

A graphical method for analyzing distance restraints using residual dipolar couplings for structure determination of symmetric protein homo-oligomers[†]

Jeffrey W. Martin,¹ Anthony K. Yan,^{1,2} Chris Bailey-Kellogg,³ Pei Zhou,² and Bruce R. Donald^{1,2*}

¹Department of Computer Science, Duke University, Durham, North Carolina 27708

²Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

³Department of Computer Science, Dartmouth College, Hanover, New Hampshire 03755

Received 1 November 2010; Revised 22 February 2011; Accepted 23 February 2011

DOI: 10.1002/pro.620

Published online 16 March 2011 proteinscience.org

Abstract: High-resolution structure determination of homo-oligomeric protein complexes remains a daunting task for NMR spectroscopists. Although isotope-filtered experiments allow separation of intermolecular NOEs from intramolecular NOEs and determination of the structure of each subunit within the oligomeric state, degenerate chemical shifts of equivalent nuclei from different subunits make it difficult to assign intermolecular NOEs to nuclei from specific pairs of subunits with certainty, hindering structural analysis of the oligomeric state. Here, we introduce a graphical method, DISCO, for the analysis of intermolecular distance restraints and structure determination of symmetric homo-oligomers using residual dipolar couplings. Based on knowledge that the symmetry axis of an oligomeric complex must be parallel to an eigenvector of the alignment tensor of residual dipolar couplings, we can represent distance restraints as annuli in a plane encoding the parameters of the symmetry axis. Oligomeric protein structures with the best restraint satisfaction correspond to regions of this plane with the greatest number of overlapping annuli. This graphical analysis yields a technique to characterize the complete set of oligomeric structures satisfying the distance restraints and to quantitatively evaluate the contribution of each distance restraint. We demonstrate our method for the trimeric *E. coli* diacylglycerol kinase, addressing the challenges in obtaining subunit assignments for distance restraints. We also demonstrate our method on a dimeric mutant of the immunoglobulin-binding domain B1 of streptococcal protein G to show the resilience of our method to ambiguous atom assignments. In both studies, DISCO computed oligomer structures with high accuracy despite using ambiguously assigned distance restraints.

Keywords: nuclear Overhauser effect; residual dipolar coupling; paramagnetic relaxation enhancement; nuclear magnetic resonance spectroscopy; alignment tensor; protein complex structure determination; symmetric homo-oligomer; configuration space; computational geometry; arrangement

[†]DISCO is freely available as open source software on our website at <http://www.cs.duke.edu/donaldlab/software.php>.

Grant sponsor: National Institutes of Health; Grant numbers: R01 GM-65982, R01 GM-079376.

*Correspondence to: Bruce R. Donald, Department of Computer Science, Duke University, Durham, NC 27708. E-mail: brd+ps10@cs.duke.edu

Introduction

A vast number of macromolecules, including many membrane proteins in higher eukaryotic cells, form symmetrical oligomeric complexes containing multiple subunits.¹ The determination of high-resolution solution structures of oligomeric protein complexes, unfortunately, remains a difficult task.² In NMR studies, the fundamental challenge for such systems

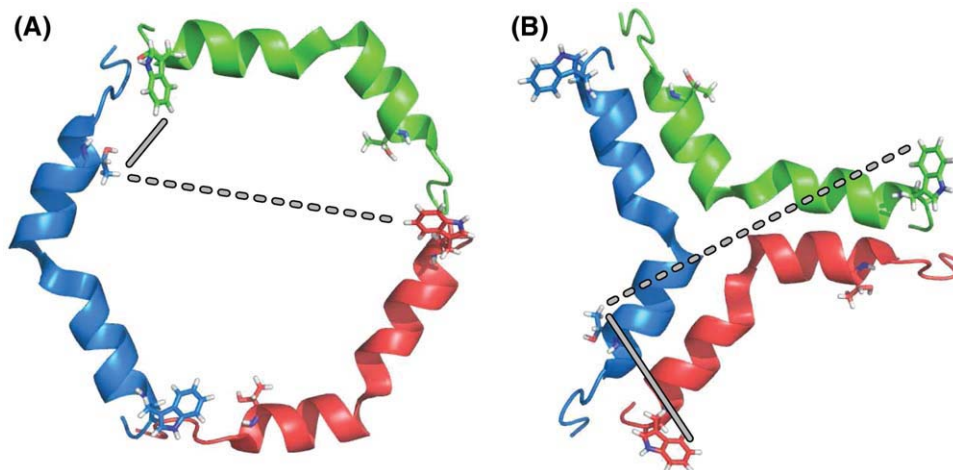


Figure 1. Subunit ambiguity: An ambiguous intermolecular NOE between an H^γ proton of a threonine residue and an $H^{\epsilon 3}$ proton of a tryptophan residue for a hypothetical symmetric trimer has two possible assignments. If we choose the H^γ proton to lie in the blue subunit, then the $H^{\epsilon 3}$ proton could lie in either the green or red subunits. Therefore, the distance restraint either relates the blue-green pair of protons or the blue-red pair of protons. The choice of assignment can potentially lead to vastly different overall folds for the trimer. A: A ring-shaped scaffold satisfies the blue-green assignment but not the blue-red assignment. B: A star-shaped scaffold satisfies the blue-red assignment but not the blue-green assignment. Satisfied restraints are shown with solid grey lines. Unsatisfied restraints are shown with dashed grey lines. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

is that the equivalent atoms from different subunits share identical chemical shifts. Therefore, even if it is possible to narrow down the observed NOEs to particular pairs of nuclei, it is still difficult to determine within which subunits these nuclei are located. This dilemma can be partially resolved by isotope-filtered experiments and elegant isotope labeling schemes, which have made it possible to isolate intermolecular NOEs. By excluding intermolecular NOEs from the complete set of NOE distance restraints, it is possible to determine the high-resolution structure of each subunit based on entirely intramolecular restraints.^{3–5} However, current techniques are not able to differentiate which pairs of subunits contribute to the observed intermolecular dipolar interactions, giving rise to *subunit ambiguity*⁶ (see Fig. 1). Subunit ambiguity hinders analysis of not only NOEs but also distance restraints derived from disulfide bonds. While identical chemical shifts for symmetric protons hinder subunit assignment, merely similar chemical shifts also complicate resonance assignment. *Atom ambiguity*⁶ characterizes an NOE that could be assigned to multiple pairs of nuclei when overlapping ranges of chemical shifts are unable to be resolved. Therefore, NOE assignment and structure determination of trimeric complexes or complexes with higher order symmetry remains an unresolved challenge for NMR spectroscopists.

Even with precise unambiguous distance restraint assignments, structure determination remains a difficult task. Structure determination protocols

that rely on distance geometry calculations are computationally expensive to perform on large proteins, and protocols that rely on simulated annealing require careful selection of annealing parameters, may not converge, or can potentially miss structures consistent with experimental restraints. However, by considering distance restraint assignment and oligomeric structure determination simultaneously, we arrive at an elegant solution. Our approach addresses both assignment and structure determination by incorporating information from residual dipolar couplings (RDCs) into the analysis.

In traditional NMR structure determination protocols for symmetric homo-oligomers, RDCs are typically saved until the structure refinement phase, after calculation of an initial fold using a combination of distance restraints and restraints on dihedral angles. Work by Nilges⁷ focused on calculating oligomer models where additional potentials guided the protein structure to satisfy the symmetry constraints. The method relied primarily on simulated annealing and molecular dynamics and has been successfully used in structure determination of homo-oligomers including a trimer⁸ and a hexamer.⁹ A non-crystallographic symmetry potential ensured subunits shared the same local conformation modulo relative placement and global orientation, while an additional potential arranged the subunits symmetrically by minimizing differences in distances for a chosen subset of the distance restraints. When it is difficult to assign distance restraints unambiguously, ARIA¹⁰ can be used to perform simultaneous

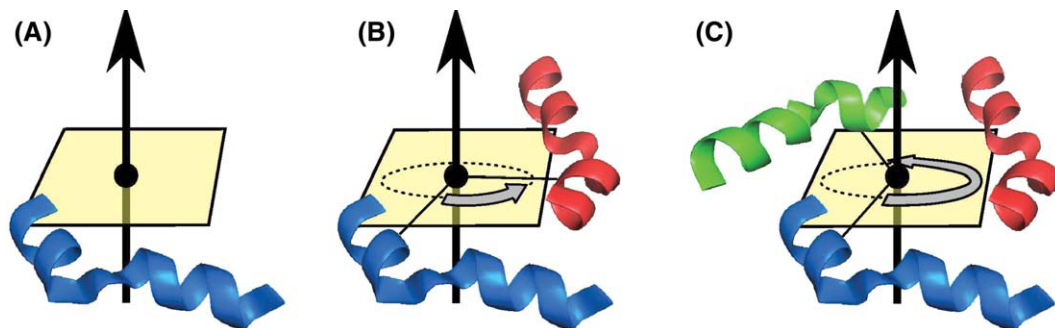


Figure 2. Generating a trimer structure using symmetry: A: Compute the position and orientation of the symmetry axis (vertical arrow) relative to the subunit structure (blue α -helix). B: Copy the subunit structure and rotate by 120° about the symmetry axis to place the second subunit (red α -helix). C: Copy the subunit structure again and rotate by 240° to place the final subunit (green α -helix). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

structure calculation and distance restraint assignment using ambiguous distance restraints and has been improved by Bardiaux *et al.*¹¹ who implemented network anchoring¹² and spin-diffusion correction¹³ into the framework.

We propose a new protocol for structure calculation of homo-oligomers with cyclic symmetry that incorporates RDCs into the beginning of the oligomeric assembly method, so that we may take advantage of the global nature of the restraint provided by the RDCs. Our RDC-first approach creates a framework in which we analyze local intermolecular distance restraints without requiring a complete oligomer structure. Instead, the oligomer structure can be represented in terms of its axis of symmetry and the structure of its subunit (see Fig. 2). Therefore, we perform structure determination in the configuration space of symmetry axes: two translational degrees of freedom (a plane, \mathbb{R}^2) and two rotational degrees of freedom (a unit sphere, \mathbb{S}^2).

Our method, DISCO, uses the observation that the symmetry axis of the oligomeric structure must be parallel to one of the eigenvectors of the alignment tensor,¹⁴ and, therefore, uses RDCs to compute the orientation (in \mathbb{S}^2) of the symmetry axis. Uncertainty in the orientation of the symmetry axis (due to experimental error) is considered by perturbing the experimental RDC values via sampling from a normal distribution, a basic technique that has been previously used to model the experimental error of RDCs for backbone structure determination of monomers.^{15–17} DISCO computes a set of sampled RDC values by drawing a single sample from a normal distribution for each recorded RDC value. Using the set of sampled RDC values and the subunit structure, DISCO computes an alignment tensor that describes a possible orientation for the symmetry axis. By computing a large number of alignment tensors from perturbed RDCs, it is possible to estimate the set of possible symmetry axis orientations, which DISCO conservatively bounds, and then approximates using

a systematic grid search (see section *Computing the uncertainty in the symmetry axis orientation*).

For each orientation on the grid (a *grid orientation*), DISCO uses experimental intermolecular distance restraints such as NOEs and disulfide bonds to compute the position of the symmetry axis—even when precise subunit assignments and atom assignments are not known. Using the computed orientation of the symmetry axis, each possible assignment for a distance restraint restricts the positions of the symmetry axis to an annulus in the plane (\mathbb{R}^2). Each annulus is the set of points lying between two concentric circles whose two radii are mathematically derived from the upper and lower bounds of the corresponding distance restraint (see section *Computing distance restraint unions of annuli* and Fig. 12). Both the inner and the outer radii of each annulus are also dependent on uncertainty in the subunit structure (see section *Calculating subunit structural uncertainty*). Multiple possible assignments for a distance restraint are encoded as a union of the annuli resulting from each possible assignment. DISCO analyzes the unions of annuli using a geometric algorithm to compute the *maximally satisfying regions* (MSRs) of the plane (see section *Computing MSRs*), each of which defines a continuous set of symmetry axis positions representing the complete set of oligomer structures that satisfy the greatest number of intermolecular distance restraints.

Specifically, let us refer to any oligomer structure whose symmetry axis orientation has been computed from RDCs as an *oriented oligomer structure*. Therefore, the space of oriented oligomer structures corresponds to the space of symmetry axis positions. In the case that all distance restraints are simultaneously satisfiable, DISCO can guarantee the MSRs describe all satisfying oriented oligomer structures without missing any of them. If all distance restraints cannot be satisfied, DISCO can guarantee that any oriented oligomer structure whose symmetry axis position has been chosen from the MSRs

will satisfy strictly more intermolecular distance restraints than oriented oligomer structures whose symmetry axis positions have been chosen from outside the MSRs.

Previous work^{6,18,19} also formulated structure determination of homo-oligomers in a symmetry configuration space. Potluri *et al.*^{6,19} computed the orientation and position of the symmetry axis without RDCs using hierarchical subdivision of the configuration space ($\mathbb{R}^2 \times \mathbb{S}^2$). The configuration space was partitioned into regions which were pruned if geometric bounds proved they did not contain any symmetry axes whose oligomer structures satisfied the intermolecular NOEs. Otherwise, the regions were subdivided and the search recursed on their children. Wang *et al.*²⁰ computed symmetry parameters for oligomer models using ambiguously assigned distance restraints by partitioning Cartesian space instead of axis configuration space. After choosing three of the distance restraints as a geometric base, AMBIPACK²⁰ computed symmetry axis parameters by computing the rigid transformation across the interface between two identical subunits. The three chosen distance restraints were used to define a coarse relative orientation between the subunits at the interface, which was iteratively refined against the remaining distance restraints. However, as random sampling and numerical optimization were used to calculate geometric bounds, AMBIPACK is unable to guarantee that all satisfying structures will be discovered. Wang *et al.*¹⁸ computed the orientation of the symmetry axis using only RDCs. The axis position was computed by generating putative dimer models on a grid over \mathbb{R}^2 and scoring the intermolecular interface using a residue-pairing molecular mechanics function. Since dimer models were ranked only according to molecular mechanics scores, van der Waals energy, and agreement with the RDCs, the method does not incorporate the structural information provided by intermolecular distance restraints into the calculation.

DISCO computes symmetry parameters explicitly by analyzing RDCs and distance restraints such as NOEs and disulfide bonds. DISCO computes dimer models and also generalizes to trimers and higher-order oligomers by considering subunit ambiguity. Possible subunit and atom assignments for an intermolecular distance restraint are encoded as a union of annuli in \mathbb{R}^2 , allowing our method to analyze all assignments simultaneously and avoid the need for explicit (and expensive) enumeration of possible assignment combinations. Furthermore, all distance restraints are given the same geometric treatment, avoiding the need to subjectively select a small number of distance restraints at the outset to bootstrap the structure determination. Representing distance restraints as planar annuli also allows us to analyze each restraint independently. We characterize a dis-

tance restraint as *inconsistent* if its corresponding union of annuli does not contain any of the MSRs. No oriented oligomer structure whose symmetry axis position is chosen from a MSR can satisfy an inconsistent restraint. Moreover, DISCO can compute the MSRs exactly without relying on random sampling or numerical optimization and, therefore, is able to guarantee that no satisfying oriented oligomer structures will be missed.

To demonstrate the ability of DISCO to perform structure determination without subunit assignments, we show results for *E. coli* diacylglycerol kinase (DAGK)²¹ using disulfide bonds as distance restraints in section *Structure determination of DAGK under subunit ambiguity*. Like intermolecular NOEs, subunit assignments for disulfide bonds are not known. In addition to considering subunit ambiguity, DISCO also considers atom ambiguity. To demonstrate the resilience of DISCO under ambiguous atom assignments for NOEs, we show results for a dimeric mutant of the immunoglobulin-binding domain B1 of Streptococcal protein G (the GB1 domain-swapped dimer)²² in section *Structure determination of the GB1 domain-swapped dimer under atom ambiguity*. The GB1 mutant differs from the wild type by the L5V/F30V/Y33F/A34F mutations resulting in a domain-swapped dimer. The methodology for our computational tests, is described in the Materials and Methods section.

Results and Discussion

Structure determination of DAGK under subunit ambiguity

To compute the oligomeric structure for the trimeric DAGK, we used the following experimental data: 67 NH RDCs and 24 disulfide bonds per subunit.²¹ We chose model 1 from PDB²³ ID: 2KDC²¹ to serve as the reference structure. The subunit structure used by DISCO was the first subunit in the reference structure, which was determined using traditional protocols. This mirrors the situation where the subunit structure can be determined with confidence,^{3–5} but the main bottleneck is subunit assignment and the assembly of subunit structures to form the oligomer structure.

To compute an initial coarse orientation for the symmetry axis, DISCO first computes an alignment tensor from the RDCs and the subunit structure (see section *Computing the central symmetry axis orientation* for more details). The rhombicity of the computed alignment tensor is 0.02, which is near zero, the value one expects for a symmetric trimer. The alignment tensor fits well to the RDCs and the subunit structure, which is shown by computing the RMS deviation of the recorded RDC values to those back-computed from the subunit structure (the RDC RMSD, 0.28 Hz for the NH RDCs). Using 10,000

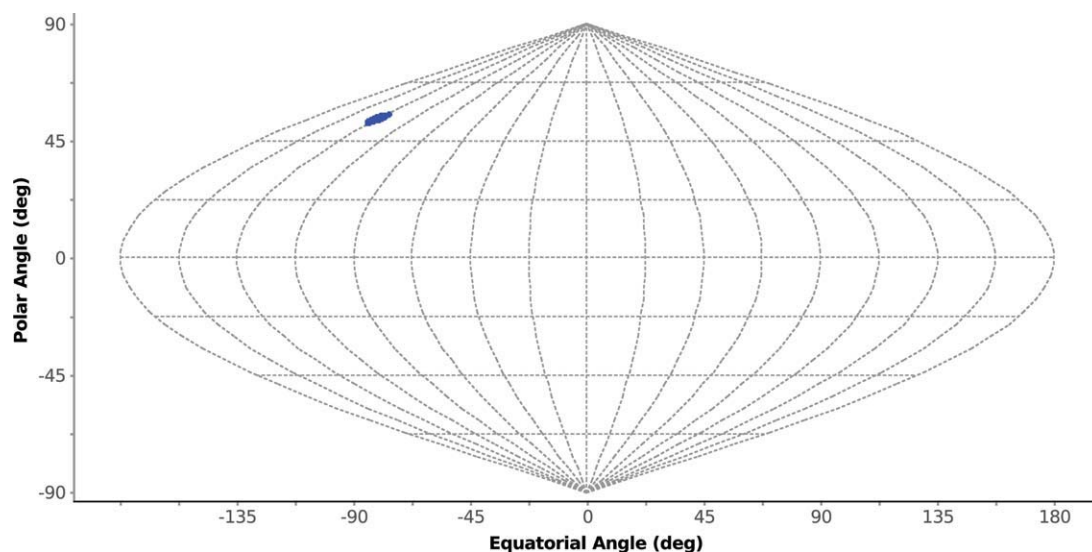


Figure 3. Sampling 10,000 sets of perturbed values using the experimental NH RDCs for DAGK resulted in the symmetry axis orientations (blue) shown on a sinusoidal or Sanson-Flamsteed projection.²⁶ [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

sets of sampled RDCs (see section *Computing the uncertainty in the symmetry axis orientation*), DISCO computed 10,000 alignment tensors whose D_{zz} eigenvectors (the z -axes, using the notation of Clore *et al.*²⁴ and Wedemeyer *et al.*²⁵) show possible symmetry axis orientations (shown in Fig. 3). The RDC values were sampled from normal distributions with standard deviations equal to 1 Hz, resulting in sampled RDC values differing from the recorded RDC values by as much as 4.7 Hz, which are significant deviations for NH RDCs. Figure 4 shows the symmetry axis orientation for the reference structure, which is within the range of the z -axes resulting from the RDC sampling, and also shows the grid of orientations used by DISCO to approximate the z -axes.

DISCO computed MSR for each of the 17-grid orientations (from Fig. 4), drawn from a uniform grid with a resolution of 1° . Figure 5 shows the MSR computed from the disulfide bonds for the most central grid orientation (at coordinates (0,0) in Fig. 4). The MSR contains the position of the symmetry axis for the reference structure, indicating that the distance restraint analysis is able to successfully recover the symmetry parameters of the reference structure. Since DISCO computes the exact set of oriented oligomer structures consistent with the distance restraints, the absence of any additional MSR farther away rules out the possibility of a satisfying oligomer structure that is dissimilar to those already discovered by the algorithm.

To perform detailed structural analysis, we generate a set of discrete structures to represent the MSR by sampling symmetry axis positions from the MSR interiors. One of the advantages of DISCO is that by computing the exact MSR, it is unnecessary

to sample the entire symmetry axis position configuration space. Instead, we can sample only within the MSR at a much finer resolution than would be possible using a grid search over the full configuration space. This is especially important when separate MSR are computed for each symmetry axis

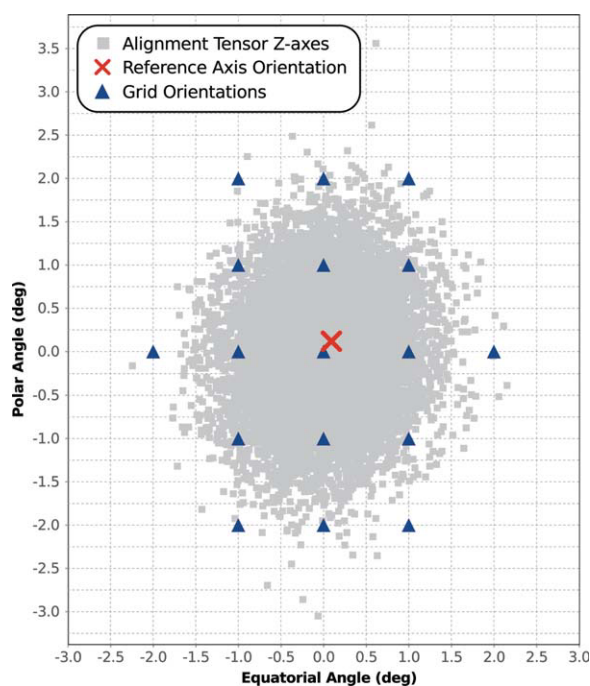


Figure 4. A comparison of the reference symmetry axis orientation (red cross) against the 10,000 z -axes resulting from RDC sampling (grey squares) for DAGK. The orientations illustrated by the blue triangles (the grid orientations) were each used, in turn, to generate constraint annuli. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

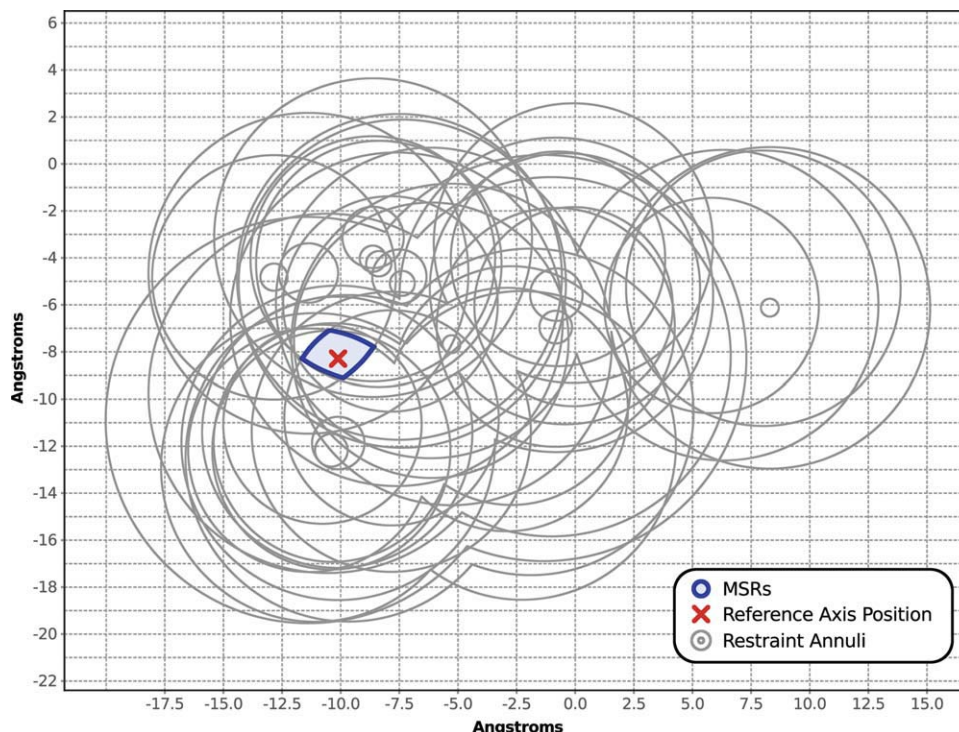


Figure 5. Distance restraint unions of annuli and MSR computed from the 24 disulfide bonds for DAGK using the central grid orientation from Figure 4. This MSR is one of 17 computed for DAGK. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

orientation represented by the grid points. Symmetry axis positions were sampled from the MSRs from the 17 grid orientations on a 0.75 Å resolution uniform grid resulting in 68 oligomer structures for DAGK. All oligomer structures computed by DISCO were within 1.5 Å backbone atom RMSD to the reference structure, with the closest at 0.3 Å.

We scored computed structures according to two criteria⁶: RMS distance restraint violation and van der Waals energy (measured in kcal/mol) after energy minimization in XPLOR-NIH²⁷ where the backbone remains fixed, but side chains are allowed to repack. The RMS distance restraint violation measure scores structural agreement with the intermolecular distance restraints, and the van der Waals energy scores the structures for intermolecular packing. Since DISCO can discriminate between consistent and inconsistent distance restraints, it is easily possible to minimize the computed oligomer structures subject to only the consistent distance restraints, ensuring that inconsistent restraints cannot influence the final minimized structures. However, all of the disulfide bonds were mutually consistent and hence, the minimization was conducted with all available distance restraints. Figure 6 plots the scores of all computed structures for DAGK as well as the score for the reference structure. The computed structures have distance restraint satisfaction scores distributed around the score of the reference structure, with computed structures scoring as much

as 0.12 Å better. Since we expect an oligomer structure to have better packing than the subunit alone, six structures with energies higher than the van der Waals energy of the subunit in isolation (−367 kcal/mol) were removed from the final computed ensemble. Figure 7 shows all 68 oligomer structures computed by DISCO aligned to the reference structure.

Since DISCO can compute the complete set of oriented oligomer structures consistent with the distance restraints, the average RMS deviation from the mean for each backbone atom of the computed structural ensemble represents uncertainty about the position of the symmetry axis inherent in the distance restraints. Structure determination methods that can fail to report satisfying oligomeric conformations (possibly due to under-sampling) can only report the RMS deviation from the mean for each atom of the computed ensemble of structures, which is unable to completely characterize uncertainty about the symmetry axis position. DISCO computed an average RMS deviation from the mean of 1.12 Å for all atoms and 1.08 Å for backbone atoms for the 68 minimized oligomer structures for DAGK.

Structure determination of the GB1 domain-swapped dimer under atom ambiguity

To compute the dimeric structure of the GB1 domain-swapped dimer, we used 56 NH RDCs and 296 experimental intermolecular NOEs (initially assigned unambiguously) per subunit.²² We chose

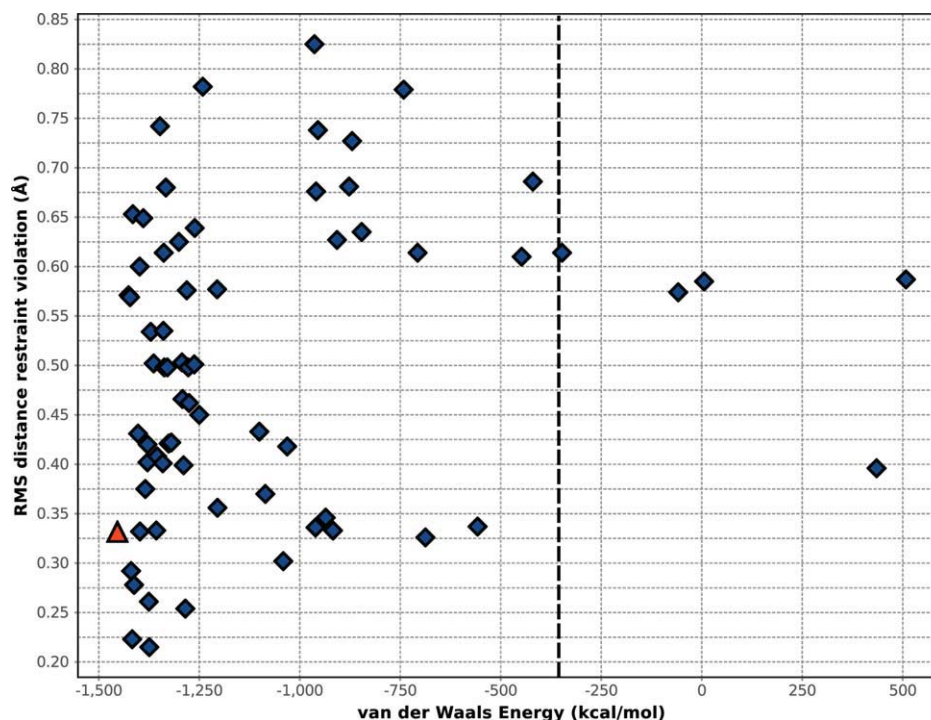


Figure 6. Distance restraint satisfaction scores (lower is better) and van der Waals energies for computed (blue diamonds) and reference (red triangle) oligomer structures for DAGK after minimization. The energy cutoff at -367 kcal/mol is shown with a black dashed line. One computed structure with a very high van der Waals energy has been omitted from the figure. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

model 1 from PDB ID: 1Q10²² to serve as the reference structure. The subunit structure used by DISCO was the first subunit in the reference structure, which was determined using traditional protocols. Again, DISCO focuses on the oligomeric assembly bottleneck, because the subunit structures can in many cases be determined with confidence.^{3–5} While the symmetry axis for a dimer must be parallel to one of the eigenvectors of the alignment tensor, which eigenvector satisfies this condition cannot be uniquely determined from RDCs alone. A search over the three possible choices revealed the D_{xx}

eigenvector as the best candidate (see section *Computing the central symmetry axis orientation* for more details about this search). The alignment tensor computed from the recorded RDCs and the subunit structure fits well, with a RDC RMSD of 0.57 Hz.

Using 10,000 sets of sampled RDCs (see section *Computing the uncertainty in the symmetry axis orientation*), DISCO computed 10,000 alignment tensors whose x -axes show possible symmetry axis orientations. The RDC values were sampled from normal distributions with standard deviations equal to 1

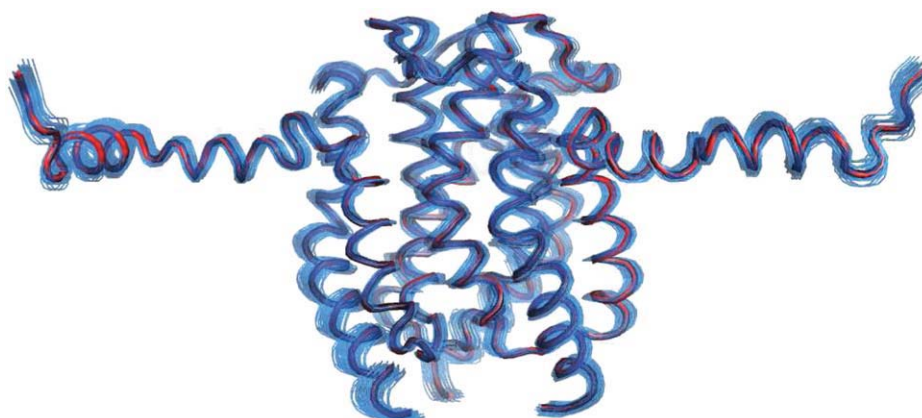


Figure 7. The 68 oligomer structures computed by DISCO (thin blue strands, including the six removed by the energy cutoff) are all within 1.5-Å backbone atom RMSD to the reference (thicker red backbone) for DAGK. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

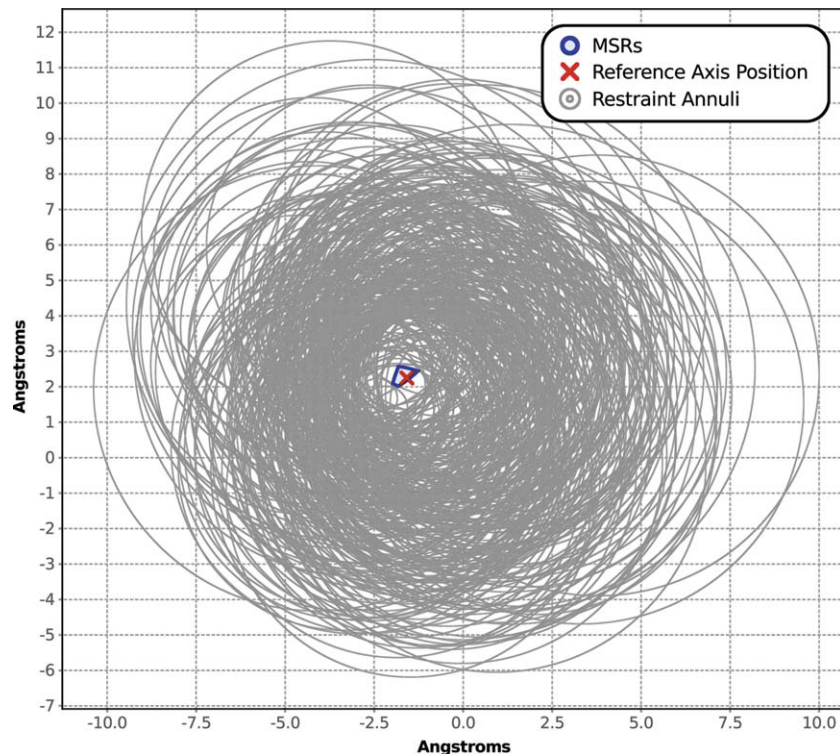


Figure 8. Distance restraint unions of annuli and MSR computed from the 296 NOEs for the GB1 domain-swapped dimer using the central grid orientation. This MSR is one of 19 computed for the GB1 domain-swapped dimer. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Hz, resulting in sampled RDC values differing from the recorded RDC values by as much as 4.6 Hz. Similar to DAGK, the symmetry axis orientation for the

reference structure for the GB1 domain-swapped dimer was also within the range of x -axes resulting from the RDC sampling. DISCO computed MSRs for

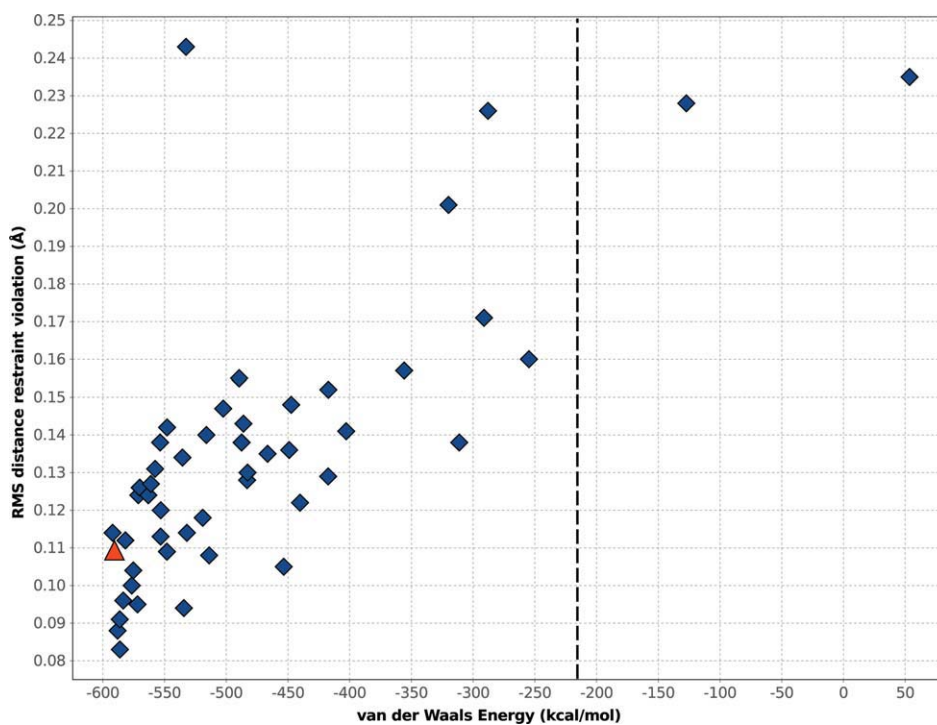


Figure 9. Distance restraint satisfaction scores (lower is better) and van der Waals energies for computed (blue diamonds) and reference (red triangle) oligomer structures for the GB1 domain-swapped dimer. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

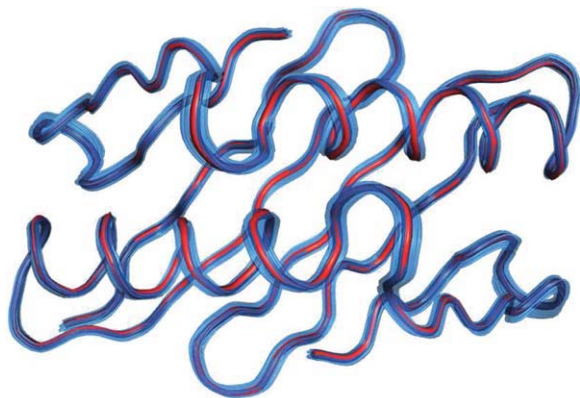


Figure 10. The 48 oligomer structures computed by DISCO (thin blue strands, including the two removed by the energy cutoff) are all within 0.72 Å backbone atom RMSD to the reference (thicker red backbone) for the GB1 domain-swapped dimer. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

each of the 19 grid orientations, which were drawn from a 0.75 Å resolution uniform grid. Figure 8 shows the single MSR computed from the NOEs for the most central grid orientation.

DISCO sampled the 19 MSRs from the grid orientations at a resolution of 0.25 Å, which produced 48 oligomer structures, all of which were within 0.72 Å backbone RMSD to the reference with the closest at 0.07 Å. Figure 9 shows the scores for the energy-minimized oligomer structures compared with a minimized version of the reference structure. The backbone was fixed during minimization, but side-

chains were allowed to repack, and all available NOEs were used to restrain the oligomer structures, because DISCO did not discover any inconsistent NOEs. The structures computed for the GB1 domain-swapped dimer have distance restraint satisfaction scores distributed around the score for the reference structure, with some computed structures scoring negligibly (almost 0.03 Å) better. A single computed structure scored with a lower van der Waals energy than the reference structure (by a narrow margin of 1.5 kcal/mol), and most of the remaining computed structures scored within 200 kcal/mol of the reference. The energy cutoff (−215 kcal/mol) was determined as in section *Structure determination of DAGK under subunit ambiguity*. Two structures whose van der Waals energies were over the energy cutoff were removed from the final computed ensemble. Figure 10 shows all 48 of the oligomer structures computed by DISCO aligned to the reference structure. The DISCO ensemble has an average RMS deviation from the mean of 0.54 Å for all atoms and 0.50 Å for backbone atoms for the 48 computed oligomer structures for the GB1 domain-swapped dimer.

DISCO analyzes distance restraints with atom ambiguity as well as subunit ambiguity. All NOEs for the GB1 domain-swapped dimer were deposited as unambiguously assigned restraints,²² so we simulated atom ambiguity by expanding the assignments to include protons with similar chemical shifts. We obtained 294 ¹H, 175 ¹³C, and 61 ¹⁵N chemical shifts from the BMRB²⁸ using the accession number 5875. We chose a window size of 0.05 ppm for hydrogen

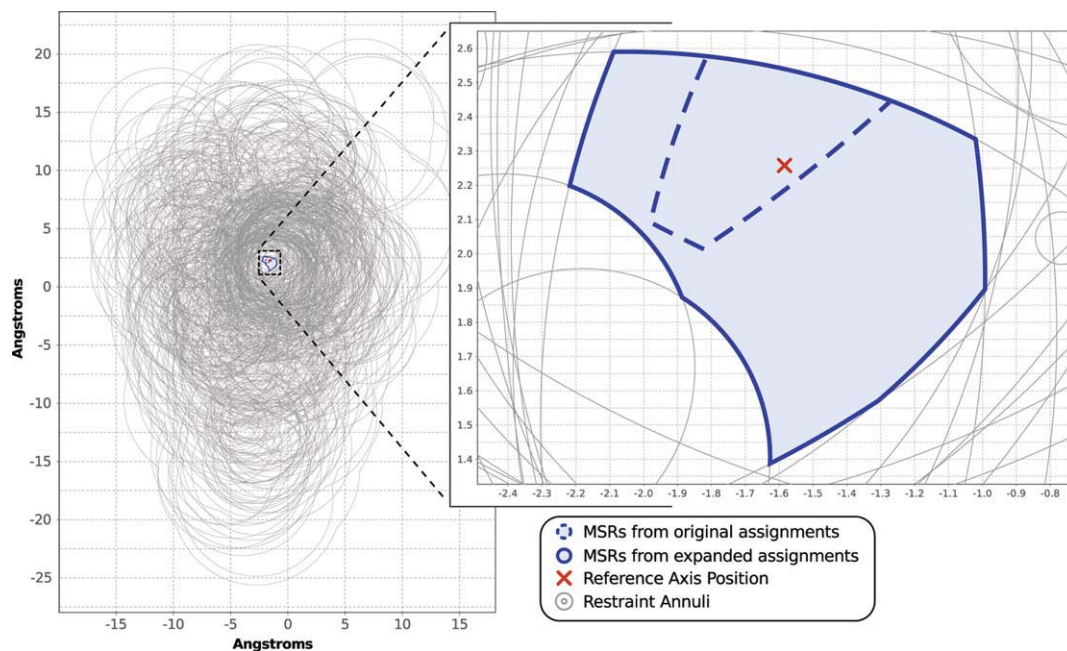


Figure 11. MSRs computed from the NOE assignments for the GB1 domain-swapped dimer with simulated atom ambiguity. Structures sampled finely from the MSR (on a 0.05 Å resolution grid) are all within 0.81 Å backbone RMSD to the reference. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

shifts (δ_H), and a window size of 0.5 ppm for the nitrogen and carbon shifts (δ_C , see section *NOE atom ambiguity simulation* for more details). This simulation increased the average number of assignments per NOE from 1 to 6.7.

To evaluate the effect of these additional NOE assignments on the range of oligomer structures computed by DISCO, we computed MSRs for only the central symmetry axis orientation, which was computed from the original recorded RDC values without perturbation. After comparison with the MSR computed from the same symmetry axis orientation, but using the original unambiguously assigned NOEs (see Fig. 11), we discovered that the MSR computed from the expanded NOE assignments completely contains the MSR computed from the original NOE assignments, as well as the symmetry axis position of the reference structure. Structures sampled finely from the single MSR computed from the expanded NOE assignments on a 0.05 Å resolution grid are all within 0.81 Å backbone RMSD to the reference. Since the MSR computed from the expanded NOE assignments is larger than the MSR computed from the original assignments but still contains the reference axis position, these results indicate that despite a high degree of ambiguity in the distance restraints, DISCO still computes the correct symmetry axis position—just at a slightly lower precision.

Of the 1909 expanded possible assignments for all the NOEs, DISCO discovered that 13.6% of them could not be satisfied by any oriented oligomer structure, indicating a conflict between the expanded possible assignments and the RDC-determined symmetry axis orientation. The annuli for this 13.6% of the expanded assignments enclosed no points (i.e., are the empty set) and therefore, no satisfying symmetry axis positions exist. The section *Computing distance restraint unions of annuli* describes in more detail the distance restraint geometry that results in no satisfying symmetry axis positions for an assignment. These expanded assignments with no satisfying symmetry axes were clearly incorrect and were eliminated immediately using DISCO's RDC-first analysis.

Materials and Methods

To perform the structure determination of DAGK and the GB1 domain-swapped dimer, we conducted a number of computational tests. The NMR data were downloaded from the PDB²³ and the BMRB²⁸; deposition IDs are given in the Results and Discussion section; the data collection is described previously.^{21,22} All computations were performed on a single core of an Intel Core i7 processor at 1.6 GHz that completed in time on the order of hours. The number and type of distance restraints are described in the Results and Discussion section. To compute

oligomer models, DISCO executes a seven-step protocol, which is outlined in the following sections.

Computing the central symmetry axis orientation

DISCO considers uncertainty in the orientation of the symmetry axis by first computing a central orientation and later perturbing it indirectly. The central orientation is computed from an alignment tensor fit to the recorded RDC values and the subunit structure using singular value decomposition.²⁹ To evaluate the fit of this alignment tensor, DISCO back-computes RDCs for the subunit structure and computes the RMSD from the experimental RDCs. For homooligomers with cyclic symmetry, if the alignment tensor has zero rhombicity, one of its eigenvectors must be parallel to the symmetry axis.¹⁴ Further details of computing the central orientation from an alignment tensor depends on the oligomeric state of the protein.

Trimers and higher order oligomers. For trimers and higher-order oligomers, we expect an alignment tensor with zero rhombicity. In this case, the central symmetry axis is parallel to the eigenvector of the alignment tensor whose eigenvalue has the largest magnitude (the principal axis, or *z*-axis). Alignment tensors with nonzero rhombicity for trimers and higher order oligomers do not reflect the symmetry of the oligomer, and we are unable to apply DISCO to the RDCs in that case.

Dimers. For dimers, any value of rhombicity is acceptable, although nonzero rhombicity is actually preferred, because it guarantees the alignment tensor will have at most three eigenvectors. Which eigenvector corresponds to the symmetry axis cannot be uniquely determined from a single set of RDCs alone, so all possibilities must be examined. If an alignment tensor has three distinct eigenvalues, as the tensor for the GB1 domain-swapped dimer does, then one needs to merely consider the three corresponding eigenvectors. DISCO evaluates each choice of eigenvector for distance restraint satisfaction by computing MSRs (see section *Computing MSRs*). The eigenvector whose MSRs satisfy the greatest number of distance restraints is selected as the central symmetry axis orientation.

Computing the uncertainty in the symmetry axis orientation

Once the central symmetry axis orientation has been computed, it is perturbed using the following method that uses the experimental error of the RDCs. For each recorded NH RDC value, define a normal distribution with mean equal to the RDC value and standard deviation equal to 1 Hz. Experimental error for RDCs corresponding to different

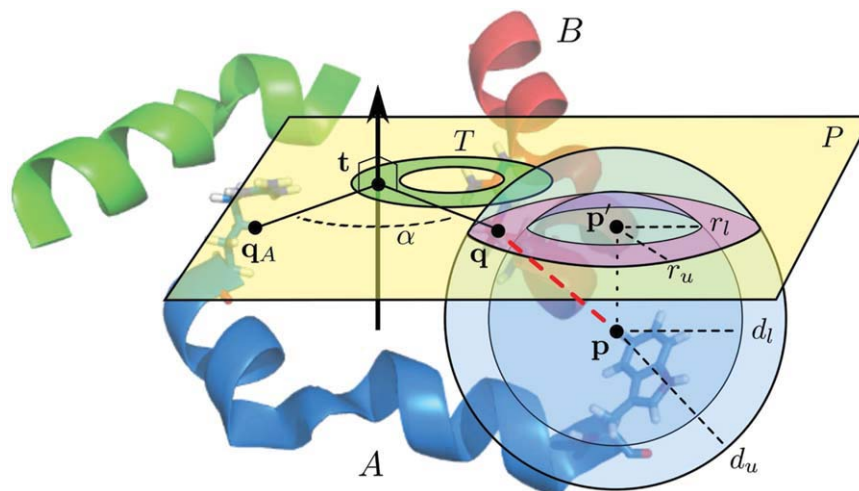


Figure 12. Symmetric distance restraint geometry for a hypothetical trimer: An intermolecular distance restraint (red dashed line) between atom \mathbf{p} (in subunit A, whose position is known) and atom \mathbf{q} (in subunit B, whose position is unknown) is satisfied when \mathbf{q} lies between two 3D spheres with radii d_l and d_u centered at \mathbf{p} . The orientation of the symmetry axis (black arrow), \mathbf{q} , and \mathbf{q}_A (the symmetric partner of \mathbf{q} in subunit A) define a plane P , which allows us to reduce the problem to two dimensions: The distance restraint is satisfied when \mathbf{q} lies between two circles with radii r_l and r_u centered at \mathbf{p}' , which is the projection of \mathbf{p} onto P along the direction of the symmetry axis. The position of the symmetry axis \mathbf{t} relates \mathbf{q} to \mathbf{q}_A by a fixed angle α . Positions of \mathbf{t} that satisfy the distance restraint compose T , the green annulus in P . [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

internuclear vectors can be modeled by varying the choice of standard deviation. Next, compute one set of sampled RDCs by sampling one value from each distribution. Then, fit an alignment tensor using the sampled set of RDCs and the subunit structure to compute one possible symmetry axis orientation. Repeat a large number of times (10,000 sufficed for our computational tests) to compute a large number of possible orientations. Next, bound the set of possible orientations within an elliptical cone. Finally, sample orientations uniformly from the elliptical cone at a desired resolution to compute the set of grid orientations. DISCO uses the grid orientations to represent uncertainty in the orientation of the symmetry axis and evaluates each grid orientation for agreement with the distance restraints in sections *Computing distance restraint unions of annuli* and *Computing MSRs*.

Calculating subunit structural uncertainty

To account for uncertainty in the subunit structure, DISCO adds a padding value α_i to the upper and lower bounds of each distance restraint similarly to how NOEs are adjusted for pseudoatoms. The upper bound for a distance restraint D_i is increased by α_i and the lower bound is decreased by α_i . In the case where the subunit structure was determined by NMR, the ensemble of structures for the subunit directly represents the uncertainty of each atom position. Alternatively (and for X-ray structures), simulations of molecular dynamics can be used to probe for structural variability in the subunit. DISCO computes a padding value for each atom involved in

each possible assignment for the distance restraint. For a unique atom a in the subunit, let $E(a)$ be the set of all instances of that atom in the ensemble. Additionally, let $M(E(a))$ be a function that returns the maximum distance of any atom in $E(a)$ to the centroid of $E(a)$. Given a distance restraint $D_i = \{(\mathbf{p}_k, \mathbf{q}_k)\}$ relating two atoms \mathbf{p}_k and \mathbf{q}_k for each assignment k , DISCO computes $\alpha_i = \max_k M(E(\mathbf{p}_k)) + \max_k M(E(\mathbf{q}_k))$. Hence, the upper bound of the distance restraint increases with the uncertainty in the positions of the two related atoms, and the lower bound decreases toward zero. The computation of padding is automated and requires no user-defined parameters or human choices.

Computing distance restraint unions of annuli

For each possible assignment of each intermolecular distance restraint, DISCO computes one annulus, which describes a continuous set of points in the configuration space of symmetry axis positions (a plane, \mathbb{R}^2). Each point in this annulus describes an oriented oligomer structure that satisfies the assignment.

To compute an annulus for a distance restraint assignment, we first define a coordinate system in which the z -axis ($\hat{\mathbf{z}}$) is parallel to a chosen grid orientation. Using the structure of a single subunit A, we define the origin of this coordinate system to be the centroid of all the atoms in the subunit. DISCO computes the position of the symmetry axis in this coordinate system using the distance restraints, which have been padded according to section *Calculating subunit structural uncertainty*.

Consider a single assignment for an intermolecular distance restraint with minimum and maximum distances d_l, d_u between atoms $\mathbf{p}, \mathbf{q} \in \mathbb{R}^3$ (see Fig. 12). Since the restraint must be intermolecular, let \mathbf{p} lie in subunit A and \mathbf{q} lie in subunit B . If we assume the position and orientation of only subunit A are known, then \mathbf{p} is known, but \mathbf{q} is unknown. Let \mathbf{q}_A be the position of the symmetric partner of \mathbf{q} in subunit A . Because of the symmetry, \mathbf{q} is related to \mathbf{q}_A by a rotation about the symmetry axis (whose orientation is parallel to $\hat{\mathbf{z}}$, but whose position \mathbf{t} is unknown):

$$\mathbf{q} = R(\mathbf{q}_A - \mathbf{t}) + \mathbf{t} \quad (1)$$

where R denotes a rotation about $\hat{\mathbf{z}}$ by an angle $\alpha = \frac{2\pi}{n}$, and n is the oligomeric number of the protein. Therefore, to compute positions of the symmetry axis whose oligomer structures satisfy the distance restraint assignment, DISCO computes values of \mathbf{t} such that the distance restraint is satisfied: $d_l \leq |R(\mathbf{q}_A - \mathbf{t}) + \mathbf{t} - \mathbf{p}| \leq d_u$.

Since we chose a coordinate system in which the symmetry axis is parallel to $\hat{\mathbf{z}}$, we can simplify this problem to two dimensions instead of three. Construct a plane P perpendicular to $\hat{\mathbf{z}}$ such that it contains \mathbf{q} and \mathbf{q}_A . Let $A_3(\mathbf{p}, d_l, d_u)$ be a three-dimensional annulus centered at \mathbf{p} whose radii d_l , and d_u are equal to the lower and upper distance bounds of the distance restraint. The intersection of P with $A_3(\mathbf{p}, d_l, d_u)$ yields a two-dimensional annulus $A_2(\mathbf{p}', r_l, r_u)$ where \mathbf{p}' is the projection of \mathbf{p} along $\hat{\mathbf{z}}$ onto P and the radii are: $r_l = \sqrt{d_l^2 - |\mathbf{p} - \mathbf{p}'|^2}$ and $r_u = \sqrt{d_u^2 - |\mathbf{p} - \mathbf{p}'|^2}$. Therefore, the distance restraint is satisfied when

$$\mathbf{q} \in A_2(\mathbf{p}', r_l, r_u). \quad (2)$$

By substituting Eq. (1) into Eq. (2), we relate the symmetry axis position \mathbf{t} to satisfying positions of \mathbf{q} :

$$R(\mathbf{q}_A - \mathbf{t}) + \mathbf{t} \in A_2(\mathbf{p}', r_l, r_u). \quad (3)$$

To solve for \mathbf{t} , we return to Eq. (1) which can be rewritten: $(R - I)\mathbf{t} = R\mathbf{q}_A - \mathbf{q}$. Next, we substitute Eq. (2) for \mathbf{q} and lift the operators to set operators to consider set membership in place of strict equality: $(R - I)\mathbf{t} \in R\mathbf{q}_A \ominus A_2(\mathbf{p}', r_l, r_u)$ where \ominus represents the Minkowski difference.³⁰ We evaluate the Minkowski difference by simply translating the annulus:

$$(R - I)\mathbf{t} \in A_2(R\mathbf{q}_A - \mathbf{p}', r_l, r_u). \quad (4)$$

Consider all solutions to Eq. (4) for \mathbf{t} as a set T , which represents the set of symmetry axis positions whose oligomer structures satisfy the distance restraint assignment:

$$T = \{\mathbf{t} \in \mathbb{R}^2 \mid (R - I)\mathbf{t} \in A_2(R\mathbf{q}_A - \mathbf{p}', r_l, r_u)\}. \quad (5)$$

To analyze T , we first note that the matrix $(R - I)$ is the composition of a 2D rotation W and a scaling h . To describe h and W , the 2D rotation matrix R can be expressed as a matrix with two orthogonal column vectors of unit length:

$$R = [\mathbf{u} \ \mathbf{v}]. \quad (6)$$

Similarly, $(R - I)$ can be expressed as

$$(R - I) = [\mathbf{u} - \hat{\mathbf{x}} \ \mathbf{v} - \hat{\mathbf{y}}] \quad (7)$$

where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the 2D unit axes. Since $\hat{\mathbf{y}} = R_{\frac{\pi}{2}}\hat{\mathbf{x}}$ and $\mathbf{v} = R_{\frac{\pi}{2}}\mathbf{u}$ where $R_{\frac{\pi}{2}}$ is a rotation in the plane of $\frac{\pi}{2}$ radians, then the following must also be true: $\mathbf{v} - \hat{\mathbf{y}} = R_{\frac{\pi}{2}}(\mathbf{u} - \hat{\mathbf{x}})$, thus, showing that $\mathbf{v} - \hat{\mathbf{y}}$ and $\mathbf{u} - \hat{\mathbf{x}}$ are orthogonal and right-handed. $\mathbf{u} - \hat{\mathbf{x}}$ and $\mathbf{v} - \hat{\mathbf{y}}$ are not of unit length but share a common scaling h that normalizes them:

$$\frac{1}{h}|\mathbf{u} - \hat{\mathbf{x}}| = 1 \quad (8)$$

$$\frac{1}{h}|\mathbf{v} - \hat{\mathbf{y}}| = 1 \quad (9)$$

$$h = |\mathbf{u} - \hat{\mathbf{x}}| = |\mathbf{v} - \hat{\mathbf{y}}|. \quad (10)$$

Together, $\mathbf{u} - \hat{\mathbf{x}}$ and $\mathbf{v} - \hat{\mathbf{y}}$ form the basis for the rotation W

$$W = \frac{1}{h}[\mathbf{u} - \hat{\mathbf{x}} \ \mathbf{v} - \hat{\mathbf{y}}]. \quad (11)$$

Using h , W , and Eq. (5), we can rewrite Eq. (4):

$$hWT = A_2(R\mathbf{q}_A - \mathbf{p}', r_l, r_u). \quad (12)$$

Since T represents a set and hW is invertible, we have replaced the set inclusion of Eq. (4) with strict equality. Solving for T , we see it must also be an annulus in two dimensions:

$$T = A_2\left(\frac{1}{h}W^{-1}(R\mathbf{q}_A - \mathbf{p}'), \frac{r_l}{h}, \frac{r_u}{h}\right). \quad (13)$$

Therefore, DISCO computes the annulus T for a single distance restraint assignment exactly and in closed form using Eq. (13). If $\mathcal{D} = \{D_i\}$ is the set of distance restraints where $D_i = (\mathbf{p}_i, \mathbf{q}_i)$, DISCO evaluates Eq. (13) for each i to compute a set of annuli $T = \{T_1, \dots, T_{|\mathcal{D}|}\}$ that lies on P . In the cases where $T_i = \emptyset$ (when $A_3(\mathbf{p}_i, d_l, d_u)$ and P do not intersect), the restraint cannot be satisfied by any oriented oligomer structure. Effectively, $T_i = \emptyset$ indicates the corresponding restraint is inconsistent with respect to the grid orientation and the symmetry.

To account for possible subunit assignments for a distance restraint (e.g., subunit ambiguity), DISCO computes an annulus for each possible subunit assignment of \mathbf{q} by varying the angle of rotation described by R in Eq. (13) to choose different

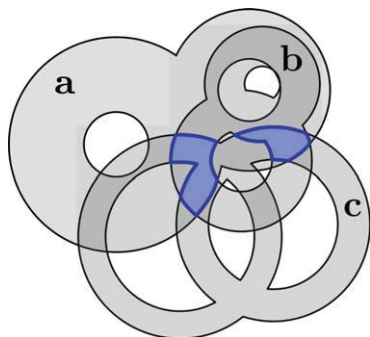


Figure 13. Unions of annuli (grey) for three hypothetical distance restraints (a, b, and c): Computing the arrangement of the unions of annuli gives all intersection points of the circles bounding the annuli, all edges between intersection points, and all faces bounded by the edges. This example shows two MSRs (blue), which are the two faces of the arrangement contained in all three unions of annuli. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

subunits. Since the restraint could be interpreted with any one of these possible assignments, and all of them are mutually exclusive, we conservatively encode the choices using a logical OR operator to avoid a combinatorial enumeration of assignment possibilities. Hence, we redefine T_i to encode the annuli from the possible subunit assignments using set union:

$$T_i = \bigcup_{j=1}^{n-1} A_2 \left(\frac{1}{h} W^{-1}(R_j \mathbf{q}_A - \mathbf{p}'), \frac{r_l}{h}, \frac{r_u}{h} \right) \quad (14)$$

where R_j is a rotation about the \hat{z} axis by an angle of $j\alpha$. T_i now represents the set of symmetry axis positions that satisfy at least one possible subunit assignment for the distance restraint.

Atom ambiguity, which characterizes a distance restraint that could be assigned to multiple pairs of atoms, often due to overlapping chemical shifts, can also be encoded using a union of annuli. We now redefine D_i to represent set of possible atom assignments $\{(\mathbf{p}_k, \mathbf{q}_k)\}$ where \mathbf{p}_k and \mathbf{q}_k are the two atoms for assignment k . Then, T_i can be defined as the union of annuli resulting from all possible atom assignments for the distance restraint:

$$T_i = \bigcup_k A_2 \left(\frac{1}{h} W^{-1}(R \mathbf{q}_A^{(k)} - \mathbf{p}_{k'}), \frac{r_l}{h}, \frac{r_u}{h} \right) \quad (15)$$

where $\mathbf{q}_A^{(k)}$ represents the symmetric partner of \mathbf{q}_k in subunit A. Whether a distance restraint possesses atom ambiguity or subunit ambiguity, DISCO represents the set of satisfying symmetry axis positions as a union of annuli in P .

For the annulus analysis to be meaningful, we require the distance restraints to be strictly intermolecular. If a distance restraint were to possess possible

intramolecular assignments, then it is possible for the true assignment of the distance restraint to be strictly intramolecular. Since an intramolecular distance restraint cannot possibly characterize the oligomeric structure of the protein, no annulus can be computed for an intramolecular assignment. Hence, the remaining intermolecular assignments must all be incorrect, resulting in an incorrect union of annuli. If the incorrect unions of annuli outnumber the correct unions of annuli, and they all happen to conspire and share some common region, then the resulting computed MSRs may not correctly describe the oligomer structure. Therefore, distance restraints that have possible intramolecular assignments, such as PREs, must not be used, unless other reasoning or data can rule out their intramolecular interpretations.

Computing MSRs

To compute positions of the symmetry axis whose oligomer structures satisfy the maximal number of distance restraints (the MSRs), DISCO simultaneously evaluates the unions of annuli for all the distance restraints computed in section *Computing distance restraint unions of annuli*. Ideally, the intersection of all unions of annuli will define a set of symmetry axis positions, but noise and incorrect assignments can result in an empty intersection. Instead, DISCO computes the *arrangement* of the unions of annuli (see Fig. 13 for explanation) using the CGAL software library³¹ and chooses as MSRs the faces from the arrangement that are contained in the greatest number of unions of annuli. CGAL is a C++ software library that implements algorithms from computational geometry (such as computing arrangements) and guarantees exact numerical precision. Further details of our algorithm and an analysis of its asymptotic complexity are presented in Ref. 32. While the arrangement can, in theory, contain multiple faces with equal restraint satisfaction, the computational tests for DAGK (Fig. 5) and the GB1 domain-swapped dimer (Fig. 8) yielded only a single simply-connected MSR in each case.

Computing discrete oligomer structures

The MSRs computed in section *Computing MSRs* define continuous sets of symmetry axis positions and consequently describe continuous sets of oligomer structures, which are difficult to analyze directly. Therefore, to perform detailed structural analyses of the oligomer structures described by the MSRs, DISCO samples discrete symmetry axis positions from the MSRs on a uniform grid at a user-specified resolution. The axis position sampling is repeated for each grid orientation resulting in a set of complete symmetry axes that vary in orientation as well as position, since each arrangement corresponds to a different symmetry axis orientation. The sampled axes define rigid transformations that,

when applied to the subunit structure, generate symmetric oligomer structures. Figure 2 illustrates an example using a trimer.

Evaluating computed structures

DISCO evaluates each computed structure for restraint satisfaction and intermolecular packing, as measured by van der Waals energy. DISCO performs local structure energy minimization using XPLOR-NIH²⁷ where oligomer structures are refined with 1000 steps of Cartesian minimization. The backbone is completely fixed, but the side chains are allowed to move. They are restrained by a van der Waals potential (with default parameters), an NOE potential (with a weight of 30), and the default chemical potentials: BOND, ANGL, and IMPR. Inconsistent distance restraints can optionally be excluded from the NOE potential. After minimization, DISCO computes the distance restraint satisfaction score by evaluating the RMSD of the distance restraints using their (unpadded) upper distances. DISCO also computes the van der Waals packing score for each minimized structure using the pairwise Lennard-Jones potential. The structural ensemble returned by DISCO is composed of the minimized structures. Finally, since we expect an oligomer structure to have better packing than the subunit in isolation, structures with van der Waals energies higher than the subunit structure are removed from the final computed ensemble.

NOE atom ambiguity simulation

We simulated atom ambiguity for the NOEs for the GB1 domain-swapped dimer [5] by expanding the assignments to include protons with similar chemical shifts. Using the ¹H, ¹³C, and ¹⁵N chemical shifts deposited in the BMRB²⁸ (1Q10: 5875) along with the intermolecular NOEs deposited in the PDB²³ (1Q10), we simulated two 3D X-filtered NOESY experiments: 3D ¹⁵N-edited-HSQC-NOESY and 3D ¹³C-edited-HSQC-NOESY. We expanded the assignments for an NOE between protons **p** and **q** in the following way.

Let $H(\mathbf{p})$ be the value of the ¹H chemical shift for proton **p**, and similarly $H(\mathbf{q})$ for **q**. Let $\Theta(\mathbf{p})$ be the chemical shift of the heavy atom covalently bonded to **p**. For example, when **p** is bonded to a N atom, $\Theta(\mathbf{p})$ is the ¹⁵N chemical shift of the bonded N atom. We can view the point $(H(\mathbf{p}), \Theta(\mathbf{p}))$ as residing in a 2D chemical shift space. In this space, finding protons with similar chemical shifts corresponds to finding neighbors of **p**. We define a proton **s** to be a neighbor of **p** if (and only if) the following criteria are satisfied:

$$|H(\mathbf{p}) - H(\mathbf{s})| \leq \delta_H \quad (16)$$

$$|\Theta(\mathbf{p}) - \Theta(\mathbf{s})| \leq \delta_\Theta \quad (17)$$

where δ_H and δ_Θ are user-specified similarity parameters. Since we are simulating 3D NOESY experiments, we must treat **q** as if we do not know $\Theta(\mathbf{q})$. Therefore, DISCO searches for neighbors of **q** using only the criterion in Eq. 16. Alternatively, we could interpret **q** as having a known $\Theta(\mathbf{q})$ instead of **p**, but with only chemical shifts, we cannot determine which interpretation was used during assignment. Therefore, we arbitrarily chose **p** to have known $\Theta(\mathbf{p})$ for all NOEs.

Conclusions

DISCO can accurately determine the oligomer structures of proteins with cyclic symmetry using RDCs and distance restraints such as NOEs and disulfide bonds. It provides a graphical analysis of the distance restraints and is able to differentiate between consistent and inconsistent distance restraints using the maximally satisfying regions. Since DAGK and the GB1 domain-swapped dimer are both high-quality solved structures, it is not surprising DISCO did not discover any inconsistent restraints. DISCO's inconsistency analysis is likely to be more useful during earlier stages of structure calculation when distance restraint assignments may be less certain. DISCO computes oligomer structures using intermolecular distance restraints even when precise atom and subunit assignments are not known, thus, reducing the need to assign distance restraints unambiguously for structure determination. However, only distance restraints with strictly intermolecular possible assignments must be used. Distance restraints with possible intramolecular assignments (such as PREs) cannot be used without first attempting to discard the distance restraints whose true assignments are intramolecular.

DISCO requires a subunit structure to build models of the oligomeric state, but computing an accurate model of the subunit structure in isolation using intramolecular distance restraints can sometimes be challenging. If intramolecular distance restraints are insufficient to adequately constrain the subunit structure, it may be necessary to record additional RDCs and use an RDC-first approach.¹⁷ As an alternative, one could model adjacent subunits during subunit structure calculation using intermolecular restraints to ensure the subunit structure presents an interface amenable to oligomerization. Additionally, for trimers and higher order oligomers, we expect that an alignment tensor computed from the RDCs and the subunit structure will have zero rhombicity. If the rhombicity is significantly greater than zero, the RDCs do not reflect the oligomeric symmetry, they may not accurately describe the orientation of the symmetry axis, and it will not be possible to apply DISCO. For dimers, the rhombicity is not able to indicate agreement between the

symmetry axis orientation and the RDCs, but DISCO is able to search for the best symmetry axis orientation among the three available possibilities; namely, the three eigenvectors of the alignment tensor.

Since DISCO can compute the exact set of oriented oligomer structures that satisfy the distance restraints for each grid orientation, the variance in atom position of the computed ensemble of structures yields a meaningful measure of the range of oligomer structures allowed by the distance restraints. DISCO's graphical analysis is easy to visualize and can find distance restraints that are inconsistent with the RDCs or are inconsistent with other distance restraints. The entire protocol has been completely automated in a software package that is freely available and open-source.

Acknowledgments

The authors thank the members of the Donald lab for comments and suggestions and Dan Halperin for insightful discussions about geometry. Finally, the authors are also grateful to the anonymous reviewers whose helpful suggestions greatly improved this work.

References

- Goodsell DS, Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29: 105–105.
- White SH (2004) The progress of membrane protein structure determination. *Protein Sci* 13:1948–1949.
- Oxenoid K, Chou JJ (2005) The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Nat Acad Sci USA* 102:10870–10875.
- Schnell JR, Chou JJ (2008) Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451:591–595.
- Wang J, Pielak RM, McClintock MA, Chou JJ (2009) Solution structure and functional analysis of the influenza B proton channel. *Nat Struct Mol Biol* 16:1267–1271.
- Potluri S, Yan AK, Donald BR, Bailey-Kellogg C (2007) A complete algorithm to resolve ambiguity for intersubunit NOE assignment in structure determination of symmetric homo-oligomers. *Protein Sci* 16:69–81.
- Nilges M (1993) A calculation strategy for the structure determination of symmetric dimers by ^1H NMR. *Proteins: Struct Funct Genet* 17:297–309.
- Kovacs H, O'Donoghue SI, Hoppe HJ, Comfort D, Reid KBM, Campbell ID, Nilges M. Solution structure of the coiled-coil trimerization domain from lung surfactant protein D. *J Biomol NMR* 24:89–102.
- O'Donoghue SI, Chang X, Abseher R, Nilges M, Led JJ (2000) Unraveling the symmetry ambiguity in a hexamer: Calculation of the R6 human insulin structure. *J Biomol NMR* 16:93–108.
- Rieping W, Habeck M, Bardiaux B, Bernard A, Malliavin TE, Nilges M (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* 23:381–382.
- Bardiaux B, Bernard A, Rieping W, Habeck M, Malliavin TE, Nilges M (2009) Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA. *Proteins: Struct Funct Bioinformatics* 75:569–585.
- Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* 319:209–227.
- Linge JP, Habeck M, Rieping W, Nilges M (2004) Correction of spin diffusion during iterative automated NOE assignment. *J Magn Reson* 167:334–342.
- Al-Hashimi HM, Bolon PJ, Prestegard JH (2000) Molecular symmetry as an aid to geometry determination in ligand protein complexes. *J Magn Reson* 142: 153–158.
- Donald BR, Martin J (2009) Automated NMR assignment and protein structure determination using sparse dipolar coupling constraints. *Prog Nuclear Magn Reson Spectrosc* 55:101–127.
- Wang L, Donald BR (2004) Exact solutions for internuclear vectors and backbone dihedral angles from NH residual dipolar couplings in two media, and their application in a systematic search algorithm for determining protein backbone structure. *J Biomol NMR* 29: 223–242.
- Zeng J, Boyles J, Tripathy C, Wang L, Yan A, Zhou P, Donald B (2009) High-resolution protein structure determination starting with a global fold calculated from exact solutions to the RDC equations. *J Biomol NMR* 45:265–281.
- Wang X, Bansal S, Jiang M, Prestegard JH (2008) RDC-assisted modeling of symmetric protein homooligomers. *Protein Sci* 17:899–907.
- Potluri S, Yan AK, Chou JJ, Donald BR, Bailey-Kellogg C (2006) Structure determination of symmetric homooligomers by a complete search of symmetry configuration space, using NMR restraints and van der Waals packing. *Proteins Struct Funct Bioinformatics* 65: 203–219.
- Wang CE, Lozano-Pérez T, Tidor B (1998) AmbiPack: a systematic algorithm for packing of macromolecular structures with ambiguous distance constraints. *Protein Struct Funct Genet* 32:26–42.
- Van Horn WD, Kim HJ, Ellis CD, Hadziselimovic A, Sulistijo ES, Karra MD, Tian C, Sönnichsen FD, Sanders CR (2009) Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* 324:1726–1729.
- Byeon In-JaL, Louis JM, Gronenborn AM (2003) A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *J Mol Biol* 333: 141–152.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2009) The protein data bank. *Nucl Acids Res* 28:235–242.
- Clore GM, Gronenborn AM, Tjandra N (1998) Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson* 131:159–162.
- Wedemeyer WJ, Rohl CA, Scheraga HA (2002) Exact solutions for chemical bond orientations from residual dipolar couplings. *J Biomol NMR* 22:137–151.
- Bugayevskiy LM, Snyder JP (1995) Map Projections: A Reference Manual. London/Bristol. Taylor & Francis.
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:65–73.
- Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao

- H, Markley L (2007) BioMagResBank. *Nucl Acids Res* 36:D402–D408.
29. Losonczi JA, Andrec M, Fischer MWF, Prestegard JH (1999) Order matrix analysis of residual dipolar couplings using singular value decomposition. *J Magn Reson* 138:334–342.
30. Lozano-Perez T (1983) Spatial planning: A configuration space approach. *Computers IEEE Trans C-32*:108–120.
31. Hanniel I, Halperin D (2000) Two-dimensional arrangements in CGAL and adaptive point location for parametric curves. *Proc. 4th Workshop on Algorithm Engineering, Lecture Notes in Computer Science, Springer-Verlag*, vol. 1982, 171–182.
32. Martin J, Yan T, Zhou P, Bailey-Kellogg C, Donald BR, A geometric arrangement algorithm for structure determination of symmetric protein homo-oligomers from NOEs and RDCs. In: Bafna V, Sahinalp SC, Eds. (2011) *Proceedings of the Fifteenth Annual International Conference on Research in Computational Molecular Biology (RECOMB)*. Springer, Berlin, pp 222–237.