

Determination of the Structures of Symmetric Protein Oligomers from NMR Chemical Shifts and Residual Dipolar Couplings

Nikolaos G. Sgourakis,[†] Oliver F. Lange,[†] Frank DiMaio,[†] Ingemar André,[§] Nicholas C. Fitzkee,[‡] Paolo Rossi,^{||} Gaetano T. Montelione,^{||} Ad Bax,[‡] and David Baker^{*,†,‡}

[†]Department of Biochemistry, University of Washington, Seattle, Washington 98195-7350, United States

[‡]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195-7350, United States

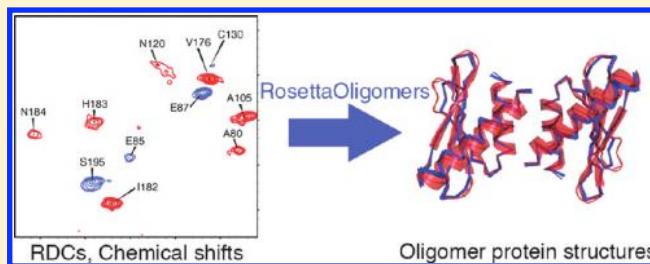
[§]Department of Biochemistry and Structural Biology, Centre for Molecular Protein Science, Chemical Centre, Lund University, PO Box 124, SE-22100 Lund, Sweden

[‡]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, United States

^{||}Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, and Northeast Structural Genomics Consortium, Rutgers University, Piscataway, New Jersey 08854, United States

S Supporting Information

ABSTRACT: Symmetric protein dimers, trimers, and higher-order cyclic oligomers play key roles in many biological processes. However, structural studies of oligomeric systems by solution NMR can be difficult due to slow tumbling of the system and the difficulty in identifying NOE interactions across protein interfaces. Here, we present an automated method (RosettaOligomers) for determining the solution structures of oligomeric systems using only chemical shifts, sparse NOEs, and domain orientation restraints from residual dipolar couplings (RDCs) without a need for a previously determined structure of the monomeric subunit. The method integrates previously developed Rosetta protocols for solving the structures of monomeric proteins using sparse NMR data and for predicting the structures of both nonintertwined and intertwined symmetric oligomers. We illustrated the performance of the method using a benchmark set of nine protein dimers, one trimer, and one tetramer with available experimental data and various interface topologies. The final converged structures are found to be in good agreement with both experimental data and previously published high-resolution structures. The new approach is more readily applicable to large oligomeric systems than conventional structure-determination protocols, which often require a large number of NOEs, and will likely become increasingly relevant as more high-molecular weight systems are studied by NMR.



INTRODUCTION

The majority of cellular proteins exist as symmetric oligomers with distinct biochemical and biophysical properties, which often provide the means for additional regulation of their function at the post-translational level.^{1–3} The study of oligomeric systems in solution is frequently hindered by their large molecular weight, which limits the resolution of NMR spectra due to fast transverse spin–spin relaxation rates, and complicates the application of NOE methods for the derivation of interface restraints. Moreover, the symmetry inherent to such protein complexes gives rise to spectral degeneracy, as the equivalent spin sites experience the same chemical environment among the protein subunits. Despite the spectral simplification resulting from symmetry, analysis of corresponding NOE spectra can be more complicated as cross peaks can represent intra- or intermolecular interactions. While more laborious isotopic filtering schemes can be used to distinguish between intra and inter subunit NOEs,^{4,5} such

measurements intrinsically offer lower sensitivity and are often not carried out.

Previous methods for the determination of the solution structure of dimeric proteins and protein complexes have often relied upon the availability of interface NOEs⁶ or highly ambiguous distance restraints, such as those obtained from chemical shift mapping experiments.^{7–9} Residual dipolar couplings (RDCs) and small-angle X-ray scattering (SAXS) data also have been used as supplementary refinement restraints to improve the convergence of the NMR ensemble.^{10–12} In the absence of distance restraints, the docking of protein complexes has proven to be a challenging task. In several recent studies, information obtained from fitting of RDC data to previously available structural models of the monomeric subunits was used to limit

Received: December 15, 2010

the degrees of freedom in a rigid-body search, according to the orientation of the alignment tensor.^{6,13–15} In systems with internal symmetry one of the axes of the alignment tensor must be collinear with the symmetry axis of the system and hence the rigid body degrees of freedom need only be sampled in the plane that is perpendicular to the symmetry axis. Although this method can successfully identify a binding interface that is consistent with the crystal structure,^{13,15} it is inherently limited by inaccuracies in the protein orientation due to lack of precision in the atomic coordinates of the models used to fit the experimental RDC data. Analysis has shown that these inaccuracies can lead to errors in the orientation of the alignment tensor on the order of 5–10°,¹⁶ which can dramatically alter the results of docking calculations. Moreover, these methods rely upon the availability of a previously determined structural model from either X-ray crystallography or conventional NMR methods.¹⁷ The difficulty in interpreting RDC data in the absence of an accurate structural model limits their use in determining the structure of dimers with unknown monomer structure, thus reducing the range of targets that can be studied in solution using RDCs as the only type of experimental data that report on the arrangement of the monomeric subunits.

Recent work from our group has shown that the structure of symmetrical assemblies of considerable size can be predicted using modeling methods as implemented in the program Rosetta. Symmetric docking in Rosetta¹⁸ can provide accurate structures of oligomers with various sizes and topologies, for which the structure of the monomeric subunit has been previously determined using X-ray crystallography. Rosetta can also provide high-resolution structures of multichain, symmetric oligomers with interleaved topologies using a protocol (fold-and-dock) in which the folding and docking degrees of freedom are explored simultaneously.¹⁹ In the current work, we extend these approaches to allow the high accuracy oligomer structure determination from chemical shifts, limited NOEs, and RDCs. We initially assume the oligomer is nonintertwined and begin by calculating the structure of the monomeric state using sparse NMR data, taking advantage of recent advances in the CS-Rosetta modeling methodology.^{20,21} We next use backbone RDC data as domain orientation restraints to dock the monomeric subunits. With such data, the Rosetta symmetric docking algorithm can effectively identify the native oligomer structure (provided that the number of monomers in the oligomer is known from experiments), without need for interface NOE restraints which are the main source of convergence in previously published protocols addressing this task. If the oligomeric structures produced are not converged, we restart the modeling calculations allowing for the possibility that the oligomer is intertwined by using the Rosetta fold-and-dock protocol guided by the chemical shift and RDC information. The method produces accurate oligomer models using backbone chemical shifts and RDCs from 1 to 2 alignment media for all of the benchmark cases studied.

RESULTS AND DISCUSSION

Overview of the Method. We have generalized both the Rosetta symmetric docking protocol¹⁸ and the Rosetta fold-and-dock protocol¹⁹ to take full advantage of NMR data to guide the conformational search. For noninterleaved oligomers, the NMR data allow fuller sampling of variations in monomer structure and guide homo-oligomer assembly. For interleaved structures, the NMR data better define the local structure and guide formation of the correct dimer interface.

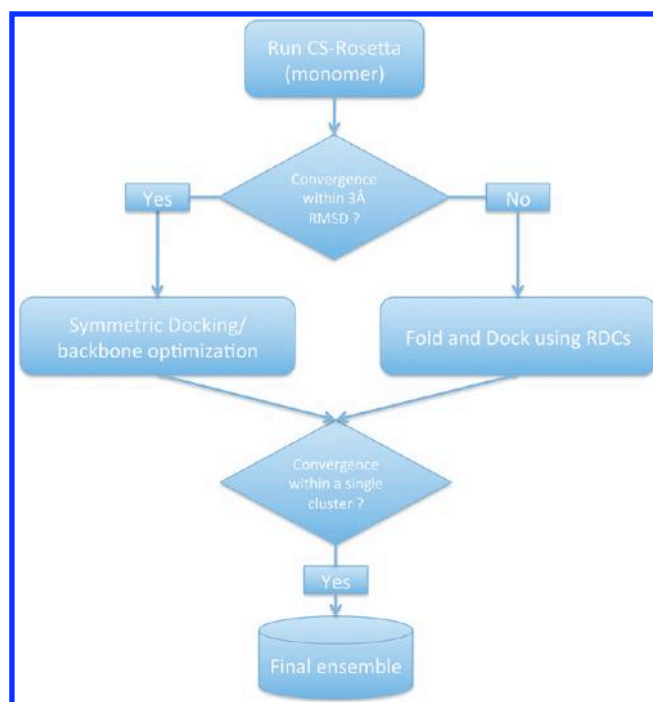


Figure 1. Flow diagram of the method. Starting from an extended protein chain, CS-Rosetta uses backbone chemical shift assignments to derive an ensemble of models for the monomer. Convergence at this step is used as an indication of a noninterleaved interface, in which case symmetric docking is performed for each monomer seed from the low-energy ensemble. If convergence is not observed, the complex may be intertwined and the internal and rigid body degrees of freedom are optimized simultaneously using the fold and dock protocol guided by the NMR data.

In the first step, models of the free monomer are generated using Rosetta supplemented with either chemical shifts alone²¹ or chemical shifts and a small number of backbone NOEs, depending on available data. Previous work has shown that by using chemical shifts alone, convergence to the correct structure is generally achieved for proteins of ~100 amino acids, while larger structures of up to ~150 amino acids often can be determined using a very limited number of NOEs. At this point, we assume that all assigned NOEs arise from intramolecular interactions. Having sampled a diversity of low-energy conformations at the monomer stage (Figure 2A), we then proceed to Rosetta symmetric docking¹⁸ supplemented with domain orientation restraints from RDC data.^{22,23} The rigid body orientation of the monomers and the side chain conformations are optimized by Monte Carlo minimization, as described in Materials and Methods.²⁴ The lowest energy dimers sampled are then subjected to simultaneous refinement of backbone, side chain, and rigid body degrees of freedom to allow for restricted (within 1 Å rmsd), local adaptations of the backbone to the final docked orientation of the dimer. The low-energy refined structures from independent docking runs starting from different monomers are pooled, and convergence is then assessed by structural clustering and by evaluation of pair wise RMSDs within the low-energy pool. The structural ensemble is considered converged if the 10 lowest-energy structures can be clustered within 3 Å backbone rmsd to the center of the cluster (a larger rmsd threshold could be useful for oligomers with more disordered loops). If there is clear convergence in the low energy population, the converged

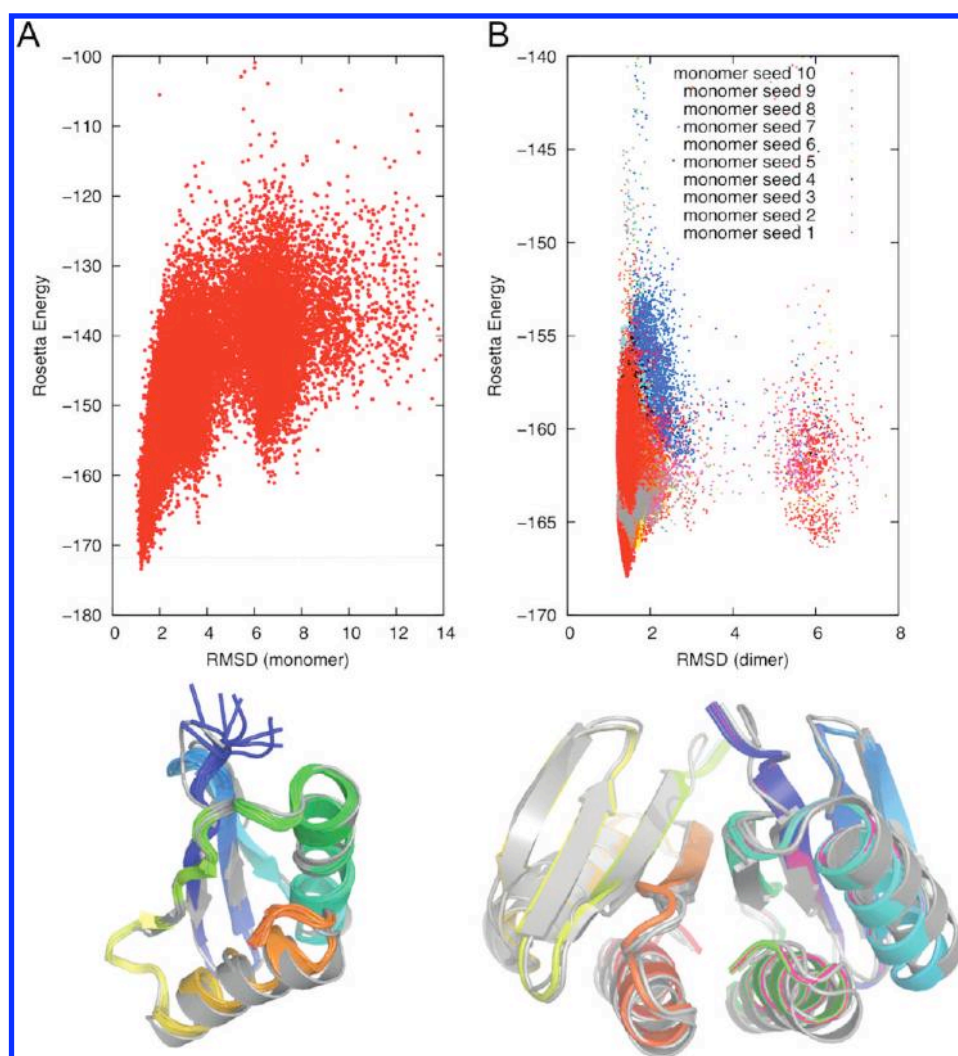


Figure 2. Overview of RosettaOligomers method. The two steps of the method are illustrated for the periplasmic protein TolR. (A) In the first step, the CS-Rosetta protocol produces a low-energy ensemble of full-atom conformations, showing a high degree of convergence to the monomer in the NMR structure (PDB ID 2JWK). An overlay of the 10 lowest-energy structures (color) on the native structure (gray) is shown in the structure diagram on the bottom. Only backbone chemical shift data are sufficient to produce this result (no RDCs were used at this step). (B) In a second step employing RDC data, the low-energy monomeric conformations from A, in this case produced in docking calculations from monomers 10 and 5, are docked in a symmetric manner to produce homodimeric structures that are within 1.5 Å from the previously reported NMR structure of the dimer, obtained using RDCs and a full set of interface NOEs.¹⁰

ensemble is taken as the computed model of the complex. Otherwise, if convergence is not observed, the oligomeric complex may be intertwined, and the Rosetta fold-and-dock protocol¹⁹ supplemented with RDCs is carried out as described below. This pipeline enables the structure determination of dimers showing a variety of interface types (Figure 1).

The structures generated by the protocol map the energy landscape of the complex, subject to the experimental constraints. Such a landscape is shown in Figure 2B for a complex of known structure. The funneling of the energy toward the native structure shows that among the diversity of backbone structures obtained at the monomer stage there exist conformations that can provide the correct backbone scaffold for convergence toward the native structure of the dimer. In this case the low energy docked conformations generated with different monomer seeds converge on the native structure of the dimer. The use of RDCs biases sampling of the rigid body degrees of freedom toward experimentally relevant regions of the conformational space and

further helps to discriminate against low-energy, non-native conformations.

Application of the Method to Symmetric Dimers and Comparison to Previously Published NMR Structure Determination Protocols. We compare the results of our method to docking results previously obtained using RDCs as the only type of interdomain restraints (Table 1). The structure of the homodimer ykuJ has been previously determined by both X-ray crystallography as well as a protocol based on a fixed symmetry axis forced to coincide with one of the principal axes of the alignment tensor.¹⁵ Using our new method we find a low-energy, converged ensemble that agrees with RDC data collected in two alignment media, as indicated by RDC Q-factors of 0.26 and 0.18 respectively, and which falls very close (0.9 Å backbone rmsd) to the X-ray structure (Figure 3A). For the side chains of most interface residues, there also is a high degree of convergence to the rotamers observed in the crystal structure (0.4 Å RMSD for all interfacial atoms), which reflects the use of Rosetta's

Table 1. Structural Statistics Reported from the Application of the RosettaOligomers Pipeline on Ten Oligomers with Experimentally Determined Structures

target name	PDB ID/method	size	fold/interface	data used ^a	Rosetta rmsd (Å) ^b	RDC Q-factor ^c
TolR	2JWK/NMR	74 × 2	αβ/αβ	CS, RDC(1,261), SAXS ^d	1.2/1.5/0.6	0.4
ykuJ	2FFG/X-ray	80 × 2	αβ/α	CS, RDC(2,59)	0.7/0.9/0.4	0.26/0.18
SeR13	2K1H/NMR ^e	86 × 2	αβ/β	CS, RDC(2,53), NOE (32)	2.0/3.4/3.4	0.26/0.24
At5g22580	1RJJ/NMR	101 × 2	αβ/β	CS, RDC(1,96)	2.2/2.2/1.7	0.4
HIV-1 CCD	1BIS/X-ray	152 × 2	αβ/αβ	CS, RDC(2,96)	1.0/1.3/0.7	0.30/0.32
yiiF	2K5J/NMR	44 × 2	αβ/αβ ^f	CS, RDC(1,24)	0.9/1.0/0.5	0.05
KR150	3OBH/X-ray ^e	74 × 2	αβ/αβ ^f	CS, RDC(1,56), NOE (67)	1.4/2.8/2.7	0.32
ATU0232	2K7I/NMR	64 × 2	αβ/αβ ^f	CS, RDC(1,46)	2.2/2.5/2.1	0.24
CA dimer	2KOD/NMR	77 × 2	α/α	CS, RDC(1,100)	1.3/1.4/1.2	0.1
P53	1C26/X-ray	31 × 4	αβ/αβ ^f	CS only	0.7/1.1/0.3	N/A

^a RDC (number of alignment media, number of RDCs per medium per monomer). All NOEs are assumed to be intramolecular. Typical estimated RDC errors are 0.5–1 Hz. ^b The rmsd calculated here for backbone atoms in the monomer/dimer/interface (defined here according to a 3.5 Å distance cutoff). ^c Q-factors calculated according to Cornilescu and co-workers⁵² for alignment Media A/B (if available). ^d When using SAXS data in addition to RDCs, the dimer structure can be determined using a limited data set of 68 NH RDCs. ^e Indicating comparison to a low-resolution structural model or structure of a homologous protein. ^f Indicating a dimer with an interleaved interface.

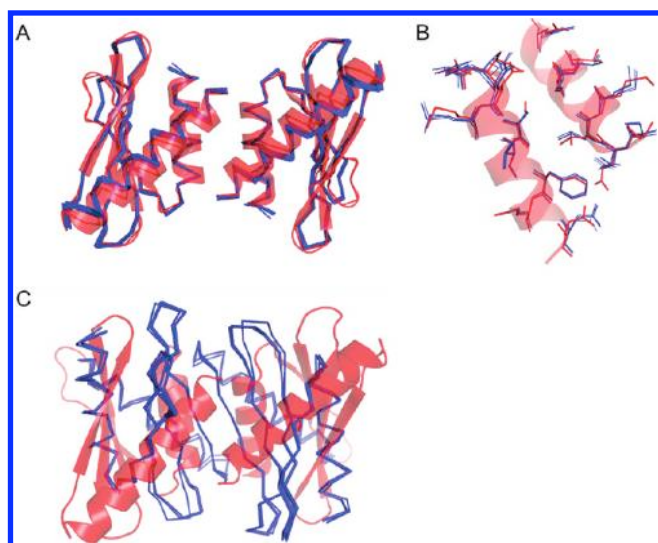


Figure 3. Structural convergence of the method for bacterial protein YkuJ. (A) Backbone of the 10 lowest energy conformations generated using RosettaOligomers (blue) superimposed on the crystal structure of the dimer in red (PDB ID 2FFG). (B) Detailed view of the side chains at the interface of the dimer, presented in a view that is perpendicular to the interface. Good convergence of the low energy conformations to the PDB structure is indicated by backbone rmsd values for the interface residues of ca. 0.4 Å (here the interface is defined as all residues within 3.5 Å from the other subunit). (C) Without the use of RDCs as orientation restraints during the search the docking algorithm converges to an alternative structural ensemble that uses a β-sheet interface to form the dimer.

full-atom energy function and the fact that all backbone and side chain degrees of freedom are optimized to accommodate structural changes arising from interactions between the monomeric subunits (Figure 3B). Such a degree of atomic detail in the absence of interface distance restraints could not be obtained using the rigid body search with the coarse energy terms presented in the earlier study.¹⁵ Using a similar protocol to fix the symmetry axis, in addition to employing paramagnetic

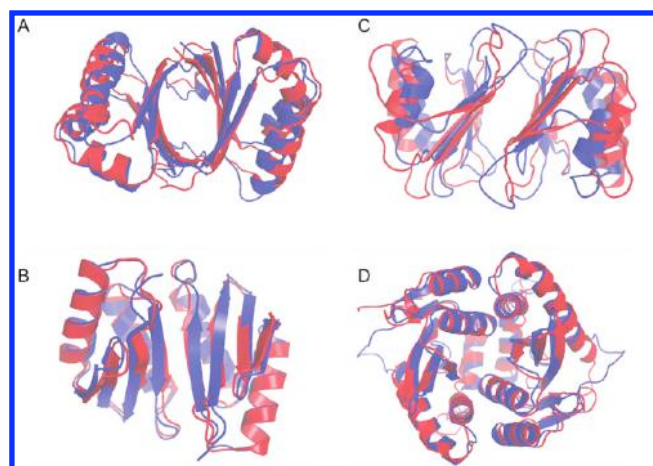


Figure 4. Comparison to previously determined structures. The lowest-energy dimer conformations obtained using RosettaOligomers are superimposed on the previously determined structures (in red). (A) *Arabidopsis thaliana* hypothetical protein and structural genomics target At5g22580, showing good agreement to the previously published X-ray structure (PDB ID 1RJJ). (B) Periplasmic protein TolR (PDB ID 2JWK), previously determined using a docking protocol in CNS²⁵ using RDCs and a set of 65 interface NOEs. (C) Weakly associating homodimer SeR13, previously modeled using a rigid body search of monomer orientations around a fixed axis of symmetry (PDB ID 2K1H). (D) Catalytic core domain of the HIV-1 integrase dimer (residues 50–212), showing 1.3 Å backbone coordinate rmsd relative to the previously determined X-ray structure (PDB ID 1BIS).

surface mapping, Lee and co-workers reported a structural model of the weakly associated homodimer SeR13.¹⁴ We applied our protocol to compute the dimer structure using only backbone chemical shifts and RDC data from two alignment media. The resulting structural ensemble is in excellent agreement with the measured RDCs (Q-factors of 0.26 and 0.24 respectively) and shows a high degree of convergence to the structural model of Lee et al. (Figure 4C).¹⁴

Taken together, these results show that our method can reproduce the results obtained using a fixed symmetry axis and

that the produced structural ensembles show a high degree of resolution and convergence to the native structure. Moreover, in the case of C_2 symmetry considered here, by searching all four independent rigid body degrees of freedom simultaneously rather than fixing the symmetry axis of the system according to the RDC alignment tensor, the present method is robust to errors in determining the axis of symmetry from the RDC data due to structural noise present in the monomer models.

We have also evaluated the performance of the new method relative to established protocols that make use of interface NOEs as well as RDCs and other data types. For the periplasmic protein TolR, previously determined with the program CNS²⁵ using a full set of distance restraints, including 65 interface NOEs, RDCs, and SAXS data,¹⁰ our method converged to a solution within 1.5 Å rmsd for the backbone and 0.6 Å rmsd for interface atoms from the published structural ensemble (Figure 4B). This was achieved using only backbone chemical shifts and RDCs, far less data than the CNS calculation. Convergence of the symmetric docking calculations improved when increasing amounts of RDC information was utilized; best results were obtained using all four internuclear vectors (N–H, C α –C', C'–N, C α –H α).

Similar results were obtained for the protein At5g22580, a 101-residue structural genomics target from *Arabidopsis thaliana*. The structure of the At5g22580 dimer was previously determined using RDCs, dihedral and hydrogen bond restraints, and to a set of 2117 assigned NOEs of which 31 were intermolecular.²⁶ In contrast, using only RDCs to dock the dimer, the current protocol converged to a 1.7 Å rmsd for the interface atoms (defined according to a 3.5 Å distance threshold between any pair of atoms on the monomeric subunits) relative to the NMR ensemble (Figure 4A). This illustrates the merit of our approach for solving the solution structure of larger-size homodimers (>100 residues), for which obtaining interface NOEs can become challenging and labor intensive.

Interleaved Dimers. The approach used in the above examples assumes only modest structural adaptations (within 1 Å rmsd) of the backbone due to the interactions between the monomers. However, the assumption that the monomers can fold as independent chains does not necessarily hold for all homo-oligomeric structures. A concern regarding the general applicability of the current strategy is its performance in the case of homo-oligomers with a significant degree of interaction between the two subunits, such as found in domain-swapped multimeric systems.²⁷ We have tested whether we can diagnose such cases without any a priori knowledge on the degree of interaction between the monomeric subunits. We considered the homodimer yif from *Shigella flexneri* with interleaved backbone topology (PDB ID 2K5J), involving the formation of a β -sheet using strands from the two monomers. As expected, CS-Rosetta calculations of the individual monomers failed to converge (Figure 5A); the native state cannot be energetically distinguished by considering only interactions within the monomer. If nevertheless the low energy, partially unfolded monomers are used as starting points in the symmetric docking protocol we obtain a converged structural ensemble, which shows a significant degree of interaction between the chains (Figure 5B and D) that further suggests an interleaved dimer. The previously published fold-and-dock protocol¹⁹ is a more suitable treatment for this system. When improved with the use of RDC data, this method converges to a 0.5 Å interface rmsd structure relative to the previously reported NMR ensemble (Figure 5C and E). The converged low-energy structures resulting from this protocol (Figure 5E) are consistently lower in energy than the

partially unfolded dimers obtained with symmetric docking (Figure 5D), providing further indication that the structure is interleaved.

We have further tested this approach for the protein ATU0232 from *Agrobacterium tumefaciens* (PDB ID 2K7I). The previously determined solution structure shows a complicated interleaved interface with an intermolecular 5-strand β -sheet in which subunit α forms strands 2–4, while subunit β forms strands 1,5. Again, the monomer CS-Rosetta calculations and symmetric docking calculations starting from the CS-Rosetta monomers do not show convergence, indicating an interleaved dimer interface. The fold-and-dock protocol, supplemented by 45 RDCs, converges to a 2.5 Å structure, which shows the correct interleaved backbone topology (Supporting Information Figure 1).

Very similar results were obtained for the structural genomics target KR150 with remote homology to the protein SP_0782 from *Streptococcus pneumoniae* for which a crystal structure of the dimer (PDB ID 3OBH) shows an interaction interface containing an exposed α -helix. Using the Rosetta fold-and-dock branch of the protocol together with the backbone chemical shifts, a set of 67 manually assigned backbone NOEs, and RDCs, we obtain a dimer structure that is within 2.6 Å backbone rmsd from the homologous structure (Supporting Information Figure 2). Furthermore, the structures are very similar in fold to the structures obtained using conventional structure calculations and a full set of NOEs. Good agreement with the RDCs is indicated by a Q-factor of 0.32 (see Materials and Methods for definition of the Q-factor). These results show that RDCs can guide accurate structure determination even for interleaved oligomers. Again, the lack of convergence for the monomer CS-Rosetta calculations are an indication of a potentially interleaved dimer and the user is guided to use the fold-and-dock branch of the protocol (Figure 1).

An intermediate case of a dimer with a semi-interleaved interface topology is the HIV-1 capsid protein (CA) C-terminal domain (CTD). Dimerization of the CTD results in the cross-linking of individual 5mer and 6mer rings formed through interactions of the N-terminal domain (NTD), which promotes the assembly of the virus capsid. The solution structure of the protein has been previously determined using RDCs, TALOS dihedral angle and hydrogen bond restraints, and a large set of NOEs, of which 210 were intermolecular.²⁸ The structure determined in solution fits well into the Cryo-EM density map and shows a semi-interleaved dimer interface in which the N-terminus of the CTD (residues 145–151 in the full-length protein) fit into a helical groove on the symmetric subunit, as confirmed by the observation of several intersubunit NOEs (PDB ID 2KOD). Here, we have used the chemical shift assignments from solid-state NMR experiments reported in reference 29 and backbone N–H and C α –H α RDCs measured in solution²⁸ as the only source of experimental information to dock the dimer into a structure that is very close (1.2 Å interface rmsd) to the previously reported NMR ensemble (Supporting Information Figure 3). The side chains of residues Trp 184 and Met 185, shown previously to be crucial for dimerization,³⁰ are found in the core of the interface forming packing interactions in the dimeric structure. Moreover, the monomeric subunit shows a kink in helix 9, which forms the core of the dimer interface, in agreement with the NMR structure and Cryo-EM density.²⁸ In this case, we used the fold-and-dock protocol followed by symmetric docking optimization of the obtained low-energy conformations using perturbation runs (see Materials and

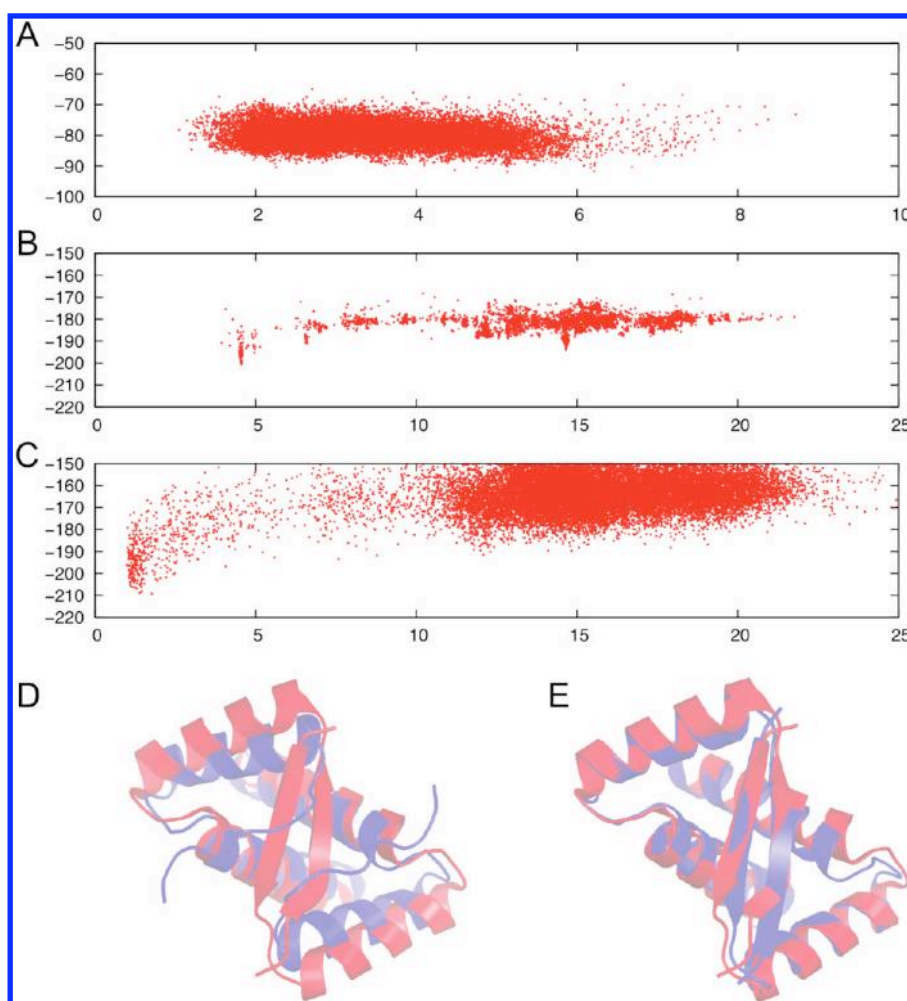


Figure 5. Results for the yif interleaved dimer using the two different modes of the RosettaOligomers structure determination pipeline. Results from the application of CS-Rosetta followed by RDC-assisted symmetric docking (A, B) are contrasted to results obtained using the fold-and-dock protocol supplemented with RDCs,¹⁹ specifically designed to predict the structure of interleaved dimers (C). (A) CS-Rosetta applied to the monomer. This does not converge to a single structure, due to the fact that interactions within the dimer are essential for folding the monomeric subunit. (B) Converged docking solutions. Using the nonconverged low-energy conformations from A, converged docking solutions are obtained (D) which contain a major part of the interaction interface but are missing the intersubunit β -strand pairing. (C) Correct native structure of the dimer. Using the previously published fold-and-dock protocol, this structure of the dimer is obtained (E). In all structure diagrams the native structure (PDB ID 2K5J) is shown in red, and the lowest-scoring Rosetta structure, in blue; X-axis backbone rmsd relative to the crystal structure (Å); Y-axis Rosetta full-atom energy, supplemented with an RDC energy term.

Methods). Although the low-energy conformations obtained from the fold-and-dock protocol showed moderate convergence, the use of symmetric docking optimization of the low-energy conformations resulted in improved convergence and recovered a docking funnel towards the native structure of the dimer.

Solution Structure of the HIV Integrase Homodimeric Catalytic Core Domain. We also evaluated the capability of RosettaOligomers to determine the structure of dimers built from larger monomers, using as a test case the 36 kDa homodimeric catalytic core domain (residues 50–212) of the HIV-1 integrase enzyme (IN^{50–212}), whose structure was originally solved by X-ray crystallography.^{31,32} Solution studies of a soluble variant of the wild-type sequence containing five point mutations have shown that it exists in a conformation for which the monomeric unit is very similar to that seen in the crystal structure; however, the data collected in solution were insufficient to determine the structure of the dimer.³³ The same study found that this variant of IN^{50–212} also exists predominantly as a symmetric dimer. We measured RDCs in two different alignment

media (see Materials and Methods for data collection details). With the use of backbone chemical shift-derived fragments, a converged structural ensemble is obtained for the monomeric subunit that falls very close to the monomeric subunit observed in the crystal structure (within 1 Å backbone rmsd calculated for residues in the well-ordered regions of the structure, secondary structure elements and structured loops). Consistent with earlier NMR data, the final ensemble shows a high degree of structural variability in the loop connecting the two C-terminal helices, spanning residues 185–195.³³ Moreover, the catalytic loop spanning residues 140–153, previously shown to be conformationally dynamic by ¹⁵N relaxation analysis, is found to be structurally variable in the low-energy monomer ensemble derived by CS-Rosetta (Supporting Information Figure 4).

Starting from the ensemble of monomers, using symmetric docking and the RDCs obtained in two alignment media, a converged dimer structural ensemble is obtained (Figure 6 inset). This ensemble is in good agreement with the crystal structure of the wild-type sequence, with a 1.3 Å backbone rmsd

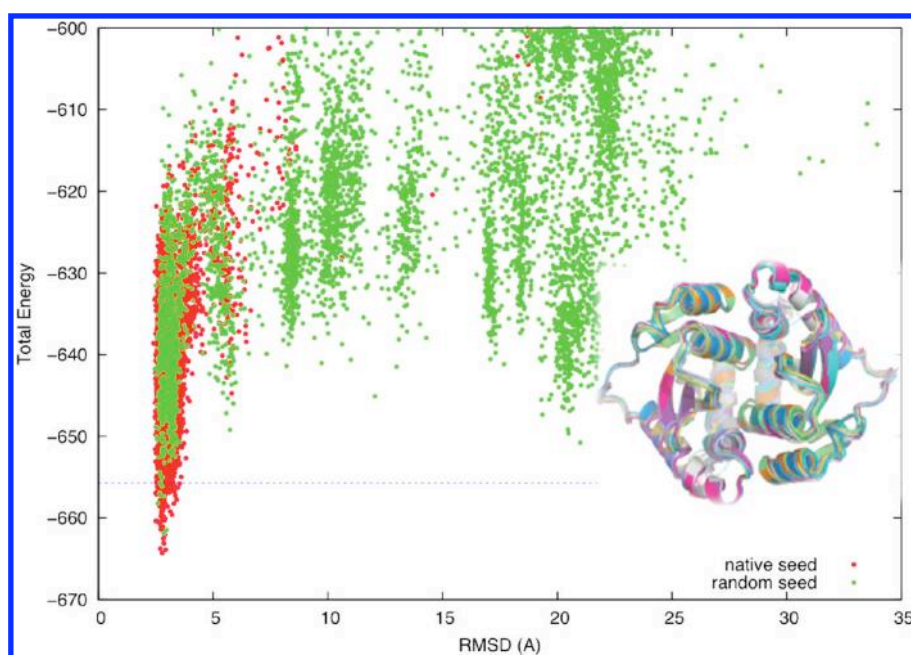


Figure 6. Solution structure of the HIV integrase dimer determined with RosettaOligomers using backbone chemical shifts and RDC data. The structure of the catalytic core domain of the HIV integrase homodimer (residues 50–212) was determined using exclusively solution NMR data. Backbone chemical shifts were used to solve the structure of the monomer, which was then docked in a symmetric manner with the use of N–H RDCs measured in two alignment media.⁵⁶ Two types of docking calculations, starting from the native dimer orientation (red) and starting from a completely randomized orientation (green) both converge to the same energy minimum, indicating global convergence of the method to the X-ray structure (PDB ID 1BIS), as shown in the structure diagram (inset). The rmsd is computed for the backbone atoms of the well-ordered regions of the molecule, as described in the main text. An ensemble of the 10 lowest-energy conformations (shown in color) is superimposed on the X-ray structure (shown in gray). An RDC Q-factor of 0.3 for both alignment media indicates good agreement of the final ensemble to the RDC data.

over the regions with well-defined electron density (excluding the disordered active site loop at residues 140–153). The C-terminal helix spanning residues 197–209, which forms part of the dimer interface, is in a very similar orientation as that in the crystal structure. The rmsd of the interface atoms is 0.7 Å, and there is a high degree of convergence of the interface side chains to the rotamers observed in the crystal structure. In the final dimer ensemble, the lack of structural variability for the catalytic loop is due to the fact that a single monomer conformation seed happened to provide most of the low-energy solutions at the symmetric docking stage.

These results are consistent with the observation of a single, ensemble-averaged resonance for each atom in the NMR spectra and show that the solution data are consistent with the dimeric structure observed in the crystal state, as also indicated by RDC quality factors of 0.3 and 0.32 for RDCs collected in liquid crystalline phage and PEG media, respectively. The high degree of consistency between experimental RDCs and the ones calculated from the dimer models presented here suggest that our method offers a reliable way to interpret limited solution data in deriving structural models of a quality that approaches high-resolution X-ray structures.

Applications to Higher-Order Oligomers. We have further tested the practical usefulness of our approach for the symmetric modeling of larger-size oligomers by applying it to determine the structure of the equine infectious anemia matrix virus protein homotrimer using previously published NMR data.³⁴ Previous solution studies have shown that the protein exists in equilibrium between a monomer and a trimer; however the structure of the trimer has not been previously determined in solution. Using

RDCs and backbone chemical shifts alone, our approach converges to a trimer structure that is in agreement with the RDC data (Q-factor of 0.4). Moreover, the structure determined here is in qualitative agreement with chemical shift mapping results from titration experiments that report on the residues that form the trimer interface (shown as red spheres in Supporting Information Figure 7). On the basis of the structure of a remote homologue, a different trimer organization was previously suggested;³⁴ the two alternative models should be distinguishable in future work using additional solution data, such as small-angle X-ray scattering and intermolecular NOEs.

The p53 oligomerization domain was the oligomer with the largest number of subunits evaluated here. Previous experimentally determined structures by both solution-state NMR^{35–37} and X-ray crystallography³⁸ have shown that it exists as a tetramer of D2 symmetry (a dimer of dimers), in which the basic dimer has an interface with an interleaved topology. Deriving the correct topology for such a system out of the large possible number of arrangements that would be compatible with identical resonance positions for the four components of the tetramer presents a difficult challenge. This problem was solved correctly by a detailed analysis of multiple isotope-edited and isotope-filtered NMR spectra. Modest differences between the original NMR structures and the subsequent 1.7 Å resolution X-ray structure in terms of backbone rmsd (1.2 and 1.9 Å to the mean coordinates of the two NMR ensembles, respectively^{35,37}) in part reflect the technical challenge in obtaining a high accuracy solution structure from NMR data. Subsequent refinement of the NMR structure³⁹ with the addition of multiple interdomain NOEs (through a more exhaustive peak-picking in the NOESY spectra) resulted in a more

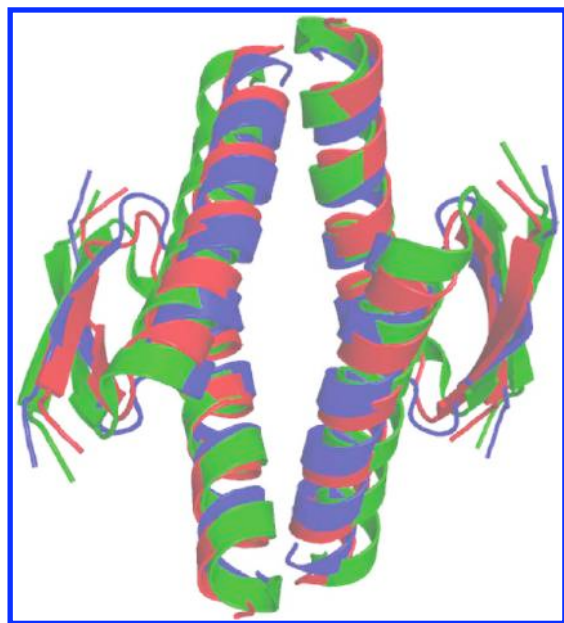


Figure 7. Comparison of structures determined using different methods. The structures of the p53 oligomerization domain tetramer (D2 symmetry) determined using X-ray crystallography³⁸ (PDB ID 1C26) (red), solution-state NMR³⁵ (PDB ID 1OLG) (green), and our method (blue) are superimposed on a same reference frame. Only backbone chemical shifts were used here to determine a highly similar structure, otherwise obtained using a full set of assigned intersubunit NOEs.

compact interface between the dimers and interhelical angles that are closer to those seen in the X-ray structure (backbone rmsd of 0.6 Å).³⁸ All these solution studies employed 2D, 3D, and 4D heteronuclear-separated and isotope-filtered NOESY techniques using samples of both uniformly labeled (¹⁵N, ¹³C) and mixed heterotetramers with equal amounts of labeled and unlabeled proteins, to distinguish between NOEs arising from interactions between the different subunit combinations. Together with other types of experimental information, such as chemical shifts, J-couplings, and hydrogen–deuterium exchange, numerous restraints (e.g., 4472 restraints in reference 39 of which 3752 were distance NOEs) were used to derive converged structural ensembles. The number of intersubunit restraints used in reference 35 was 864, including 840 NOEs and 24 hydrogen bond restraints.

Using our method and backbone chemical shifts alone, we obtain a converged structural ensemble that falls very close to the X-ray structure in terms of backbone rmsd (1.1 Å for all backbone atoms, 0.3 Å for all interface atoms, defined according to a 3.5 Å distance cutoff). A comparison with the crystal³⁸ and conventional NMR structures³⁵ (Figure 7) illustrates that the structure determined here is of comparable quality to the one determined using standard (and more laborious) NMR structure determination protocols, using a much more limited set of data (only N, H, C α , C β , and CO assignments were sufficient to obtain a converged structural ensemble), which is typically the starting point in data collection for NMR structure determination. Although we assumed D2 symmetry to obtain this result, C4 symmetry (the only other alternative for a 4-subunit protein) was excluded on the basis of separate calculations: with this type of symmetry, the calculations do not converge to a single structure and result in average energies that are far greater than when using D2 symmetry. This shows that our method also has the potential

to distinguish between different point groups in simple cases. Taken together, these results indicate the practical use of our approach for determining the structures of symmetric oligomers of various numbers of subunits and symmetry groups.

Use of Small-angle X-ray Scattering Data. We have evaluated the use of small-angle X-ray scattering (SAXS) data in our approach for the protein TolR using previously published data.¹⁰ To calculate SAXS curves from the coordinates of the sampled conformations we have implemented a method that uses a coarse-grained representation of the protein with residue-specific form factors that have been parametrized using a database of high-resolution protein structures.⁴⁰ A score term that is proportional to the rms from the experimental data is used in both the low-resolution search and full-atom refinement stages of the symmetric docking calculation (Supporting Information Figure 6c). When supplemented by SAXS data alone, the symmetric docking calculations converge to two local minima, showing that the use of SAXS data effectively eliminates the search in many additional false minima of the docking energy landscape otherwise observed in an unbiased calculation (Supporting Information Figure 6b, green versus red points). Inspection of representative dimer structures from each minimum shows that one corresponds to the native structure, while the other is a dimer in which one of the monomeric subunits is inverted relative to its native orientation (Supporting Information Figure 6d, e). This results in very similar SAXS profiles (Supporting Information Figure 6a) and Rosetta energies, suggesting that additional data types, such as RDCs, are needed for full convergence to a single structure. In fact, with the use of RDCs for the NH bond vectors alone in addition to the SAXS data, our method converges to the native dimer structure (Supporting Information Figure 6b, blue points). This indicates that the use of SAXS data can complement RDCs in dimer structure determination, by reducing the amount of RDC data required to achieve convergence (68 vs 261 RDCs required to achieve convergence in the absence of SAXS data).

CONCLUSIONS

The strategy presented here enables the determination of the higher-order structure of protein dimers using exclusively NMR data, such as backbone chemical shifts and amide ¹⁵N–¹H RDCs, without the need for any prior structures of the monomeric subunits. In all cases tested here, the method converges on structures similar to previously published high-resolution dimer structures obtained by X-ray crystallography or by conventional NMR structure determination protocols making use of interface NOEs. Moreover, the computed structures show details in terms of side chain orientations at the interface that are very similar to those determined using high-resolution methods. It is perhaps surprising that accurate models of oligomers can be generated from chemical shift and RDC data alone. The success of our approach illustrates the power of molecular symmetry in confining the search space and making modeling more tractable. Even with the constraints provided by symmetry, it is expected that for larger systems, inaccuracies in determining the monomeric structure from chemical shifts alone and the existence of many local minima in the docking energy landscape would make additional data necessary to unambiguously converge on the native structure of the oligomeric complex.

RosettaOligomers provides an automated pipeline for deriving accurate dimer structures by NMR that can be readily applied in high-throughput structural genomics initiatives. The approach

described here for homodimers can be readily extended to trimers and other size homo-oligomers of various symmetry groups, as illustrated for the equine anemia virus matrix protein trimer and the p53 tetramer. However, it is expected that in larger systems, inaccuracies in determining the monomeric structure from chemical shifts alone and the existence of many local minima in the docking energy landscape would require additional data to unambiguously converge on the native structure of the oligomeric complex. It is anticipated that incorporation of additional data types that report on the interface (sparse NOEs) and shape (SAXS) of the protein complexes will enable the structures of much larger oligomeric systems with internal symmetry to be solved, further expanding the range of biologically important systems amenable to solution NMR. Our method is ready to support such data types, thus providing a powerful tool for determining the solution structure of symmetric protein assemblies.

MATERIALS AND METHODS

Structural Ensemble Generation Using CS-Rosetta/Symmetric Docking. We have used the CS-Rosetta method as described previously^{20,21} to determine the ensemble of the monomeric subunit. All protocols used here can be downloaded as part of the standard Rosetta 3.0 distribution⁴¹ (SVN version 39640 can be obtained at <http://www.rosettacommons.org/>). To run CS-Rosetta starting from an extended polypeptide sequence, we first select backbone conformations of all possible overlapping residue fragments of three and nine residue lengths that are consistent with the recorded backbone chemical shifts. Also, 200 fragments are selected for each residue position from a chemical shift-annotated database.^{21,42} To perform this task, we are using a new fragment search method (manuscript in preparation), with information such as backbone secondary structure prediction using the TALOS+ program⁴³ and sequence profile information provided by the program PSI-BLAST.⁴⁴ The weights for the different types of selection criteria are given in a separate weights parameter file (see the Supporting Information). This method is robust to incomplete assignments to as low as one atom type per residue,⁴⁵ and, prior to running the structure calculations in CS-Rosetta, can be executed as a stand-alone application, using the command line options shown in the Supporting Information. In all cases attempted here, good convergence of the CS-Rosetta protocol was obtained by running 10 000–20 000 calculations on a Linux-based cluster.

In all cases used here to benchmark the method, homologues present in the fragment database were excluded from our analysis according to a sequence similarity criterion (PSI-BLAST score of 0.05 or less) and by manual exclusion from the fragment database of the structures that were highly represented in the selected fragments (present in more than 10% of all sequence positions) and showed structural similarity to the target proteins.

Having obtained fragments of lengths three and nine residues using chemical shift information, we proceed to the CS-Rosetta monomer calculation using the recently implemented minirosetta application,⁴¹ which also supports the inclusion of any available NOE and RDC constraints. The command-line options for this step are included in the Supporting Information.

The symmetric docking protocol,¹⁸ that was adapted to use RDC data, was then used to dock the low-scoring monomers extracted from the CS-Rosetta runs. In this protocol, the individual subunits are assumed to be perfectly symmetric about a user-defined axis. A detailed description of the implementation of symmetry used here is included in the original publication.¹⁸ For the C2 symmetry used here, or any type of cyclic symmetry, the orientation of the symmetry axis is defined in an input symmetry definition file. By default, the *z*-axis is set as the

symmetry axis of the system. Alternatively, the program can take an arbitrary symmetry axis as the axis of symmetry, which can be extracted in the form of a symmetry definition file from a pdb input file containing the coordinates of the dimer chains (A, B) using an in-house script, which is part of the standard Rosetta SVN distribution (an example of running the script is shown in the Supporting Information).

Using the symmetry definition file prepared in this manner and the input PDB files for the monomer conformations, we then perform symmetric docking calculations using the SymDock application (SVN version 39640 can be obtained at <http://www.rosettacommons.org/>) as described in the Supporting Information. In all cases attempted here, good convergence of the symmetric docking protocol was obtained by running 10 000 calculations on a Linux-based cluster.

Starting from a completely randomized orientation between the monomeric subunits around the symmetry axis, the symmetric docking protocol performs iterations of Monte Carlo-based optimization of the rigid body and side chain degrees of freedom in two steps: In a first, low resolution step using a coarse energy function, the rigid body orientation of the monomeric subunits is randomly perturbed and the two subunits are translated into contact along an axis that is perpendicular to the symmetry axis. At this step, side chains are represented using a single, residue-specific pseudoatom, positioned at the C^β carbon. Monte Carlo trials of the total energy of the system are used to find a local energy minimum of the rigid-body orientation of the symmetric subunits. In the second, more time-consuming high-resolution step, Rosetta's full-atom energy function is used with a soft-repulsive term for van der Waals interactions. During this stage, the side chains are combinatorially optimized and the rigid body and side chain degrees of freedom are subjected to quasi-Newton minimization after which the trial is accepted or rejected according to a Metropolis criterion.⁴⁶ Up to this point, the backbone is kept fixed to that of any of the lowest energy conformers, obtained by the CS-Rosetta structure determination for the monomer. A final relaxation step of all degrees of freedom, including the backbone dihedrals and side chains, was implemented in this study to account for local structural changes due to the interactions between the monomeric subunits, according to the algorithms described previously.⁴⁷ The adaptation in backbone rmsd during this step was found to be less than 1 Å for all proteins tested here. This step also allows for improved discrimination of the native docking funnel in Rosetta's full-atom energy.

To evaluate the robustness of our docking approach and to test the presence of a clear energetic signature of the native state in the Rosetta full-atom energy, for all test cases, we performed independent docking perturbation studies as previously described in reference 24. Starting from a symmetry definition file prepared using the native structure of the dimer as input, the orientation of the monomeric subunits was randomly perturbed by a displacement and a rotation around each one of the three axes drawn from Gaussians centered at 3 Å and 5°, respectively. The correct orientation between the monomeric subunits is consistently recovered in docking calculations using a small perturbation of the native dimer structure (red scatter plot in Figure 6) as well as using a completely randomized orientation of the two monomers (green scatter plot in Figure 6). This indicates a high degree of convergence of the docking algorithm to the lowest-energy structure and further shows that Rosetta's all-atom energy function enhanced by the RDC energy term is able to discriminate the native structure from the many non-native local energy minima.

Structural Ensemble Generation Using Fold-and-Dock. In the cases of dimers with interleaved interfaces, the fold-and-dock protocol simultaneously explores the folding and docking degrees of freedom, as described previously.¹⁹ The protocol consists of four low-resolution stages of increasing complexity in the energy function, in which symmetric fragment insertions are interleaved with symmetric rigid-body trials. Finally, symmetric repacking of the side chains and gradient-based minimization of the side chain, rigid body, and backbone

degrees of freedom are applied. In the current implementation, the protocol is run using the minirosetta application.

First, a symmetry definition file is constructed in the same manner as described previously for the symmetric docking protocol. Overlapping residue backbone fragments are selected according to the same methods used in CS-Rosetta.^{21,45} This application is included in the Rosetta 3.0 software suite⁴¹ and can be run as described in the Supporting Information. In one case (see CA dimer), the use of symmetric docking runs starting from the low-energy conformations obtained from the fold-and-dock protocol by perturbing the orientation of the individual subunits in the dimer was found to greatly improve convergence and native fold-discrimination. To perform such perturbation runs, a symmetric definition file is prepared using as input a conformation from the fold-and-dock low-energy ensemble, followed by symmetric docking as previously described. In all cases attempted here, good convergence of the fold-and-dock protocol was obtained by running 20 000–30 000 calculations on a Linux-based cluster.

Use of RDCs in Docking and Structure Refinement. During the symmetric docking or fold-and-dock protocols, RDC-based restraints were constructed by duplicating the measured RDC values for each subunit of the dimer. The singular value decomposition method as described by Losonczi and co-workers⁴⁸ was used to determine the elements of the alignment tensor that best fit the experimental data in the least-squares sense and to calculate RDC values given a structural model. The Jacobi method was implemented to calculate the eigenvalues of the order matrix for subsequent analysis.⁴⁹ Using this treatment, one of the axes of the order matrix is collinear with the symmetry axis of the system. Finally, a term that is proportional to the rmsd between experimental and calculated RDCs was used during the Monte Carlo trials and gradient-based minimization, according to the implementation previously described by Hess and Scheek, which allows for gradient-based optimization of the RDC target function.⁵⁰ Initial estimates for the magnitude of the alignment tensor, for the purpose of rescaling data sets from multiple alignment media, were obtained from a powder pattern distribution of the RDC data.⁵¹ For the purpose of validation of final structural models, we have calculated Q -factors, defined as

$$Q = \sqrt{\sum (D_{\text{calc}} - D_{\text{obs}})^2} / \text{RMS}(D_{\text{obs}})$$

after Cornilescu and co-workers.⁵²

RDC Measurements. The catalytic core domain (residues 50–212) of HIV-1 integrase (strain NL4-3) was expressed recombinantly and purified as described previously.³³ The soluble Q53E C56S W131E F185K Q209E variant was used for all experiments. Perdeuterated, ¹⁵N-, ¹³C-labeled protein was concentrated by centrifugal ultrafiltration to 500 μ M monomer concentration (250 μ M dimer) in 100 mM NaCl, 20 mM PIPES buffer (pH 6.5), 40 mM MgCl₂, 0.5 mM Tris (2-Carboxyethyl) phosphine (TCEP), 0.02% (w/v) NaN₃, 6% (v/v) D₂O. The sample was separately aligned in two media, bacteriophage Pfl^{53,54} obtained from ASLA Biotech (Riga, Latvia) and 4% (w/v) C12E5 polyethylene glycol (PEG)/*n*-hexanol.⁵⁵ For the Pfl-aligned sample, Pfl was added to a final concentration of 12 mg/mL, and the NaCl concentration was increased to 200 mM to reduce nonspecific interactions between the protein and the phage. The ²H quadrupolar splittings were 8.3 and 19.6 Hz in the Pfl and PEG-aligned samples, respectively. RDCs were measured at 25 °C on a Bruker Avance-III 900 MHz spectrometer, equipped with a triple-resonance cryogenic probe. Couplings were obtained from 2D ¹⁵N–¹H TROSY-HSQC spectra using the ARTSY technique.⁵⁶ The ¹⁵N acquisition time was 80 ms (250 complex points), and the ¹H acquisition time was 110 ms (1784 complex points).

NMR sample preparation and backbone assignments of KR150 using standard triple resonance experiments were performed as described previously.⁵⁷ RDCs were measured in 4% (w/v) C12E5 polyethylene glycol (PEG)/*n*-hexanol using a J-modulated experiment.^{58,59}

■ ASSOCIATED CONTENT

S Supporting Information. An example of all input files and commands used in the different steps of RosettaOligomers referenced in Materials and Methods; a table with the HIV integrase CCD RDCs; results of the method for targets KR150, ATU0232, and the HIV CA CTD dimer; structural superposition of the monomer seeds used for calculation of the CCD dimer; results using synthetic RDC data; results using SAXS data; complete reference 41. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

dabaker@u.washington.edu

■ ACKNOWLEDGMENT

This work was supported by NIH (grant 5R01GM092802 – 02) and HHMI (to D.B.), the Intramural Research Program of the NIDDK, NIH, by the Intramural AIDS-Targeted Antiviral Program of the Office of the Director, NIH (to A.B.), and by an NIH Intramural AIDS Research Postdoctoral Fellowship (to N.C.F.). A fellowship by Knut and Alice Wallenberg foundation (to I.A.) and the Human Frontiers of Science Program (to O.F.L.).

■ REFERENCES

- (1) Andre, I.; Strauss, C. E. M.; Kaplan, D. B.; Bradley, P.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16148.
- (2) Goodsell, D. S.; Olson, A. J. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105.
- (3) Wolynes, P. G. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 14249.
- (4) Ikura, M.; Bax, A. *J. Am. Chem. Soc.* **1992**, *114*, 2433.
- (5) Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 6711.
- (6) Clore, G. M. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9021.
- (7) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. *J. Am. Chem. Soc.* **2003**, *125*, 1731.
- (8) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2003**, *125*, 2902.
- (9) Lingel, A.; Weiss, T. M.; Niebuhr, M.; Pan, B.; Appleton, B. A.; Wiesmann, C.; Bazan, J. F.; Fairbrother, W. J. *Structure* **2009**, *17*, 1398.
- (10) Parsons, L. M.; Grishaev, A.; Bax, A. *Biochemistry—US* **2008**, *47*, 3131.
- (11) Schwieters, C. D.; Suh, J. Y.; Grishaev, A.; Ghirlando, R.; Takayama, Y.; Clore, G. M. *J. Am. Chem. Soc.* **2010**, *132*, 13026.
- (12) Pons, C.; D'Abramo, M.; Svergun, D. I.; Orozco, M.; Bernado, P.; Fernandez-Recio, J. *J. Mol. Biol.* **2010**, *403*, 217.
- (13) Dam, J.; Baber, J.; Grishaev, A.; Malchiodi, E. L.; Schuck, P.; Bax, A.; Mariuzza, R. A. *J. Mol. Biol.* **2006**, *362*, 102.
- (14) Lee, H. W.; Wylie, G.; Bansal, S.; Wang, X.; Barb, A. W.; Macnaughtan, M. A.; Ertekin, A.; Montelione, G. T.; Prestegard, J. H. *Protein Sci.* **2010**, *19*, 1673.
- (15) Wang, X.; Bansal, S.; Jiang, M.; Prestegard, J. H. *Protein Sci.* **2008**, *17*, 899.
- (16) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127.
- (17) Simon, B.; Madl, T.; Mackereth, C. D.; Nilges, M.; Sattler, M. *Angew. Chem., Int. Ed. Engl.* **2010**, *49*, 1967.
- (18) Andre, I.; Bradley, P.; Wang, C.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17656.
- (19) Das, R.; Andre, I.; Shen, Y.; Wu, Y. B.; Lemak, A.; Bansal, S.; Arrowsmith, C. H.; Szyperki, T.; Baker, D. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 18978.

- (20) Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G. H.; Ramelot, T. A.; Eletsky, A.; Szyperki, T.; Kennedy, M. A.; Prestegard, J.; Montelione, G. T.; Baker, D. *Science* **2010**, 327, 1014.
- (21) Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperki, T.; Montelione, G. T.; Baker, D.; Bax, A. *Proc. Natl. Acad. Sci. USA* **2008**, 105, 4685.
- (22) Tjandra, N.; Bax, A. *Science* **1997**, 278, 1111.
- (23) Tolman, J. R.; Flanagan, J. M.; Kennedy, M. A.; Prestegard, J. H. *Proc. Natl. Acad. Sci. USA* **1995**, 92, 9279.
- (24) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. *J. Mol. Biol.* **2003**, 331, 281.
- (25) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr. D Biol. Crystallogr.* **1998**, 54, 905.
- (26) Cornilescu, G.; Cornilescu, C. C.; Zhao, Q.; Frederick, R. O.; Peterson, F. C.; Thao, S.; Markley, J. L. *J. Biomol. NMR* **2004**, 29, 387.
- (27) Liu, Y.; Eisenberg, D. *Protein Sci.* **2002**, 11, 1285.
- (28) Byeon, I. J. L.; Meng, X.; Jung, J. W.; Zhao, G. P.; Yang, R. F.; Ahn, J. W.; Shi, J.; Concel, J.; Aiken, C.; Zhang, P. J.; Gronenborn, A. M. *Cell* **2009**, 139, 780.
- (29) Han, Y.; Ahn, J.; Concel, J.; Byeon, I. J. L.; Gronenborn, A. M.; Yang, J.; Polenova, T. *J. Am. Chem. Soc.* **2010**, 132, 1976.
- (30) Gamble, T. R.; Yoo, S. H.; Vajdos, F. F.; vonSchwedler, U. K.; Worthylake, D. K.; Wang, H.; McCutcheon, J. P.; Sundquist, W. I.; Hill, C. P. *Science* **1997**, 278, 849.
- (31) Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Engelman, A.; Craigie, R.; Davies, D. R. *Science* **1994**, 266, 1981.
- (32) Goldgur, Y.; Dyda, F.; Hickman, A. B.; Jenkins, T. M.; Craigie, R.; Davies, D. R. *Proc. Natl. Acad. Sci. U S A* **1998**, 95, 9150.
- (33) Fitzkee, N. C.; Masse, J. E.; Shen, Y.; Davies, D. R.; Bax, A. *J. Biol. Chem.* **2010**, 285, 18072.
- (34) Chen, K.; Bachtair, I.; Piszczek, G.; Bouamr, F.; Carter, C.; Tjandra, N. *Biochemistry—US* **2008**, 47, 1928.
- (35) Clore, G. M.; Omichinski, J. G.; Sakaguchi, K.; Zambrano, N.; Sakamoto, H.; Appella, E.; Gronenborn, A. M. *Science* **1994**, 265, 386.
- (36) Clore, G. M.; Omichinski, J. G.; Sakaguchi, K.; Zambrano, N.; Sakamoto, H.; Appella, E.; Gronenborn, A. M. *Science* **1995**, 267, 1515.
- (37) Lee, W.; Harvey, T. S.; Yin, Y.; Yau, P.; Litchfield, D.; Arrowsmith, C. H. *Nat. Struct. Biol.* **1994**, 1, 877.
- (38) Jeffrey, P. D.; Gorina, S.; Pavletich, N. P. *Science* **1995**, 267, 1498.
- (39) Clore, G. M.; Ernst, J.; Clubb, R.; Omichinski, J. G.; Kennedy, W. M.; Sakaguchi, K.; Appella, E.; Gronenborn, A. M. *Nat. Struct. Biol.* **1995**, 2, 321.
- (40) Stovgaard, K.; Andreetta, C.; Ferkinghoff-Borg, J.; Hamelryck, T. *BMC Bioinf.* **2010**, 11, 429.
- (41) Leaver-Fay, A.; et al. *Methods Enzymol.* **2011**, 487, 545.
- (42) Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D. *J. Mol. Biol.* **1997**, 268, 209.
- (43) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, 44, 213.
- (44) Schaffer, A. A.; Aravind, L.; Madden, T. L.; Shavirin, S.; Spouge, J. L.; Wolf, Y. I.; Koonin, E. V.; Altschul, S. F. *Nucleic Acids Res.* **2001**, 29, 2994.
- (45) Shen, Y.; Vernon, R.; Baker, D.; Bax, A. *J. Biomol. NMR* **2009**, 43, 63.
- (46) Bradley, P.; Misura, K. M. S.; Baker, D. *Science* **2005**, 309, 1868.
- (47) Tyka, M. D.; Keedy, D. A.; Andre, I.; Dimaio, F.; Song, Y.; Richardson, D. C.; Richardson, J. S.; Baker, D. *J. Mol. Biol.* **2011**, 405, 607.
- (48) Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. *J. Magn. Reson.* **1999**, 138, 334.
- (49) Golub, G. H.; Van Loan, C. F. *Matrix Computations*; Johns Hopkins University Press: Baltimore, 1996.
- (50) Hess, B.; Scheek, R. M. *J. Magn. Reson.* **2003**, 164, 19.
- (51) Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Magn. Reson.* **1998**, 133, 216.
- (52) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, 120, 6836.
- (53) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, 120, 10571.
- (54) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Biol.* **1998**, 5, 1065.
- (55) Otting, G.; Ruckert, M.; Levitt, M. H.; Moshref, A. *J. Biomol. NMR* **2000**, 16, 343.
- (56) Fitzkee, N. C.; Bax, A. *J. Biomol. NMR* **2010**, 48, 65.
- (57) Raman, S.; Huang, Y. J.; Mao, B.; Rossi, P.; Aramini, J. M.; Liu, G.; Montelione, G. T.; Baker, D. *J. Am. Chem. Soc.* **2010**, 132, 202.
- (58) Tjandra, N.; Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1996**, 118, 6264.
- (59) Chou, J. J.; Gaemers, S.; Howder, B.; Louis, J. M.; Bax, A. *J. Biomol. NMR* **2001**, 21, 377.