

# A Geometric Arrangement Algorithm for Structure Determination of Symmetric Protein Homo-Oligomers from NOEs and RDCs

JEFFREY W. MARTIN,<sup>1</sup> ANTHONY K. YAN,<sup>1,2</sup> CHRIS BAILEY-KELLOGG,<sup>3</sup>  
PEI ZHOU,<sup>2</sup> and BRUCE R. DONALD<sup>1,2</sup>

## ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is a primary tool to perform structural studies of proteins in physiologically-relevant solution conditions. Restraints on distances between pairs of nuclei in the protein, derived from the nuclear Overhauser effect (NOE), provide information about the structure of the protein in its folded state. NMR studies of symmetric protein homo-oligomers present a unique challenge. Using X-filtered NOESY experiments, it is possible to determine whether an NOE restrains a pair of protons across different subunits or within a single subunit, but current experimental techniques are unable to determine in which subunits the restrained protons lie. Consequently, it is difficult to assign NOEs to particular pairs of subunits with certainty, thus hindering the structural analysis of the oligomeric state. Computational approaches are needed to address this subunit ambiguity, but traditional solutions often rely on stochastic search coupled with simulated annealing and simulations of simplified molecular dynamics, which have many tunable parameters that must be chosen carefully and can also fail to report structures consistent with the experimental restraints. In addition, these traditional approaches rarely provide guarantees on running time or solution quality. We reduce the structure determination of homo-oligomers with cyclic symmetry to computing geometric arrangements of unions of annuli in a plane. Our algorithm, DISCO, runs in expected  $O(n^2)$  time, where  $n$  is the number of distance restraints, potentially assigned ambiguously. DISCO is guaranteed to report the exact set of oligomer structures consistent with the distance restraints and also with orientational restraints from residual dipolar couplings (RDCs). We demonstrate our method using two symmetric protein complexes: the trimeric *E. coli* diacylglycerol kinase (DAGK) and a dimeric mutant of the immunoglobulin-binding domain B1 of streptococcal protein G (GB1). In both cases, DISCO computes oligomer structures with high precision and also finds distance restraints that are either mutually inconsistent or inconsistent with the RDCs. The entire protocol DISCO has been completely automated in a software package that is freely available and open-source at [www.cs.duke.edu/donaldlab/software.php](http://www.cs.duke.edu/donaldlab/software.php).

**Key words:** algorithms, computational molecular biology, protein structure.

---

<sup>1</sup>Department of Computer Science, Duke University, Durham, North Carolina.

<sup>2</sup>Department of Biochemistry, Duke University Medical Center, Durham, North Carolina.

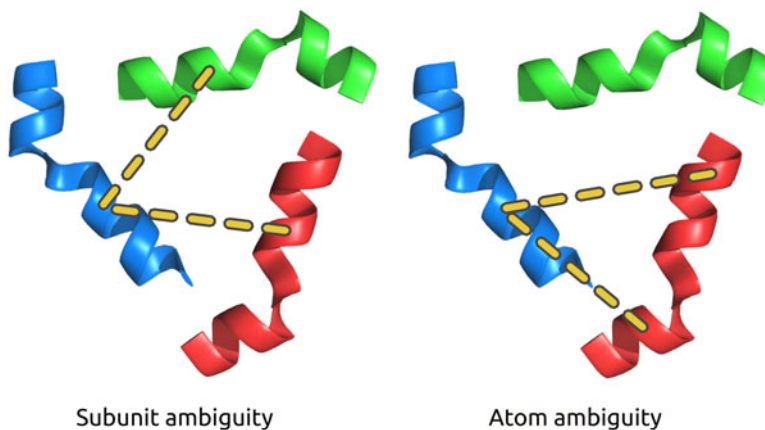
<sup>3</sup>Department of Computer Science, Dartmouth College, Hanover, New Hampshire.

## 1. INTRODUCTION

**S**TRUCTURAL CHARACTERIZATION OF PROTEINS YIELDS INSIGHT into their biological functions, which has become increasingly important for understanding the biochemical basis of human disease. Once the mechanism by which pathogens affect a host is better understood, one can begin to ask how it might be possible to alleviate the effects of infection, or prevent infection altogether. Determining the high-resolution three-dimensional (3D) structures of proteins can enable design of molecules (drugs) to inhibit the native function of a pathogenic protein, or modify helpful proteins to perform a novel function to help stave off infection. One such protein redesign study modified a phenylalanine adenylation domain of the nonribosomal peptide synthetase enzyme gramicidin S synthetase A, an enzyme that originally manufactured the decapeptide gramicidin S, a strong antibiotic, to incorporate different substrates into the molecular assembly line (Chen et al., 2009), thus showing it may be possible to use computational algorithms to engineer enzymes to produce new molecules of potential pharmacological interest. In addition, the same protein design methodology can help predict antibiotic resistance mutations in harmful pathogens such as methicillin-resistant *Staphylococcus aureus* (MRSA) (Frey et al., 2010), giving drug discovery the opportunity to keep one step ahead of its bacterial adversaries. Computational protein redesign is an increasingly popular tool for efficiently exploring possible modifications to protein sequence, but usually requires a structural model of the enzyme or protein of interest.

The majority of proteins assemble as symmetric homo-oligomers (Goodsell and Olson, 2000; Levy et al., 2008), including many membrane proteins, yet the symmetry complicates assignment of inter-subunit distance restraints, and hence oligomeric structure determination by NMR. The pace of structure determination of membrane proteins has lagged significantly behind soluble globular proteins (White, 2004), in part due to these challenges arising from symmetry. Structure determination of symmetric trimers and higher-order homo-oligomers is hindered by *subunit ambiguity* (Potluri et al., 2007): even if an NOE between two protons can be assigned as intra-subunit or inter-subunit through X-filtered NOESY (Ikura and Bax, 1992), current experimental techniques are still unable to determine precisely in which subunits the restrained protons lie. Ambiguity can also arise from other sources of experimental uncertainty. When overlapping chemical shifts prove difficult to separate, distance restraints can be assigned to multiple atoms, i.e., *atom ambiguity* (Potluri et al., 2007). Figure 1 illustrates both types of ambiguity.

Even with precise unambiguous distance restraint assignments and without the complications of symmetry, structure determination of monomeric proteins by NMR remains a difficult task. The formulation of the structure determination problem using only a network of local distance restraints has been proven strongly NP-Hard (Saxe, 1979) and therefore vitiates guarantees of efficiency, accuracy, and completeness. Remarkably, the addition of global orientational constraints on internuclear vectors from residual dipolar couplings (RDCs) and a reduction to *sparse* distance restraints enabled a polynomial-time algorithm for monomeric structure determination (Wang et al., 2006). For symmetric homo-oligomers of at least three subunits, subunit ambiguity complicates assignment of inter-subunit distance restraints, and hence, calculation of oligomer structures, since naïvely enumerating possible assignment combinations requires exponential time. Potluri et al. (2006, 2007) employed a branch-and-bound search algorithm which



**FIG. 1.** An inter-subunit distance restraint (yellow dashed lines) for a hypothetical trimeric protein complex can be assigned to multiple subunits (subunit ambiguity), or multiple atoms within a single subunit (atom ambiguity).

computed symmetric oligomer structures using ambiguously-assigned inter-subunit distance restraints and was guaranteed to return a superset of all oligomer structures satisfying the restraints. The algorithm avoided computing explicit distance restraint assignments, but provided no bound on running time. In this article, we show how the addition of RDCs allows polynomial-time algorithms for structure determination of symmetric homo-oligomers, but with guarantees on solution quality as well as running time.

In practice, approaches such as simulated annealing coupled with simplified molecular dynamics, which lack both combinatorial precision and guarantees on running time and solution quality, are used routinely for structure determination. These approaches require careful selection of annealing parameters, may not always converge, or can potentially miss structures consistent with the experimental restraints due to under-sampling. Simulated annealing protocols begin with an extended model of the oligomeric protein structure which is then folded via simulation of highly simplified molecular dynamics at successively decreasing temperatures. The geometry of nascent protein structure is guided to satisfy restraints on internuclear distance and dihedral angles by potential functions incorporated into the simulation. Since for every internuclear vector orientation that satisfies an RDC value, there exists an equally satisfying inverse vector, restraints on internuclear vector orientation from RDCs are typically not included in the first annealing run. After an initial fold has been calculated from complementary restraints, RDCs are used to refine the structure with further annealing runs at lower temperatures.

We instead propose to incorporate RDCs into the beginning of the structure determination method, thereby creating a framework in which we analyze inter-subunit distance restraints without requiring a complete oligomer structure. When the subunit structure can be determined from intramolecular distance restraints by NMR (Oxenoid and Chou, 2005; Schnell and Chou, 2008; Wang et al., 2009), or by x-ray crystallography (Kuzin et al., 2005), oligomeric structure determination can proceed sequentially, by (1) computing the orientation of the axis of symmetry, (2) computing its position relative to the subunit, and then (3) assembling the subunits together according to the computed symmetry (Fig. 2). Our method, DISCO, computes the orientation of the symmetry axis by analyzing the alignment tensor computed from the RDCs. Then, by using the symmetry axis orientation, the distance restraints are analyzed geometrically by computing the arrangement of unions of annuli in the plane. Faces of the arrangement corresponding to oligomer structures with the greatest distance restraint satisfaction are selected by analyzing the dual graph of this arrangement.

Algorithms from computational geometry (such as the computation of arrangements in the plane) have previously found applications in structural biology outside of direct structure calculation. To characterize structure at protein-protein interfaces, Headd et al. (2007) computed Voronoi diagrams to rigorously define not only which residues of the proteins contribute to the interface, but also provided a method for unambiguously partitioning the interface into core and boundary regions. Liang et al. (1998a) developed an algorithm relying on alpha shapes to compute surface areas and volumes of molecular sphere models analytically, using symbolic perturbation to guarantee robust computations. Alpha shape theory has also been applied via the CAST software (Liang et al., 1998b) to compute volumes of surface pockets and occluded cavities for macromolecules.

Structure determination of homo-oligomers using a symmetry configuration space has also been previously studied (Wang et al., 2008; Potluri et al., 2006, 2007). Potluri et al. (2006, 2007) computed the orientation and position of the symmetry axis using only inter-subunit distance restraints and a hierarchical subdivision of the configuration space ( $\mathbb{R}^2 \times \mathbb{S}^2$ ). Regions of the space were pruned if geometric bounds

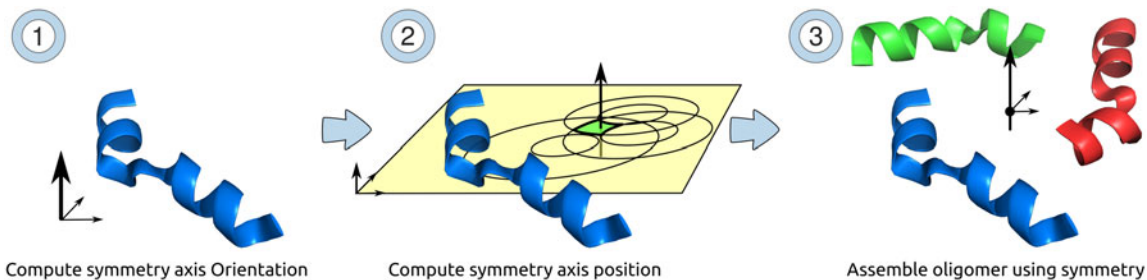


FIG. 2. Overview of protein structure determination using DISCO.

proved they did not contain any symmetry axes whose oligomer structures satisfied the inter-subunit NOEs. Otherwise, the regions were subdivided and the search recursed on their children, continuing in this fashion until a termination criterion was met. Although the method had no bound on time complexity, it was guaranteed to return a superset of all structures satisfying the distance restraints.

Wang et al. (1998) 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, AMBIPACK (Wang et al., 1998) 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. Due to the reliance on random sampling and local numerical optimization, the method may potentially miss structures that satisfy the distance restraints. Wang et al. (2008) computed the orientation of the symmetry axis using just RDCs. The axis position was computed by generating putative dimer models on a grid over  $\mathbb{R}^2$  and scoring the inter-subunit interface using a residue-pairing molecular mechanics function. Since dimer models were ranked only according to molecular mechanics scores, van der Waals energies, and RDC satisfaction, it was not necessary to assign or use inter-subunit NOEs. However, in doing so, the method misses the opportunity to incorporate the structural information provided by these distance restraints.

Nilges (1993) calculated oligomer models without explicit knowledge of the symmetry axis. Instead, structure calculation relied on symmetry potentials during runs of simulated annealing, and has been successfully employed in structure determination of homo-oligomers, including a trimer (Kovacs et al., 2002) and a hexamer (O'Donoghue et al., 2000). 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. Building on the ambiguous distance restraint approach, Bardiaux et al. (2009) implemented network anchoring into ARIA to simultaneously perform distance restraint assignment and oligomeric structure calculation.

To avoid the pitfalls of structure determination methods based on stochastic search, we instead propose algorithms that provide guarantees on the quality of the computed structures. In this article, we describe a novel algorithm, DISCO, that computes oligomeric structures of protein complexes with cyclic symmetry ( $C_n$ ) using RDCs and distance restraints such as NOEs, disulfide bonds, and distance restraints derived from paramagnetic relaxation enhancement (PREs). Along with returning the computed structural ensemble, DISCO guarantees the complete set of oligomer structures satisfying the RDCs and inter-subunit distance restraints can be computed exactly and in polynomial time. The following contributions are made in this article:

- A novel geometric arrangement algorithm, DISCO, is presented to compute structures of homo-oligomeric protein complexes with  $C_n$  symmetry from RDCs, distance restraints such as NOEs, PREs, and disulfide bonds, and a structure of the subunit;
- DISCO guarantees all symmetric homo-oligomers satisfying the RDCs and the distance restraints are discovered, computed exactly, and computed in expected  $O(n^2)$  time, where  $n$  is the number of distance restraints;
- DISCO characterizes the uncertainty in the position of the symmetry axis by computing the variance in atomic coordinates of oligomers sampled uniformly from the exact set of oligomer structures satisfying the RDCs and distance restraints;
- We introduce a technique to analyze ambiguous distance restraints that can discriminate between mutually consistent and inconsistent restraints; and
- We present results on the performance of DISCO on two symmetric proteins: *E. coli* diacylglycerol kinase (DAGK) (Van Horn et al., 2009) and a dimeric mutant of the immunoglobulin-binding domain B1 of streptococcal protein G (GB1) (Byeon et al., 2003).

## 2. METHODS

The atomic structure of an oligomeric protein with cyclic symmetry can be parameterized by the structure of its subunit and the orientation and position of its axis of symmetry (Potluri et al., 2006, 2007).

When it is possible to determine the high-resolution structure of the subunit based on entirely intramolecular restraints (Oxenoid and Chou, 2005; Schnell and Chou, 2008; Wang et al., 2009), the oligomer structure can be assembled from identical subunit structures by computing the parameters of the axis of symmetry: the *orientation* (in  $\mathbb{S}^2$ ) and the *position* (in  $\mathbb{R}^2$ ).

### 2.1. Computation of the symmetry axis orientation

A single scalar RDC value  $r$ , measured experimentally, probes the orientation of an internuclear vector  $\mathbf{v}$  through the following tensor equation:

$$r = D_{\max} \mathbf{v}^T S \mathbf{v} \quad (1)$$

where  $D_{\max}$  is the dipolar interaction constant, and  $S$  is the Saupe order matrix which represents an alignment tensor describing the average weak alignment of a protein in solution (Donald and Martin, 2009). Our algorithm, DISCO, uses the observation that, for a  $C_n$  oligomer, one of the eigenvectors of the alignment tensor must be parallel to the symmetry axis of the complex (Al-Hashimi et al., 2000; Bewley and Clore, 2000), and hence computes the orientation of the symmetry axis (in  $\mathbb{S}^2$ ) from the RDCs. The alignment tensor can be least squares fit to the RDCs and the subunit structure using singular value decomposition (Losonczi et al., 1999), yielding, in the non-degenerate case, three distinct eigenvectors. Which eigenvector to choose from the alignment tensor depends on the oligomeric state of the protein complex.

For trimers and higher-order oligomers, we expect an alignment tensor with zero rhombicity. In this case, the symmetry axis is parallel to the  $D_{zz}$  eigenvector, the eigenvector whose eigenvalue has the largest magnitude. Alignment tensors with non-zero 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. For dimers, *which* eigenvector corresponds to the symmetry axis cannot be uniquely determined from RDCs recorded in a single medium alone, so all eigenvector possibilities must be examined. If the alignment tensor has three distinct eigenvalues, then each corresponding eigenvector is evaluated by executing DISCO three times. We call any oligomer structure whose symmetry axis orientation has been computed from RDCs an *oriented oligomer structure*. Hence, the space of oriented oligomer structures corresponds to the space of symmetry axis positions,  $\mathbb{R}^2$ . To assemble full oligomer structures, all that remains is to compute the set of satisfying symmetry axis positions, thereby building a subset of the full symmetry configuration space,  $\mathbb{S}^2 \times \mathbb{R}^2$ .

### 2.2. Geometric analysis of inter-subunit distance restraints

Once the orientation of the symmetry axis has been computed, each possible assignment for an intermolecular distance restraint is represented as an annulus in a plane encoding the configuration space of the symmetry axis position (Martin et al., 2011). Briefly, let atoms  $\mathbf{p}$  and  $\mathbf{q}$  be restrained with lower and upper distances of  $d_l$  and  $d_u$  respectively. Since the distance restraint is inter-subunit, let  $\mathbf{p}$  lie in subunit  $A$  whose structure is known and  $\mathbf{q}$  lie in subunit  $B$  whose position and orientation relative to  $A$  are unknown. Let  $\hat{\mathbf{z}}$  be the  $z$ -axis of the coordinate system which is parallel to the axis of symmetry. The distance restraint confines the possible positions of the symmetry axis to an annulus in two dimensions perpendicular to  $\hat{\mathbf{z}}$  with center  $\mathbf{c}$ , inner radius  $l$ , and outer radius  $u$ :

$$\mathbf{c} = (R - I)^{-1} (R \mathbf{q}_A - \mathbf{p}') \quad (2)$$

$$l = \frac{\sqrt{d_l^2 - |\mathbf{p} - \mathbf{p}'|^2}}{h} \quad (3)$$

$$u = \frac{\sqrt{d_u^2 - |\mathbf{p} - \mathbf{p}'|^2}}{h} \quad (4)$$

where  $R$  is a rotation about  $\hat{\mathbf{z}}$  of  $\frac{2\pi}{m}$  radians,  $m$  is the number of subunits,  $I$  is the identity matrix,  $\mathbf{q}_A$  is the symmetric partner of atom  $\mathbf{q}$  in subunit  $A$ , and  $\mathbf{p}'$  is the projection of  $\mathbf{p}$  onto the plane containing  $\mathbf{q}$  and  $\mathbf{q}_A$  that is perpendicular to  $\hat{\mathbf{z}}$ . Each annulus is computed exactly and in closed-form using Eqns. (2–4).

Distance restraints with ambiguous assignments yield a set of annuli—one for each possible assignment. After projection onto the  $x$ - $y$  plane, the annuli are combined using set union. Hence, each distance restraint is encoded as a union of annuli. If the possible assignments possess subunit ambiguity, the rotation  $R$  can be varied to select different subunits by changing the angle of rotation  $\frac{j2\pi}{m}$  for  $j \in \{1, \dots, m-1\}$ . If the

possible assignments possess atom ambiguity, the annuli are calculated from the pairs of atoms ( $\mathbf{p}_k, \mathbf{q}_k$ ) corresponding to each assignment  $k$ .

### 2.3. Analysis of the arrangement of unions of annuli

Once a union of annuli is computed for each distance restraint, DISCO computes the arrangement  $\mathcal{A}$  of all the circular curves bounding the unions of annuli.  $\mathcal{A}$  is computed using a randomized incremental algorithm (Halperin, 1997), implemented in the CGAL library (Hanniel and Halperin, 2000).  $\mathcal{A}$  represents all intersection points of the circular curves, all edges bounded by the intersection points, and all faces bounded by the edges. We refer to the faces of  $\mathcal{A}$  contained in the greatest number of unions of annuli as the *maximally satisfying regions* (MSRs). These faces represent symmetry axis positions that satisfy the greatest number of inter-subunit distance restraints.

Formally, let each face  $f$  in  $\mathcal{A}$  have an associated *depth*,  $d(f)$ , equal to the number of unions of annuli that contain  $f$ . The MSRs are the faces in  $\mathcal{A}$  with the maximum depth, and the unbounded face  $f_u$  has a depth of zero. The MSRs are found by analyzing the dual graph  $G = (V, E)$  of  $\mathcal{A}$  (Fig. 3), where  $V$  contains one node for each face in  $\mathcal{A}$ , including  $f_u$ . Let  $f'$  represent the dual of  $f$ , which is the vertex in  $V$  corresponding to  $f$ .  $E$  contains an edge  $(f'_1, f'_2)$  when two faces  $f_1$  and  $f_2$  in  $\mathcal{A}$  share an edge. To annotate the faces of  $\mathcal{A}$  with depths, DISCO performs breadth-first search (BFS) beginning at the vertex  $f'_u$ . When BFS traverses an edge in  $E$ , this corresponds to crossing an edge  $h$  from the previous face  $f_p$  to the next face  $f_n$  in  $\mathcal{A}$ . Therefore,  $d(f_n)$  is assigned  $d(f_p) - 1$  if crossing  $h$  leaves a union of annuli,  $d(f_p) + 1$  if crossing  $h$  enters a union of annuli, or  $d(f_p)$  if  $h$  lies in the interior of a union of annuli (i.e., the depth remains the same; Fig. 4). Once the faces of  $\mathcal{A}$  have been annotated with depths, DISCO returns the MSRs by enumerating the faces with maximum depth.

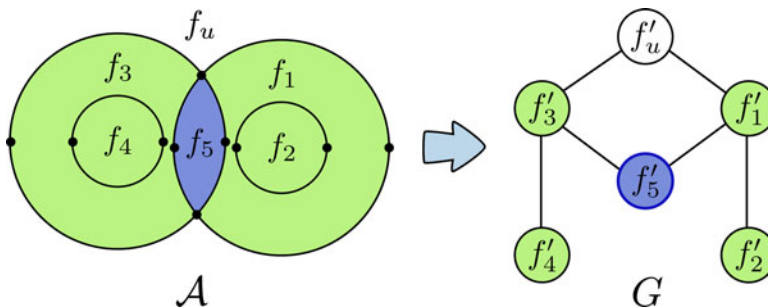
To construct  $\mathcal{A}$ , the circles bounding the unions of annuli are decomposed into  $x$ -monotone circular arcs, which are restricted to be monotonic in the  $x$ -direction (i.e., no vertical line intersects the curve more than once). The two  $x$ -monotone circular arcs resulting from the circle decomposition are subsequently divided into smaller arcs resulting from intersections with other  $x$ -monotone circular arcs during construction of  $\mathcal{A}$ . We will say the resulting circular arcs are all *supported* by the original circle. To decide whether a crossing enters/leaves or remains within a union of annuli, we determine if the edge  $h$  lies in the interior of the union of annuli whose constituent circles support  $h$ . Since each edge in  $\mathcal{A}$  can only be supported by a circle from a single union of annuli, the supporting union of annuli can be referenced by a pointer stored at the edge, and this pointer can be set during the construction of the arrangement. If the midpoint of the edge  $h$  lies in the interior of its union of annuli, then  $h$  is an *interior edge* and the crossing remains within the union of annuli. Any point along  $h$  (except for the endpoints) can be used to test if  $h$  is an interior edge, but the midpoint is the most numerically-stable choice. If  $h$  lies on the boundary of its union of annuli (i.e., is not an interior edge), the crossing enters or leaves a union of annuli. The orientation of the circle supporting  $h$  is used to encode whether or not the circle defines the interior or exterior boundary of the union of annuli.

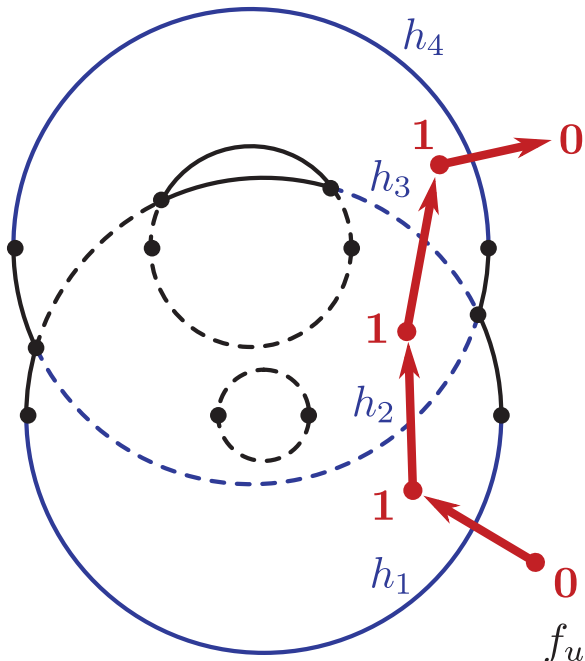
### 2.4. Analysis of complexity

In this section, we prove bounds on the time and space complexity of the computation of the MSRs from ambiguously assigned inter-subunit distance restraints.

**Lemma 1.** *For an oligomeric protein complex with cyclic symmetry and  $n$  distance restraints assigned ambiguously, each having at most  $s$  possible assignments, the MSRs can be computed in expected  $O(s^3 n^2)$  time and  $O(s^2 n^2)$  space.*

**FIG. 3.** A sample dual graph ( $G$ ) for a hypothetical arrangement ( $\mathcal{A}$ ) of  $x$ -monotone curves bounding two annuli, showing the single MSR (blue), the remaining bounded faces (green), and the unbounded face ( $f_u$ ). The bounded faces in  $\mathcal{A}$ , which are labeled  $f_1, \dots, f_5$ , map to vertices in  $G$ , which are labeled  $f'_1, \dots, f'_5$ . The unbounded face maps to the vertex  $f'_u$ .





**FIG. 4.** A path (red arrows) visits faces in a union of annuli (blue/black curves) starting with the unbounded face ( $f_u$ ) and crosses four edges:  $h_1$ ,  $h_2$ ,  $h_3$ , and  $h_4$  (blue curves). Interior edges of the union of annuli are shown with dashed curves. Since  $f_u$  is initialized with a depth of zero, crossing  $h_1$  increases the depth to one, crossing the interior edges  $h_2$  and  $h_3$  do not modify the depth, and crossing  $h_4$  returns the depth to zero.

**Proof.** DISCO computes a single annulus for each assignment of each distance restraint which results in  $n$  unions of annuli, each having at most  $s$  annuli. Hence, there are  $sn$  annuli in total and each union of annuli has a complexity of  $O(s)$ . In the next step, DISCO decomposes the boundaries of the unions of annuli into  $x$ -monotone circular arcs, four for each annulus, resulting in  $O(sn)$  circular arcs. The computation of the arrangement  $\mathcal{A}$  can be accomplished using a randomized incremental algorithm requiring expected  $O(sn \log(sn) + k)$  time and  $O(s^2n^2)$  space, where  $k$  is the number of intersections in the arrangement (Halperin, 1997). The depths can be stored using constant space at each face, thus leaving the  $O(s^2n^2)$  space requirements of the original algorithm unchanged. To test if an edge  $h$  lies on the interior of its supporting union of annuli, we must determine whether its midpoint lies in the interior. Since the single union of annuli supporting  $h$  can be found in constant time by following the pointer at  $h$ , and the complexity of each union of annuli is  $O(s)$ , the interior predicate for  $h$  can be evaluated in  $O(s)$  time. The complexity of  $\mathcal{A}$  is bounded by  $O(s^2n^2)$ , so the dual graph  $G$  has  $O(s^2n^2)$  nodes and  $O(s^2n^2)$  edges, hence the interior predicate will be evaluated  $O(s^2n^2)$  times. Therefore, BFS on  $G$  can be performed in  $O(s^3n^2)$  time using  $O(s^2n^2)$  space.

Finally, to find the MSRs, the faces of  $\mathcal{A}$  (which have been annotated with depths) can be enumerated in  $O(f)$  time, where  $f$  is the number of faces in the arrangement. Thus, the time required to compute the MSRs from  $n$  distance restraints, each having at most  $s$  possible assignments, is expected  $O(sn \log(sn) + k) + O(s^3n^2) + O(f)$  and is output-sensitive. Since the total complexity of the arrangement is bounded by  $O(s^2n^2)$ ,  $f + k$  is also bounded by  $O(s^2n^2)$ , and therefore we can simplify the total time to expected  $O(s^3n^2)$ . The overall space requirements depend on the size of  $\mathcal{A}$ , which is  $O(s^2n^2)$ , and the size of  $G$ , which is  $O(s^2n^2)$ . Therefore, the total space required is  $O(s^2n^2)$  as well.  $\square$

Next, we will use biophysical facts to place bounds on the number of possible assignments for an inter-subunit distance restraint and simplify the complexity bounds.

**Lemma 2.** *For an oligomeric protein complex with cyclic symmetry and  $n$  distance restraints assigned with subunit and/or atom ambiguity, the MSRs can be computed in expected  $O(n^2)$  time and  $O(n^2)$  space.*

**Proof.** For subunit ambiguity, the number of possible assignments is bounded by the number of subunits in the complex. In proteins for *E. coli*, only 2.2% of proteins annotated with subunit designations are composed of more than 12 subunits (Goodsell and Olson, 2000). Therefore, we assume the oligomeric number of protein complexes is bounded by a constant. For atom ambiguity, the number of possible assignments is bounded by the spectral overlap of neighboring resonances and peaks in NMR spectra. In practice, for proteins of up to around 200 residues, 3D NOESY experiments (Marion et al., 1989) are



sufficient to limit spectral overlap to a constant amount per peak. For larger proteins, 4D NOESY (Kay et al., 1990), which uses an extra dimension to resolve cross peaks (similar in ways to a lifting transform), may be required to limit spectral overlap to a constant amount. Even higher-dimensional NMR experiments are possible (Kim and Szyperski, 2003). Since the oligomeric number of protein complexes and the amount of spectral overlap per peak can be bounded by a constant, the number of possible assignments for a distance restraint assigned with subunit and/or atom ambiguity is also bounded by a constant. Consequently, each union of annuli has a constant number of annuli. Therefore,  $s$  is  $O(1)$ , and the bounds of Lemma 1 simplify to expected  $O(n^2)$  time and  $O(n^2)$  space.  $\square$

### 2.5. Evaluation of oligomer structures

Once MSR’s have been computed, DISCO evaluates the distance restraints using the continuous set of oligomer models described by the MSR’s. We characterize a distance restraint as *inconsistent* if its corresponding union of annuli does not contain any of the MSR’s. No oriented oligomer structure whose symmetry axis position was chosen from an MSR could satisfy an inconsistent restraint.

Since the MSR’s computed by DISCO represent continuous sets of symmetry axis positions, the corresponding oligomer structures are also continuous sets. To perform detailed structural analysis and for visualization, the MSR’s are sampled on a uniform grid at a fine resolution to generate a discrete set of symmetry axis positions. One of the advantages of DISCO is that, by computing the exact MSR’s, 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. DISCO combines the sampled axis positions with the symmetry axis orientation computed from the RDC’s to define a set of rigid transformations that, when applied to the subunit structure, generate symmetric oligomer structures (Fig. 2, step 3). Each resulting structure is energy-minimized in XPLOR-NIH (Schwieters et al., 2003) using a fixed backbone, but flexible side chains to relieve minor steric clashes, and to remove over-packed structures.

### 2.6. Extensions to DISCO

One restriction of DISCO as described above is that distance restraints must not have possible intra-subunit assignments. Since PRE’s have potential intra-subunit assignments as well as inter-subunit assignments, if the true assignment for a PRE is intra-subunit, then the annulus analysis presented in Section 2.2 will yield a decoy union of annuli. This union of annuli does not truly constrain the symmetry axis since an intra-subunit distance restraint cannot possibly describe the symmetry of the oligomer structure. One might hope to resolve the intra-/inter-subunit assignment ambiguity directly, but no experimental or computational methods are currently known to perform such an assignment for PRE’s. However, DISCO’s restriction can be relaxed if a set of distance restraints with no possible intra-subunit assignments are also available.

We therefore divide the available distance restraints into two classes: distance restraints with no possible intra-subunit assignments are considered *trusted*, and the remaining distance restraints are considered *untrusted*, since their analysis may yield decoy unions of annuli. DISCO processes the trusted and untrusted distance restraints in two different phases of the algorithm. Phase one uses only the trusted restraints to compute MSR’s (see Section 2.2), which we refer to as *trusted MSR’s*. In phase two, DISCO computes unions of annuli for the untrusted distance restraints, but does not immediately compute their arrangement. Instead, DISCO compares each of the untrusted unions of annuli to the trusted MSR’s. If an untrusted union of annuli does not intersect the trusted MSR’s, that union of annuli is discarded. The remaining unions of annuli that intersect the trusted MSR’s are used along with the original trusted unions of annuli to compute a new arrangement, from which the *final MSR’s* are selected. The final MSR’s represent oligomer structures that are guaranteed to satisfy the trusted distance restraints, and also a subset of the untrusted distance restraints. This two-phase approach ensures all distance restraints contribute to the structure determination (despite some restraints having possible intra-subunit assignments), while avoiding the need to choose explicit intra-/inter-subunit assignments.

## 3. RESULTS

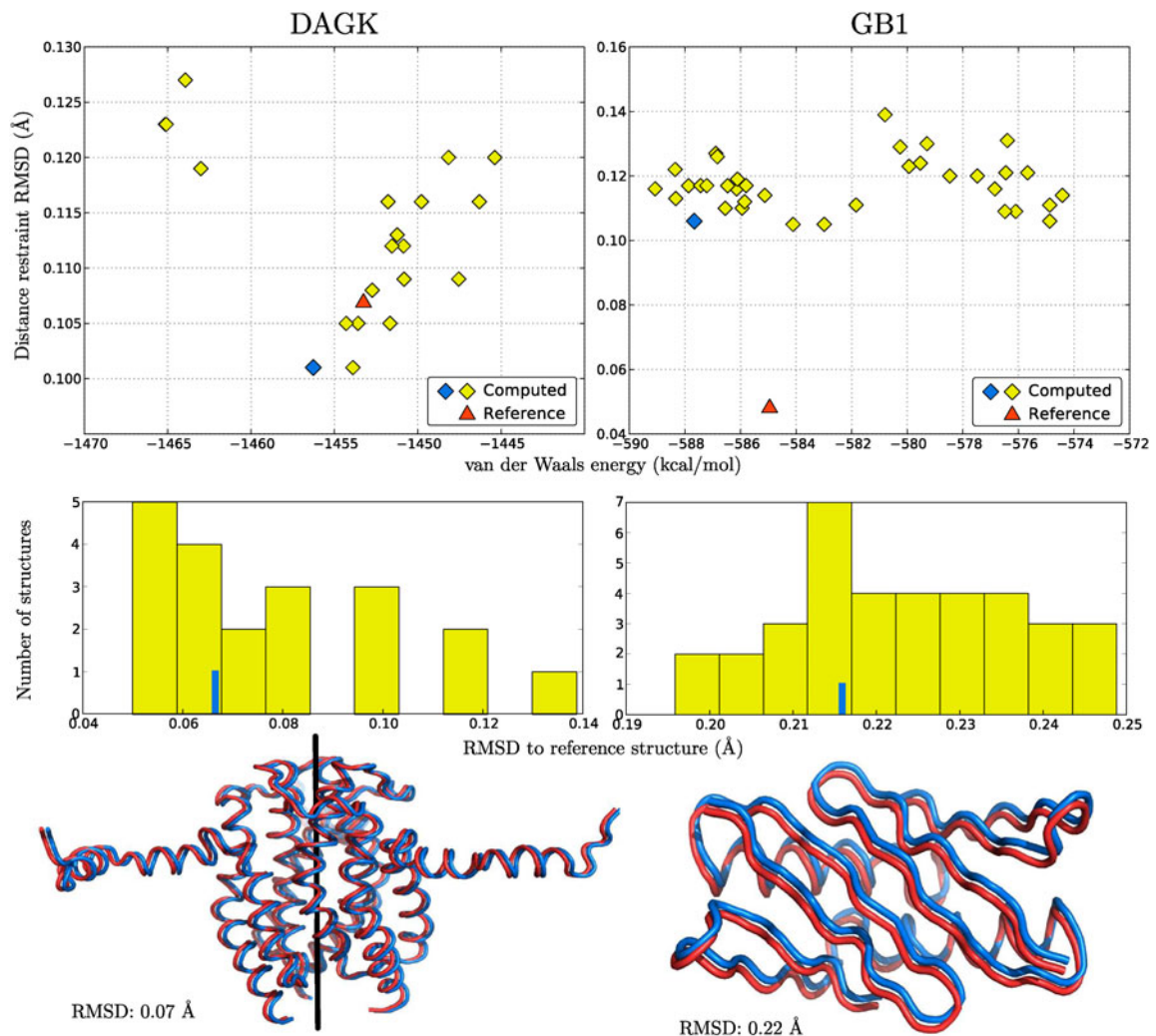
We evaluated the performance of DISCO on two proteins: DAGK (Van Horn et al., 2009) and a dimeric mutant of GB1 (Byeon et al., 2003), henceforth referred to simply as GB1. Due to the difficulty of solving



symmetric protein structures, particularly structures of membrane proteins such as DAGK, there are few experimentally collected NMR datasets with both RDCs and inter-subunit distance restraints available for testing new methodology. DAGK and GB1 are merely two examples, but cover a range of different scenarios, and are good representative test cases.

We compared structures computed by DISCO to known structures (i.e., *reference structures*) from the PDB (Berman et al., 2000) for DAGK (2kdc, model 1) and for GB1 (1q10, model 1). The subunit structure used by DISCO was the first subunit in the reference structure, which was determined using traditional protocols. This mirrors the experimental situation where the subunit structure can be determined with confidence (Oxenoid and Chou, 2005; Schnell and Chou, 2008; Wang et al., 2009), but the main bottleneck is subunit assignment and the assembly of subunit structures to form the oligomer structure.

We measured the structural similarity between structures computed by DISCO and the reference structures using the RMS deviation in backbone atom position. All computed structures were within 0.14 Å of the reference for DAGK and 0.25 Å for GB1. After energy-minimization, we evaluated the RMS distance restraint violation and van der Waals energy (using the pairwise Lennard-Jones potential) of each oligomer structure according to previous methodology (Potluri et al., 2007). Figure 5 shows the scores of the



**FIG. 5.** (Top) Distance restraint satisfaction scores (lower is better) and van der Waals energies for structures computed by DISCO in comparison to the reference structures. (Middle) Histogram (yellow bars) of backbone RMSDs of computed structures to the reference structures. The blue line represents the backbone RMSD of the blue diamond structure from the top row. (Bottom) Backbone alignment between the blue diamond structure from the top row (blue) and the reference structure (red). For this illustration, the two structures are offset from each other by 1.0 Å so they appear distinct. For DAGK, the symmetry axis is represented using a black line. For GB1, the symmetry axis is normal to the page.

computed structures, which are comparable to those of the reference structures. Since DISCO can compute the MSRs exactly and discrete structures are sampled uniformly from the MSRs, the variance in backbone atom position of the computed structural ensemble accurately represents uncertainty about the position of the symmetry axis inherent in the distance restraints. Statistics of the computed ensembles, including backbone RMSDs and the variance, are summarized in Table 1. Sections 3.1 and 3.2 describe in more detail the results for DAGK and GB1, respectively.

### 3.1. DAGK with subunit ambiguity

DAGK is a  $C_3$  homo-trimeric membrane protein with 121 residues per subunit for which 67 NH RDCs, 200 PREs, and 24 disulfide bonds have been recorded (Van Horn et al., 2009). The PREs and disulfide bonds are inter-subunit distance restraints whose assignments are complicated by subunit ambiguity and therefore have two possible assignments each. Additionally, it was not known whether the two PRE-related atoms were in the same subunit, or different subunits. Therefore, we used the two-phase extension to DISCO, (Sec. 2.6), labeling the disulfide bonds as trusted, and the PREs as untrusted.

The alignment tensor computed for DAGK had a rhombicity of 0.019, indicating the RDCs display the symmetry of the oligomer structure. During phase one, DISCO computed the arrangement of unions of annuli from the trusted distance restraints derived from disulfide bonds. DISCO allows sidechains to move during energy-minimization, but uses the rigid subunit structure to compute the annuli from the distance restraints. To account for motions of the sidechains during minimization that could potentially relieve violated distance restraints, we slightly increased the distances allowed by each restraint. We chose a *padding percentage*  $\beta$  such that the lower distance bound of each restraint is multiplied by  $(1 - \frac{\beta}{100})$  and the upper by  $(1 + \frac{\beta}{100})$ . For these tests with DAGK, we chose  $\beta = 3$ .

Figure 6A shows the arrangement and MSRs from phase one. The symmetry axis position of the reference structure is contained within the trusted MSR indicating the annulus analysis is able to correctly describe the satisfying symmetry axis positions of the oligomer structure for DAGK. Since DISCO computes the MSRs exactly (and thus, the set of satisfying oriented oligomer structures exactly), the absence of any additional MSRs farther away rules out the possibility of a satisfying oligomer structure that is dissimilar to those already discovered by the algorithm. The MSRs for DAGK are sensitive to the padding percentage  $\beta$  chosen. With  $\beta = 5$ , oligomer structures sampled from the final MSRs differed from the reference structure by as much as 2.7 Å backbone RMSD, but had a distance restraint RMSD of no worse than 0.38 Å. Section 3.3 further discusses the effect of different choices of  $\beta$  on structure calculation.

The single trusted MSR computed satisfied 21 of the 24 disulfide bond restraints. The remaining three disulfide bond restraints were labeled inconsistent by DISCO, each resulting in small violations in the oligomer structures (Table 2). For comparison, the same three disulfide bond restraints were also unsatisfied in the reference structure. In general, an inter-subunit distance restraint with two possible subunit assignments will result in two distinct (but possibly overlapping) annuli, which are later combined into a union of annuli. However, in the case when the same atoms in the same residues, but in two different subunits are restrained, the union of annuli will contain two identical annuli. We will refer to such a

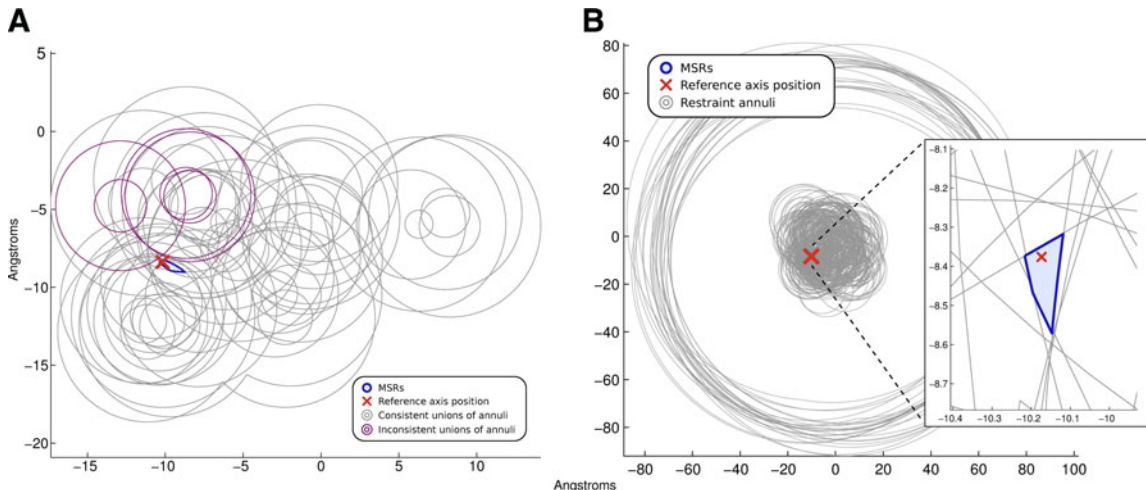
TABLE 1. STATISTICS OF STRUCTURE DETERMINATION FOR DAGK AND GB1

	DAGK	GB1
Symmetry	$C_3$	$C_2$
Rhombicity $R \in [0, \frac{2}{3}]$	0.019	0.486
Orientation difference <sup>1</sup>	0.16°	0.66°
MSR sample resolution <sup>2</sup>	0.025 Å	0.005 Å
Computed ensemble size	20	36
Average all-atom variance	$9.2 \times 10^{-3} \text{ \AA}^2$	$2.7 \times 10^{-2} \text{ \AA}^2$
Average backbone variance	$5.8 \times 10^{-3} \text{ \AA}^2$	$7.9 \times 10^{-4} \text{ \AA}^2$
Backbone atom RMSD <sup>3</sup>	0.05–0.14 Å	0.20–0.25 Å

<sup>1</sup>Difference in orientation between computed and reference symmetry axes.

<sup>2</sup>Grid resolution at which symmetry axis positions were sampled from MSRs.

<sup>3</sup>Range of RMSDs in the computed ensemble versus reference.



**FIG. 6.** (A) Unions of annuli from 24 disulfide bonds for DAGK, three of which are inconsistent. (B) Unions of annuli from 24 disulfide bonds and 154 PREs. The outer ring represents outer boundaries for the 27 PREs with large upper distances ( $\sim 125$  Å), for which only the lower bounds were meaningful. (Inset) Close-up of MSR and reference axis position. The symmetry axis position configuration space is  $\mathbb{R}^2$ , so the units for the plots in Figures 6, 7, 8, and 10 are Angstroms on the  $x$  and  $y$  axes.

distance restraint as *singular*. Of the 24 disulfide bonds for DAGK, eight are singular. Of the eight singular restraints, four disulfide bonds restrain pairs of  $\alpha$  carbons and four disulfide bonds restrain pairs of  $\beta$  carbons. The three inconsistent distance restraints selected by DISCO are all singular disulfide bonds restraining pairs of  $\beta$  carbons (Fig. 6A).

Phase two of DISCO discarded 46 of the 200 PREs (also padded by  $\beta = 3$ ) since their unions of annuli did not intersect the trusted MSR. The remaining 154 untrusted PREs were combined with the original 24 trusted disulfide bonds to compute a new arrangement and the final MSRs which are shown in Figure 6B. As with the trusted MSR, DISCO also computed a single final MSR which again contains the symmetry axis position of the reference structure.

If, by chance, decoy unions of annuli (representing distance restraints whose possible inter-subunit assignments are all incorrect) intersect the trusted MSRs, they will not be discarded and will still influence the computation of the final MSRs. If a large enough number of decoy unions of annuli remain, it is theoretically possible to compute final MSRs that are disjoint from the trusted MSRs, and hence the final MSRs will not correctly describe the oligomer structure of the protein. Although this is unlikely to occur by random chance, having a much greater number of untrusted restraints than trusted restraints increases the likelihood of computing incorrect MSRs. Even though DISCO was supplied with 200 PREs (untrusted) and only 24 disulfide bonds (trusted), the presence of the final MSRs within the trusted MSRs, and the presence of the reference axis position within the final MSRs, shows that the two-phase analysis is able to remove

TABLE 2. DISTANCE RESTRAINT VIOLATIONS FOR DAGK

Violation <sup>1</sup>	Annulus distance <sup>2</sup>	Assignment <sup>3</sup>	Minimum distance <sup>4</sup>	Maximum distance <sup>5</sup>
0.39	0.22	Ile75:C <sup><math>\beta</math></sup> – Ile75:C <sup><math>\beta</math></sup>	2.91	7.21
0.40	0.23	Ala74:C <sup><math>\beta</math></sup> – Ala74:C <sup><math>\beta</math></sup>	2.91	7.21
0.60	0.34	Ala52:C <sup><math>\beta</math></sup> – Ala52:C <sup><math>\beta</math></sup>	2.91	7.21

<sup>1</sup>Distance in Angstroms between the two atoms. Since the MSR represents a set of structures, the violation is computed using the single oligomer structure from the MSR that minimizes the violation.

<sup>2</sup>Minimum distance in Angstroms between the MSR and the annulus computed from the disulfide bond.

<sup>3</sup>Assignments are shown as a pair of atoms, each in the following format: residue type and number, atom name. For these three distance restraints, the same C <sup>$\beta$</sup>  atoms are restrained on both ends of each distance restraint, but the two atoms lie in different subunits.

<sup>4,5</sup>Distances are shown in Angstroms and include a padding of 3%.

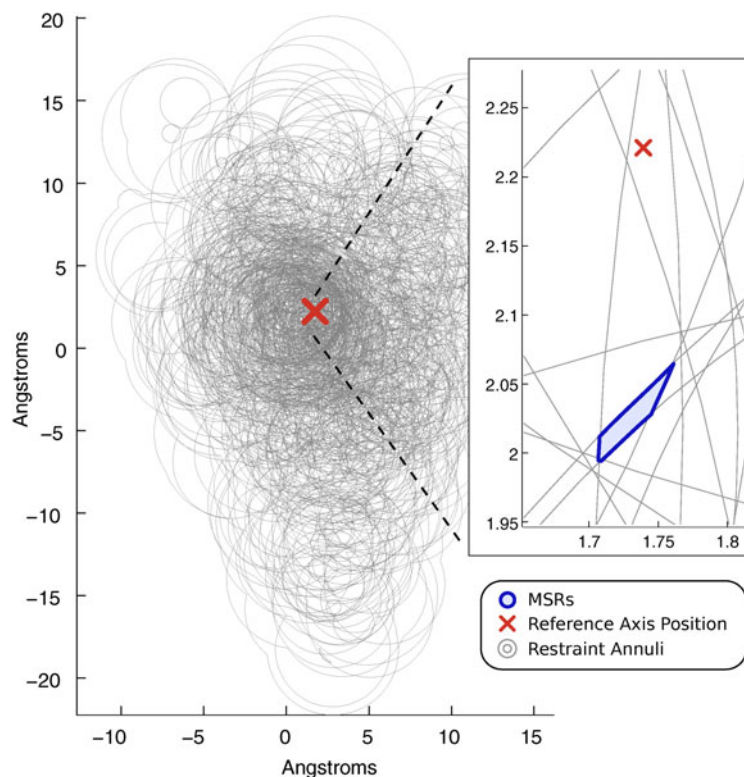
enough decoy unions of annuli to prevent the decoys from conspiring to increase support for an incorrect answer.

### 3.2. GB1 with simulated atom ambiguity

GB1 is a  $C_2$  homo-dimer with 56 residues per subunit for which 56 NH RDCs and 296 experimental inter-subunit NOEs (assigned without subunit ambiguity, since GB1 is a dimer) have been recorded (Byeon et al., 2003). In lieu of subunit ambiguity, we simulated atom ambiguity by expanding the published NOE assignments to include nuclei with similar chemical shifts, resulting in an average of 6.7 possible atom assignments per restraint. Window sizes of 0.05 ppm and 0.5 ppm were used for  $^1\text{H}$  and  $^{13}\text{C}/^{15}\text{N}$  shifts, respectively. 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 recorded in a single alignment medium alone.

A search over the three alignment tensor eigenvectors revealed that MSRs computed from the  $D_{xx}$  eigenvector resulted in the greatest distance restraint satisfaction. Figure 7 shows the single MSR computed for GB1 (with  $\beta = 0$ ) in comparison to the position of the symmetry axis of the reference structure. Even though the minimum distance between the reference symmetry axis position and the MSRs was 0.157 Å, the difference in backbone RMSDs among the structures in the computed ensemble is at least 0.20 Å and not more than 0.25 Å. The ambiguity simulation expanded the average number of assignments per NOE from 1 to 6.7, resulting in 1993 total annuli from the 296 inter-subunit NOEs. Remarkably, 32% of these annuli enclosed no points (i.e., are the empty set), and therefore, no satisfying symmetry axis positions exist, indicating these possible assignments are inconsistent with the computed symmetry axis orientation and ultimately the RDCs. DISCO also found six inconsistent inter-subunit NOEs whose unions of annuli did not intersect the MSR, each resulting in small violations in the oligomer structures (Table 3). The reference structure violates 12 of its inter-subunit NOEs, although each to a lesser degree.

Figure 8 shows a comparison of the two MSRs computed from the original unambiguous NOE assignments (left) and the simulated ambiguous NOE assignments (right). Remarkably, the MSR from the ambiguous assignments is nearly identical to the MSR computed from the unambiguous assignments, but contains an extra region comprising 2.9% of the original total area. The extra region was created when the



**FIG. 7.** Distance restraint unions of annuli for GB1 using 296 NOEs with simulated atom ambiguity. The six inconsistent unions are not labeled. **(Inset)** Close-up of the MSR and reference axis position.

TABLE 3. DISTANCE RESTRAINT VIOLATIONS FOR GB1

<i>Violation</i> <sup>1</sup>	<i>Annulus distance</i> <sup>2</sup>	<i>Assignment</i> <sup>3</sup>	<i>Minimum distance</i> <sup>4</sup>	<i>Maximum distance</i> <sup>5</sup>
0.23	0.12	Asn8:Q <sup>δ</sup> – Thr53:M <sup>γ</sup>	1.8	5.8
0.41	0.24	Lys10:H <sup>α</sup> – Glu56:H	1.8	6.0
0.46	0.29	Ile6:M <sup>γ</sup> – Thr53:H	1.8	6.5
0.52	0.35	Thr55:M <sup>γ</sup> – Asn8:H <sup>α</sup>	1.8	6.5
0.63	0.32	Thr53:M <sup>γ</sup> – Asn8:H <sup>α</sup>	1.8	6.5
0.82	0.43	Thr11:M <sup>γ</sup> – Phe33:Q <sup>β</sup>	1.8	8.5

<sup>1</sup>Distance in Angstroms between the two atoms (or pseudoatoms). Since the MSR represents a set of structures, the violation is computed using the single oligomer structure from the MSR that minimizes the violation.

<sup>2</sup>Minimum distance in Angstroms between the MSR and the annulus computed from the NOE.

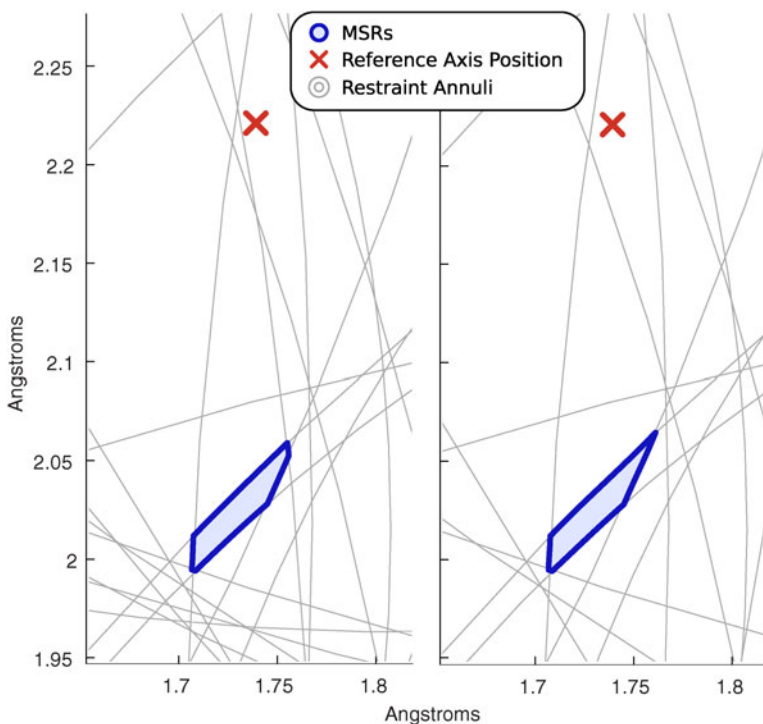
<sup>3</sup>Assignments are shown as a pair of protons (or pseudoatoms), each in the following format: residue type and number, proton name or pseudoatom name. Since the NOEs are inter-subunit, the two protons (or pseudoatoms) for each restraint lie in different subunits.

<sup>4,5</sup>Distances are shown in Angstroms and have been adjusted for pseudoatoms.

additional assignments for an NOE replaced a single annulus with a union of annuli. The boundary that defined that edge of the previous MSR had been removed since it now lies in the interior of the new union of annuli.

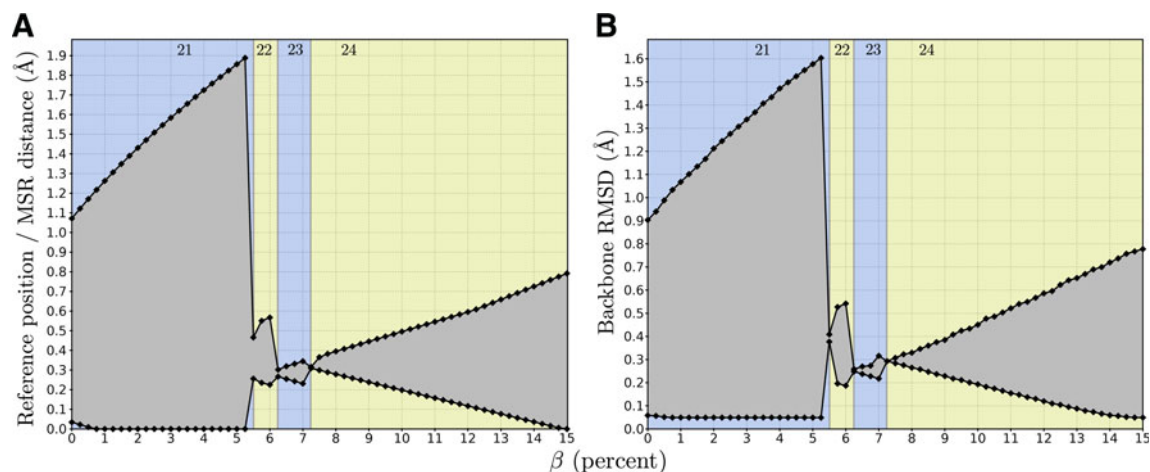
### 3.3. Perturbation analysis of $\beta$

To explore the effect of the parameter  $\beta$  on structure calculation, we computed MSRs from the disulfide bonds for DAGK for values of  $\beta$  varying from 0 to 15 in increments of 0.25. With  $\beta = 0$ , DISCO computed MSRs from the original distance restraints without any padding. For  $\beta > 0$ , the distances were padded as described in Section 3.1. For each of the resulting 61 sets of MSRs, we computed the minimum and maximum distances from the MSRs to the position of the symmetry axis of the reference structure. Figure 9A shows these distance ranges plotted against  $\beta$ . To estimate the quality of oligomer structures represented by these MSRs, we sampled symmetry axis positions from the MSRs (and hence, oligomer structures) at a



**FIG. 8.** A comparison of the MSRs for GB1 computed under two different sets of NOEs. **(Left)** Original unambiguously assigned NOEs. **(Right)** Ambiguously assigned NOEs computed from a simulation resulting in an average of 6.7 possible assignments per NOE.





**FIG. 9.** (A) Range of distances between the MSRs and the reference symmetry axis position for varying values of  $\beta$ . The top series shows the maximum distances, and the bottom series shows the minimum distances. The four yellow and blue regions show the intervals of  $\beta$  sharing the same number of satisfied disulfide bond distance restraints. The number of satisfied distance restraints is shown at the top of the interval. (B) Range of backbone RMSDs between the reference structure and oligomer structures sampled very finely from the MSRs.

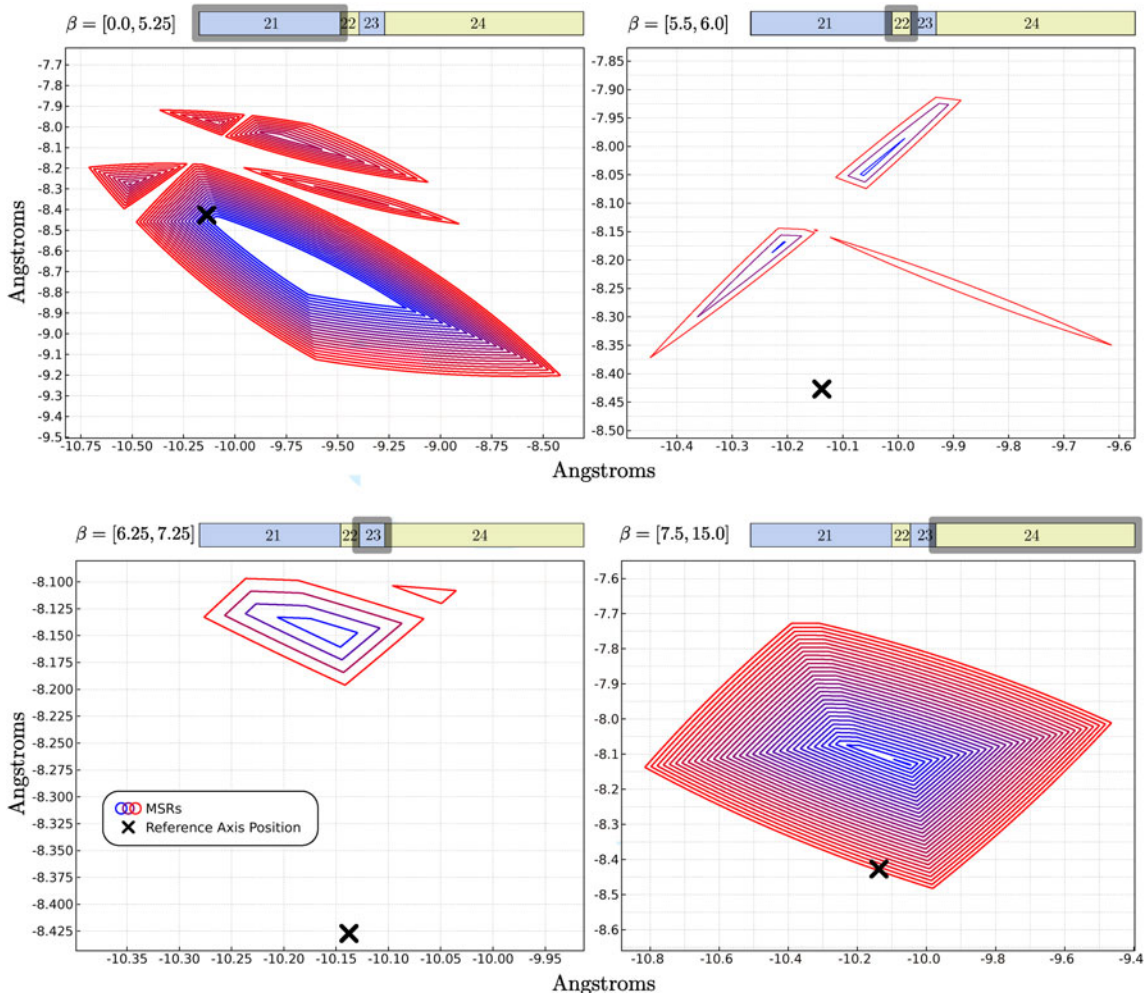
very fine resolution (0.0125 Å) and computed their backbone RMSDs to the reference structure for DAGK. The resulting ranges of backbone RMSDs are shown in Figure 9B. The ranges of reference axis position/MSR distances and the ranges of backbone RMSDs closely resemble each other, indicating that DISCO’s geometric analysis is able to accurately represent differences in oligomer structures using differences in the symmetry parameters. Interestingly, even though the MSRs for  $\beta \in [1, 5]$  allow for sampling arbitrarily close to the reference symmetry axis position, the minimum achievable backbone RMSD was 0.0490 Å. Since oligomer structures computed by DISCO are symmetric by construction, comparisons with the reference structure for DAGK (which has slight deviations from perfect symmetry) will not yield perfect matches.

In general, the size of the ranges in Figure 9 increase with  $\beta$ , but there are three interesting exceptions. With  $\beta = 0$ , 21/24 of the disulfide bond distance restraints are satisfied by the MSRs. The MSRs that are computed when  $\beta = 5.5$  are able to satisfy an additional distance restraint, increasing the count to 22. The number of satisfied distance restraints increases again at  $\beta = 6.25$ , and once more at  $\beta = 7.5$ , where all 24 distance restraints are satisfied by the MSRs. These three values of  $\beta$ , where the number of satisfied distance restraints increases, define four  $\beta$  intervals over which the number of satisfied distance restraints remains constant. In the four different  $\beta$  intervals, the geometry of the MSRs (Fig. 10) is markedly different.

As can be seen from Figure 10, the MSRs grow regularly in size in the first  $\beta$  interval of  $[0, 5.25]$ , since the outer radii of the distance restraint annuli also grow regularly with  $\beta$ . However, at  $\beta = 5.5$ , the MSRs “jump” to a new position, since the arrangement now defines a deeper face corresponding to the satisfaction of the additional disulfide bond distance restraint. As  $\beta$  increases over the next  $\beta$  interval of  $[5.5, 6.0]$ , the MSRs grow regularly again until  $\beta = 6.25$ , where the satisfaction of an additional distance restraint causes another “jump.” This grow-then-jump pattern continues until all distance restraints are satisfied. Afterwards, the MSRs simply grow regularly with  $\beta$  since there are no more distance restraints to satisfy.

#### 4. CONCLUSION

DISCO can accurately determine the oligomer structures of proteins with  $C_n$  symmetry using RDCs and distance restraints. DISCO analyzes inter-subunit distance restraints even when they are assigned ambiguously, but avoids enumerating explicit assignment combinations. The geometric annulus analysis of the distance restraints can discriminate between consistent and inconsistent distance restraints using the MSRs.



**FIG. 10.** An overlay of the MSRs computed from the disulfide bonds for DAGK for varying values of  $\beta$ . Each of the four  $\beta$  intervals is shown as a separate plot. The  $\beta$  interval itself is shown at the top of each plot using the yellow/blue colors from Figure 9. Within each plot, the reference symmetry axis position is marked with a black X, and the MSRs for all the  $\beta$  values of the interval are shown together as curves using a gradient of colors, from blue to red as  $\beta$  increases.

Additionally, the MSRs are computed exactly and in expected  $O(n^2)$  time, thus ensuring no satisfying oriented oligomer structures are missed by the algorithm. Using the two-phase protocol, DISCO incorporates structural constraint provided by distance restraints with possible intra-subunit assignments in phase two. As a prerequisite, phase one of the protocol requires restraints with strictly inter-subunit assignments.

For trimers and higher-order oligomers, the  $D_{zz}$  eigenvector of the alignment tensor is parallel to the axis of symmetry, but for dimers, a search must be conducted over the possible eigenvectors—three in the non-degenerate case. In practice, the chosen eigenvector of the alignment tensor computed from the subunit structure and the RDC values may differ slightly from the orientation of the true symmetry axis of the oligomer complex. To account for uncertainty in the symmetry axis orientation (possibly due to protein dynamics or experimental uncertainty), one can estimate the distribution of symmetry axis orientations described by the RDCs by considering the uncertainty of each experimental RDC value. Symmetry axis orientations sampled from this distribution can be analyzed by DISCO to select for the orientations whose resulting oligomer structures best satisfy the distance restraints (Martin et al., 2011).

Since DISCO computes the exact set of oriented oligomer structures that satisfy the distance restraints, the variance in atom position of the computed ensemble of structures (Table 1) yields a meaningful measure of the range of oligomer structures allowed by the distance restraints, whereas in methods that rely on stochastic search, the variance is merely an artifact of the sampling. The entire protocol DISCO has been



completely automated in a software package that is freely available and open-source at [www.cs.duke.edu/donaldlab/software.php](http://www.cs.duke.edu/donaldlab/software.php).

## ACKNOWLEDGMENTS

We would like to thank the members of the Donald lab for their helpful comments and suggestions, and Dan Halperin for insightful discussions about geometry. This work was supported by the National Institutes of Health (grants R01 GM-65982 to B.R.D. and R01 GM-079376 to P.Z.).

## DISCLOSURE STATEMENT

No competing financial interests exist.

## 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.
- Bardiaux, B., Bernard, A., Rieping, W., et al. 2009. Influence of different assignment conditions on the determination of symmetric homodimeric structures with ARIA. *Proteins* 75, 569–585.
- Berman, H.M., Westbrook, J., Feng, Z., et al. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242.
- 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.
- Byeon, I.-J.L., Louis, J.M., and Gronenborn, A.M. 2003. A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *J. Mol. Biol.* 333, 141–152.
- Chen, C.-Y., Georgiev, I., Anderson, A.C., et al. 2009. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. USA* 106, 3764–3769.
- Donald, B.R., and Martin, J. 2009. Automated NMR assignment and protein structure determination using sparse dipolar coupling constraints. *Prog. Nucl. Magn. Reson. Spectrosc.* 55, 101–127.
- Frey, K.M., Georgiev, I., Donald, B.R., et al. 2010. Predicting resistance mutations using protein design algorithms. *Proc. Natl. Acad. Sci. USA* 107, 13707–13712.
- Goodsell, D.S., and Olson, A.J. 2000. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* 29, 105–105.
- Halperin, D. 1997. Arrangements, 529–562. In Goodman, J.E., and O’Rourke, J., eds. *Handbook of Discrete and Computational Geometry*, 2nd ed. CRC Press, Boca Raton, FL.
- Hanniel, I., and Halperin, D. 2000. Two-dimensional arrangements in CGAL and adaptive point location for parametric curves. *Lect. Notes Comput. Sci.* 1982, 171–182.
- Headd, J.J., Ban, Y.E.A., Brown, P., et al. 2007. Protein–protein interfaces: properties, preferences, and projections. *J. Proteome Res.* 6, 2576–2586.
- 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.
- Kay, L.E., Clore, G.M., Bax, A., et al. 1990. Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. *Science* 249, 411–414.
- Kim, S., and Szyperski, T. 2003. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J. Am. Chem. Soc.* 125, 1385–1393.
- Kovacs, H., O’Donoghue, S.I., Hoppe, H.-J., et al. 2002. Solution structure of the coiled-coil trimerization domain from lung surfactant protein D. *J. Biomol. NMR* 24, 89–102.
- Kuzin, A., Abashidze, M., Forouhar, F., et al. 2005. Novel x-ray structure of the YkuJ protein from *Bacillus subtilis*. Northeast Structural Genomics target SR360. Northeast Structural Genomics Consortium.
- Levy, E.D., Erba, E.B., Robinson, C.V., et al. 2008. Assembly reflects evolution of protein complexes. *Nature* 453, 1262–1265.
- Liang, J., Edelsbrunner, H., Fu, P., et al. 1998a. Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins* 33, 1–17.
- Liang, J., Woodward, C., and Edelsbrunner, H. 1998b. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* 7, 1884–1897.

- Losonczi, J.A., Andrec, M., Fischer, M.W.F., et al. 1999. Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Reson.* 138, 334–342.
- Marion, D., Kay, L.E., Sparks, S.W., et al. 1989. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J. Am. Chem. Soc.* 111, 1515–1517.
- Martin, J.W., Yan, A.K., Bailey-Kellogg, C., et al. 2011. A graphical method for analyzing distance restraints using residual dipolar couplings for structure determination of symmetric protein homo-oligomers. *Protein Sci.* 20, 970–985.
- Nilges, M. 1993. A calculation strategy for the structure determination of symmetric dimers by  $^1\text{H}$  NMR. *Proteins* 17, 297–309.
- O'Donoghue, S.I., Chang, X., Abseher, R., et al. 2000. Unraveling the symmetry ambiguity in a hexamer: calculation of the R6 human insulin structure. *J. Biomol. NMR* 16, 93–108.
- Oxenoid, K., and Chou, J.J. 2005. The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc. Natl. Acad. Sci. USA* 102, 10870–10875.
- Potluri, S., Yan, A.K., Chou, J.J., et al. 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.
- Potluri, S., Yan, A.K., Donald, B.R., et al. 2007. A complete algorithm to resolve ambiguity for intersubunit NOE assignment in structure determination of symmetric homo-oligomers. *Protein Sci.* 16, 69–81.
- Saxe, J. 1979. Embeddability of weighted graphs in k-space is strongly NP-hard. *Proc. 17th Allerton Conf. Commun. Control Comput.* 480–489.
- Schnell, J.R., and Chou, J.J. 2008. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 451, 591–595.
- Schwieters, C.D., Kuszewski, J.J., Tjandra, N., et al. 2003. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* 160, 65–73.
- Van Horn, W.D., Kim, H.-J., Ellis, C.D., et al. 2009. Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase. *Science* 324, 1726–1729.
- Wang, C.-S.E., Lozano-Pérez, T., and Tidor, B. 1998. AmbiPack: a systematic algorithm for packing of macromolecular structures with ambiguous distance constraints. *Proteins* 32, 26–42.
- Wang, J., Pielak, R.M., McClintock, M.A., et al. 2009. Solution structure and functional analysis of the influenza B proton channel. *Nat. Struct. Mol. Biol.* 16, 1267–1271.
- Wang, L., Mettu, R.R., and Donald, B.R. 2006. A polynomial-time algorithm for de novo protein backbone structure determination from nuclear magnetic resonance data. *J. Comput. Biol.* 13, 1267–1288.
- Wang, X., Bansal, S., Jiang, M., et al. 2008. RDC-assisted modeling of symmetric protein homo-oligomers. *Protein Sci.* 17, 899–907.
- White, S.H. 2004. The progress of membrane protein structure determination. *Protein Sci.* 13, 1948–1949.

Address correspondence to:  
Dr. Bruce R. Donald  
Department of Computer Science  
Duke University  
Durham, NC 27708

E-mail: brd+jcb11@cs.duke.edu

