as a 10° error in symmetry axis position. Weak dimer cases are not rare and may have significant biological implications. Legume lectins, for instance, are known to modulate their multivalency by adopting different oligomeric states in response to environmental changes. Many proteins involved in signal transduction are also known to regulate their activity by forming transient oligomers. One important example is the protease caspase-9, which activates

apoptosis only as a dimer (Renatus et al. 2001). Similarly, there is substantial evidence pointing to the oligomerization of many cell surface receptors as a method to stabilize their interactions with the ligand and transmit signals downstream. Our methodology offers an avenue to study these more transient oligomers by NMR.

Materials and Methods

Protein expression and purification

The YkuJ protein was expressed in *Escherichia coli* strain BL21(DE3) using the pET-21b expression vector prepared in the Rutgers laboratory of the NESG. The sample used for RDC measurements was expressed in minimal media supplemented with 1 g/L 98% ^{15}N ammonium chloride and 3 g of natural abundance glucose supplemented with 5% uniform ^{13}C-labeled glucose. The sample used for resonance assignments was expressed in minimal media supplemented with 1 g/L 98% ^{15}N ammonium chloride and 3 g of 98% uniform ^{13}C-labeled glucose. In both cases, the protein contained a His tag and was purified using a Ni chelate column. Samples were prepared to contain ~1 mM protein in 5% D_2O, 20 mM MES, pH 6.5, 100 mM NaCl, and 10 mM DTT buffer. The RDC samples were diluted by approximately a factor of two on dissolution in alignment media.

NMR data acquisition

Amide nitrogen and proton assignments of YkuJ were obtained through routine sequential assignment procedures using HNCACB and CBCA(CO)NNH experiments distributed in the Biopack package from Varian Inc. These assignments have been deposited in the BMRB, accession number 15529. For RDC measurements, ^{15}N-labeled YkuJ protein was initially aligned in 4% $C_{12}E_5$ (PEG, Sigma Aldrich) using previously published protocols (Ruckert and Otting 2000). The one-bond ^{1}H–^{15}N couplings for isotropic and aligned samples were measured using ^{15}N–IPAP–HSQC experiments (Ottiger et al. 1998). To obtain a second set of RDCs, the protein was aligned in 10 mg/mL Pf1 phage (ASLA biotech; Hansen et al. 1998). Alignment was monitored by splitting of the resonance from 5% D_2O added to the sample buffer. In the PEG and phage media, this was 13 and 7 Hz, respectively. H–N splittings under these conditions ranged from −15 to 11 Hz and −22 to 26 Hz for the two media.

RDC analysis

Extracted NH RDCs were fit to chain A of the X-ray structure of YkuJ (PDB access code 2FFG), modified by the addition of protons in standard geometries, using REDCAT (Valafar and Prestegard 2004). REDCAT repetitively finds order matrix solutions using singular value decomposition and diagonalizes the resulting matrix to determine alignment axis directions. Results from 10,000 trials using values randomly chosen from within error bounds entered for each dipolar coupling (typically ± 3 Hz) were obtained for each medium, and directions of the axes were plotted on a Sauson–Flamsteed plot to assess the extent of overlap for each axis.

Search for allowed geometries

Once the orientation of the dimer symmetry axis is known, models of the dimer having this symmetry axis were built using the VMD software package (Humphrey et al. 1996) and evaluated based on several criteria. First, the monomeric structure was replicated and the copy rotated by 180° around the proposed symmetry axis. Since translations in the plane perpendicular to the symmetry axis do not break the symmetry, all dimer models having the same symmetry axis can be generated by fixing one monomer at the center of the grid and moving the rotated duplicate to each grid point. This algorithm is illustrated schematically in Figure 3. Because YkuJ is a globular molecule with each side being ~35 Å, a grid step of 1 Å was used in each dimension for 70 steps during this study. Models having subunits that were too close (at least one intermolecular backbone distance <4 Å) or too far apart (shortest intermolecular atomic distance >2 Å) were discarded. Finally, the side chains of interfacial residues of the remaining models were relaxed with 500 ps of molecular dynamics simulation in implicit solvent, followed by 100 steps of energy minimization using the molecular dynamics simulation program NAMD (Phillips et al. 2005) and the CHARMM22 force field.

Evaluation of the models

Several measures were used to rank the models generated by the grid search. The most important measure was the agreement between experimental RDC values and the theoretically calculated RDC values for the model. To calculate the simulated RDC for each model, PALES (Zweckstetter and Bax 2000) in the bicelle mode with a media concentration of 4% and a rM of 35 Å was used. The correlation between the experimental RDC collected in PEG and simulated RDCs was measured in terms of the Pearson correlation coefficient between experimental and predicted RDCs. The substitution of the Pearson correlation coefficient with the RDC quality factor (Cornilescu and Bax 2000) produced no difference in the overall result. The second measure used was a residue-pairing score for each model (Moont et al. 1999). A variety of other metrics, including change in solvent accessible surface area and shape complementarity (Lawrence and Colman 1993), have been tested as evaluation criteria, but a residue-pairing score was found to be the most consistent and selective. A residue-pairing score evaluates the quality of the interfacial area by summing the likelihood that each pair of amino acids found at the interface (residues whose intermolecular Cβ distances is within 7 Å) would contribute to a stable interface. The likelihoods come simply from a statistical analysis of pairs found at interfaces of a set of high-resolution X-ray structures in the PDB. Van der Waals energy was also used in the evaluation as an indicator of close contacts and bad packing. Models were considered valid only if they possessed a van der Waals energy that was below the median value of the entire set, a Pearson correlation coefficient >0.85, and a residual pairing score >0. A combined score was then assigned to each valid model. This score was the product of the correlation coefficient with a normalized residue-pairing score, which consisted of the ratio between the residue-pairing score of the model and the maximum residue-pairing score for the entire set of models.

Acknowledgments

This work was supported by a grant from the NIH in support of the Northeast Structural Genomics Consortium (U54-GM074958, G. Montelione, PI) and fellowships to X.W. from the Alberta Heritage Foundation for Medical Research and Canadian Institutes of Health Research. We also thank Dr. Gaetano Montelione of Rutgers University for insightful discussions.

References

Al-Hashimi, H.M., Bolon, P.J., and Prestegard, J.H. 2000. Molecular symmetry as an aid to geometry determination in ligand protein complexes. *J. Magn. Reson.* **142:** 153–158.

Al-Hashimi, H.M., Majumdar, A., Gorin, A., Kettani, A., Skripkin, E., and Patel, D.J. 2001. Field- and phage-induced dipolar couplings in a homodimeric DNA quadruplex: Relative orientation of G · (C-A) triad and G-tetrad motifs and direct determination of C2 symmetry axis orientation. *J. Am. Chem. Soc.* **123:** 633–640.

Ali, M.H. and Imperiali, B. 2005. Protein oligomerization: How and why. *Bioorg. Med. Chem.* **13:** 5013–5020.

Bachhawat, P., Swapna, G.V.T., Montelione, G.T., and Stock, A.M. 2005. Mechanism of activation for transcription factor PhoB suggested by different modes of dimerization in the inactive and active states. *Structure* **13:** 1353–1363.

Bax, A. and Grishaev, A. 2005. Weak alignment NMR: A hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.* **15:** 563–570.

Bewley, C.A. and Clore, G.M. 2000. Determination of the relative orientation of the two halves of the domain-swapped dimer of cyanovirin-N in solution using dipolar couplings and rigid body minimization. *J. Am. Chem. Soc.* **122:** 6009–6016.

Cornilescu, G. and Bax, A. 2000. Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **122:** 10143–10154.

de Vries, S.J., van Dijk, A.D.J., and Bonvin, A.M.J.J. 2006. WHISCY: What information does surface conservation yield? Application to data-driven docking. *Proteins* **63:** 479–489.

Delaglio, F., Kontaxis, G., and Bax, A. 2000. Protein structure determination using molecular fragment replacement and NMR dipolar couplings. *J. Am. Chem. Soc.* **122:** 2142–2143.

Folkers, P.J.M., Nilges, M., Folmer, R.H.A., Konings, R.N.H., and Hilbers, C.W. 1994. The solution structure of the Tyr41 → His mutant of the single-stranded DNA binding protein encoded by gene V of the filamentous bacteriophage M13. *J. Mol. Biol.* **236:** 229–246.

Goodsell, D.S. and Olson, A.J. 2000. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29:** 105–153.

Hansen, M.R., Mueller, L., and Pardi, A. 1998. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5:** 1065–1074.

Huang, P.S., Love, J.J., and Mayo, S.L. 2005a. Adaptation of a fast Fourier transform-based docking algorithm for protein design. *J. Comput. Chem.* **26:** 1222–1232.

Huang, Y.J., Powers, R., and Montelione, G.T. 2005b. Protein NMR recall, precision, and F-measure scores (RPF scores): Structure quality assessment measures based on information retrieval statistics. *J. Am. Chem. Soc.* **127:** 1665–1674.

Humphrey, W., Dalke, A., and Schulten, K. 1996. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14:** 33–38.

Ikura, M. and Bax, A. 1992. Isotope-filtered 2D NMR of a protein peptide complex—study of a skeletal-muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. Chem. Soc.* **114:** 2433–2440.

Jain, N.U., Wyckoff, T.J.O., Raetz, C.R.H., and Prestegard, J.H. 2004. Rapid analysis of large protein–protein complexes using NMR-derived orientational constraints: The 95 kDa complex of LpxA with acyl carrier protein. *J. Mol. Biol.* **343:** 1379–1389.

Jensen, M.R., Lauritzen, C., Dahl, S.W., Pedersen, J., and Led, J.J. 2004. Binding ability of a HHP-tagged protein towards Ni^{2+} studied by paramagnetic NMR relaxation: The possibility of obtaining long-range structure information. *J. Biomol. NMR* **29:** 175–185.

Jones, S. and Thornton, J.M. 1996. Principles of protein–protein interactions. *Proc. Natl. Acad. Sci.* **93:** 13–20.

Katchalskikatzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. 1992. Molecular-surface recognition—determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Natl. Acad. Sci.* **89:** 2195–2199.

Kuzin, A.P., Abashidze, M., Forouhar, F., Vorobiev, S.M., Ho, C.K., Janjua, H., Cunningham, K., Conover, K., Ma, L.C., Xiao, R., et al. 2005. Novel X-ray structure of the YkuJ protein from *Bacillus subtilis*. Northeast Structural Genomics target SR360. Northeast Structural Genomics Consortium.

Lawrence, M.C. and Colman, P.M. 1993. Shape complementarity at protein–protein interfaces. *J. Mol. Biol.* **234:** 946–950.

Levy, E.D., Pereira-Leal, J.B., Chothia, C., and Teichmann, S.A. 2006. 3D complex: A structural classification of protein complexes. *PLoS Comput. Biol.* **2:** 1395–1406.

Losonczi, J.A., Andrec, M., Fischer, M.W.F., and Prestegard, J.H. 1999. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* **138:** 334–342.

Mayer, K.L., Qu, Y., Bansal, S., LeBlond, P.D., Jenney, F.E., Brereton, P.S., Adams, M.W.W., Xu, Y., and Prestegard, J.H. 2006. Structure determination of a new protein from backbone-centered NMR data and NMR-assisted structure prediction. *Proteins* **65:** 480–489.

Moont, G., Gabb, H.A., and Sternberg, M.J.E. 1999. Use of pair potentials across protein interfaces in screening predicted docked complexes. *Protein Struct. Funct. Genet.* **35:** 364–373.

Nabuurs, S.B., Spronk, C.A.E.M., Vuister, G.W., and Vriend, G. 2006. Traditional biomolecular structure determination by NMR spectroscopy allows for major errors. *PLoS Comput. Biol.* **2:** 71–79.

Nakasako, M., Matsuoka, D., Zikihara, K., and Tokutomi, S. 2005. Quaternary structure of LOV-domain containing polypeptide of *Arabidopsis* FKF1 protein. *FEBS Lett.* **579:** 1067–1071.

Ottiger, M., Delaglio, F., and Bax, A. 1998. Measurement of *J* and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131:** 373–378.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26:** 1781–1802.

Potluri, S., Yan, A.K., Chou, J.J., Donald, B.R., and Bailey-Kellogg, C. 2006. Structure determination of symmetric homo-oligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing. *Proteins* **65:** 203–219.

Prestegard, J.H., Mayer, K.L., Valafar, H., and Benison, G.C. 2005. Determination of protein backbone structures from residual dipolar couplings. *Methods Enzymol.* **394:** 175.

Renatus, M., Stennicke, H.R., Scott, F.L., Liddington, R.C., and Salvesen, G.S. 2001. Dimer formation drives the activation of the cell death protease caspase-9. *Proc. Natl. Acad. Sci.* **98:** 14250–14255.

Ruckert, M. and Otting, G. 2000. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.* **122:** 7793–7797.

Rumpel, S., Becker, S., and Zweckstetter, M. 2008. High-resolution structure determination of the CylR2 homodimer using paramagnetic relaxation enhancement and structure-based prediction of molecular alignment. *J. Biomol. NMR* **40:** 1–13.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160:** 65–73.

Sun, Z., Reid, K.B.M., and Perkins, S.J. 2004. The dimeric and trimeric solution structures of the multidomain complement protein properdin by X-ray scattering, analytical ultracentrifugation and constrained modelling. *J. Mol. Biol.* **343:** 1327–1343.

Tjandra, N. and Bax, A. 1997. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278:** 1111–1114.

Valafar, H. and Prestegard, J.H. 2003. Rapid classification of a protein fold family using a statistical analysis of dipolar couplings. *Bioinformatics* **19:** 1549–1555.

Valafar, H. and Prestegard, J.H. 2004. REDCAT: A residual dipolar coupling analysis tool. *J. Magn. Reson.* **167:** 228–241.

Wunderlich, Z., Acton, T.B., Liu, J.F., Kornhaber, G., Everett, J., Carter, P., Lan, N., Echols, N., Gerstein, M., Rost, B., et al. 2004. The protein target list of the Northeast Structural Genomics Consortium. *Proteins* **56:** 181–187.

Yang, D., Yamamoto, K., Kanaya, E., Kanaya, S., and Nagayama, K. 1996. Characterization of an artificial dimer of ribonuclease H using ^{1}H NMR spectroscopy. *J. Biomol. NMR* **7:** 29–34.

Zhang, X.J., Wozniak, J.A., and Matthews, B.W. 1995. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J. Mol. Biol.* **250:** 527–552.

Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. 1997. Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage λ N-peptide/boxB RNA complex. *J. Am. Chem. Soc.* **119:** 6711–6721.

Zweckstetter, M. and Bax, A. 2000. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J. Am. Chem. Soc.* **122:** 3791–3792.