

# Structure Determination of Symmetric Homo-Oligomers by a Complete Search of Symmetry Configuration Space, Using NMR Restraints and van der Waals Packing

Shobha Potluri,<sup>1</sup> Anthony K. Yan,<sup>1</sup> James J. Chou,<sup>2</sup> Bruce R. Donald,<sup>1,3,4\*</sup> and Chris Bailey-Kellogg<sup>1\*</sup>

<sup>1</sup>Department of Computer Science, Dartmouth College, Hanover, New Hampshire 03755

<sup>2</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115

<sup>3</sup>Department of Chemistry, Dartmouth College, Hanover, New Hampshire 03755

<sup>4</sup>Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03755

**ABSTRACT** Structural studies of symmetric homo-oligomers provide mechanistic insights into their roles in essential biological processes, including cell signaling and cellular regulation. This paper presents a novel algorithm for homo-oligomeric structure determination, given the subunit structure, that is both complete, in that it evaluates all possible conformations, and data-driven, in that it evaluates conformations separately for consistency with experimental data and for quality of packing. Completeness ensures that the algorithm does not miss the native conformation, and being data-driven enables it to assess the structural precision possible from data alone. Our algorithm performs a branch-and-bound search in the symmetry configuration space, the space of symmetry axis parameters (positions and orientations) defining all possible  $C_n$  homo-oligomeric complexes for a given subunit structure. It eliminates those symmetry axes inconsistent with intersubunit nuclear Overhauser effect (NOE) distance restraints and then identifies conformations representing any consistent, well-packed structure to within a user-defined similarity level.

For the human phospholamban pentamer in dodecylphosphocholine micelles, using the structure of one subunit determined from a subset of the experimental NMR data, our algorithm identifies a diverse set of complex structures consistent with the nine intersubunit NOE restraints. The distribution of determined structures provides an objective characterization of structural uncertainty: backbone RMSD to the previously determined structure ranges from 1.07 to 8.85 Å, and variance in backbone atomic coordinates is an average of 12.32 Å<sup>2</sup>. Incorporating vdW packing reduces structural diversity to a maximum backbone RMSD of 6.24 Å and an average backbone variance of 6.80 Å<sup>2</sup>. By comparing data consistency and packing quality under different assumptions of oligomeric number, our algorithm identifies the pentamer as the most likely oligomeric state of phospholamban, demonstrating that it is possible to determine the oligomeric number

directly from NMR data. Additional tests on a number of homo-oligomers, from dimer to heptamer, similarly demonstrate the power of our method to provide unbiased determination and evaluation of homo-oligomeric complex structures. *Proteins* 2006;65:203–219. © 2006 Wiley-Liss, Inc.

**Key words:** Nuclear magnetic resonance (NMR); Structure determination; Symmetric homo-oligomer; Membrane protein; Configuration space

## INTRODUCTION

Symmetric homo-oligomers play pivotal roles in complex biological processes, including ion transport and regulation, signal transduction, and transcriptional regulation. We are particularly interested in phospholamban, a symmetric homopentameric membrane protein that regulates the calcium levels between cytoplasm and sarcoplasmic reticulum, and hence aids in muscle contraction and relaxation.<sup>1</sup> Ion conductance studies<sup>2</sup> also suggest that phospholamban might have a separate role as an ion channel. To understand the dual function of phospholamban and other such symmetric homo-oligomers, we have developed a combined experimental–computational ap-

*Abbreviations:* NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RMSD, root-mean-square deviation; SCS, symmetry configuration space;  $S^2$ , space of orientations represented on a two-sphere; vdW, van der Waals; WPS, well-packed satisfying.

The Supplementary Material referred to in this article can be found at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/>

Grant sponsor: National Institutes of Health; Grant number: R01 GM 65982; Grant sponsor: National Science Foundation; Grant numbers: EIA-9802068, EIA-0305444; Grant sponsor: National Science Foundation; Grant numbers: IIS-0444544, IIS-0502801. Grant sponsor: Smith Family award.

\*Correspondence to: Bruce R. Donald, 6211 Sudikoff Laboratory, Hanover, NH 03755. E-mail: brd@cs.dartmouth.edu or Chris Bailey-Kellogg, 6211 Sudikoff Laboratory, Hanover, NH 03755. E-mail: cbk@cs.dartmouth.edu

Received 22 December 2005; Revised 2 May 2006; Accepted 15 May 2006

Published online 8 August 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21091

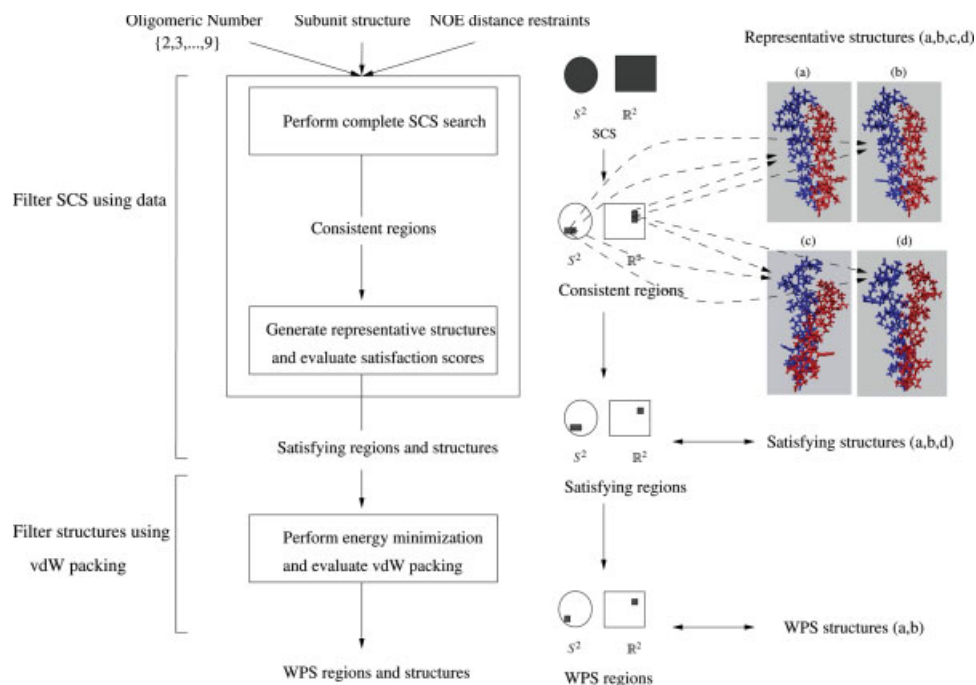


Fig. 1. The overall protocol for determining structures of symmetric homo-oligomers. The input to the protocol includes the subunit structure, the NOE restraints, and the oligomeric number. The first phase of our two-phase approach involves a complete data-driven search using a branch-and-bound algorithm in the SCS, the space of symmetry axis parameters. The 4D SCS is represented as two 2D regions: a sphere representing the orientation space  $S^2$  and a square representing the translation space  $R^2$ . The output from the branch-and-bound algorithm is a set of consistent regions in SCS that represent all structures satisfying the restraints. Representative structures are chosen from the consistent regions. These structures are evaluated for quality of restraint satisfaction, and those that are good enough are identified as satisfying structures. Regions corresponding to the satisfying structures in the consistent regions form the satisfying regions (represented by the dark regions). In the second phase, each of the satisfying structures is energy-minimized and evaluated for the quality of vdW packing. The output from this step is a set of WPS structures and WPS regions that represent conformations that are consistent with data and have high-quality vdW packing. The entire protocol is repeated for different oligomeric numbers, and the WPS structures of each are evaluated, in order to identify the most probable oligomeric number. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

proach that, given the subunit structure, determines the complex structure from a combination of sparse intersubunit nuclear Overhauser effect (NOE) distance restraints and van der Waals (vdW) packing. Our approach is complete, in that we test all possible conformations, and is data-driven, in that we first test conformations for consistency with data and only then evaluate each of the consistent conformations for vdW packing. Completeness in a structure determination method is a key requirement, since it ensures that no conformation consistent with the data is missed. This avoids any bias in the search, as well as any potential for becoming trapped in local minima, both of which are problems inherent in energy minimization-based approaches. The data-driven nature of our method allows us to independently quantify the amount of structural constraint provided by data alone, versus both data and packing. This avoids over-reliance on subjective choices of parameters for energy minimization, and consequent false precision in determined structures. Our method also allows us to determine the oligomeric number of the complex. To our knowledge, this is the first approach that determines the oligomeric number of a

symmetric homo-oligomer from intersubunit NOE distance restraints. The confidence in the determined oligomeric number depends on the information content in the available data.

The key to our approach is the observation that, given the structure of a subunit, the structure of a symmetric homo-oligomer is completely determined by the parameters (position and orientation) of its symmetry axis. Thus, we can formulate structure determination in terms of a search in the space of all symmetry axis parameters, which we call the symmetry configuration space (SCS). Figure 1 shows our overall protocol, which consists of two phases. In the first phase, we perform a complete, data-driven search in SCS and return consistent regions in SCS—a superset of the regions containing all conformations consistent with the data. We generate representative structures from the consistent regions, such that every structure in the consistent regions is within a user-defined backbone RMSD,  $\tau_0$  Å, to some representative structure. Each of the representative structures is evaluated for restraint satisfaction, and those that are good enough form a set of satisfying structures. In the

second phase, each of the satisfying structures is evaluated for high-quality vdW packing. We ultimately return a set of well-packed satisfying (WPS) structures that are consistent with data and have high-quality vdW packing. We quantify the uncertainty in determined structures in terms of the size of the regions in SCS and the variations in atomic coordinates in the structures. The difference in uncertainty between the satisfying structures and the WPS structures illustrates the relative precision that is possible from data alone, versus data and packing together.

Our approach allows us to test whether the available data suffices to determine the oligomeric number with high confidence. For each possible oligomeric number, we determine a set of WPS structures. We place higher confidence in the oligomeric number that has WPS structure with better vdW packing. Thus we determine the oligomeric number using the NMR data and vdW packing. Our approach provides for an independent verification of the oligomeric state, which is typically determined using experiments, such as chemical cross-linking, followed by SDS-PAGE, or by equilibrium sedimentation.

Traditional protocols (e.g., Refs. 3, 4) for structure determination of protein complexes from NMR data use simulated annealing and molecular dynamics. A variant of this protocol was used to determine the structure of the phospholamban pentamer.<sup>1</sup> However, the simulated annealing/molecular dynamics mechanism is not complete and could get trapped in local minima. Furthermore, the precision in the determined structure is strongly affected by temperature. Our approach avoids temperature dependence and local minima problems, and does not suffer from “false precision” in characterizing the diversity of determined structures.

Previous computational techniques<sup>5</sup> for predicting the structures of complexed proteins fall into two categories: homology modeling and docking. One can confidently predict protein interactions by homology modeling<sup>6,7</sup> only if the structure of a homologous complex is available.<sup>8</sup> Docking strategies usually involve a two-stage approach: generate a set of possible docked structures, and then score them. The possible structures are typically generated by sampling the space of rotations and translations of the docked subunit with respect to the fixed subunit.<sup>9–15</sup> Alternatively, geometric hashing<sup>16</sup> samples conformations that are consistent with shape complementarity. Scoring is usually based on shape and chemical complementarity.<sup>17–20</sup> Docking-based methods are immediately applicable to homo-oligomers, by docking two subunits of the homo-oligomer. Pierce and Weng,<sup>21</sup> Duhovny et al.,<sup>22</sup> and Comeau and Camacho<sup>23</sup> have all used docking-based approaches, followed by filtering based on symmetry and scoring by minimizing energy functions, to predict structures of homo-oligomers.

Several minimalist experiments, though not comprehensive enough to determine structures by themselves, have been used in conjunction with docking approaches to score predicted models.<sup>24–27</sup> Proximity information between adjacent units in a complex can be obtained from

experiments such as mutagenesis,<sup>28–30</sup> hydrogen-deuterium exchange,<sup>26,31</sup> chemical cross-linking,<sup>32–34</sup> fluorescence energy transfer,<sup>35,36</sup> Fourier transform infra-red spectroscopy,<sup>37</sup> and sparse intermolecular NOE data.<sup>38</sup>

Existing docking-based approaches are not complete, because they generate possible docked structures by sampling, and might miss conformations close to the native structure. This is evident from our tests on phospholamban, when we used several available docking-based methods<sup>21–23</sup> to predict the complex structure from the subunit structure. A significant number of the models provided by these approaches satisfy none of the experimental intersubunit NOE restraints determined by Oxenoid and Chou.<sup>1</sup> The backbone RMSD of the model that best satisfies the experimental restraints from each of the approaches in Refs. 21–23 to a reference structure from Ref. 1 are, respectively, 3.25, 2.52, and 10.19 Å (data not shown). The high RMSD value for the last approach is due to a parallel arrangement of the helices in the subunits, whereas phospholamban in reality has a left-handed twist. Our approach identifies structures with a left-handed twist and that have RMSD as low as 1.07 Å to the reference structure.

Docking-based approaches also focus on putting together a single pair of subunits without any consideration of the effect on the other subunits. Our search in the SCS takes advantage of the “closed-ring” constraint of a symmetric homo-oligomer. Wang et al.<sup>38</sup> propose a branch-and-bound algorithm to compute rigid body transformations satisfying potentially ambiguous intersubunit distance restraints. In contrast, our algorithm uses the oligomeric number to enforce an *a priori* symmetry constraint. In this sense, it is analogous to the manner in which noncrystallographic symmetry is handled in molecular replacement for X-ray crystallography.<sup>39</sup> By formulating the structure determination problem in the SCS rather than in the space of atom positions, we are able to exploit directly the kinematics of the “closed-ring” constraint, and thereby derive an analytical bound for pruning, which is tighter and more accurate than previous randomized numerical techniques.<sup>38</sup>

In contrast to previous techniques, our method separately quantifies the amount of structural constraint provided by data alone, versus data and packing. For phospholamban, we show that the average backbone variance in the set of satisfying structures is 12.32 Å<sup>2</sup>, but is reduced to 6.80 Å<sup>2</sup> after incorporating packing. Our approach also allows us to characterize information content as a function of the number of independent restraints. Using simulated restraints for influenza haemagglutinin, we found that, as the number of independent restraints increases from 8 to 64, the satisfying region in the SCS determined by our approach shrinks, and the average variance in atomic positions decreases from 0.9 to 0.2 Å<sup>2</sup>. We also show, by applying our approach to different test cases, that both NOE data and vdW packing are required to determine the oligomeric number.

## METHODS

We determine the 3D structure of symmetric homo-oligomers given a set of NOE restraints, the subunit structure (i.e., the “bound” structure of a monomer within the complex), and the oligomeric number. We assume that the NOE restraints are correct and the backbone symmetry is exact, and thus consider invalid any conformation violating even one restraint. Our approach does allow for asymmetry and flexibility in the side chains. Experimentally, it is possible to determine the bound subunit structure prior to computing the oligomeric assembly. Phospholamban is one example of a system satisfying these assumptions. Nine intersubunit NOE restraints were obtained<sup>1</sup> by using a mixture of labeled and unlabeled subunits and filtering NOE signals appropriately. Only those restraints with no chemical shift degeneracy were used. The subunit structure was determined by a simulated annealing protocol, using intramolecular distance restraints from NOEs, backbone orientation restraints from residual dipolar couplings, and side-chain  $\chi_1/\chi_2$  restraints from three-bond scalar couplings.<sup>1</sup>

As shown in Figure 1, the inputs to our approach are a set of intersubunit NOE restraints, the subunit structure, and the oligomeric number. Given the structure of a subunit, identifying the position ( $\mathbf{t} \in \mathbb{R}^2$ ) and orientation ( $\mathbf{a} \in S^2$ ) of the symmetry axis determines the structure of a homo-oligomer with  $C_n$  symmetry. Geometrically, the axis of symmetry is a line parallel to the unit vector  $\mathbf{a}$ , which intersects the  $xy$ -plane at the point  $\mathbf{t}$ . For a given axis, rotating the subunit  $n$  times by the angle of symmetry ( $360^\circ/n$ ) around the symmetry axis yields the structure. Each possible conformation of the symmetric homo-oligomer is represented by a point in the 4D SCS,  $S^2 \times \mathbb{R}^2$ . Hence, a search in SCS allows us to identify all WPS conformations. Each region in the SCS represents a set of symmetry axes, and each individual axis represents a unique conformation of the homo-oligomer. We perform a complete data-driven search in the 4D space of symmetry axis parameters, and prune out regions representing conformations that are inconsistent with the data. We ultimately return regions in SCS, the consistent regions, which contain all conformations that are consistent with the data. SCS is too large to search naively or exhaustively. Therefore, we have developed a novel branch-and-bound algorithm that is efficient and provably conservative, in that it examines and conservatively eliminates nonsatisfying regions. Without this algorithm, a complete, data-driven search would not be computationally feasible. We next choose representative structures from the consistent regions, such that every conformation in the consistent regions is within an RMSD of  $\tau_0 \text{ \AA}$  to at least one representative structure. Owing to the conservative bounds used in our search, the representative structures might contain conformations that are inconsistent with the data. The set of satisfying structures includes only those representative structures with restraint satisfaction scores below a chosen threshold. We then evaluate each

of the satisfying structures by energy-minimizing and scoring them, based on vdW packing. The set of WPS structures includes those energy-minimized satisfying structures with vdW packing scores below a chosen threshold. The remainder of this section describes each step in detail.

### Complete Search of SCS

We must identify all possible symmetry axis parameters,  $(\mathbf{a}, \mathbf{t}) \in (S^2 \times \mathbb{R}^2)$ , such that corresponding structures satisfy all the restraints. Here,  $\mathbf{a}$  represents the orientation of the axis and  $\mathbf{t}$  represents the relative position between the axis and one of the subunits. An exact algebraic formulation for identifying all possible values of  $(\mathbf{a}, \mathbf{t})$  is possible, but it would yield high-degree polynomials that are expensive to solve exactly. Hence, we develop here a branch-and-bound algorithm that searches the SCS and conservatively eliminates regions that probably cannot satisfy all the restraints. The branch-and-bound approach facilitates a complete search over all possible conformations and ultimately returns consistent regions in SCS.

As Figure 2 shows, the branch-and-bound search follows a tree structure, and performs a recursive search through regions in SCS. Each node in the tree is a SCS grid cell—a 4D hypercuboid defined by values representing extrema along each of the four dimensions. At each node of the branch-and-bound search, we test whether any point in the grid cell represents a consistent conformation. If such a point possibly exists, we branch and partition the cell into smaller subcells. We continue branching, until we can either eliminate or accept each grid cell. We eliminate a cell by computing bounds on the conformations it represents and determining that they all violate at least one restraint or contain significant steric clashes. We conservatively accept a grid cell as part of the consistent regions, when all the structures it represents either provably satisfy all the restraints or are within an RMSD of  $\tau_0 \text{ \AA}$  of each other, and when each restraint is satisfied by at least one conformation represented by the cell.

### Bounding

We evaluate a grid cell for potential steric clash only when the cell is “small” enough ( $\leq 5^\circ$  in  $S^2$  and  $\leq 0.5 \text{ \AA}$  in  $\mathbb{R}^2$ ). A grid cell is pruned only when there is a severe steric clash (see Results section) between atoms in the 16 structures represented by the corners of the 4D grid cell. This soft pruning allows for conformations with a few steric clashes in side chains that can be overcome through energy minimization performed later.

To test whether we can eliminate a grid cell  $G$  due to restraint violation, we independently consider each restraint,  $\|\mathbf{p} - \mathbf{q}'\| \leq d$ , where  $\mathbf{p}$  and  $\mathbf{q}'$  are positions of atoms in adjacent subunits in clockwise order. Let position  $\mathbf{q}$  correspond to  $\mathbf{q}'$  in the same subunit to which  $\mathbf{p}$  belongs. We then compute  $G\mathbf{q}$ , the set of all possible

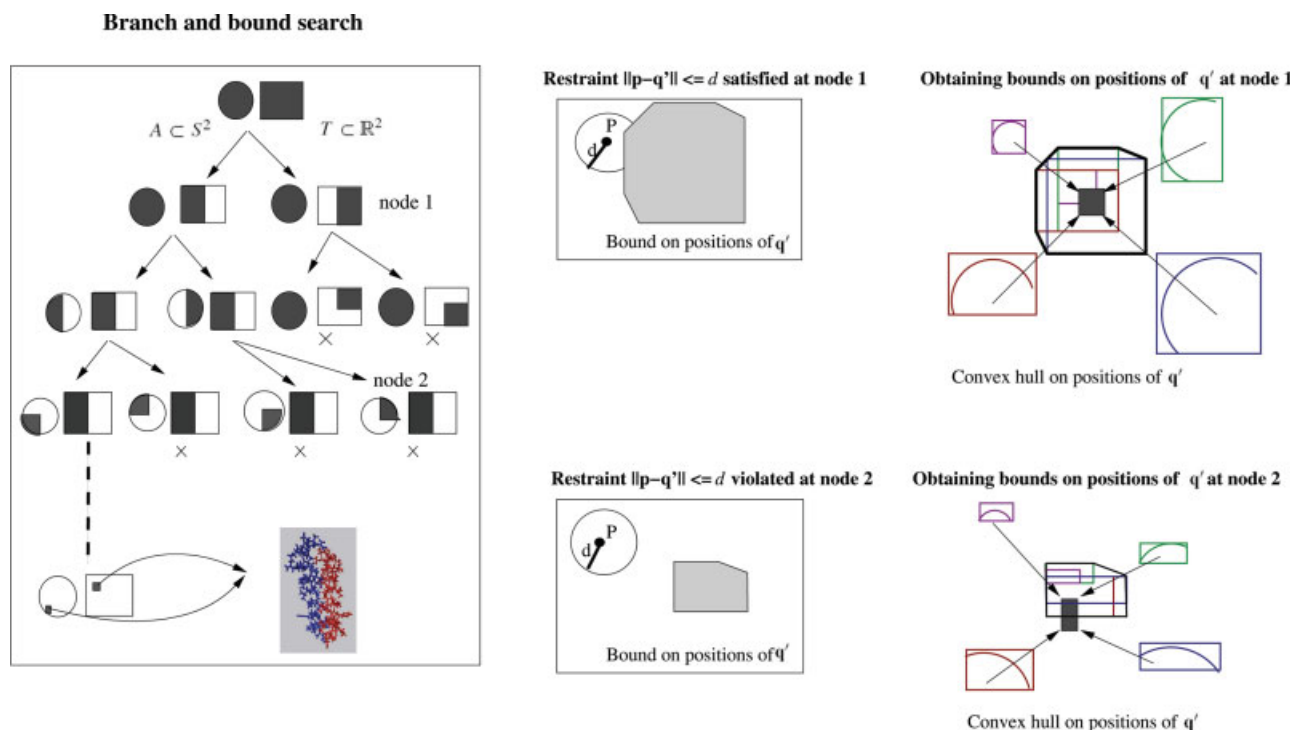


Fig. 2. The branch-and-bound algorithm proceeds as a tree search in SCS, the space of orientations and translations of the symmetry axes ( $S^2 \times \mathbb{R}^2$ ). (left) The 4D SCS is represented as two 2D regions: a sphere representing the orientation space  $S^2$  and a square representing the translation space  $\mathbb{R}^2$ . The dark shaded regions at each node of the tree represent the region in SCS being explored ( $A \times T \subset S^2 \times \mathbb{R}^2$ ). Ultimately (bottom left of the tree), the branch-and-bound algorithm returns regions in 4D space representative of structures that possibly satisfy all the restraints. (middle) At each node, we test satisfaction of each restraint of form  $\|\mathbf{p} - \mathbf{q}'\| \leq d$ , by testing intersection between the ball of radius  $d$  centered at  $\mathbf{p}$  and the convex hull bounding possible positions of  $\mathbf{q}'$ . If there exists an intersection between the ball and the convex hull for each restraint, further branching is done (as for node 1); otherwise, the entire node and its subtree are pruned (as for node 2). (right) The orientations and translations for a node restrict the possible positions of  $\mathbf{q}'$ . For each node, the four colored curves represent all possible positions of  $\mathbf{q}'$ , when considering the node's orientation space at each of the four corners of the node's translation space. The colored boxes represent the corresponding AABBs. The convex hull of the four AABBs (the black box) is the bound on positions of  $\mathbf{q}'$  for the node. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

positions of  $\mathbf{q}'$  under the symmetries defined by  $G$ . If there is an empty intersection between  $G\mathbf{q}$  and the ball of radius  $d$  centered at  $\mathbf{p}$ , then none of the structures represented by  $G$  satisfy the restraint. Formally, let  $B(\mathbf{p}, d)$  be the solid ball in  $\mathbb{R}^3$ , which has radius  $d$  and is centered at the point  $\mathbf{p}$ . If  $G\mathbf{q} \cap B(\mathbf{p}, d) = \emptyset$  then  $G$  is eliminated. Under the assumption of exact symmetry, each restraint  $\|\mathbf{p} - \mathbf{q}'\| \leq d$  implies another restraint  $\|\mathbf{p} - \mathbf{q}\| \leq d$ , where  $\mathbf{q}$  and  $\mathbf{p}'$  are atoms on the adjacent subunits in counterclockwise order. The satisfaction of this restraint is tested in a similar manner.

The region  $G\mathbf{q}$  is characterized by high-degree polynomials, and it is computationally expensive to test for intersections with  $G\mathbf{q}$ . Hence, we approximate  $G\mathbf{q}$  by a conservative bounding region,  $W(G, \mathbf{q})$ , that completely contains  $G\mathbf{q}$  (i.e.,  $G\mathbf{q} \subset W(G, \mathbf{q})$ ), but is simpler to compute than  $G\mathbf{q}$ . The conservative nature ensures that intersection tests between  $W(G, \mathbf{q})$  and  $B(\mathbf{p}, d)$  provably prune out only structures inconsistent with the data. If  $W(G, \mathbf{q}) \cap B(\mathbf{p}, d) = \emptyset$ , then we know  $G\mathbf{q} \cap B(\mathbf{p}, d) = \emptyset$ . We compute  $W(G, \mathbf{q})$  by first deriving a bounding region from the orientation space ( $S^2$ ) and then finding the bound from SCS ( $S^2 \times \mathbb{R}^2$ ). The details are as follows.

**Bound from orientation space.** Let  $A \subset S^2$  be a part of the unit sphere representing the set of orientations in  $G$ . Let  $A\mathbf{k}$  represent the new positions of  $\mathbf{k} \in \mathbb{R}^3$  after rotation around each axis in  $A$  by the angle of symmetry,  $\alpha = 360^\circ/n$ . Since rotations preserve distances,  $A\mathbf{k}$  must lie on a sphere of radius  $\|\mathbf{k}\|$  centered at the origin. We bound  $A\mathbf{k}$  by a spherical cap (region of a sphere which lies above or below a given plane) formed by the intersection of the sphere and a ball. The center and radius of the ball are obtained as follows. Let  $\mathbf{a}$  be an axis in  $A$  passing through the origin, such that all other axes in  $A$  lie within a ball of radius  $r_a$  centered at  $\mathbf{a}$ . Hence,  $A \subset S^2 \cap B(\mathbf{a}, r_a)$ . Let  $\mathbf{k}'$  be the position of  $\mathbf{k}$  rotated by  $\alpha$  radians around axis  $\mathbf{a}$ . The position  $\mathbf{k}'$  is the center of the ball approximating  $A\mathbf{k}$  and is obtained as follows:

$$\mathbf{k}' = (\mathbf{k} \cdot \mathbf{a})\mathbf{a} + (\sin \alpha)(\mathbf{a} \times \mathbf{k}) + (\cos \alpha)(\mathbf{k} - (\mathbf{k} \cdot \mathbf{a})\mathbf{a}). \quad (1)$$

The radius,  $r_k$ , of the ball approximating  $A\mathbf{k}$  is computed as

$$r_k = r_a \left( \sqrt{(\sin \alpha)^2 \|\mathbf{k}\|^2 + (1 - \cos \alpha)^2 (\mathbf{k} \cdot \mathbf{a})^2} + |1 - \cos \alpha| \|\mathbf{k}\| \right). \quad (2)$$

Finally, we define our bounding region to be the spherical cap  $C = B(\mathbf{k}', r_k) \cap S(\mathbf{0}, \|\mathbf{k}\|)$ , where  $S(\mathbf{0}, \|\mathbf{k}\|)$  denotes the sphere of radius  $\|\mathbf{k}\|$  that is centered at the point  $\mathbf{0}$ , the origin. By construction,  $A\mathbf{k} \subset C$ . We approximate  $C$  with a bounding box that is aligned along the  $x$ -,  $y$ -, and  $z$ -axes—an axis-aligned bounding box (AABB), which we refer to as  $V(A, \mathbf{k})$ . We use  $V(A, \mathbf{k})$  to help us perform a quick test for intersection. If this quick test is passed, we then perform a second, more careful and expensive test for intersection, using a tighter bounding region for  $A\mathbf{k}$ . This tighter bound is the smallest AABB that contains  $A\mathbf{k}$ . The dimensions of the box are found by performing a numerical global minimization (and maximization) on the  $x$ -,  $y$ -, and  $z$ -coordinates of  $\mathbf{k}'$ . The global optimization is done by gridding  $A$  and starting a gradient descent from each of the grid points. The details of both our intersection tests, and the Eqs. (1) and (2) are provided in the Supplementary Material.

**Bound from SCS.** Let  $T \subset \mathbb{R}^2$  and  $A \subset S^2$  denote the sets of translations and orientations in  $G$ . The region  $G\mathbf{q}$  represents the positions of  $\mathbf{q}$  on rotation around each axis in  $A \times T$  by the angle of symmetry  $\alpha$ . To bound  $G\mathbf{q}$  and determine  $W(G, \mathbf{q})$ , we need to find the orientation-based bound (as mentioned earlier) for each translation  $\mathbf{t} \in T$ . We choose our bounding region  $W(G, \mathbf{q})$  as an approximation of the convex hull of  $G\mathbf{q}$ . Using the properties of convex hulls and the fact that  $T$  is convex, we are able to prove that the convex hull of  $G\mathbf{q}$  is determined by just the corners of  $T$  (see Supplementary Material for proof). Let  $H(U)$  be the convex hull of  $U \subset \mathbb{R}^3$ . It can be shown that  $H(G\mathbf{q}) = H(\cup_{i=1}^4 \{A\mathbf{q}_i + \mathbf{t}_i\})$  where  $\mathbf{t}_i$  represents the four corners of  $T$  and  $\mathbf{q}_i$  denotes the position  $(\mathbf{q} - \mathbf{t}_i)$ . We use our bounds on regions of  $A\mathbf{k}$ , the AABB  $V(A, \mathbf{k})$ , to bound  $A\mathbf{q}_i$ . We then bound  $G\mathbf{q}$  by finding the convex hull of the AABBs at the corners of  $T$ . This convex hull is our bounding region  $W(G, \mathbf{q})$ . It can be proven that  $W(G, \mathbf{q})$  is a conservative bounding region for  $G\mathbf{q}$  (see Supplementary Material). That is,  $G\mathbf{q} \subset W(G, \mathbf{q})$ . Hence, testing satisfaction of a restraint  $\|\mathbf{p} - \mathbf{q}'\| \leq d$  requires testing for the intersection of the convex hull of the AABBs at the corners of  $T$  (which is a bounding region for  $G\mathbf{q}$ ), with a solid ball centered at  $\mathbf{p}$  with radius  $d$ . We test the intersection between a ball and a convex polyhedron, using the method described in Ref. 40. For details on the bounding regions and pruning criteria, please see the Supplementary Material.

### Branching

In partitioning a grid cell into subcells, we seek to divide the cell into two regions along the dimension that will cause one of the restraints to be violated by all the conformations represented by one of the subcells. This kind of a division will allow us to efficiently eliminate the subcell. We use the following heuristic to achieve this. For each restraint  $\|\mathbf{p} - \mathbf{q}'\| \leq d$ , we compute  $\mathbf{q}'$  for each of the corners of the grid cell. We then identify the dimension that has the largest difference in  $\|\mathbf{p} - \mathbf{q}'\|$  distances for its pair of corners. We partition along that dimension.

### Determining Satisfying Structures

Our algorithm guarantees that the consistent regions it returns represent every conformation of the symmetric homo-oligomer that is consistent with the data. Because we prune conservatively, they might also represent additional structures inconsistent with the data. To identify the most consistent structures and evaluate packing quality, we generate representative structures from the consistent regions. Choosing good representatives helps ensure that we do not miss native structures (as can occur with sampling-based docking approaches; see Introduction section). We choose the set of representative structures as follows. A grid cell is accepted as part of the consistent regions only when all structures it represents are within  $\tau_0$  Å of each other. We first obtain the structures from the centroids of the grid cells in the consistent regions. We then cluster these structures using an agglomerative complete linkage hierarchical clustering<sup>41</sup> that allows two structures to be within a cluster only if their backbone RMSD is within  $\tau_0$  Å. The centroids of the clusters then form a set of representative structures. This procedure ensures that every structure of the consistent region is within  $\tau_0$  Å of at least one representative structure, and hence ensures that the representative structures form a good representation of the consistent regions.

Some of representative structures might be inconsistent with data, because of the conservative bounds used when pruning regions. We define the satisfaction score for each structure as the sum over the violated NOE restraints of the difference in expected and observed distance.

The set of satisfying structures are those representative structures with satisfaction scores below a threshold (chosen as 1 Å). Note that each satisfying structure represents a set of grid cells in the consistent regions. The union of all such grid cells forms the satisfying regions.

### Determining WPS Structures

Having obtained the set of satisfying structures, we now evaluate each of them for packing quality. We first energy-minimize them with the LBFGS conjugate-gradient minimization algorithm (10,000 minimization steps) in CNS.<sup>42</sup> The energy function being minimized includes the NOE restraint energy terms as well as the modeling energy terms of vdW (6-12 Lennard-Jones potential), bond length, bond angle, dihedral, and improper energies.<sup>42</sup> We harmonically restrain the backbone and the NOE restraints to ensure that we maintain the symmetry and satisfy the restraints. The minimization accounts for flexibility and asymmetry in the side chains and should help obtain conformations of the side chains that aid in good vdW packing.

We define the packing score of an energy-minimized structure as the difference between the vdW energy of a subunit in the structure when it is in the complex, and the vdW energy of the subunit when it is not in the complex.

We then define the set of WPS structures as those satisfying structures with packing scores below a threshold, chosen here as 0 kcal/mol, since well-modeled structures should have negative packing scores. We refer to the union of grid cells in the consistent regions corresponding to the WPS structures as the WPS regions.

### Evaluating Uncertainty

We evaluate uncertainty in two ways. One way is by calculating the uncertainty in the conformations as the variance in the position of each atom across the set of satisfying (or WPS) structures. The second way is by quantifying uncertainty in SCS as the spread in the areas of the satisfying (or WPS) regions in the translation space and the orientation space. The 4D volume of the satisfying (or WPS) regions also indicates uncertainty. By comparing these measures for satisfying versus WPS structures and regions, we evaluate the constraint on structure provided by data alone versus the constraint provided by both data and packing together.

### Handling Ambiguity

The spectral overlap inherent in NMR spectra for symmetric homo-oligomers leads to intersubunit distance restraints that have a two-fold ambiguity. By two-fold ambiguity, we mean that the relative order of the restraints (whether they have a clockwise or a counterclockwise orientation) is not apparent. Each restraint has two possible orientations. For instance, when an intersubunit restraint is observed between residues 32 and 38, the restraint could be between residue 32 on the first subunit and residue 38 on the adjacent subunit in clockwise order or between residue 32 on the first subunit and residue 38 on the adjacent subunit in counterclockwise order.

We handle this two-fold ambiguity by extending our branch-and-bound search of SCS. If we have  $r$  restraints, each with two orientations, then there are  $2^{r-1}$  possible combinations (the first orientation assignment can be made arbitrarily, since the measurements are relative). However, the structure of our search allows us to avoid testing all the combinations explicitly. Recall that, for each grid cell, we test each of the restraints individually using our conservative bounds (Bounding section). Now we simply individually test each of the "oriented restraints" (i.e., a restraint with either a clockwise or counterclockwise orientation). This yields  $1 + 2(r - 1)$  independent tests, rather than the exponential blow-up required to consider all combinations. If an oriented restraint is violated, then all combinations including it are implicitly eliminated from further consideration. If both orientations for any restraint are eliminated for some grid cell, then the cell is completely pruned. Otherwise, we continue branching as usual. At the end of the search, we can enumerate the combinations represented in the satisfying regions.

## RESULTS

### Results on Phospholamban

We tested our protocol on phospholamban, using the subunit structure and nine experimental NOE restraints from the Chou lab. We chose as the reference structure, the best representative conformer (as indicated in Ref. 1) among the deposited 20 NMR structures (PDB id: 1ZLL). For all our test cases, we chose the structural similarity threshold,  $\tau_0$ , to be 1 Å. Also, we declare a structure to have a steric clash when there are at least five pairs of atoms, such that each pair is separated by less than 1.5 Å. Energy minimization of structures allows side-chain flexibility such that one to two steric clashes can be eliminated. Our analysis is conservative and declares a clash only when at least five atoms collide.

Figure 3(a) plots the packing scores versus the satisfaction scores for phospholamban. The set of satisfying structures has an overall range of 1.07–8.85 Å backbone RMSD to the reference structure. This range indicates the diversity in structures possible, using just the nine experimental intersubunit NOE restraints. The average variance in the positions of the atoms in the set of satisfying structures is 12.32 Å<sup>2</sup>. The area of the translation space in the satisfying regions is  $\sim 290$  Å<sup>2</sup> and that of the orientation space is  $\sim 0.40$  radian<sup>2</sup>. The volume of the 4D region is  $\sim 1.24$  Å<sup>2</sup> radian<sup>2</sup>. These values indicate the constraint on structure provided by the data alone.

Figure 3(a) also shows the set of WPS structures (in green). The reference structure has a satisfaction score of around 0.8 Å and packing score of  $-17$  kcal/mol, and it lies in the WPS region. Figure 3(b) shows the backbone RMSD of the reference structure to each of the WPS structures. Incorporating packing quality reduces the maximum RMSD to the reference structure from 8.85 to 6.24 Å. The area of the translation space reduces from 290 to  $\sim 135$  Å<sup>2</sup> and that of the orientation space reduces from 0.40 to  $\sim 0.23$  radian<sup>2</sup>. The volume of the 4D space reduces from 1.24 Å<sup>2</sup> radian<sup>2</sup> in the set of satisfying structures to  $\sim 0.51$  Å<sup>2</sup> radian<sup>2</sup>. All these values indicate the additional constraint that packing quality imposes on the structure of phospholamban. The average variance in the positions of the atoms is reduced from 12.32 to 6.80 Å. Figure 4(a) illustrates all the WPS structures, and Figure 4(b) illustrates the variance of each of the backbone atoms. The figures show that there is more uncertainty in the amphipathic helices than in the transmembrane helices. The average variance in the amphipathic helices is 10.75 Å<sup>2</sup>, whereas that in the transmembrane helices is 2.96 Å<sup>2</sup>. This is because the experimental restraints are between residues in the transmembrane helices. This agrees with the observation in Ref. 1 that the amphipathic helices are less well determined. From this we conclude that we need more restraints in the amphipathic helices to determine the structure with greater precision.

We resolved the ambiguity in the nine intersubunit restraints, as described in the Handling Ambiguity sec-

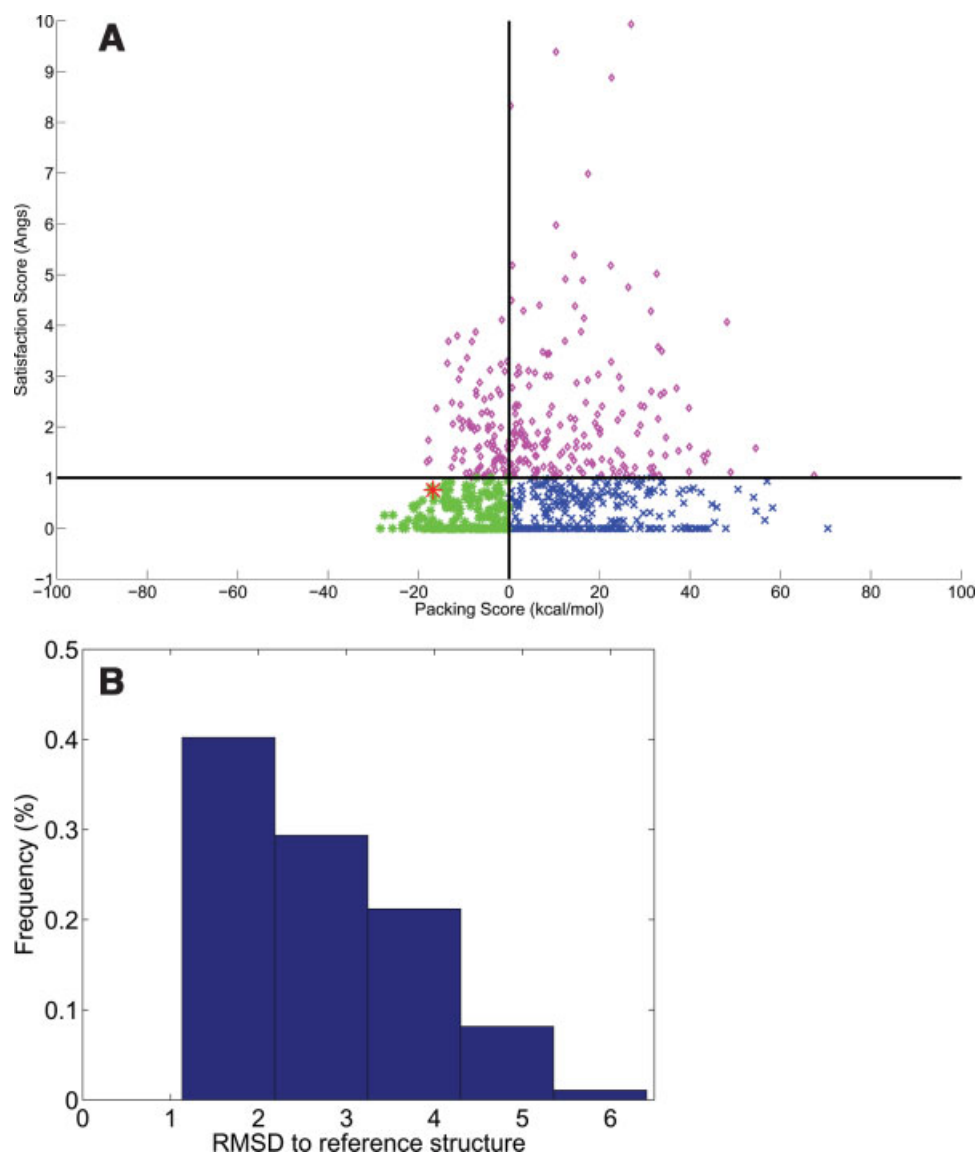


Fig. 3. Phospholamban results: (a) Restraint satisfaction score versus packing score for all representative structures. The vertical and horizontal lines indicate the chosen cutoffs for WPS structures: 1 Å for the satisfaction score and 0 kcal/mol for the packing score. The green stars and the blue crosses indicate the set of satisfying structures. The magenta points indicate the set of nonsatisfying structures that have been pruned. The green stars indicate the set of WPS structures, and the red star indicates the reference structure. (b) Histogram of backbone RMSD to the reference structure for the WPS structures.

tion. This was done by considering satisfaction of only a linear number of oriented restraints ( $1 + 2(9 - 1) = 17$ ) rather than the exponential number of possible combinations ( $2^{9-1} = 256$ ). Only one of the 256 combinations actually remained in the satisfying regions. This is the same combination that was reported by Oxenoid and Chou.<sup>1</sup>

We further tested whether the experimental data available are sufficient to choose one oligomeric number over others with reasonable confidence. We expect structures obtained from the correct oligomeric number to not only satisfy the data but also to have good vdW packing.

Hence, restraint satisfaction and vdW packing should help discriminate among putative oligomeric numbers. We tested this possibility by searching in the extended SCS (ESCS),  $\mathbf{Z}_9 \times S^2 \times \mathbb{R}^2$ , where  $\mathbf{Z}_9$  is a set of possible oligomeric numbers from 2 to 9. We first obtain the set of WPS structures for each oligomeric number. We immediately prune out those oligomeric numbers that have no WPS structures. This allows us to determine the oligomeric number with high certainty, when only a single oligomeric number has WPS structures. When several oligomeric numbers have WPS structures, we determine the oligomeric number as follows. Let  $E_1(m)$  and  $E_1(n)$



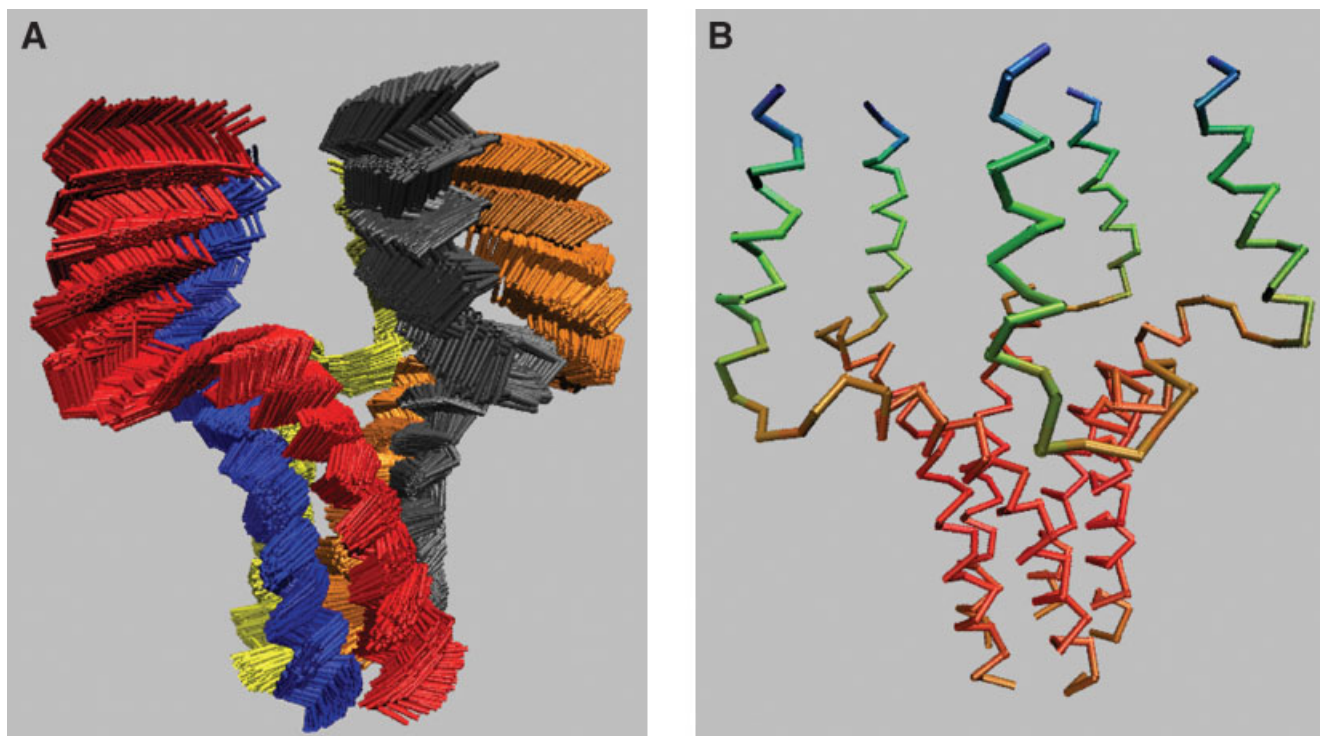


Fig. 4. Phospholamban structures: (a) The set of WPS structures after alignment to the structure with the lowest packing score. Different chains are in different colors. (b) Variance of the backbone atoms is illustrated by a color scale, with blue indicating maximum variance and red minimum variance.

TABLE I. Satisfying Structure Results

Protein	PDB id	No. of restraints	RMSD (Å)		Uncertainty in SCS			Variance in atoms (Å <sup>2</sup> )		
			Min	Max	$T$ (Å <sup>2</sup> )	$A$ (rad <sup>2</sup> )	Volume (Å <sup>2</sup> rad <sup>2</sup> )	Min	Max	Mean
Glycophorin A (2) <sup>a</sup>	1AFO	6 (expt)	0.61	1.77	4.72	0.06	0.06	0.34	2.62	0.97
Haemagglutinin (3)	1HTM	85 (simulated)	0.86	1.08	1.10	0.001	4e-4	0.07	0.64	0.22
Kv1.2 potassium channel (4)	1QDV	32 (simulated)	0.92	2.85	7.47	0.07	0.13	0.24	4.91	1.24
Phospholamban (5)	1ZLL	9 (expt)	1.07	8.85	289.53	0.40	1.24	2.87	43.97	12.32
Gp31 co-chaperonin (7)	1G31	85 (simulated)	0.40	2.72	21.20	0.07	0.15	0.36	7.73	1.66

Backbone RMSD of the set of satisfying structures to the reference structure, the uncertainty in SCS represented by the area of the translation ( $T$ ) and orientation ( $A$ ) space for the satisfying region, and the 4D volume of the satisfying region, and finally the variance in the position of atoms for the set of satisfying structures.

<sup>a</sup>Values in parentheses indicate symmetry.

represent the lowest packing scores of the WPS structures from oligomeric numbers of  $m$  and  $n$ , respectively. If  $E_1(m) < E_1(n)$ , the difference  $E_1(n) - E_1(m)$  indicates the confidence we have in preferring  $m$  versus  $n$  as the oligomeric number.

When we applied this approach to determine the oligomeric number of phospholamban, WPS structures were present only for oligomeric numbers of tetramer, pentamer, hexamer, and heptamer. The lowest packing scores obtained were  $E_1(4) = -21.80$  kcal/mol,  $E_1(5) = -28.44$  kcal/mol,  $E_1(6) = -19.28$  kcal/mol, and  $E_1(7) = -15.52$  kcal/mol. Since  $E_1(5)$  is the lowest packing score, we correctly conclude that the pentamer is the most feasible

oligomeric number. We expect that with the availability of more experimental data we can determine the oligomeric number with greater confidence.

### Results on Other Proteins

We tested the performance of our approach on several other homo-oligomers whose structures have been solved either by NMR or X-ray crystallography. Tables I and II summarize the results. For the case of human glycoporphin A (1AFO), six experimental NOE restraints were available. The authors reported 20 NMR structures, and we chose as the subunit structure the first chain from

TABLE II. WPS Structure Results

Protein	PDB id	No. of restraints	RMSD (Å)		Uncertainty in SCS			Variance in atoms (Å <sup>2</sup> )		
			Min	Max	$T$ (Å <sup>2</sup> )	$A$ (rad <sup>2</sup> )	Volume (Å <sup>2</sup> rad <sup>2</sup> )	Min	Max	Mean
Glycophorin A (2) <sup>a</sup>	1AFO	6 (expt)	0.61	1.77	2.97	0.045	0.04	0.07	1.51	0.47
Haemagglutinin (3)	1HTM	85 (simulated)	0.86	1.08	1.10	0.001	4e-4	0.07	0.64	0.22
Kv1.2 potassium channel (4)	1QDV	32 (simulated)	0.92	2.85	7.47	0.07	0.12	0.24	4.91	1.24
Phospholamban (5)	1ZLL	9 (expt)	1.07	6.24	135.50	0.23	0.51	1.52	24.96	6.80
Gp31 co-chaperonin (7)	1G31	85 (simulated)	0.40	2.72	21.20	0.07	0.15	0.36	7.73	1.66

Backbone RMSD of the set of WPS structures to the reference structure, the uncertainty in SCS represented by the area of the translation ( $T$ ) and orientation ( $A$ ) space for the WPS region, and the 4D volume of the WPS region, and finally the variance in the position of atoms for the set of WPS structures.

<sup>a</sup>Values in parentheses indicate symmetry.

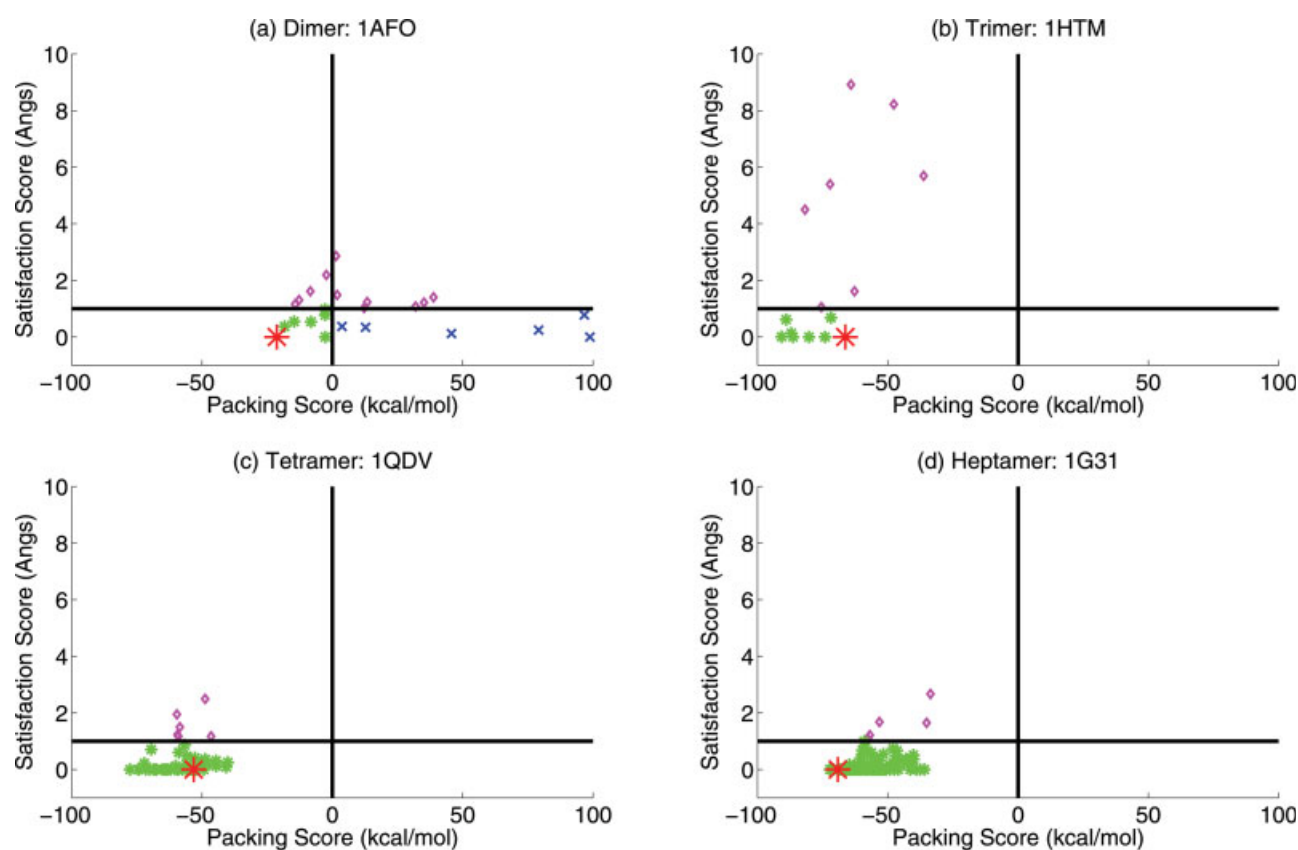


Fig. 5. Restraint satisfaction score versus packing score for all satisfying structures of (a) human glycophorin A (dimer:1AFO), (b) influenza haemagglutinin (trimer:1HTM), (c) Kv1.2 potassium channel (tetramer:1QDV), and (d) Gp31 co-chaperonin (heptamer:1G31). The vertical and horizontal lines indicate the chosen cutoffs for WPS structures: 1 Å for the satisfaction score, and 0 kcal/mol for the packing score. The green stars and the blue crosses (when present) indicate the set of satisfying structures. The magenta points indicate the set of nonsatisfying structures that have been pruned. The green stars indicate the set of WPS structures, and the red star indicates the reference structure.

the structure that best satisfies the restraints. For all of the remaining test cases, the structures were obtained by X-ray crystallography (protons were added using CNS), and we chose as subunit structure the first chain from the respective complexes as the subunit structure. We simulated the NOE restraints from the X-ray reference structures by finding pairwise distances between protons on adjacent monomers. Every pair that had a

distance less than 5 Å and almost exact symmetry (the mean of difference in distances across adjacent subunits is within 0.5 Å) was chosen as a restraint and an uncertainty of  $\pm 1$  Å was added. Choosing restraints with almost exact symmetry simulates the scenario of choosing intersubunit restraints that have significant signal overlap. Our simulations yielded 85 simulated restraints for haemagglutinin (1HTM), 32 for the Kv1.2 potassium

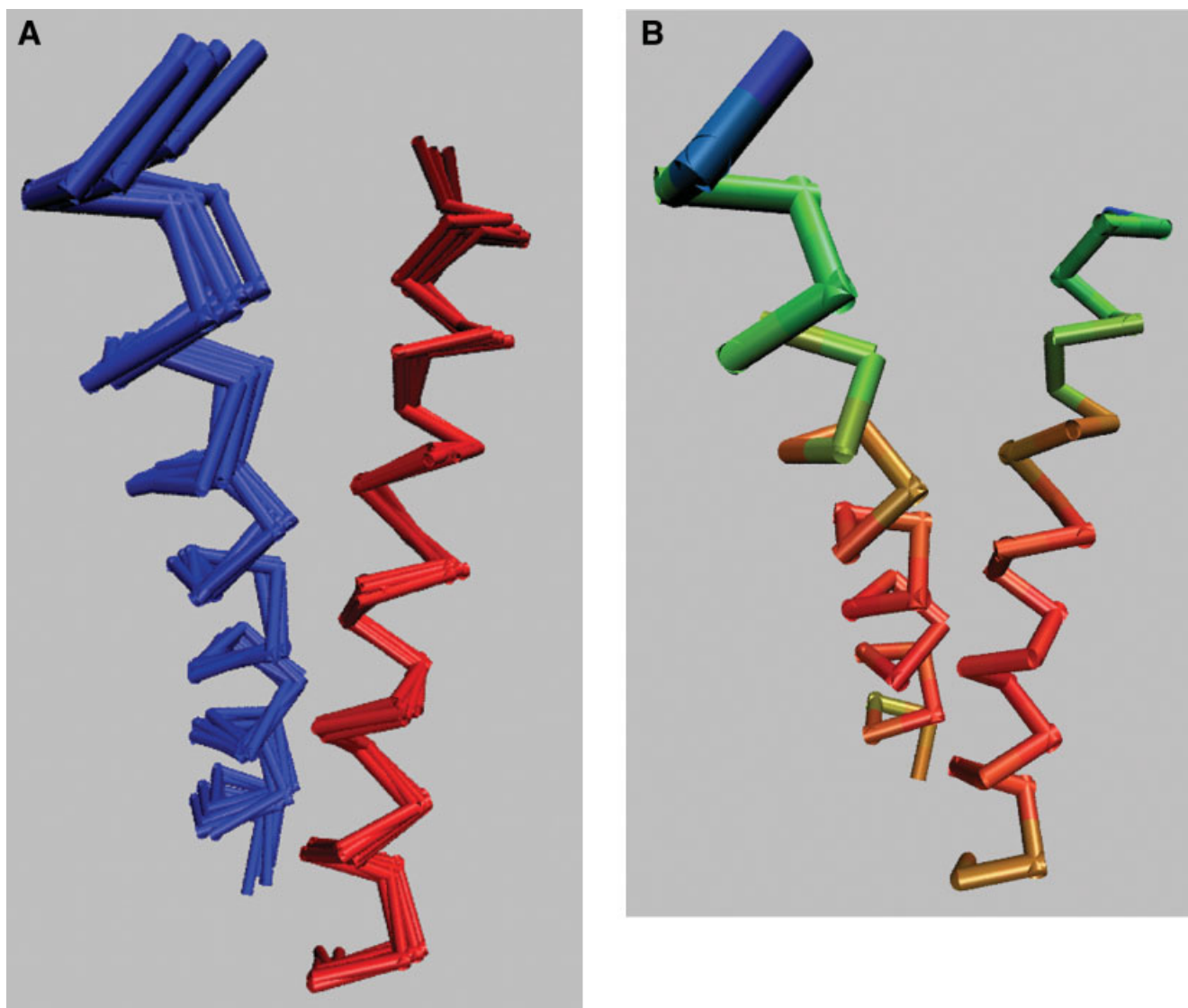


Fig. 6. Glycophorin A (1AFO) structures: (a) The set of WPS structures after alignment to the structure with the lowest packing score. Different chains are in different colors. (b) Variance of the backbone atoms is illustrated by a color scale, with blue indicating maximum variance and red minimum variance.

channel (1QDV), and 85 for the Gp31 co-chaperonin (1G31).

Figure 5 shows for each test case a plot of packing scores versus satisfaction scores. Table I shows the results on the satisfying regions and satisfying structures, and Table II shows the results on WPS regions and WPS structures. The tables and figure clearly show that except for human glycophorin A, the remaining test cases have identical sets of satisfying structures and WPS structures. Further, the spread in the satisfying region and the WPS region is almost the same. The reason for this similarity might be because we have used all possible restraints (85, 32, and 85) for these test cases. This use of all restraints causes almost all the satisfying structures returned by the branch-and-bound

algorithm to have high-quality vdW packing, and hence to belong to the set of WPS structures.

Figures 6–9 illustrate the uncertainty in the set of WPS structures for each of the test cases. Despite using 32 and 85 restraints, the potassium channel and co-chaperonin have considerable spread in the translation space. The variance in the position of the atoms is also high, with a maximum as high as  $4.9 \text{ \AA}^2$  and  $7.7 \text{ \AA}^2$ , respectively. The higher uncertainty is because the chosen restraints (restricted to those with exact symmetry) are not distributed all along the intersubunit surface, but are concentrated toward one end of the surface [Fig. 10(a)]. On the other hand, haemagglutinin is a long helical trimer, and the 85 restraints are distributed across the entire intersubunit surface, thereby yielding

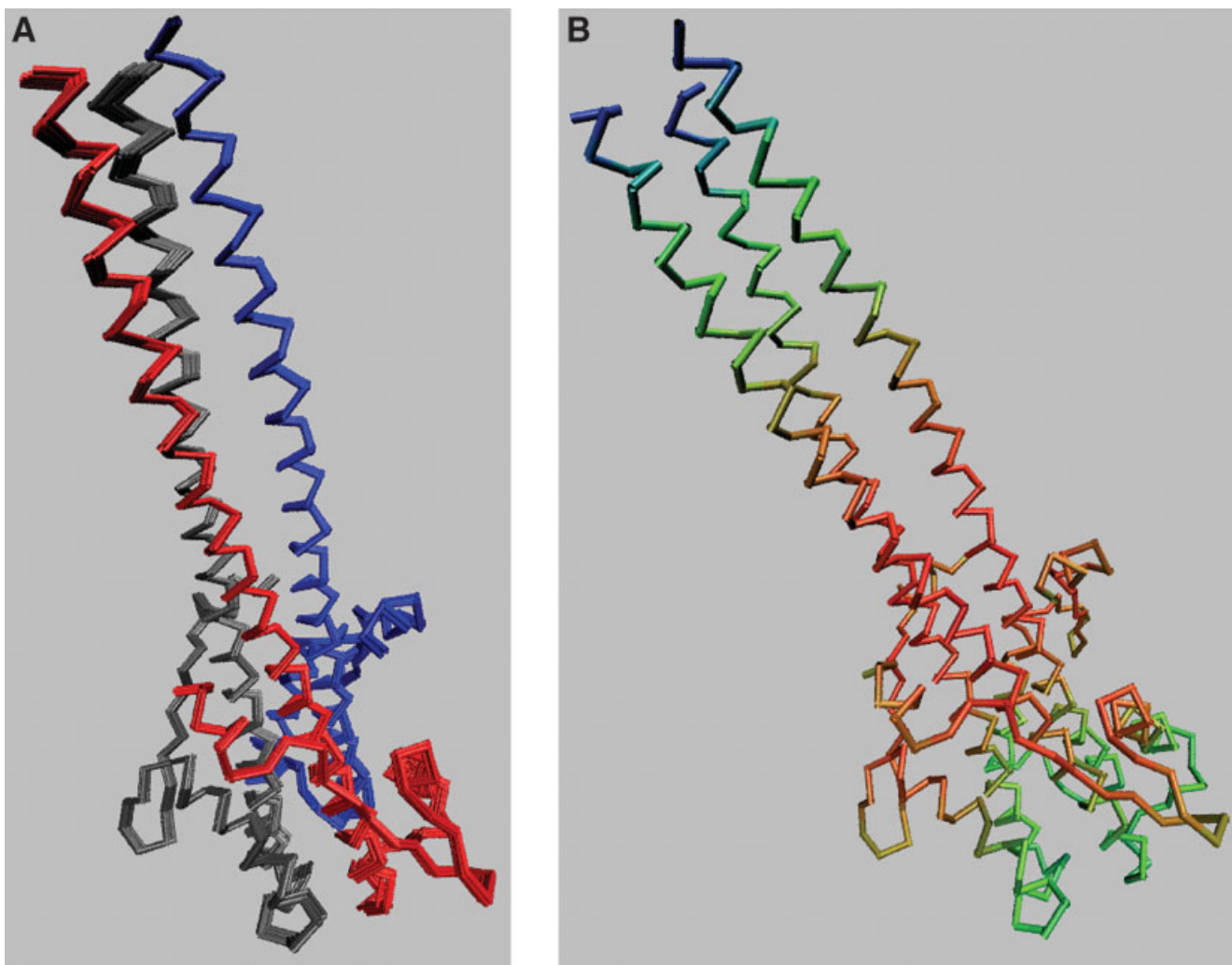


Fig. 7. Haemagglutinin (1HTM) structures: (a) The set of WPS structures after alignment to the structure with the lowest packing score. Different chains are in different colors. (b) Variance of the backbone atoms is illustrated by a color scale, with blue indicating maximum variance and red minimum variance.

less uncertainty [Fig. 10(b)]. This indicates the effect of independence of restraints on the uncertainty.

The reference structure lies in the WPS region for all cases, and Table II indicates the range of RMSDs to the reference structure. Figure 11 plots histograms of the backbone RMSD of the set of WPS structures to the reference structure. The histograms peak at 1 Å RMSD, which indicates that the structures obtained are close to the reference structure. The potassium channel (1QDV) and co-chaperonin (1G31) have larger ranges. It is interesting to note that the dimer, with as few as six experimental restraints, provides for comparatively much less uncertainty. This smaller uncertainty might be because the restraints are spread out across the subunit.

We also tested the change in uncertainty of the structure for the trimer, haemagglutinin, as the number of relatively independent restraints varies. We refer to a set of restraints as relatively independent if the

restraints have no common atoms and are far apart from each other. We choose the set of independent restraints by the following heuristic. We consider midpoints of each possible restraint, and represent distances between the restraints by distances between their midpoints. We choose as the first two members of the set, the two restraints that have their midpoints farthest apart. We choose the third restraint as the one whose midpoint has the largest minimum distance from the first two members. We continue this process iteratively to choose the remaining restraints. Figure 12 illustrates the change in the translation spread, orientation spread, and average variance in position of atoms with an increasing number of independent restraints. The average variance in the atom positions decreases significantly as the number of independent restraints increases. This analysis helps us quantify the minimum number of independent restraints required to determine

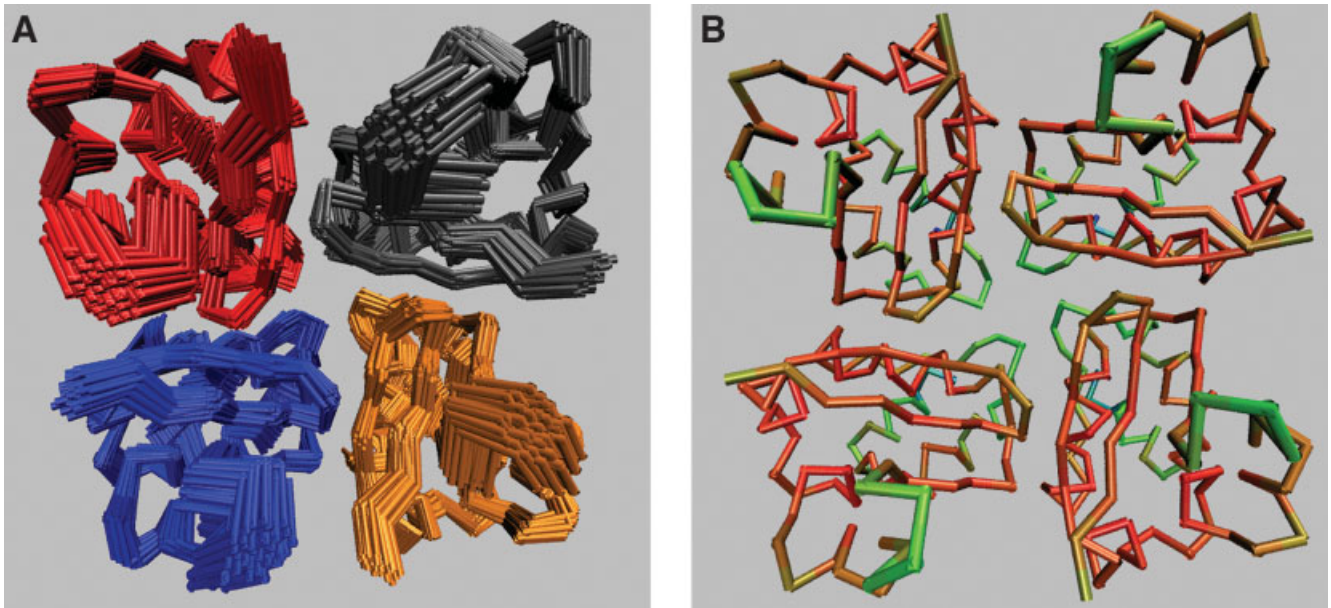


Fig. 8. Kv1.2 potassium channel (1QDV) structures: (a) The set of WPS structures after alignment to the structure with the lowest packing score. Different chains are in different colors. (b) Variance of the backbone atoms is illustrated by a color scale, with blue indicating maximum variance and red minimum variance.

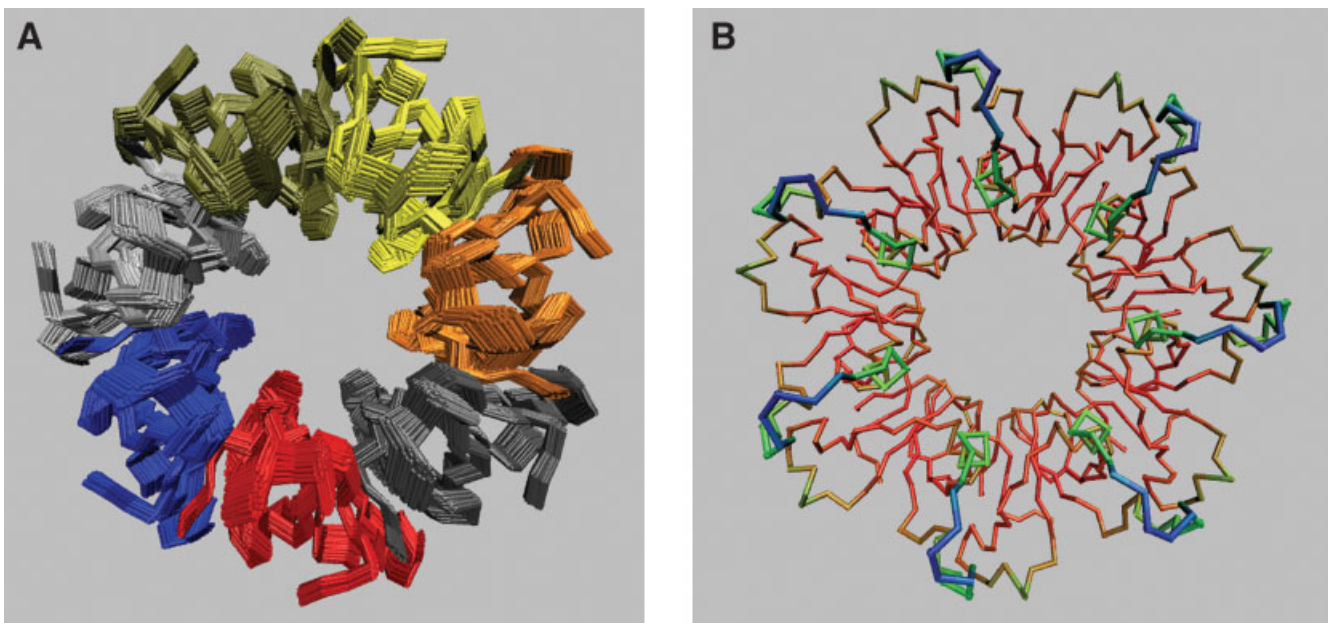


Fig. 9. Gp31 co-chaperonin (1G31) structures: (a) The set of WPS structures after alignment to the structure with the lowest packing score. Different chains are in different colors. (b) Variance of the backbone atoms is illustrated by a color scale, with blue indicating maximum variance and red minimum variance.

the structure of the homo-oligomer up to a specified precision. For example, if we are willing to tolerate  $\sim 1 \text{ \AA}^2$  uncertainty in the positions of the atoms, about eight independent restraints will be sufficient. On the other hand, if we want high precision and are not willing to tolerate uncertainty above  $0.3 \text{ \AA}^2$  in the positions of atoms, we need as many as 64 independent restraints.

We also applied our method for determination of oligomeric number (Results on Phosphalamban) to the test cases. For glycophorin A, haemagglutinin, and the Kv1.2 potassium channel, we could determine the oligomeric number with high certainty, because the set of WPS structures was empty for oligomeric numbers other than the correct one. Glycophorin A is especially interesting,

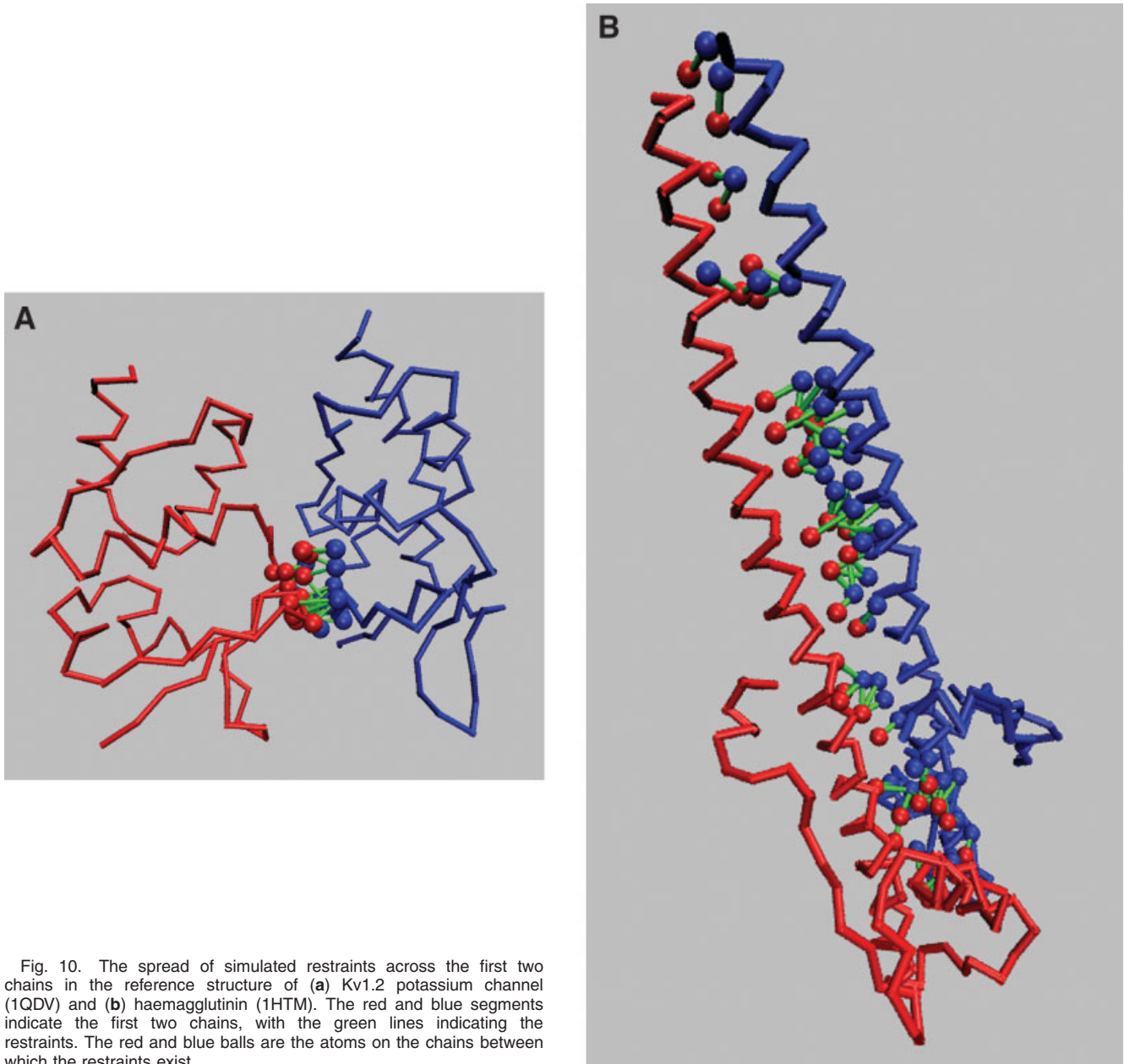


Fig. 10. The spread of simulated restraints across the first two chains in the reference structure of (a) Kv1.2 potassium channel (1QDV) and (b) haemagglutinin (1HTM). The red and blue segments indicate the first two chains, with the green lines indicating the restraints. The red and blue balls are the atoms on the chains between which the restraints exist.

since we determined the correct oligomeric number using just six restraints. For the co-chaperonin, oligomeric numbers of hexamer, heptamer, and octamer have WPS structures with  $E_1(6) = -69.09$  kcal/mol,  $E_1(7) = -72.19$  kcal/mol, and  $E_1(8) = -68.92$  kcal/mol. In this case too, since  $E_1(7)$  is the lowest packing score, we correctly conclude that the oligomeric number is 7.

We continued our study of the effect of number of independent restraints for haemagglutinin (1HTM). When the number of restraints chosen was 85, 64, or 32, WPS structures were absent for oligomeric numbers other than 3, allowing us to determine with high confidence that haemagglutinin is a trimer. With only 16 or 8 re-

straints, we obtain WPS structures for oligomeric numbers other than 3. As expected, as the available experimental data increases, our confidence in determining the oligomeric number increases.

## DISCUSSION

We have developed a novel approach that performs a complete, data-driven search to identify all conformations of a homo-oligomeric complex that are consistent with NOE restraints and display high-quality vdW packing. This approach is particularly important in sparse-data

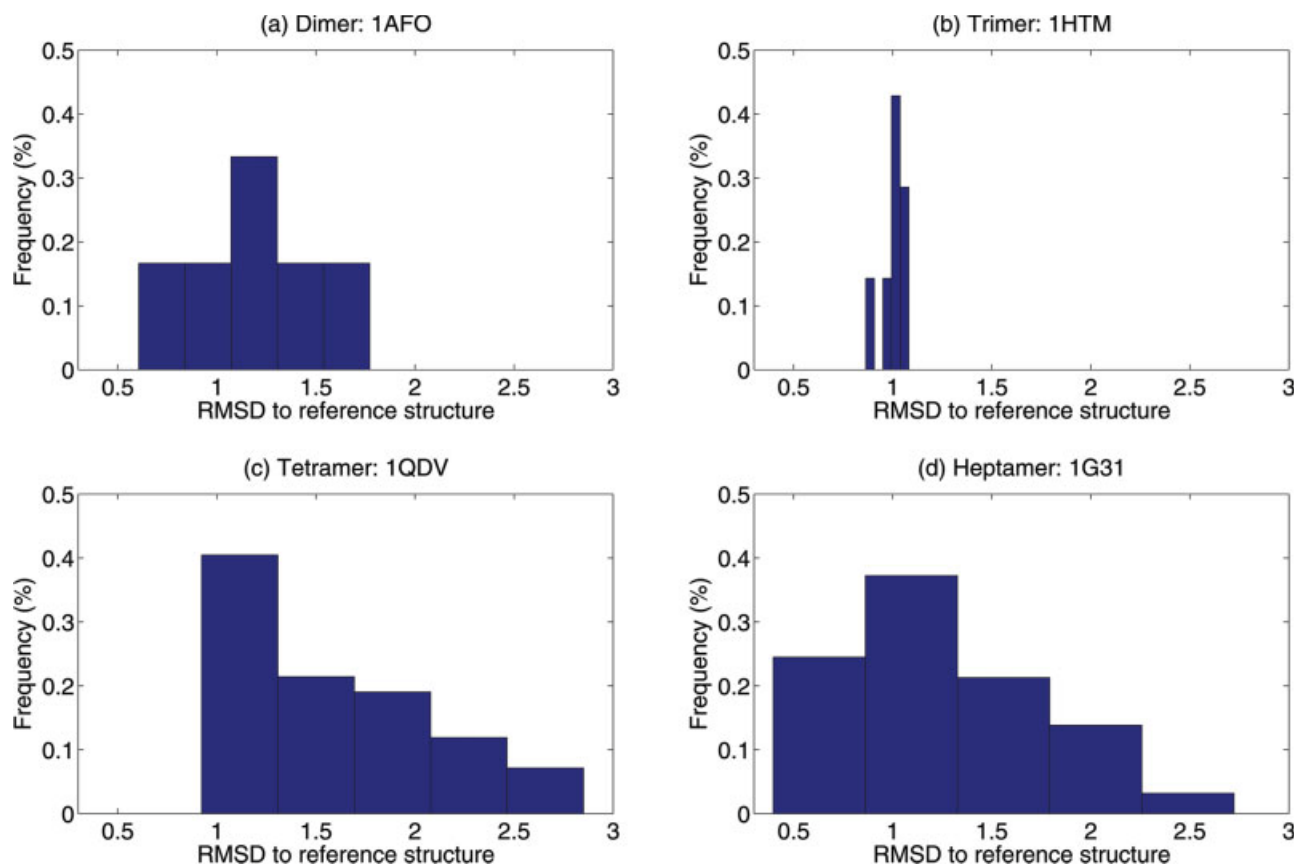


Fig. 11. Histogram of backbone RMSD to the reference structure for the WPS structures returned by our approach for (a) human glycoprotein A (dimer:1AFO), (b) influenza haemagglutinin (trimer:1HTM), (c) Kv1.2 potassium channel (tetramer:1QDV), and (d) Gp31 co-chaperonin (heptamer:1G31). [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

cases, where relying on an incomplete, biased search may result in missing WPS conformations. In contrast with typical sampling-based approaches, our algorithm provides provable guarantees on all the points in each grid cell, including those not at the grid points. Examination of the entire solution space further enables objective evaluation of the amount of structural uncertainty. Since data terms and packing quality terms are kept separate, our approach can evaluate how much information is provided by data versus by modeling, how much the determined structures depend on each, and how in general to treat them uniformly, consistently, and transparently. By first searching for regions of conformation space consistent with the NOE restraints, and then filtering these regions according to predicted quality of packing, our approach makes good use of the relatively greater discriminatory information content in sparse NOEs to focus on plausible conformations for subsequent analysis by relatively finer-grained packing metrics.

Our branch-and-bound approach currently uses a conservative analytical bound and a simple branching technique. Tighter bounds and better partitioning techniques will allow a more efficient search of SCS. We intend to replace our global optimization-based bound with an analytical bound. At each node in our search, our current

tests for restraint satisfaction are conservative, in which we individually test whether for each restraint, there exists at least one point in the grid cell, such that the structure represented by the point satisfies the restraint. A better bound would be to test for simultaneous restraint satisfaction by checking whether there exists at least one point in the grid cell being tested, which simultaneously satisfies all the restraints. Currently, our check for steric clash assumes that checking for clashes at the corners of a grid cell suffices. Provably correct tests for steric clash would be more accurate. Our conservative bounds on possible positions of a point for a grid cell (see Bounding section), when applied to atoms of the subunit that could be in a steric clash, will allow us to prune a grid cell only when all the conformations represented by the grid cell have clashes. Finally, currently we evaluate packing using energy functions available in CNS. Using sophisticated energy functions to evaluate packing might provide better constraint to the structure and help determine the oligomeric number with higher confidence. At this time, our approach assumes that all the restraints obtained from the experiment are correctly assigned. To handle noisy data, we can extend our search to eliminate a cell only when several restraints are violated.

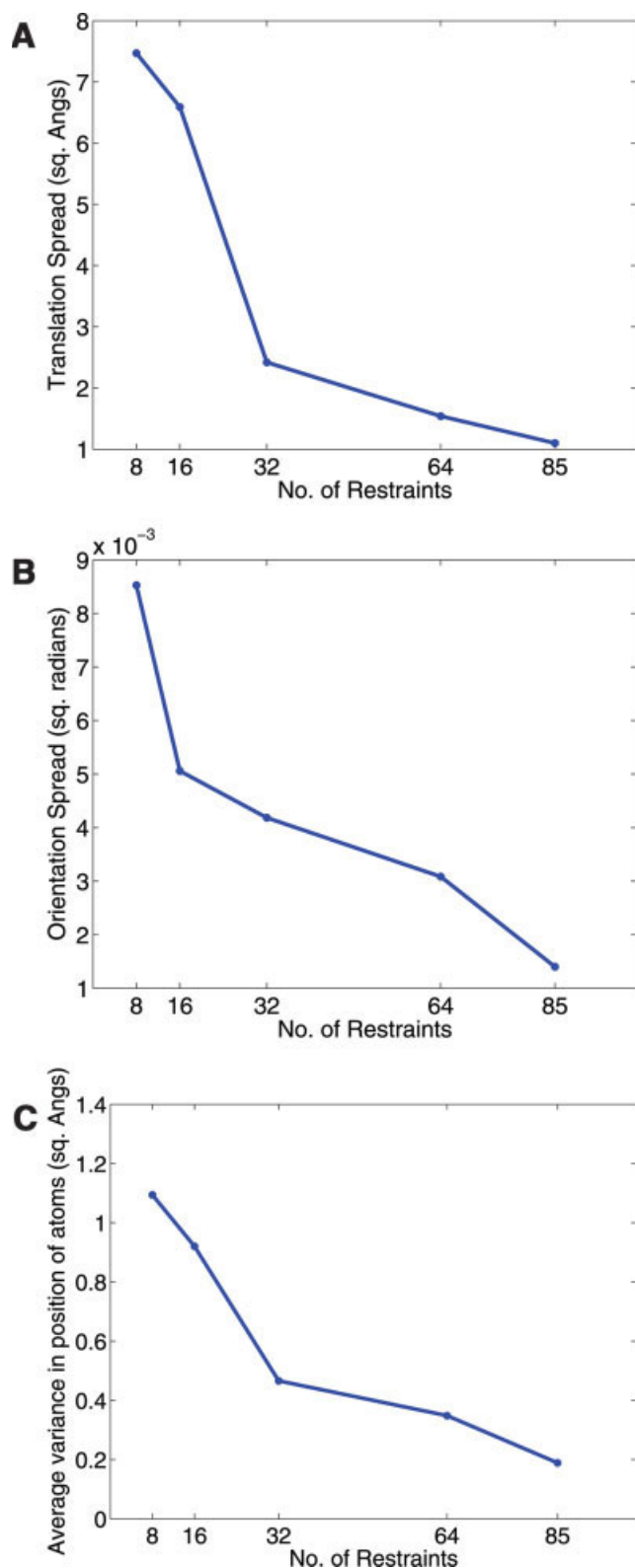


Fig. 12. Change with number of simulated restraints of (a) translation spread of WPS regions, (b) orientation spread of WPS regions, and (c) average variance in atomic positions for influenza haemagglutinin (1HTM). [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

Our approach is applicable to any  $C_n$  symmetric homooligomer irrespective of its size, given the subunit structure. While our scoring functions and bounds allow for some uncertainty in the subunit structure, we are currently working on extensions, to explicitly account for both side-chain and backbone uncertainty. Also, other kinds of symmetry (such as a dimer of dimers or a trimer of dimers), can be handled by defining appropriate configuration spaces and searching them in an analogous manner. In our current method to determine oligomeric number, we run our branch-and-bound search on each different putative oligomeric number. We intend to explore the possibility of incorporating the oligomeric number into the search space and perform our search in the ESCS,  $\mathbf{Z}_9 \times S^2 \times \mathbb{R}^2$  rather than using eight sequential searches of the SCS,  $S^2 \times \mathbb{R}^2$ . We also assume that the structure of a subunit when it is in complex is known. It is possible to extend the approach to account for flexibility in the subunit structure by establishing additional dimensions that represent concerted motions.

The only experimental data we currently use are from intersubunit NOEs. We intend to explore the possibility of using proximity information from other experiments, such as mutagenesis and chemical shift perturbation. The proximity information can be formulated as a set of “or” distance restraints—every atom will have a distance restraint to one of several atoms rather than just to one atom. We then eliminate grid cells based on violation of any of the “or” restraints. Residual dipolar couplings are another type of experimental information that we intend to use. When the dynamics of a symmetric homo-oligomer allows determination of high-quality RDCs, the orientation of the symmetry axis lies along the director of the alignment tensor and can be easily obtained. The structure determination problem then reduces from a 4D search in  $S^2 \times \mathbb{R}^2$  to a search in the 2D translation space,  $\mathbb{R}^2$ , and can be efficiently solved.

Our software can be freely obtained for academic use by request from the corresponding authors. The ensemble of structures for phospholamban determined by our algorithm will be deposited in the protein databank.

## ACKNOWLEDGMENTS

We would like to acknowledge members of the CBK lab and Ivelin Georgiev from the BRD lab for helpful discussions. JJC is a Pew scholar.

## REFERENCES

- Oxenoid K, Chou JJ. The structure of phospholamban pentamer reveals a channel-like architecture in membranes. *Proc Natl Acad Sci USA* 2005;102:10870–10875.
- Kovacs RJ, Nelson MT, Simmerman HK, Jones LR. Phospholamban forms  $\text{Ca}^{2+}$ -selective channels in lipid bilayers. *J Biol Chem* 1988;263:18364–18368.
- Nilges M. A calculation strategy for the structure determination of symmetric dimers by  $^1\text{H}$  NMR. *Proteins* 1993;17:297–309.
- O'Donoghue SI, Chang X, Abseher R, Nilges M, Led JJ. Unraveling the symmetry ambiguity in a hexamer: calculation of the R6 human insulin structure. *J Biomol NMR* 2000;16:93–108.



5. Russell RB, Alber F, Aloy P, Davis FP, Korkin D, Pichaud M, Topf M, Sali A. A structural perspective on protein-protein interactions. *Curr Opin Struct Biol* 2004;14:313-324.
6. Aloy P, Bottcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin AC, Bork P, Superti-Furga G, Serrano L. Structure-based assembly of protein complexes in yeast. *Science* 2004;303:2026-2029.
7. Aloy P, Russell RB. The third dimension for protein interactions and complexes. *Trends Biochem Sci* 2002;27:633-638.
8. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 2000;29:291-325.
9. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. *Proteins* 2000;39:178-194.
10. Gabb HA, Jackson RM, Sternberg MJ. Modeling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 1997;272:106-120.
11. Vakser IA. Protein docking for low-resolution structures. *Protein Eng* 1995;8:371-377.
12. Fernandez-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. *Protein Sci* 2002;11:280-291.
13. Mandell JG, Roberts VA, Pique ME, Kotlovi V, Mitchell JC, Nelson E, Tsigelny I, Eyck LFT. Protein docking using continuum electrostatics and geometric fit. *Protein Eng* 2001;14:105-113.
14. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 2003;331:281-299.
15. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* 2003;52:80-87.
16. Norel R, Petrey D, Wolfson HJ, Nussinov R. Examination of shape complementarity in docking of unbound proteins. *Proteins* 1999;36:307-317.
17. Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini A, Barzilay A, Dror O, Haspel N, Nussinov R. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* 2003;52:107-112.
18. Smith GR, Sternberg MJ. Prediction of protein-protein interactions by docking methods. *Curr Opin Struct Biol* 2002;12:28-35.
19. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, Vakser I, Wodak SJ. CAPRI: a critical assessment of predicted interactions. *Proteins* 2003;52:2-9.
20. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov A, Baker D. Protein-protein docking predictions for the CAPRI experiment. *Proteins* 2003;52:118-122.
21. Pierce B, Weng Z. M-ZDOCK: a grid-based approach for  $C_n$  symmetric multimer docking. *Bioinformatics* 2005;21:1472-1476.
22. Duhovny D, Nussinov R, Wolfson HJ. Efficient unbound docking of rigid molecules. In *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI)*, Rome, Italy, Sep 17-21, 2002. *Lecture Notes in Computer Science*, Vol. 2452, (Guigo R, Gusfield D, editors). Berlin: Springer-Verlag, 2002. pp 185-200.
23. Comeau SR, Camacho CJ. Predicting oligomeric assemblies: N-mers a primer. *J Struct Biol* 2005;150:233-244.
24. Dobrodumov A, Gronenborn AM. Filtering and selection of structural models: combining docking and NMR. *Proteins* 2003;53:18-32.
25. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125:1731-1737.
26. Anand GS, Law D, Mandell JG, Snead AN, Tsigelny I, Taylor SS, Eyck LFT, Komives EA. Identification of the protein kinase A regulatory RI $\alpha$ -catalytic subunit interface by amide H/2H exchange and protein docking. *Proc Natl Acad Sci USA* 2003; 100:13264-13269.
27. Morillas M, Gomez-Puertas P, Rubi B, Clotet J, Arino J, Valencia A, Hegardt FG, Serra D, Asins G. Structural model of a malonyl-CoA-binding site of carnitine octanoyltransferase and carnitine palmitoyltransferase I: mutational analysis of a malonyl-CoA affinity domain. *J Biol Chem* 2002;277:11473-11480.
28. Cunningham BC, Jhurani P, Ng P, Wells JA. Receptor and antibody epitopes in human growth hormone identified by homolog-scanning mutagenesis. *Science* 1989;243:1330-1336.
29. Adams PD, Engelman DM, Brünger AT. Improved prediction for the structure of the dimeric transmembrane domain of glycoporphin A obtained through global searching. *Proteins* 1996;26:257-261.
30. Adams PD, Arkin IT, Engelman DM, Brünger AT. Computational searching and mutagenesis suggest a structure for the pentameric transmembrane domain of phospholamban. *Nat Struct Biol* 1995; 2:154-162.
31. Lanman J, Lam TT, Barnes S, Sakalian M, Emmett MR, Marshall AG, Prevelige PE, Jr. Identification of novel interactions in HIV-1 capsid protein assembly by high-resolution mass spectrometry. *J Mol Biol* 2003;325:759-772.
32. Serino G, Su H, Peng Z, Tsuge T, Wei N, Gu H, Deng XW. Characterization of the last subunit of the arabidopsis COP9 signalosome: implications for the overall structure and origin of the complex. *Plant Cell* 2003;15:719-731.
33. Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* 2003;331:303-313.
34. Trester-Zedlitz M, Kamada K, Burley SK, Fenyo D, Chait BT, Muir TW. A modular cross-linking approach for exploring protein interactions. *J Am Chem Soc* 2003;125:2416-2425.
35. Truong K, Ikura M. The use of FRET imaging microscopy to detect protein-protein interactions and protein conformational changes in vivo. *Curr Opin Struct Biol* 2001;11:573-578.
36. Yan Y, Marriott G. Analysis of protein interactions using fluorescence technologies. *Curr Opin Chem Biol* 2003;7:635-640.
37. Kariakin A, Davydov D, Peterson JA, Jung C. A new approach to the study of protein-protein interaction by FTIR: complex formation between cytochrome P450BM-3 heme domain and FMN reductase domain. *Biochemistry* 2002;41:13514-13525.
38. Wang CE, Pérez TL, Tidor B. AmbiPack: a systematic algorithm for packing of macromolecular structures with ambiguous distance constraints. *Proteins: Struct Funct Genet* 1998;32:26-42.
39. Lilien RH, Bailey-Kellogg C, Anderson AA, Donald BR. A subgroup algorithm to identify cross-rotation peaks consistent with non-crystallographic symmetry. *Acta Crystallogr D Biol Crystallogr* 2004;60:1057-1067.
40. Ericson C. Real-time collision detection. San Francisco: Morgan Kaufmann, 2005.
41. Han J, Micheline K. Data mining concepts and techniques. San Francisco: Morgan Kaufmann, 2001.
42. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 1998;54:905-921.